



Superb supervision: A pilot study on training supervisors to convey responsible research practices onto their PhD candidates

Tamarinde Haven, Dr^{a,*}, Lex Bouter, Professor Dr^{b,c}, Louise Mennen, Dr Ir^d, and Joeri Tjink, Dr^{b,e}

^a QUEST Center for Responsible Research, Berlin Institute of Health at Charité – Universitätsmedizin, Berlin, Germany ^b
Department of Philosophy, Faculty of Humanities, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands ^c
Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, The Netherlands ^d
Mennen Training & Consultancy, Haarlem, The Netherlands ^e Department of Ethics, Law and Humanities, Amsterdam
University Medical Centers, Amsterdam, The Netherlands

ABSTRACT

One way to strengthen research integrity, is through supervision. According to previous research, a supervisor should be well-versed in responsible research practices (RRPs) and possess the necessary interpersonal skills to convey RRP. We developed a 3-day pilot training for PhD supervisors that combined RRP and interpersonal skills. Our aim was to assess: perceptions regarding supervision skills (before and after the pilot) and participants' views on combining RRP and interpersonal skills. Before and after the pilot, we sent the Research Supervision Quality Evaluation survey to the participating PhD supervisors and their PhD candidates. The pilot was concluded with a focus group where participants deliberated over the combination of training in interpersonal skills and RRP and whether such training should become compulsory. Both supervisors and PhD candidates were more positive about the supervisor's interpersonal skills and the ability to foster RRP after the training. Participants were enthusiastic about the training's dual focus but believed that making the training compulsory would be undesirable. The results highlight the potential of RRP training for supervisors. However, caution is warranted, as the results regard a small sample of volunteering supervisors, underscoring the need for larger programs to foster responsible supervision that are rigorously evaluated.

KEYWORDS

research integrity, responsible conduct of research, supervision, mentoring, interpersonal skills

CONTACT Tamarinde Haven  tamarinde.haven@bih-charite.de QUEST Center for Responsible Research, Berlin
Institute of Health at Charité – Universitätsmedizin, Berlin, Germany

* Corresponding author conducted this study when employed by the Vrije Universiteit Amsterdam

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Introduction

Against increasing awareness of research misbehavior (Fanelli 2009; Xie, Wang, and Kong 2021), there is a growing interest in factors that foster research integrity. One of those factors is supervision or mentoring of PhD candidates (Bird 2001; Bouter 2015). Bird (2001) distinguishes between the mentor and the supervisor. She regards the latter as a more formal role that is assigned to a researcher and holds the responsibility for guiding the PhD candidate to obtaining the PhD degree. The mentor is thought to be a personal guide in the PhD candidate's professional development. In The Netherlands, PhD candidates often get assigned a supervisory team consisting of one or two full professors and a day-to-day supervisor. This day-to-day

supervisor is supposed to be the first point of contact for the PhD candidates and our focus in this paper is on these supervisors.

The relation between supervision and research integrity is bidirectional; on the one hand, poor supervision may increase the chances of the PhD candidate engaging in research misbehavior (Anderson et al. 2007). We have previously labeled insufficient supervision a research misbehavior (Bouter et al. 2016), and researchers across academic ranks and disciplinary fields seem concerned about it (Haven et al. 2019b). When analyzing research misconduct cases of PhD candidates, it became apparent that in most cases the supervision had been inadequate (Davis, Riske-Morris, and Diaz 2007; Wright, Titus, and Cornelison 2008). Some authors have argued that, in those cases, the supervisor should be (partially) held responsible for the misconduct (Alfredo and Hart 2011).

On the other hand, supervisors well versed in responsible research practices (RRPs) may foster research integrity among their PhD candidates (Anderson et al. 2007; Krishna and Peter 2018). Supervisors can then be considered exemplars and could function as role models for their PhD candidates (Gray and Jordan 2012). Their engagement and role may help prevent questionable research practices or even misconduct in their PhD candidates.

This led various authors to call for research integrity training for PhD supervisors (Eisen and Berry 2002; Motta 2002; Titus and Ballou 2014), which in some countries has become mandatory (LeRouge & Hol, 2020). Efforts like these are welcomed, as education is thought to play a key role in preventing research misconduct and fostering responsible research practices (Academies, 2017; Resnik and Dinse 2012; Steneck 2013). Besides, there may be a lack of knowledge among supervisors regarding how to supervise according to the highest research integrity standards (Van Noorden 2018), or lack of awareness about supervisors' role (Löfström et al. 2014) in conveying RRP (Ripley et al. 2012).

On top of that, PhD supervisors that were considered exemplars in RRP, stressed the importance of good interpersonal skills (Antes and DuBois 2018; Antes, Kuykendall, and DuBois 2019a, 2019b). They spoke of providing regular feedback, seeing PhD candidates first and foremost as “human beings” and of how this is critical for creating an atmosphere where the PhD candidates develop a reflective stance and dare to speak up and ask questions. Supervisors also need to be attentive to stress or a fear of failure among their PhD candidates, which has been linked to detrimental research practices (Kornfeld 2012; Wright, Titus, and Cornelison 2008). Recipients from the *Nature Mentoring in Science* awards gave similar advice and reported how they invested in their listening skills and in fostering an open atmosphere where failures can be discussed (Clynes, Corbett, and Overbaugh 2019). Supervisors should do well to equip themselves with these interpersonal skills to fulfil their critical role (Muthanna and Alduais 2020; Woolderink et al. 2015).

The picture that emerges is that supervisors should be encouraged to explicitly discuss RRP with their PhD candidates. As a prerequisite, the supervisor needs to be well-versed in RRP and possess the necessary interpersonal skills to create an atmosphere where mistakes and dilemmas can be discussed. It is important to stress that we consider these skills as mutually reinforcing, rather than separate components. Interpersonal skills can be instrumental in conveying RRP and discussing research integrity dilemmas and vice versa. It is against this background that we developed and piloted a training for junior PhD supervisors, called *Superb Supervision – Mentoring your PhD candidate toward research integrity*. Our research question was threefold: 1) How do PhD supervisors evaluate their supervision skills (i.e., interpersonal and RRP) before and after the *Superb Supervision* training, 2) how do the PhD candidates of the participating supervisors evaluate their supervisor's supervision skills, and 3) how do participating PhD supervisors evaluate the combination of learning interpersonal skills and RRP dexterity?

Materials and methods

Ethical considerations

The Scientific and Ethical Review Board (VCWE) of the Faculty of Behavior & Movement Sciences, VU University Amsterdam has reviewed our study entitled “Academic Research Culture in Amsterdam” (henceforth: ARCA). Their positive advice was valid for 5 years after the approval date (March 30th, 2017). Approval Number: VCWE-2017-017R1.

Development of the training curriculum

The curriculum was developed as part of the Academic Research Climate in Amsterdam project (ARCA, see: <https://abrahamkuypercenter.nl/portfolio/academic-research-climate-amsterdam-2017-2021/>). We tried to integrate key findings from the ARCA project into the training’s content (Haven et al. 2020, 2019a, 2019b). Key findings relevant to this training included that insufficient supervision of PhD candidates and junior coworkers is the most impactful research misbehavior that researchers across academic ranks and disciplinary fields are concerned about and that discussing expectations between supervisors and supervisees matters.

Between October 2017 and September 2018, we teamed up with an established trainer in academia (<https://www.mennen-tc.com>) to craft our training curriculum. We aimed to intertwine research integrity and interpersonal skill development and show the synergy between these two pillars. We wanted to emphasize the connection between interpersonal skills and RRP by continuously connecting them in the program. As can be seen from Figure 1, the topics were interwoven throughout the three days. The leading trainer rotated with the topics, i.e., when the topic was mostly related to interpersonal skills, LM was in charge and TH supported. When the topic primarily concerned RRP, TH was in charge and LM supported.

Day	Theme	RRPs	IPSS	Topic & brief explanation
1	An optimal relationship with your PhD candidate	✓	✗	<i>Research integrity: dilemmas & research misbehaviour</i> * Play the Erasmus Dilemma Game * Reflect in small groups on misbehaviour in own lab and impact on PhD candidate
	(cont.)	✓	✓	<i>Responsible working climate & expectations management</i> * Practice recognising emotions, focusing on body language & reflect on how vulnerability fosters connection * Reflect on expectations, practice management expectations
	(cont.)	✗	✓	<i>True listening, giving feedback & coaching</i> * Practice different ways of listening, explain bad ways of giving feedback and practice good ways, short role play of when and how to do coaching, practice with observer
2	Conveying research integrity	✓	✗	<i>Tools for transparency</i> * Practice how to upload a preprint and how this could help PhD candidates, virtual tour PubPeer, discussing checklist for transparent preregistration
	(cont.)	✓	✓	<i>Good practices for assigning authorship</i> * Reflect on different reasons for assigning authorship, and exchanging tips to assure authorship is handled adequately on PhD candidate's paper, discuss publication plan
	(cont.)	✓	✓	<i>Role play with actors</i> * Participants are coupled with an actor who acts out a challenging case where boundaries were crossed, the trainer and other participants support and give feedback
3	You as a responsible supervisor	✓	✓	<i>Own pitfalls regarding research integrity & recognising imposter syndrome in PhD candidates</i> * Reflect on personal drivers in supervision * Practice tools to detect imposter syndrome
	(cont.)	✓	✓	<i>Stress and its relation to research integrity</i> * Practice self observation to detect stress in PhD candidate * Practice preventing conveying stress onto PhD candidates
	(cont.)	✓	✓	<i>Learning from role models & becoming your own role model</i> * World café around questions related to role modelling, making a mind map on responsible supervision

Figure 1. Overview of pilot training curriculum overview of the pilot training per topic, indicating which pillar was in focus (✓ = in focus, ✗ = not in focus). RRP = responsible research practices, IPS = interpersonal skills.

Setting

The training was delivered in person on an external location (i.e., not part of the premises of the university or university medical center that employed the participants).

Approach

The training was designed to be interactive. We used a learning-by-doing approach (Reese 2011), meaning that each topic started with a brief introduction of the context and theory, intended to answer the “what is this” and “why does this matter” questions. Participants were instructed to think of a situation with their PhD candidate where they wanted to learn about the topic at issue. For example, when it regarded listening, participants were asked to think of a situation where they had struggled with listening to their PhD candidate. Participants then proceeded to practice with the topic in pairs or small groups, e.g., when it regarded listening, participants practiced in groups of three (one talker, one listener, one observer, multiple rounds with roles changing between participants). The trainer then brought the participants together for reflection and invited them to share their experiences. Where necessary, the trainer tried to deepen the reflection and encouraged other participants to provide suggestions (“Participant Y indicates she struggles to not fill in for her PhD candidate whilst listening, and her buddy also commented on this. What strategies do you use here?”). Each topic concluded with a few minutes for reflection where participants were invited to think about how this topic could support their PhD candidate. Participants then wrote down their personal take-aways in their notebooks (provided by the trainers at the beginning of the training) and shared these with the group.

Recruitment

We recruited PhD supervisors in three different academic institutions in three ways. First, we put an advertisement on the local intranet and in the newsletter of the Vrije Universiteit. Second, the graduate school of the Amsterdam University Medical Centers sent a targeted e-mail to the PhD supervisors in their database. Third, we approached policy advisors charged with professionalization of staff of the different research institutes of the University of Amsterdam.

Procedure

Before the pilot training (December 2018), we sent an informational letter to supervisors that had signed up. The letter explained that we would like to receive the e-mail addresses of their current PhD candidates to enquire the PhD candidates’ perceptions of the supervisor’s skills before and after the start of the study. We only approached PhD candidates that consented to have their e-mail address shared. We sent out the first Research Supervision Quality Evaluation (henceforth: ReSQuE)survey (see: <https://professionalperformance-amsterdam.com/academy/projecten/>) between December 2018 and January 2019 to both the supervisors (to evaluate themselves) and their PhD candidates (to evaluate their supervisor).

At the end of each training day, participants completed a pen-and-paper evaluation form where we inquired their overall perceptions of the topics that day. At the end of the full pilot training (September 2019), we conducted two focus group interviews with the two groups of participating supervisors. Three months after the pilot training (December 2019 – January 2020), we sent out the same survey to the supervisors and their PhD candidates. See [Figure 2](#) for our training and its evaluation.

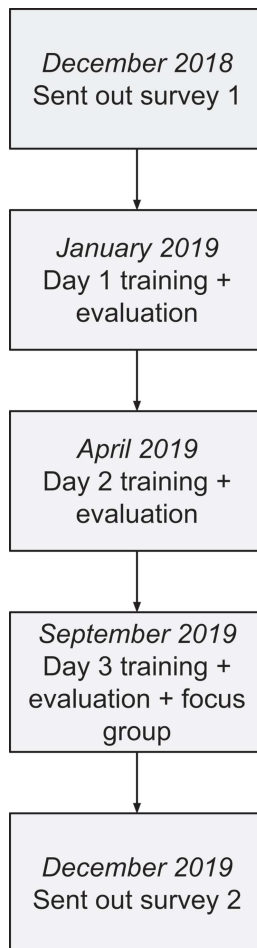


Figure 2. Study flow chart.

Measurement instruments

Survey

We asked the participating supervisors to evaluate their own supervision skills and asked their PhD candidates to rate their supervisor's supervision skills. For this we used a predecessor of the ReSQuE questionnaire. Participants rated their agreement with a particular statement on a Likert scale (1 = "fully disagree," 5 = "fully agree"), see Appendix 1.

Because the questionnaire has not been validated yet, we decided to indicate overall item changes (i.e., X number of items increased, Y number of items decreased) and to focus on those items that most directly enquired about the skills explored in our training. The research team discussed all items and collectively decided on two items per domain: a voice of reason for the PhD candidate and providing structured feedback (interpersonal skills), and on preregistration and the use of preprints (RRPs), respectively. This meant that we have focused on 4 items in total, displayed for both supervisors and PhD candidates. Since we did an explorative pilot evaluation with a small number of participants, we display descriptive statistics only. We focused on mean differences (MD) at the group level (i.e., group mean of participants before, group mean of participants after) and interpreted an increase in MD of at least .5 as a meaningful improvement.

Process evaluations

We handed out process evaluation forms at the end of each training day. Participants had to rate how much they learned in each session, whether the duration of each session was sufficient, and how relevant they considered the topics addressed (see Appendix 2). We also asked participants to rate the pilot training overall and enquired to what extent it met their expectations.

Training evaluation in focus groups

The focus group interview was based on a topic guide that involved three broad topics: 1) overall perception of training, 2) the combination of interpersonal skill development and research integrity, and 3) what their thoughts were on making the supervision training compulsory. We concluded with asking final suggestions for improvement and suggestions for sustainable implementation (see Appendix 3 for our focus group interview guide). After transcription, we applied inductive content analysis. Two authors independently coded the transcripts and resolved differences in interpretation through discussion.

Results

Demographics

Two groups of supervisors took part in our pilot training ($n = 21$). Participants came primarily from biomedicine (95%) while 5% of the supervisors worked in the social sciences. See Table 1 below.

Table 1. Demographics of participants (Table view)

Gender	<i>n</i>	Academic rank	<i>n</i>
Male	8	Postdoc	2
Female	13	Assistant professor	18
		Associate professor	1

Process evaluations

All participants completed the process evaluation forms that were handed out after each day (response rate: 100%). Overall, participants evaluated the pilot training with an 8.4 ($SD = .76$) on a scale from 1–10 (where 1 = “terrible” and 10 = “excellent”). When asked about the extent to which the training fulfilled their expectations, participants rated it 8.4 ($SD = .68$). Sessions that participants considered most relevant and where they learned most were the role play with actors and the coaching sessions, see Appendix 4.

Survey instrument

Supervisors’ skills before and after the pilot training

Before the pilot training, 20 (95%) supervisors and 23 PhD candidates participated in the survey. After the pilot training, 11 (55% – one supervisor quit the pilot as he moved to another institution) supervisors, and 11 PhD candidates participated in the survey. Both supervisors and PhD candidates seemed more positive about their interpersonal skills and the ability to foster responsible research practices after the training, see Table 2 where higher ratings indicate a greater agreement with the statement. Overall, PhD candidates rated their supervisors higher on 17 items, lower on 7 and the same on 1 item. Ratings for each item are shown in Appendix 5.

Table 2. Perceptions of supervisors and PhD candidates (Table view)

ReSQuE items	Before		After		MD ¹	
	M	SD	M	SD		
<i>Supervisor about their supervision</i>						
<i>Interpersonal skills</i>						
You act as a mentor, a confidante, an advisor, and a voice of reason for your PhD candidate	4.47	.697	4.75	.463	.28	

ReSQuE items	Before		After		<i>MD</i> ¹
	M	SD	M	SD	
You provide feedback in an open, clear, and structured communication, based on a discussion of both positive and negative elements of the research	3.74	.562	4.38	.518	.64
<i>Responsible research practices</i>					
You stimulate your PhD candidate to preregister their study	3.43	.938	4.00	.816	.57
You encourage your PhD candidate to use preprints	2.83	1.115	3.00	1.414	.17
<i>PhD candidate about the supervisor</i>					
<i>Interpersonal skills</i>					
Your supervisor acts as a mentor, a confidante, an advisor, and a voice of reason for you	4.48	.593	4.55	.688	.07
Your supervisor provides feedback in an open, clear, and structured communication, based on a discussion of both positive and negative elements of the research and the supervision	4.36	.581	4.55	.522	.19
<i>Responsible research practices</i>					
Your supervisor stimulates you to preregister your study	3.56	1.199	3.70	.949	.14
Your supervisor encourages you to use preprints	3.45	1.508	4.00	1.069	.55

¹ Mean difference, MDs printed in **bold** indicate an improvement of at least .5, which we considered “meaningful” improvement.

Items were scored on a 5-point Likert scale with 1 = “fully disagree,” 5 = “fully agree.”

The higher a score; the better participants rated themselves or the other at the behavior the item described.

Focus groups

Using inductive content analysis, we identified three themes that are closely connected to our research questions. We start with how participants described their experience participating in the training, then proceed to their views on the combined focus on RRP and interpersonal skills, and close with views on the issue of mandating participation in trainings like ours for PhD supervisors.

Eye opening, intense, peer-learning

Various participants described the pilot as eye-opening in terms of what kind of effects the interpersonal skills could have, and in terms of what they could improve in their own supervision, but they also described the pilot as intense. They appreciated the different tools that the course had to offer and the possibility to reflect on their own behavior and the effects their behavior may have on their PhD candidates. Some explained that the intensity was because the training involved a different type of learning compared to scholarly courses. Others indicated that the intensity could be related to the topics that dealt with emotions of themselves and their PhD candidates, that are rarely discussed in scientific meetings. Many participants appreciated the diversity of the other participants in terms of background, academic rank, and the different perspectives they could bring, indicating that they also learned a lot from each other.

“Considering most of what we learn in science, we learn auto-didactically, it is really nice that there is a course and I think because it is a pilot of course there are points of improvement but it is really, really helpful and [I] truly recommend it.” P. 3, female assistant professor

“... because you don’t get asked about these questions a lot of times, like, and all kinds of emotional questions as well, like the first morning I remember talking and listening to someone, so these exercises I never did that before so it was pretty intense for me, which is why it was draining but I learned a lot of things. And I will start

supervising candidates soon and it is really good to know from the beginning, like how you should behave and like find your way doing this during supervision.” P. 7, male postdoctoral researcher.

Synergy, but focus on making connections tangible

Overall, participants appreciated that they needed interpersonal skills to convey research integrity to their PhD candidates, and that it is in turn responsible to develop these skills. Hence, participants recognized the synergy between the interpersonal skill development and RRP, but also emphasized that the two could be integrated better by drawing links between the different topics. Many participants felt that interpersonal skills were overrepresented but expressed that they enjoyed the greater emphasis on interpersonal skill development and perceived this as a prerequisite for fruitful discussions about research integrity. They understood how possessing good interpersonal skills can help when discussing precarious topics related to research integrity.

“[I] did not really know what to expect of a supervision course and uhm, maybe my first idea was, oh, we learn about what it is to be a PhD candidate, we’ve all been a PhD candidate, but I liked it that we focused on the PhDs themselves, but also our role, and also science, so these three pillars, I think they come nicely together and they all interact so that for me was also, well, I knew it intuitively I think but, it is nice to, uhm, talk about that in the course.” P. 8, female assistant professor

“I agree that you can’t do only research integrity or only soft skills and I think it is nice that there is a bit more about soft skills and a bit less about the research integrity, the balance is nice, this balance.” P. 18, female assistant professor.

One participant thought this applied the other way around as well:

“I liked the combination with the open sciences practices because I think that can like alleviate a lot of stress for PhD candidates already and open up to good research practices in general and these things like preregistration may enhance your chance of getting published. So I think that is really important because it can take away a some of stress, like if you did this properly it will take away some of the stress that PhD candidates will have to deal with, which is going to take away some of the emotional shit that you will have to deal with as a supervisor, right? So I think that by combining these open sciences practices and good people skills you basically alleviate on one side the shit you have to deal with and you become better at the shit you have to deal with.” P. 4, male assistant professor.

Highly recommended for supervisors, but don’t make it compulsory

Various participants indicated that, although they enjoyed the pilot training, they feared that making the training compulsory would diminish its value. They were afraid that by making it compulsory, you could risk bringing in participants that are not able or willing to critically inspect their own behavior and with their counterproductive attitude would decrease the space for others to learn and reflect. Some played with the idea of making the training a prerequisite for departmental promotion, or something that is explicitly endorsed by the head of the department. Other participants raised that by making a course compulsory, it should also be possible to fail the course, otherwise it becomes a formality.

“Not compulsory, but I think there should be strong emphasis by the department head and others that you should follow this course too, and also, if you want to make promotion in the department, that you should have followed this course.” P. 11, female postdoctoral researcher.

“Mentoring, you can, you cannot pass or fail these things, that is the problem, it is a soft skill.” P. 5, male assistant professor

Discussion

We reported participating supervisors' perceptions regarding a novel training to supervise their PhD candidates toward research integrity. Overall, it seems that supervisors rate themselves higher on supervision skills after the training compared to before, which is mirrored by ratings from their PhD candidates. The perceptions of the PhD candidates suggest that the session where supervisors explored different tools for transparency such as preprints, have been put into practice. Perceptions of the supervisors suggest this as well but focus on preregistration. Supervisors' perceptions also suggest that the session on feedback provided them with useful tools, as is echoed in the focus groups and the daily evaluations. This suggests that supervisors have tried to apply the skills learned in our training in practice, highlighting the usefulness of our learning-by-doing approach. In addition, participants were positive about the training's dual focus on interpersonal skills and RRP, although some focus groups participants emphasized that the connections between RRP and interpersonal skills could be drawn more explicitly. Participants worried that making the training compulsory could undermine its effects but thought that explicit encouragement from university and departmental leadership could be fruitful in persuading more PhD supervisors to follow similar trainings.

Contextualization

Not all of the learning takes place within the one-on-one setting of supervision that we focused on. PhD candidates may also learn about RRP through their peers or senior staff in their research group. These environments or climates differ, and may foster particular ways of supervising (Roberts, Kavussanu, and Sprague 2001). In a climate that is focused on learning, supervisors may engage in a supportive style of supervising where the professional development of the supervisee is considered important. In a competitive climate where individual success is considered to be the ultimate goal, supervisors may focus on supervising a project with the view to obtain prestige or go after something that is good for their career (such as high-impact publications) but may be less inclined to invest in the scientific and interpersonal skills of the PhD candidate (Roberts, Kavussanu, and Sprague 2001). This points to the potential to supplement training for PhD supervisors with group mentoring like Whitbeck (Whitbeck 2001) described where PhD candidates and faculty together deliberate on dilemmas related to research integrity. Engaging in RRP reflection as a group could be helpful to foster a climate focused on learning where responsible supervision according to the practices we focused on becomes the norm.

Besides involving the research group, some authors have argued that RRP training should take place in the laboratory or research environment itself as this is the setting where researchers will run into challenges related to RRP (Peiffer, Laurienti, and Hugenschmidt 2008; Plemmons and Kalichman 2013). For example, Peiffer and colleagues developed a small discussion-based program on research integrity and related topics (Peiffer, Laurienti, and Hugenschmidt 2008). They deliberately chose to let these discussions take place in the laboratory to empower researchers across different career stages in developing a reflective stance and expanding their RRP related knowledge. More recent mentoring programs are also situated in the research environment itself (Kalichman 2014; Kalichman and Plemmons 2018; Plemmons and Kalichman 2018). Our training was conducted outside the lab which made it easy for the participants to be open and share their experiences. At the same time, we see value in continuing RRP-related discussion in the actual research environment (as we tried to mimic with peer discussions in between training days) and believe that both are necessary.

Interpretation

Our focus group participants remarked that the training was intense, in part because it also included discussing emotions, a topic that is rarely explicitly discussed during academic meetings. We believe that if

supervisors are better equipped to deal with their emotions, this may be helpful not just for the supervisory relationship, but also for other relationships in academia. Skills to work openly and effectively with others should be valued. It is promising in this regard that there are various movements to recognize and reward good supervision and mentoring, for example in the Hong Kong Principles for assessing researchers (Moher et al. 2020).

Focus group participants' hesitancy with regards to making the training program compulsory may seem puzzling, especially because participants were generally enthusiastic about the training. It makes one wonder whether our participants had specific motivations to take part in the pilot that distinguished them from other supervisors. Zooming in on participants' motivations for applying for the training indicates two broad tenets. Novice supervisors indicated they wanted to gain more confidence and to improve their supervisory skills, as they often had just started supervising. More senior faculty indicated that they wanted to learn more about giving constructive feedback, especially in cases where the PhD candidate's work was not progressing as expected. Future work should try to assess the generalizability of these motivations.

That said, leaving participation in training to improve supervision skill voluntarily means that some supervisors would never attend a training to improve RRP or interpersonal skills. Some universities have implemented compulsory research integrity training for PhD candidates and supervisors, for example the University of Copenhagen and Leuven University (LeRouge & Hol, 2020). In addition, various Nordic countries made interpersonal skills training compulsory for supervisors. As far as we are aware, these trainings often address research integrity and interpersonal skills as separate topics. The research integrity training is often focused on how the supervisor can do responsible research, not so much on how the supervisor can facilitate and support the PhD candidate to do responsible research with interpersonal skills (which was our focus).

It is our understanding that the effectiveness of compulsory training relies heavily on the goal of the training; if the training's goal is to increase knowledge or raise awareness, then a compulsory approach might be fruitful. However, if the training intends to increase (interpersonal) skills, then it is less clear whether mandating attendance would help as participants should be open to change their behavior, for which a certain degree of willingness to learn seems essential. An unwilling participant could interfere with group dynamics and could negatively influence the learning environment for the rest of the group. Future research should consider developing better guidance for institutions how to deal with unwilling supervisors, not wanting to improve their supervision skills despite signs of malfunctioning. Alternatives to making the training compulsory could include an ombudsperson for PhD candidates, the possibility and facilitation to change supervisors if mediating the relationship between PhD candidate and supervisor does not help, or finally stripping a PhD supervisor of the right to supervise after serious complaints.

Strengths and limitations

This is, to our knowledge, the first training program for PhD supervisors that combines interpersonal skills and RRP. The reflections from our pilot participants underscore that this combination has great potential. In addition, we included perceptions from the PhD candidates that our participants supervised in the evaluation, strengthening the validity of the findings.

The major limitation of our study is that it's a pilot as its design and small sample size prevent conclusions on the training's potential effectiveness. It should be noted that the purpose of our pilot study was to examine the feasibility of this new training method for implementation (Leon, Davis, and Kraemer 2011). We restricted our measures to perceptions, not behaviors in practice. The latter would be an interesting target for future studies on the effectiveness of supervision training.

Related, the questionnaire we used to measure the quality of supervision has not been validated yet. When reviewing the responses to the questionnaire, we focused on mean differences at the group level. We

chose to include all responses whereas focusing on only those supervisors and PhD candidates that filled in both the before- and after questionnaire could have been more valid but at the same time would decrease the numbers even more. All this warrants more large-scale evaluations of training programs intended to improve supervision skills using rigorous designs (see e.g., Plemmons and colleagues that conducted a randomized controlled trial to evaluate the effect of their ethics communication intervention (Plemmons et al. 2020)).

Finally, there was a small gender imbalance among our participants with 13 female and 8 male supervisors participating. However, national data on female researchers in university medical centers (Vogelezang and Vennekens 2021) also indicate more women in non-medical specialists' positions. Hence, the overrepresentation of women was expected.

Conclusion

We investigated participating supervisors' perceptions of a novel pilot training program to foster their supervision skills that combined RRP and interpersonal skill development. The results of a survey we sent out to participating supervisors as well as their PhD candidates seem promising; perceptions from both supervisors and PhD candidates seemed to indicate improvement in supervision skills. When asked about the program's set-up combining RRP and interpersonal skill development, participants were positive about the combination, but indicated that the connection between RRP and interpersonal skills could be drawn more explicitly. Participants would recommend the course to others but thought that making the training compulsory for all supervisors would be undesirable. The results highlight the potential of dedicated RRP training programs for supervisors. However, they should be interpreted with caution, as they regard a small sample of volunteering predominantly biomedical PhD supervisors, underscoring the need for larger scale programs that are fit for most disciplinary fields to foster responsible supervision that are evaluated using rigorous designs.

Acknowledgments

We are grateful for Marlies Stouthard who designed the initial ReSQuE survey and who helped recruit participating supervisors in her role as head of the graduate school.

Disclosure statement

The authors are co-developers of the Superb Supervision program and support the development of responsible research practices related training for researchers. LM currently offers this training commercially at several universities in The Netherlands.

Funding

During the development of the training and the pilot, LB, JT and TH were supported by the Templeton World Charity Foundation (<https://www.templetonworldcharity.org/>) under the grant #TWCF0163/AB106. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Alfredo, K., and H. Hart. 2011. "The University and the Responsible Conduct of Research: Who Is Responsible for What?" *Science and Engineering Ethics* 17 (3): 447–457. .
- All European Academies (2017). "*The European Code of Conduct for Research Integrity*."
- Anderson, M. S., A. S. Horn, K. R. Risbey, E. A. Ronning, R. De Vries, and B. C. Martinson. 2007. "What Do Mentoring and Training in the Responsible Conduct of Research Have to Do with Scientists' Misbehavior? Findings from a National Survey of NIH-Funded Scientists." *Academic Medicine* 82 (9 853–860).
- Antes, A. L., and J. M. DuBois. 2018. "Cultivating the Human Dimension in Research." *Molecular Cell* 72 (2): 207–210. .
- Antes, A. L., A. Kuykendall, and J. M. DuBois. 2019a. "Leading for Research Excellence and Integrity: A Qualitative Investigation of the Relationship-building Practices of Exemplary Principal Investigators." *Accountability in Research* 26 (3): 198–226. .

- Antes, A. L., A. Kuykendall, and J. M. DuBois. 2019b. "The Lab Management Practices of "Research Exemplars" that Foster Research Rigor and Regulatory Compliance: A Qualitative Study of Successful Principal Investigators." *PLoS One* 14 (4): e0214595. .
- Bird, S. 2001. "Mentors, Advisors and Supervisors: Their Role in Teaching Responsible Research Conduct." *Science and Engineering Ethics* 7 (4): 13. .
- Bouter, L. M. 2015. "Commentary: Perverse Incentives or Rotten Apples?" *Accountability in Research* 22 (3): 148–161. .
- Bouter, L. M., J. Tjldink, N. Axelsen, B. C. Martinson, and G. Ter Riet. 2016. "Ranking Major and Minor Research Misbehaviors: Results from a Survey among Participants of Four World Conferences on Research Integrity." *Research Integrity and Peer Review* 1 (1): 17. .
- Clynes, M., A. Corbett, and J. Overbaugh. 2019. "Why We Need Good Mentoring." *Nature Reviews Cancer* 19 (9): 489–493. .
- Davis, M. S., M. Riske-Morris, and S. R. Diaz. 2007. "Causal Factors Implicated in Research Misconduct: Evidence from ORI Case Files." *Science and Engineering Ethics* 13 (4): 395–414. .
- Eisen, A., and R. M. Berry. 2002. "The Absent Professor: Why We Don't Teach Research Ethics and What to Do about It." *The American Journal of Bioethics* 2 (4): 38–49. .
- Fanelli, D. 2009. "How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-analysis of Survey Data." *PLoS One* 4 (5): e5738. .
- Gray, P. W., and S. R. Jordan. 2012. "Supervisors and Academic Integrity: Supervisors as Exemplars and Mentors." *Journal of Academic Ethics* 10 (4): 299–311. .
- Haven, T., H. R. Pasman, G. Widdershoven, L. Bouter, and J. Tjldink. 2020. "Researchers' Perceptions of A Responsible Research Climate: A Multi Focus Group Study." *Science and Engineering Ethics* 26 (6): 3017–3036. .
- Haven, T. L., J. K. Tjldink, B. C. Martinson, and L. M. Bouter. 2019a. "Perceptions of Research Integrity Climate Differ between Academic Ranks and Disciplinary Fields: Results from a Survey among Academic Researchers in Amsterdam." *PLoS One* 14 (1): e0210599. .
- Haven, T. L., J. K. Tjldink, H. R. Pasman, G. Widdershoven, G. Ter Riet, and L. M. Bouter. 2019b. "Researchers' Perceptions of Research Misbehaviours: A Mixed Methods Study among Academic Researchers in Amsterdam." *Research Integrity and Peer Review* 4 (1): 25. .
- Kalichman, M. 2014. "A Modest Proposal to Move RCR Education Out of the Classroom and into Research." *Journal of Microbiology & Biology Education* 15 (2): 93–95. .
- Kalichman, M. W., and D. K. Plemmons. 2018. "Intervention to Promote Responsible Conduct of Research Mentoring." *Science and Engineering Ethics* 24 (2): 699–725. .
- Kornfeld, D. S. 2012. "Perspective: Research Misconduct: The Search for a Remedy." *Academic Medicine* 87 (7): 877–882. .
- Krishna, A., and S. M. Peter. 2018. "Questionable Research Practices in Student Final Theses – Prevalence, Attitudes, and the Role of the Supervisor's Perceived Attitudes." *PLoS One* 13 (8): e0203470. .
- Leon, A. C., L. L. Davis, and H. C. Kraemer. 2011. "The Role and Interpretation of Pilot Studies in Clinical Research." *Journal of Psychiatric Research* 45 (5): 626–629. .
- LeRouge, I., and A. Hol, On behalf of LERU. (2020). "Towards a Research Integrity Culture at Universities."
- Löfström, E., T. Trotman, M. Furnari, and K. Shephard. 2014. "Who Teaches Academic Integrity and How Do They Teach It?" *Higher Education* 69 (3): 435–448. .
- Moher, D., L. Bouter, S. Kleinert, P. Glasziou, M. H. Sham, V. Barbour, and U. Dirnagl. 2020. "The Hong Kong Principles for Assessing Researchers: Fostering Research Integrity." *Plos Biology* 18 (7): ARTN e3000737 e3000737. .
- Motta, M. M. 2002. "Mentoring the Mentors: The Yoda Factor in Promoting Scientific Integrity." *The American Journal of Bioethics* 2 (4): W17. .
- Muthanna, A., and A. Alduais. 2020. "A Thematic Review on Research Integrity and Research Supervision: Relationships, Crises and Critical Messages." *Journal of Academic Ethics* 19 (1): 95–113. .
- Peiffer, A. M., P. J. Laurienti, and C. E. Hugenschmidt. 2008. "Fostering a Culture of Responsible Lab Conduct." *Science* 322 (5905): 1186. .
- Plemmons, D. K., E. N. Baranski, K. Harp, D. D. Lo, C. K. Soderberg, T. M. Errington, B. A. Nosek, and K. M. Esterling. 2020. "A Randomized Trial of A Lab-embedded Discourse Intervention to Improve Research Ethics." *Proceedings of the National Academy of Sciences* 117 (3): 1389–1394. .
- Plemmons, D. K., and M. W. Kalichman. 2013. "Reported Goals of Instructors of Responsible Conduct of Research for Teaching of Skills." *Journal of Empirical Research on Human Research Ethics* 8 (2): 95–103. .
- Plemmons, D. K., and M. W. Kalichman. 2018. "Mentoring for Responsible Research: The Creation of a Curriculum for Faculty to Teach RCR in the Research Environment." *Science and Engineering Ethics* 24 (1): 207–226. .
- Reese, H. W. 2011. "The Learning-by-doing Principle." *Behavioral Development Bulletin* 17 (1): 20. .
- Resnik, D. B., and G. E. Dinse. 2012. "Do U.S. Research Institutions Meet or Exceed Federal Mandates for Instruction in Responsible Conduct of Research? A National Survey." *Academic Medicine* 87 (9): 1237–1242. .

- Ripley, E., M. Markowitz, A. Nichols-Casebolt, L. Williams, and F. Macrina. 2012. "Guiding the Next Generation of NIH Investigators in Responsible Conduct of Research: The Role of the Mentor." *Accountability in Research* 19 (4): 209–219. .
- Roberts, G. C., M. Kavussanu, and R. L. Sprague. 2001. "Mentoring and the Impact of the Research Climate." *Science and Engineering Ethics* 7 (4): 12. .
- Steneck, N. H. 2013. "Research Ethics. Global Research Integrity Training." *Science* 340 (6132): 552–553. .
- Titus, S. L., and J. M. Ballou. 2014. "Ensuring PhD Development of Responsible Conduct of Research Behaviors: Who's Responsible?" *Science and Engineering Ethics* 20 (1): 221–235. .
- Van Noorden, R. 2018. "Some hard numbers on science's leadership problems." *Nature* 557: 3 .
- Vogelezang, S., and A. Vennekens (2021). "Staff at University Medical Centers." *Rathenau*. Retrieved 4th April 2022 from <https://www.rathenau.nl/nl/wetenschap-cijfers/wetenschappers/personeel-aan-de-universiteiten-en-umcs/het-personeel-bij-de>
- Whitbeck, C. 2001. "Group Mentoring to Foster the Responsible Conduct of Research." *Science and Engineering Ethics* 7 (4): 17 .
- Woolderink, M., K. Putnik, H. van der Boom, and G. Klabbers. 2015. "The Voice of PhD Candidates and PhD Supervisors. A Qualitative Exploratory Study Amongst PhD Candidates and Supervisors to Evaluate the Relational Aspects of PhD Supervision in the Netherlands." *International Journal of Doctoral Studies* 10: 217–235. .
- Wright, D. E., S. L. Titus, and J. B. Cornelison. 2008. "Mentoring and Research Misconduct: An Analysis of Research Mentoring in Closed ORI Cases." *Science and Engineering Ethics* 14 (3): 323–336. .
- Xie, Y., K. Wang, and Y. Kong. 2021. "Prevalence of Research Misconduct and Questionable Research Practices: A Systematic Review and Meta-Analysis." *Science and Engineering Ethics* 27 (4): 41. .

Appendices

Appendix 1. ReSQuE questionnaires (version sent to supervisors and version sent to PhD candidates)

Appendix 2. Process evaluation questionnaires

Appendix 3. Topic guide focus groups

Appendix 4. Evaluations per course topic

Appendix 5. Ratings on all ReSQuE items