Routledge
Taylor & Francis Group

Ethics & Behavior

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hebh20

Learners' self-assessment as a measure to evaluate the effectiveness of research ethics and integrity training: can we rely on self-reports?

Anu Tammeleht & Erika Löfström

To cite this article: Anu Tammeleht & Erika Löfström (14 Oct 2023): Learners' self-assessment as a measure to evaluate the effectiveness of research ethics and integrity training: can we rely on self-reports?, Ethics & Behavior, DOI: 10.1080/10508422.2023.2266073

To link to this article: https://doi.org/10.1080/10508422.2023.2266073

Published online: 14 Oct 2023.

ETHICS & BEHAVIOR
https://doi.org/10.1080/10508422.2023.2266073

Routledge
Taylor & Francis Group

Learners' self-assessment as a measure to evaluate the effectiveness of research ethics and integrity training: can we rely on self-reports?

Anu Tammeleht[a,b] and Erika Löfström[a]

[a]Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland; [b]Centre for Ethics, University of Tartu, Tartu, Estonia

ABSTRACT

While the most prevalent means to measure the effectiveness of research ethics and integrity training formats is using learners' self-assessment, there is a need for reliable and feasible self-assessment tools to evaluate the level of understanding. The aim of the study was to design a reliable tool and test its accuracy in various training contexts. The current study utilized a design-based research (DBR) approach. Data were collected from 401 participants in training sessions and ten experts were involved in tool evaluation. The results indicate that in the context of research ethics and integrity learners can quite accurately (more than 80%) evaluate their level of understanding with the designed self-reflection form. The research revealed that participants who assessed themselves to have higher levels of understanding or who used the tool several times were more likely to provide well-elaborated descriptions of their learning that matched their indicated level of understanding.

KEYWORDS

Research ethics and integrity training; effectiveness; effectiveness; design-based research (DBR); self-report; accuracy of self-assessment

INTRODUCTION

In the recent decades, research ethics and integrity (REI) education has gained increasing attention because of a range of questionable research practices and the efforts to uphold the quality of research (Kalichman, 2013; Katsarov et al., 2022). Learning and training have been considered necessary to facilitate ethical behavior, thus we may ask what sort of training would support ethical behavior. There is a plethora of training courses and formats, and research is emerging about the effectiveness of the programs (cf. Katsarov et al., 2022; Stoesz & Yudintseva, 2018; Watts et al., 2017). Research has revealed that the most prevalent means to measure the effectiveness of REI training formats is using learners' self-assessment which is used twice the amount than any other means of evaluation (see Steele et al., 2016; Stoesz & Yudintseva, 2018; Turner et al., 2018).

We have chosen the term research ethics and integrity, REI, to combine both "research ethics" and "research integrity" as the tool we are developing should be applicable for various training formats that address both research ethics (guidelines, procedures and field-specific practices) and research integrity (behaviors guided by ethical principles and common values). In addition, combining or using either of the terms to signify both often happens in European context due to the lack of different terms for the concepts in several European languages, e.g. Finnish and Estonian.

Measuring the effectiveness of REI training is problematic as there seems to be no consensus of what constitutes "effective" in REI and it is difficult to find a feasible measure for this context (Steele et al., 2016; Turner et al., 2018). Previous studies have pointed out (e.g. a literature review by Watts et al., 2017) that it is difficult to measure the effectiveness of REI training as the aims of interventions

CONTACT Anu Tammeleht anu.tammeleht@helsinki.fi Faculty of Educational Sciences, University of Helsinki, Siltavuorenpenger 5 A, Helsinki 00014, Finland

as well as measurement means vary, so it is difficult to compare different training formats. Often the evaluation is based on participant reactions to the training, but also the improvement in knowledge and skills can be measured (as pre- and posttests) (Stoesz & Yudintseva, 2018). Occasionally, changes in attitudes, behavior and motivation follow the training; or reduced cases of misconduct and improvements in the research community can be observed (Stoesz & Yudintseva, 2018). Generally, the analysis of the effectiveness of REI training can predict the impact on the trainees, but there is rarely proof of the changes in the actual behavior (Watts et al., 2017). The current study focused on short-term measures of REI training effectiveness. Behavioral changes in learners require a longitudinal research approach that monitors individuals in their research practice.

As indicated by previous studies (Steele et al., 2016; Stoesz & Yudintseva, 2018; Turner et al., 2018), self-assessment or collecting participant reactions is used the most in evaluating training effectiveness. By self-assessment in this study we mean reaction responses (as self-evaluation/self-report/self-reflection) collected from trainees during or after the training. The format of collecting trainee reactions dates back to the late 1950s in the form of Kirkpatrick's training evaluation model (Stoesz & Yudintseva, 2018; Turner et al., 2018). The trainees' self-assessment is considered to be the most basic level of evaluation for the training, and usually reflects the perceptions about the content of the course as well as satisfaction with the training (Turner et al., 2018).

The current paper aims to tackle a gap in the evaluation of REI training programs – while self-assessment is the most popular measure with which to evaluate the effectiveness of the training, how "learning" as a measure of effectiveness of REI training be approached in a reliable and feasible way? With "feasible" we mean a measure that does not burden facilitators with excessive time demands, supports learner reflection while providing an accurate description of their learning, and is easily adaptable to a variety of trainings.

In line with the aims of design-based research, we produced a tool, in this case a self-assessment tool, which can be utilized in various REI training formats to evaluate the trainees' level of understanding as one measure of learning outcomes. The current paper outlines the background and development process of the tool and displays the initial findings of its accuracy.

SELF-ASSESSMENT AS A MEASURE TO EVALUATE COURSE EFFECTIVENESS

To evaluate what works in REI education various measures are used, the most common ones are self-assessment, posttest, pre- and posttests and combinations of the above (Steele et al., 2016; Stoesz & Yudintseva, 2018). Indeed, it is also possible to monitor researchers' behavior and the way they work, and count the cases of misconduct to make inferences of improved culture of integrity, but these are not very common (Stoesz & Yudintseva, 2018) as these measures can be difficult to implement.

The most used measure is the self-assessment reactions of trainees (Steele et al., 2016). Turner et al. (2018) outline that self-assessment most commonly asks about content satisfaction (i.e., how useful the participants found the content to be), and affective satisfaction (i.e., satisfaction with the course). In addition, self-assessment can also inquire about usability of the learned material, knowledge, motivation to apply the knowledge, confidence, attitude, and content relevance (Turner et al., 2018). Prior research shows that course satisfaction is not a strong predictor of training effectiveness as satisfaction may be influenced by the trainer, the environment and other factors contributing to momentary satisfaction, but the relationship to actual performance is limited (Turner et al., 2018). Satisfaction and learning should not be confused with each other. Learning often requires learners to step outside their comfort zones, something which can create discomfort and dissatisfaction. In measurements of satisfaction the learner is also often conceptualized as the passive recipient of the training, reflecting a dated notion of learning (Winstone et al., 2022). In the context of research ethics and integrity, active involvement in deliberation and reflection are crucial in developing ethical sensitivity (Mustajoki & Mustajoki, 2017), which is a prerequisite for ethical decision-making (Rissanen & Löfström, 2014). The strongest predictor of training success seems to be self-reported content relevance – this increases the motivation to go in detail with the topic and apply the knowledge in practice (Turner et al., 2018).

Turner et al. (2018) provide a set of recommendations for utilizing self-assessment to enhance REI training evaluation. Firstly, the content of the self-assessment questionnaire should be aligned with the goals and learning outcomes of the training. Second, combining closed and open questions, and collecting more than one piece of data are recommended. Thirdly, combining self-assessment with other kinds of data is beneficial, as it provides a more accurate picture of the learning process and outcomes.

There are several issues we noticed when considering implementation of self-assessment in the form of the tool designed as part of the DBR, and which we propose as solutions to previously identified gaps. Firstly, alignment with specific course outcomes is not possible as this way the tool would not be applicable to be implemented in different REI courses and comparisons between courses would not be possible either. Second, the self-reflection tool relies on adult learning theories (Knowles, 1975) and expects the user to take responsibility for their own learning and improvement. Thirdly, the self-assessment tool should not rely on specific course content. We used the SOLO taxonomy as an indicator of learning that can be captured with the tool, irrespective of the content of the course by focusing on the learners' perceived level of understanding. We describe this in more depth below.

The SOLO (*Structure of Observed Learning Outcomes*) taxonomy is hierarchical and based on research of learning outcomes in higher education settings (Biggs & Tang, 2007). Biggs and Tang (2007) argue that the structure of understanding becomes more complex by first quantifying details and then becoming qualitatively more complex. The SOLO taxonomy levels are related to the learning outcomes the learners display (Biggs & Tang, 2007) and it has been used to evaluate understanding in the REI context (Löfström, 2012; Tammeleht, 2022). See the descriptions of SOLO levels in Table 1.

We are aware of possible biases in self-reported data (like providing desirable responses, individual and cultural differences) as also pointed out by Ross (2006). Nevertheless, research suggests that participants in REI training can assess their level of understanding relatively accurately when compared to the facilitator's assessment (i.e. combination of various data on learning outcomes) (Tammeleht et al., 2022). In addition, recent study by Mol and van den Hoven (2022) outlines various studies regarding responsible conduct of research and conclude that many studies (including their own comparison of male and female students) are inconclusive pertaining gender differences in responsible conduct of research.

As research suggests, participants in REI training can assess their level of understanding relatively accurately when compared to the facilitator's assessment (Tammeleht et al., 2022). However, more research is needed in order to develop measurement based on self-report which is trustworthy and feasible and could be used in long-term perspective and follow-up settings. This tool would help learners monitor their learning process as well as provide important insights about the uptake of REI course content to facilitators.

For the current research, the following research questions were posed:

RQ 1: How could self-reported assessment of learning be developed as a reliable and feasible form of assessment of REI training?

RQ 2: How accurately are learners able to assess their learning in REI?

METHOD

As the research was focused on developing a tool for self-assessment as well as evaluation of the accuracy of users' responses, a DBR (design-based research) approach was used. DBR is a systematic research approach focused on improving existing practices in realistic contexts through design, development, iterations and implementation, and leading to contextually sensitive design principles and new theories (Anderson & Shattuck, 2012; Bakker, 2018; Barab, 2014). DBR involves an evolving iterative process that results in a product or process that addresses a perceived challenge. With DBR, during multiple iterations various methods can be used, including collaboration between researchers and practitioners, as well as formative evaluation by experts (Anderson & Shattuck, 2012; Edelson, 2002; Van den Akker et al., 2006). It is also noteworthy that the type or format of the "product"

Table 1. SOLO taxonomy explanation and adaptation for the REI training context (Tammeleht et al., 2019).

SOLO taxonomy level	General explanation (Biggs, 1999; Biggs & Tang, 2007)	Coding/interpretation description (Löfström, 2012; Tammeleht et al., 2019)
Extended ABSTRACT (4)	The coherent whole is generalised or reconceptualised to a higher level of abstraction - the response goes beyond conceptualising; higher level of abstraction is present with application to new and broader domains. Displaying the ability to theorise, generate, generalise, hypothesise, create, or reflect.	The learner goes beyond conceptualising the present ethical issue making steps towards relating the issues to applications beyond the present case. Displaying the ability to theorise, generate, generalise, hypothesise, create, or reflect.
Relational (3)	Relevant aspects are integrated into an overall coherent structure - ability to address the point and provide explanations, give details, and connect to the whole, giving relevant examples.	The learner displays an ability to address the most relevant ethical issues and provide explanations pointing out interrelations and providing examples demonstrating own reasoning.
Multi-structural (2)	Several aspects have been understood, but not relating them to one another - ability to enumerate, describe, illustrate, sequence, select, combine, follow procedures, but struggle to make the connections between them or draw conclusions based on interrelations.	The learner demonstrates that concepts had been understood appropriately but struggles to make connections between them and to draw conclusions based on interrelations. Knowledge-telling approach, no structuring.
Unistructural (1)	Knowledge enables learner to identify, recognise, count, find, label, match, name, perform simple procedures. One relevant aspect is understood - dealing with terminology, meeting part of the task, defining concepts while some aspects still missing.	The learner identifies one relevant aspect displaying some familiarity with relevant concepts but failing to address some more pertinent dimensions of the case.
Pre-structural (0)	The issue is not approached in a meaningful way - repeating the words in the question/case/code; no understanding.	The learner fails to identify a relevant ethical perspective or does not approach it in a meaningful way, repeating the words in the case without displaying evidence of own processing.

Figure 1. Iterations of the DBR process.

changes across the iterations, often starting with less defined tools, which are gradually defined and refined. In this process, the initial tool may change substantially. Four iterations were conducted during the current research (Figure 1).

Each iteration used different samples and methods depending on the goal of the iteration (Table 2).

The training formats that were included during different stages of DBR are described in Table 3.

Iteration 1

When designing an intervention to help bachelor's and master's students, and doctoral candidates in higher education to acquire and revise foundational knowledge about REI, we collected data about the learning process based on student work and audio recordings (see Tammeleht et al., 2019). In addition, we collected learner feedback (on paper) to evaluate how learners perceived the training and how the training could be improved, and to distinguish satisfaction with the course from the evaluation of the learning process. The feedback drew content from previous examples available to the authors and was also negotiated with practicing REI teachers (in spring 2018). The feedback form (see Appendix I) inquired about the level of difficulty of the tasks (there was one case per group followed by group discussion of questions and presentation

Table 2. Methods and foci of iterations during DBR.

Focus	Iteration 1 (N = 51)	Iteration 2 (N = 306)	Iteration 3 (N = 44)	Iteration 4 (N = 10)
Literature review; practitioners' input	Using an existing format of self-assessment			
Tool development, aligned with the learning outcomes		Design and testing of a new tool for self-assessment		
Validating the tool			Implementing the self-assessment tool in new contexts	Expert evaluation Improving the form with expert input
Accuracy of self-assessment	not measured	measured	measured	Expert feedback given to tool developers

Table 3. Outline of training descriptions, target groups and learning outcomes.

Iteration	Training	Method	Target group	Learning outcomes
1	Foundations of RE/RI	case-based collaborative learning	bachelor/master's/doctoral students	• raising awareness of ethical issues during the research process; • learning together as a team, collaborating with peers. • practicing utilising the codes of conduct, being familiar with central topics;
2	Foundations of RE/RI	case-based, computer-supported collaborative learning	master's level students	• raising awareness of ethical issues during the research process; • practicing utilising the codes of conduct, being familiar with central topics; • learning together as a team, collaborating with peers.
3	1-day REI training session	lecture, value game; ethical analysis of cases	doctoral students	• understanding one's role as a researcher; • contemplating the values in research; • learning the ethical analysis; • applying the ethical analysis on one's own research
3	RE course (1 semester)	online independent learning combined with seminars	doctoral students	• familiarising with the central topics in RE; • guidelines and laws; • discussion sessions on various cases
3	Microcredential programme for REI leadership (1 semester)	online independent learning combined with seminars	mixed group (students and professionals, e.g. museum researcher, research coordinator, academics)	• preparing experts to solve various REI issue in their field and institution; • preparing to guide others

of group results), the clarity of material, the role of the group and the impact of the training (in multiple-choice questions). It also included two open-ended questions to collect feedback on what the learners found surprising and to give them an opportunity to comment on any other aspects of the course or their own learning. After analysis of Iteration 1 feedback form results

we saw the need for a more comprehensive self-report tool and discarded the then-current version of the form. Nevertheless, it provided interesting insights (see Iteration I results below) to the next iterations.

Iteration 2

Following an analysis of an REI intervention (meant for master's level students but also open for everyone as an open online educational resource), results based on group reports and group discussion recordings indicated that learners gradually improve in their understanding, based a taxonomy of learning, which in our case was the *Structure of Observed Learning Outcomes* (the SOLO taxonomy: Biggs, 1999) (Tammeleht, 2022; Tammeleht et al., 2019, 2022). We monitored the level of group understanding that was achieved but lacked information about the perceived level of individual learners and whether the achieved and perceived levels of understanding matched. To compare the achieved and perceived levels, we used the SOLO taxonomy to design a self-reflection form (SRF). SOLO was also used to evaluate the learning process by the facilitators of the course. We generated an online questionnaire in Google Forms in which learners could express what their level of understanding was (see Figure A1 in Appendix II). The learners first had to pick a statement describing the SOLO level (Table 1) that they felt best matched their own perceived level of understanding, and briefly explain what they were learning, how they thought they were progressing and where they would like to aim for next.

Iteration 3

Iteration 3 aimed to implement the SRF in different REI training formats and to do so repeatedly to get a sense of the capacity of the measurement to capture learning as a process. The first trial took place in autumn 2022 when doctoral candidates who participated in a 1-day REI training session were asked to use the SRF. The form was also included (online learning environment) in two 1-semester long courses on research ethics and integrity. In the SRF the participants were asked twice during their training day to indicate which activity or topic they wished to reflect upon. They then picked a SOLO level statement that they felt matched their level of understanding, and finally provided a short explanation of why they thought this was their level of understanding. The SRF is general enough to be applicable in various REI training formats (during longer training sessions it can be used repeatedly and for each topic, for shorter training sessions it can be used once or twice).

Iteration 4

For the final iteration, previously collected responses were analyzed and an alternative version was created (see content of the form in Appendix III), which asks learners first to describe their understanding and then pick the SOLO level. An REI training expert group was invited to provide feedback and evaluate the potential implementation suggestions for the tool.

DATA COLLECTION AND PARTICIPANTS

Data collection spanned from spring 2018 (for Iteration 1) to 2023 (for Iteration 4). Initially data were accumulated from one university in Estonia, but once the online versions (in several languages) were ready, international responses were collected. For Iteration 1 feedback forms, informed consent was sought from training participants (see more in Tammeleht et al., 2019). In total, 51 paper-based feedback forms were received.

For Iteration 2 an online form was created in Google Forms (platform name reported because we used two platforms, and these are associated with different technical options), which provided

information about data collection and asked learners to indicate whether they would allow their filled-in forms to be analyzed anonymously for research purposes. The SRF was attached to an open online REI resource intended for master's level students. Still, as the resource is open access and was disseminated within Estonia as well as internationally, it is not possible to know exactly who used the resource, whether they did it alone or in a group, or whether they did it only once or repeatedly (some descriptions indicated that some learners had worked alone, some in groups, some indicated doing the training several times). Out of 380 people who had filled in the online SRF, 342 had provided consent for research and after cleaning up the data 306 responses qualified for analysis. Responses with missing data were omitted. Sometimes the respondents had not given a description of their understanding related to research ethics, in which case the response was also omitted. No personal data were collected. The data were accumulated from December 2020 to spring 2022.

As part of Iteration 3, the SRF was implemented in various training formats. The first implementation took place during a 1-day REI training session for doctoral candidates in 2022. They were asked to fill in the SRF twice: after a more formal lecture-type of session and one practical activity (value game) ($N = 24$), and then after a fully practical session (using ethical analysis with cases) ($N = 21$). The form contained the same questions as the one used in Iteration 2 ("evaluate your level of understanding and describe what you are learning, how is it going and what should be done next to improve your understanding"). No personal data were collected, filling in the form was voluntary and had no impact on successful completion of the training. Consent for using data was collected retrospectively as this data collection was initially intended for internal development but was later included in the reporting of the DBR. All participants were contacted about using the forms for research purposes, and we received no objections at this point. Had we received objections, we would not have reported the data for Iteration 3 at all, as we would not have been able to determine which forms were from objecting participants.

The SRF was also included in two research ethics courses (both 1-semester long courses, one was Research Ethics (RE) for doctoral candidates, the other as part of a micro-credential program focusing on research integrity (RI) for researchers) to be filled in at any stage of the course (chosen by the learner). The link and description of the SRF was added into an online learning environment for both courses, filling in the form was optional and voluntary. Throughout the semester, 20 responses were submitted (both courses combined, micro-credential program learners were more active in using the form, and more than three quarters of the responses came from them). The form included the information and a consent section. All respondents consented to the researchers using their data. No personal data were collected.

During Iteration 4, ten REI experts from six countries were invited to participate in group discussions to provide feedback and to evaluate the usability of the tool. Two sessions were conducted from which feedback was not audio-recorded. The experts' ideas were written down by the authors. Participation was voluntary and no personal data were collected.

We deliberately did not collect demographic data and are aware that certain individual factors may affect responses of self-reflection (e.g., age, discipline, gender) (e.g. Ross, 2006). The point of departure for the development work has been that research ethics and research integrity competencies and ethical sensitivity can be developed (Bebeau et al., 1985; Jordan, 2013; Rest, 1986), and all learners (irrespective of their background) should uphold the highest quality while engaging in research work. In the recommendations we provide ideas on how to alleviate possible accuracy issues.

RESEARCH ETHICS

The research followed the European Code of Conduct for Research Integrity (ALLEA, 2017), the Estonian National Code of Conduct (Teadustava, 2017) and Finnish National Board on Research Integrity guidelines (2009 and 2023). The participants were asked for their informed consent. Participation was voluntary, and the decision not to participate in the research was not to influence the participants' status during their studies. Participants had the right to withdraw their consent at any

time. The data were anonymized before analyses (in case any identifiable details were provided by respondents in the open questions). According to the above mentioned national guidelines, a study such as the one presented here does not require an ethics review. Ethics review is required when a study involves an intervention in the physical integrity of research participants; deviates from the principle of informed consent; involves underage participants being studied without parental consent; exposes participants to exceptionally strong stimuli; causes long-term mental harm beyond the risks encountered in normal life; or signifies a security risk to subjects (Finnish Advisory Board on Research Ethics, 2009). None of these conditions were met in this study.

RESULTS

Iteration 1

Feedback forms included multiple-choice and open-ended questions. The form was filled in by bachelor's, master's and doctoral candidates after participating in the same training format. Reaction categories, namely utility of training, attitude about the training, motivation to apply knowledge, knowledge gained, content relevance, and course and content satisfaction (based on Turner et al., 2018) were used. Data from the multiple-choice answers was counted and the answers to open-ended questions were grouped and generalized.

While most learners considered the level of the training to be average, the material was evaluated as being unclear by some learners, mostly doctoral candidates (Table 4). Working in a group was considered to be an asset, and almost all learners believed that it would be easier to notice ethical issues in the future. Most participants said that the training fulfilled their expectations.

The first open-ended question inquired about unexpected or surprising aspects of the training and its content. The reaction criterion was "knowledge gained" (Turner et al., 2018). The most prevalent response was the surprise that there were so many aspects to think about in REI that the

Table 4. Respondents and responses to multiple-choice questions.

Topic	Reaction category	Bachelor students (*N*=27)	Master's students (*N*=16)	Doctoral candidates (*N*=8)	Total (*N*=51)
The level of tasks	Utility	Easy: 4 (15%) Complicated: 3 (11%) Average: 20 (74%)	Easy: 5 (31%) Complicated: 1 (6%) Average: 10 (63%)	Easy: 0 Complicated: 1 (12%) Average: 7 (88%)	Easy: 9 (18%) Complicated: 5 (9%) Average: 37 (73%)
Clarity of material generally	Utility	Unclear: 9 (partially) (33%) Too simple: 0 OK: 18 (67%)	Unclear: 1 (6%%) Too simple: 2 (12%) OK: 13 (82%)	Unclear: 6 (inventing the context) (75%) Too simple: 0 OK: 2 (25%)	Unclear: 16 (31%) Too simple: 2 (4%) OK: 33 (65%)
Role of the group	Attitude	Helped: 26 (96%) Confused: 1 (4%) Neutral: 0	Helped: 16 (100%) Confused: 0 Neutral: 0	Helped: 6 (75%) Confused: 2 (25%) Neutral: 0	Helped: 48 (94%) Confused: 3 (6%) Neutral; 0
Is it easier to notice ethical issues now/apply the knowledge?	Motivation to apply knowledge; knowledge gained; content relevance	Yes: 23 (85%) No: 0 Already knew: 1 (4%) Maybe: 3 (11%)	Yes: 14 (88%) No: 0 Already knew: 1 (6%) Maybe: 1 (6%)	Yes: 8 (100%) No: 0 Already knew: 0 Maybe: 0	Yes: 45 (88%) No: 0 Already knew: 2 (4%) Maybe: 4 (8%)
Fulfilled expectations	Content/course satisfaction	Yes: 21 (78%) No expectations: 2 (7%) Was better: 4 (15%)	Yes: 12 (76%) No expectations: 2 (12%) Was better: 2 (12%)	Yes: 7 (88%) No expectations: 1 (12%) Was better: 0	Yes: 40 (78%) No expectations: 5 (10%) Was better: 6 (12%)

participants had not known or had not thought about before. For instance, the role of the university, the availability of guidelines and rules, data management, ethics review process, and authorship issues were such aspects, which had presented the participants with new insights. Doctoral candidates were surprised about the positive impact of group work and the variety of opinions within the group.

In open comments learners pointed out that working in a group had helped with understanding and various points of views had also contributed to a better result. Based on the responses, the reaction criterion could be interpreted as "course satisfaction" (Turner et al., 2018). The level of the tasks was neither too difficult nor easy, doctoral candidates pointed out that inventing the context had made the task complicated at first. Learners also pointed out that the workshop format and case-based learning had helped them implement the national code of conduct and other regulations, hence put theory to practice, and that had improved understanding. The case-based approach was considered to be better than the lecture format.

Iteration 2

Analysis of online SRF was based on respondents' marked SOLO level: the learners had to pick a statement that was closest to their understanding. In addition, the SOLO level was deduced (by authors) from the respondents' descriptions on their learning (their described level). The authors used deductive content analysis where the responses were compared to the SOLO criteria (outlined in Table 1). The first author analyzed all the responses. The second author analyzed 37% of the responses. The agreement rate between the two raters was 89%, which suggests that there was a high degree of agreement. A detailed description of deductive content analysis based on the SOLO taxonomy and utilized in the REI context can be found in Tammeleht et al. (2019). An example with data analysis criteria is also brought out in Table A1 in Appendix IV.

Most respondents had marked their level of understanding as being at the unistructural, multi-structural or relational level (Figure 2). Descriptions indicated mostly a multi-structural level. There was a substantial number of learners who had underestimated their level of understanding as they had marked their level at the pre-structural or unistructural level while their descriptions actually indicated higher levels of understanding. Overestimating one's understanding was not as common (difference between marked and described levels for relational and extended abstraction levels was relatively small). Learners indicating higher levels of understanding appeared to be quite accurate in self-assessment.

To measure accuracy of marked and described levels, a match between the results was sought. For this analysis the respondents' marked and described responses were compared. The match between the marked and described SOLO levels was coded as following:

(1) Match – the marked and described levels matched exactly
(2) Close match (up) – there was a one level difference between the marked and described levels, the described level was higher than the marked level (e.g., the participant had marked SOLO level "unistructural" while the description was on the "multi-structural" level).
(3) Close match (down) – there was a one level difference between the marked and described levels, the described level was lower than the marked level.
(4) Mismatch (up) – there was at least a two-level difference between the marked and described levels, the described level was higher than the marked level.
(5) Mismatch (down) – there was at least a two-level difference between the marked and described levels, the described level was lower than the marked level.

Mismatch, especially overestimating one's level of understanding was not very common (13% in total) (Figure 3). It was more common to underestimate one's proficiency – 29% of participants

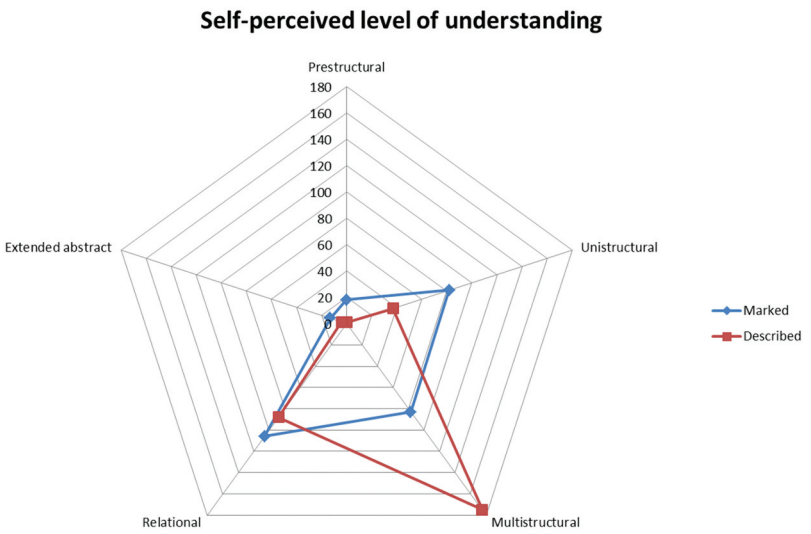

Figure 2. Self-perceived level of understanding during Iteration 2 (numbers indicate the number of respondents).

thought their level was lower by one level than what they displayed, and for 8% of respondents they had underestimated their level by 2 steps in comparison to the facilitator's assessment. Overall, 87% of respondents could quite accurately evaluate their level of understanding and support it descriptively.

Iteration 3

In Iteration 3 the SRF was first implemented during a 1-day REI training session for doctoral candidates in the humanities. The participants were asked to fill in the form twice during the day; after a more formal lecture-type of session and one practical activity, namely a value-game, and after a fully practical session using ethical analysis with cases. The form still asked the same questions as the one used during Iteration 2, i.e., "evaluate your level of understanding on the SOLO scale and describe what you are learning, how it is going and what to do next to improve your understanding."

Not all participants took the effort to describe their learning. Some descriptions were missing and sometimes the answer provided could not be analyzed because of too little information being provided. We removed such cases from the analysis as we could not interpret these reliably. From reflection I, 24 responses were collected, five descriptions were absent, thus 19 responses were included in the analysis. From reflection II, 22 responses were collected, three descriptions were absent, leaving 19 responses in the analysis.

Data from the first round indicates that learners tended to overestimate their level of understanding (Figure 4). While the marked level was mostly relational, the descriptions mostly displayed multistructural levels.

Matches between marked and described SOLO levels during reflection round I were relatively frequent (37% of responses), however, equally frequent were close mismatches down in which the described level was lower than the marked level (37%) (Figure 5). Mismatch by two levels was present in 15% of the responses.

Still, the match between the marked level and the described level was achieved by 84% of the responses including the complete and close matches. 11% of responses indicated a serious overestimation of the self-assessed level.

The second reflection round indicates a more realistic evaluation of the learners' understanding if we take a match between self-assessed and self-described understanding to be an indication of

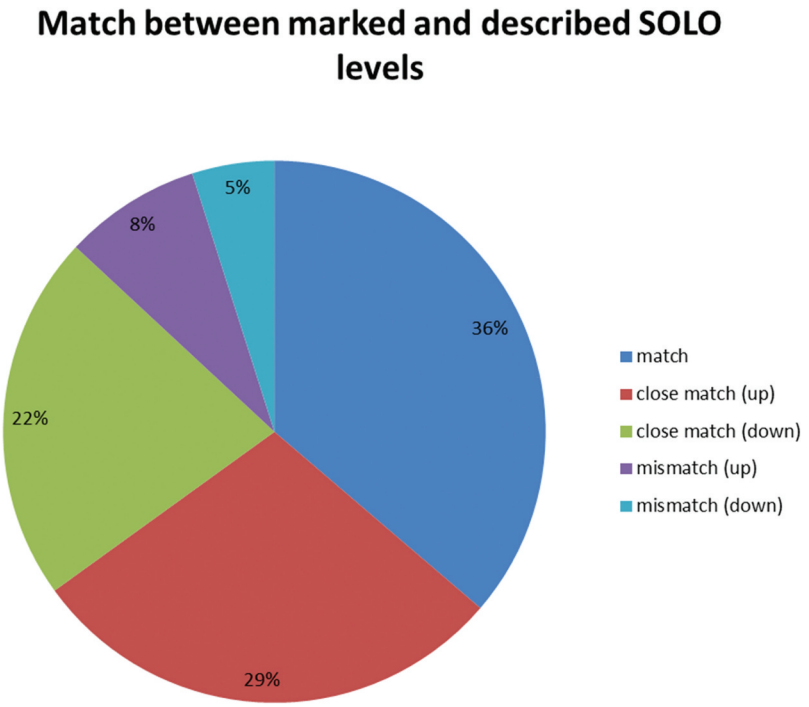

Figure 3. Match between marked and described SOLO levels during Iteration 2.

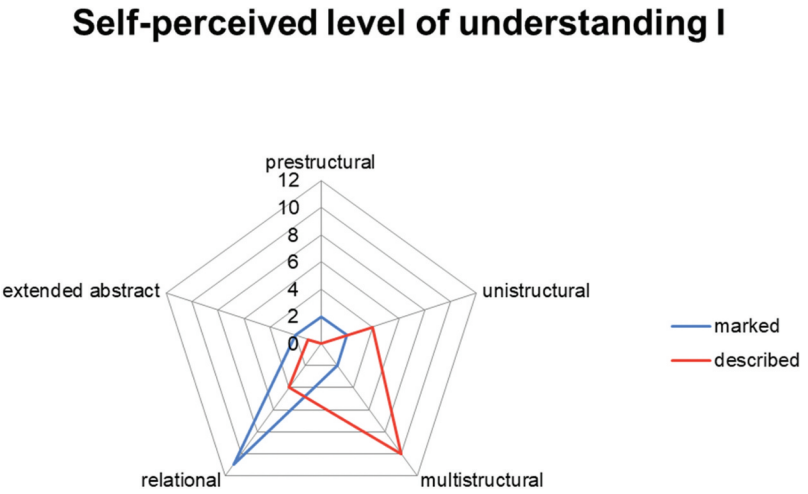

Figure 4. Self-perceived level of understanding during reflection round I of Iteration 3.

accuracy. For the pre-structural, unistructural and multi-structural levels, the evaluation was almost entirely unison (Figure 6). For the relational level, some respondents overestimated their understanding (or did not describe their learning in sufficient detail to be properly analyzed). Nevertheless, some respondents marked their level being at the extended abstraction level, but only one respondent displayed this level in the description part.

In addition, the comparison of match between marked and described SOLO levels showed improvement (Figure 7): 47% of marked and described levels matched completely, 16%

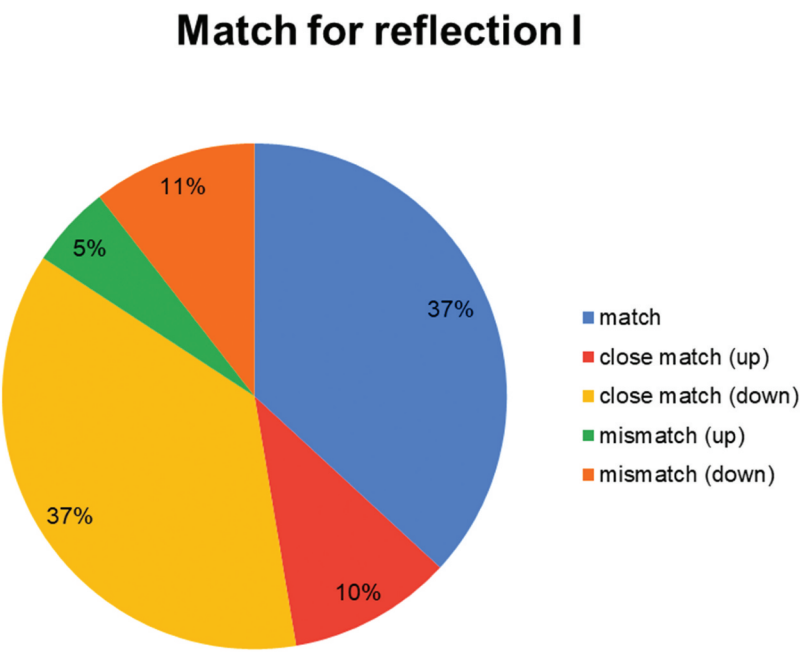

Figure 5. Match between marked and described SOLO levels during reflection round I of Iteration 3.

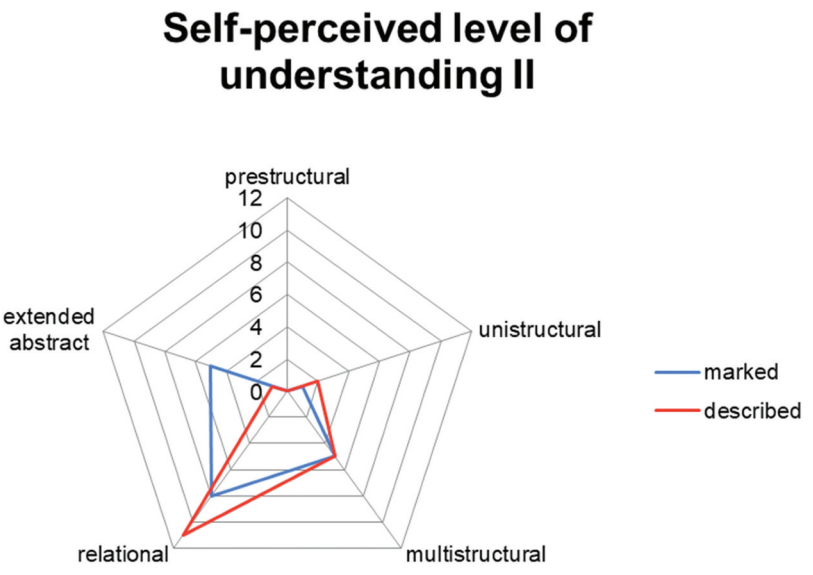

Figure 6. Self-perceived level of understanding during reflection round II of Iteration 3.

underestimated their level of understanding and a quarter slightly overestimated their level of understanding. Mismatch by two levels was present in 11% of the responses. Overall, 89% of respondents had a complete or close match in their responses.

For Iteration 3, the SRF was also included in an online learning environment for two courses in spring 2023, namely a RE course for doctoral candidates and a RI micro-credential program ($N = 20$). Participants in the micro-credential program were probably more motivated to reflect on their learning, so most (17) answers were provided by them. They also took

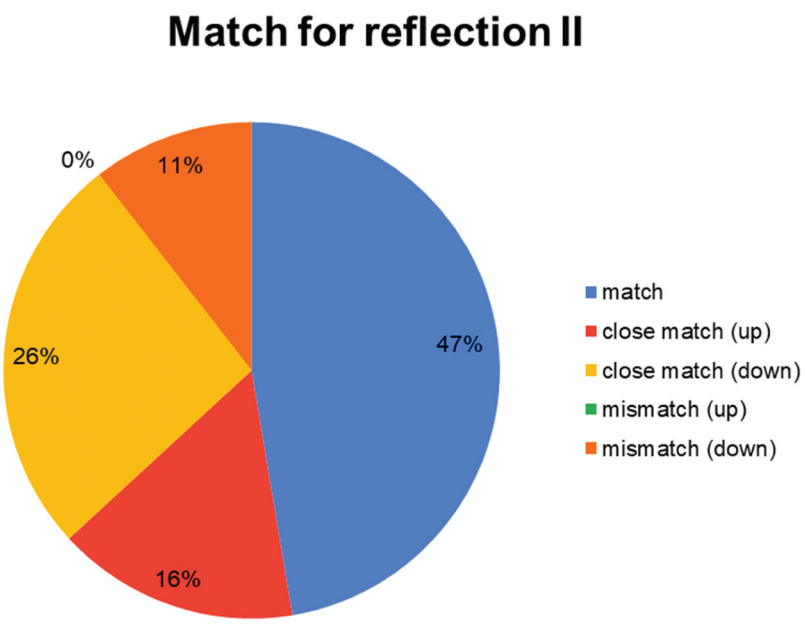

Figure 7. Match between marked and described SOLO levels during reflection round II of Iteration 3.

initiative to reflect on their learning several times during the semester. They occasionally provided specific descriptions of their learning, but they also used the form to reflect on their emotions and to provide feedback on the learning process. Learners usually evaluated their level of understanding to be on the relational level (50% of responses), 40% of responses indicated the multi-structural level and 25% extended abstraction. Still, 50% of descriptions were a reflection on their individual progress and feedback to the course. The other 50% described understanding on the topic mentioned and here the descriptions matched completely the indicated level.

Even though doctoral candidates participating in the RE course did not submit many answers ($N = 3$), their responses indicated that they had used the form purposefully. Namely, they chose a very specific topic they had just covered during their course, they evaluated their level of understanding and then provided a relatively detailed description of their learning process. Two participants indicated their level to be on the extended abstraction level and one description matched this level. One participant was evaluated as being on the unistructural level while the description indicated a relational level.

Iteration 4

For the fourth iteration, the lessons learned from the previous iterations resulted in a new and improved version of the SRF (Microsoft Forms was used this time). The SOLO levels were modified to help produce more accurate self-evaluations. The description section was also modified to provide support and encourage more details to be included. During Iteration 4, ten experts were involved in the tool development. The tool and initial results were introduced to them. In the next two months, two meetings were organized for mutual discussions and feedback. During the first meeting the experts were provided with a small training intervention in which they discussed an ethics case following the value-game format and were then asked to fill in the SRF. Then a discussion followed, and improvements were proposed. The main insights from the session with the experts were that the

form is probably more suitable for short training sessions or specific activities; and that instant feedback for the learners on how to develop one's understanding would be helpful.

For the second meeting, the SRF was improved further, and the form also provided pre-written level-based feedback that would be displayed to respondents after selecting the SOLO level. This way the feedback was not dependent on an analysis of the description provided by the participant but pertained to the indicated SOLO level more generally. Each SOLO level piece of feedback provided a short description of what the respondent (possibly) knows and how to improve their understanding further (how to reach the next SOLO level). This time the experts only received the SRF link without the ethics case, and they tested filling in the form. The experts pointed out that the SOLO level descriptions may be too complex for bachelor's and master's level students, but would be suitable for doctoral candidates and researchers, and suggested that the wording of the feedback could be tailored according to the target group. The experts also suggested the use of visualizations or illustrations to indicate the SOLO levels.

Expert feedback resulted in a new version of the form with question wordings specified and visuals were considered. Unfortunately, it was not possible to add visuals due to technical limitations of the platform used. It was also discussed whether there should be alternative forms for learners of different levels, and this we have taken as a future development stage.

DISCUSSION

Collecting learner reactions with a feedback form is common and often follows Kirkpatrick's training evaluation model (Stoesz & Yudintseva, 2018; Turner et al., 2018), whereby the usability of content and satisfaction with the training is measured. This design-based research aimed to identify how reliable self-assessment is as a measure of training effectiveness, and to develop a tool for this purpose. Iteration 1 focused on feedback forms and indeed the results indicated that the training content and format was considered adequate. Based on the high scores of the reaction criteria of content relevance, satisfaction with content as well as course, motivation to apply knowledge, and knowledge gained (see Turner et al., 2018), it can be assumed that the training format is likely to improve future performance. Turner et al. (2018) have indicated that if the learner finds the content of the course to be useful and helpful, there is more motivation to apply the knowledge and skills acquired. Turner et al. (2018) also found a strong correlation between content relevance, course satisfaction and performance. The responses in our study seemed to be quite similar, i.e., content was perceived as being good, learning took place, and the group was considered an asset. Still, the feedback form did not provide information about the alignment with the learning outcomes and the levels of understanding, so it was discarded.

A new self-reflection tool was designed based on the SOLO taxonomy. There was a need for a simple online form for helping learners reflect on their understanding as a form of self-assessment, and to provide facilitators with a chance to compare the learners' achieved and perceived levels of understanding (see more about the development of the SRF in Tammeleht et al., 2022). Initial findings, based only on the SOLO level chosen by the learners, indicated an 83% accuracy of the achieved and perceived levels of understanding by learners (Tammeleht et al., 2022). Nevertheless, this conclusion can only be drawn if there is more than one piece of evidence of learning. In this case there were e-portfolios that were evaluated by the facilitator and SOLO level chosen by the learners. To implement the self-reflection format in different contexts, analysis of SOLO levels and descriptions provided by learners was necessary.

It should be noted that validity (i.e., agreement between learner and facilitator evaluation) of self-assessment usually receives controversial findings (cf. González-Betancor et al., 2019; Ross, 2006), namely, there are no conclusive results of accuracy of self-assessment. Still, there are indicators that can make self-assessment more accurate. Iteration 2 results indicated that higher levels of understanding are usually quite accurately evaluated (see Figure 2). The learners who indicated relational or extended abstract levels were also mostly able to support this assertion in their descriptions. We can conclude that self-report becomes more reliable with more advanced

level learners or with learners who have gained some experience/expertise in research ethic and integrity topics. This interpretation was also supported by González-Betancor et al. (2019). The slight mismatch between unistructural and multi-structural levels may be caused by the limited input to descriptions – the learner may have understood the content of the training at the multi-structural level (and they also chose that SOLO level), but if the description was superficial and there was no other evidence of the achieved level, it was difficult to conclude if that level was actually achieved. The limited description may also be the result of refusal or inability to reflect on one's learning (e.g., Nevgi & Löfström, 2015).

In Iteration 2 we identified a close match between the chosen and described level in most of the assessments. In cases of mismatch, learners tended to underestimate their level of understanding. Some studies (González-Betancor et al., 2019; Ross, 2006) indicate that learners tend to overestimate their knowledge or understanding. Still, the Dunning-Kruger effect (Dunning, 2011) indicates that beginners and people who lack expertise tend to overestimate their performance, at the same time when people have already gained some expertise, they start underestimating themselves, possibly because of a broadening understanding of the extent of the "territory" of knowledge yet unknown to them. In the current study we conclude that learners had most likely participated actively in the training and had already gained some expertise.

During Iteration 3, the SRF was tested in several contexts: once during a longer training day when it was used twice, and during longer courses in which learners could choose when and how to use it. Repeated use of the SRF showed that when filling in the form for the first time, learners tended to overestimate their level of understanding. Their marked level of understanding was mostly relational, while their descriptions mostly displayed multi-structural levels. This may have been a result of participants not putting much effort into the descriptions or the lack of experience in the topic (thus the Dunning-Kruger effect). The second round indicated a more accurate evaluation of understanding. Not only did the learners try to evaluate their understanding (more even distribution of responses between unistructural, multi-structural, relational and extended abstraction levels), but the descriptions also displayed the level that they had indicated on the SOLO scale. Indeed, as also indicated by Ross (2006), reflecting over one's learning takes practice and should be taught to learners. The context of research ethics and integrity requires learner to become reflective practitioners as making ethically sound choices and developing one's ethical sensitivity takes conscious deliberation (Musatjoki & Mustajoki, 201). The repeated use of SRF as a part of a learning process supports reflective thinking, which is crucial for REI competence development. If participants' ethical sensitivity and competencies (levels of understanding) improve, we may consider a training to be effective.

The fact that in some courses not many SRFs were returned (during Iteration 3) suggest that the implementation of self-reflection may need to be supported through prompts and encouragement. Also, the topics the participants choose to reflect on may be side-tracks and not about specific ethical issues. Utilizing self-assessment takes careful planning as emphasized by Ross (2006). Self-assessment becomes more valid and accurate when learners are aware of the criteria of evaluation and the alignment of evaluation and learning outcomes, they are taught about how to use self-assessment and feedback is provided on their progress as well as how to improve their understanding (Ross, 2006).

Even though Iteration 3 did not accumulate much data, those ethical topics that were included and described had an almost complete match between the SOLO level the learners chose and the description they provided. Since the participants filled in the form anonymously, we have no way of checking, but based on the number of returned forms, we may assume that several learners also used the SRF multiple times, which made their responses more accurate. Ross (2006) supports this by outlining that practicing self-assessment makes it more accurate. Also, expertise increases, which makes learners more prone to accurately evaluate their understanding (González-Betancor et al., 2019).

Expert evaluation during Iteration 4 was in line with the DBR methodology (Anderson & Shattuck, 2012). Expert evaluation can be implemented at various stages of intervention development, and the expert can be a person with experience in the field but also a potential user of a tool (Vesper et al., 2011). Experts in this study were all experts in REI training, so their insights were considered valuable

and reliable. The main suggestions by the experts were that the form should provide feedback to learners (also suggested by Ross, 2006) and the form is probably more suitable for short training sessions or specific activities. This was also supported by Ross (2006) who has claimed that self-assessment is more reliable across tasks, specific items, and shorter periods of time.

CONCLUSIONS AND RECOMMENDATIONS

The current study focused on development of a self-reflection tool as a measure of self-assessment for research ethics and integrity training and aimed to develop an accurate tool. The DBR methodological approach with four iterations was used for research and development including background search, tool development and testing, and expert evaluation. Originally a feedback form was used to collect learner reactions, but it did not provide information about the learning process and alignment with learning outcomes. A self-reflection form was designed and tested during several iterations to evaluate its usability and accuracy.

The results from testing iterations show that most learners can evaluate quite accurately their level of understanding in the context of research ethics and integrity. There were occasional instances of overestimating and underestimating one's level of understanding, but these were not prevalent. While self-reported assessments of various sorts are commonly used in the context of REI training (see Steele et al., 2016; Stoesz & Yudintseva, 2018; Turner et al., 2018), to our knowledge this is the first time that the reliability of self-reported data has been systematically tested with an intention to develop a reliable tool.

The results indicate that repeated reflection appears to improve accuracy of self-reflection. More participants can evaluate their actual level of understanding and the match between the selected and described level is better if the SRF is used several times. In addition, as ethical sensitivity and competence, i.e., expertise in REI, grows, self-assessment also improves by becoming more accurate.

We recognize that there are limitations in this study. While the DBR approach made use of the input of different target groups, it did not include all target groups in all iterations. In some iterations, the number of participants from training sessions in which the SRF was used, was relatively small, and may not accurately reflect those target groups. The respondents who chose to return the SRF may be particularly interested in REI, perhaps more so than the average learners of each target group. We did not assess how participants react to the instant feedback, which we added based on expert suggestions in Iteration 4. Future research may target how participants respond to feedback in terms of attitude, emotions, and reflection.

We are also aware of possible bias during self-report responses. The respondents may be inclined to indicate a higher level of understanding depending on individual difference. Nevertheless, based on adult learning theories (Knowles, 1975) and the requirement that all researchers should uphold the highest quality in research (ALLEA, 2017), we did not look into individual differences of responses. This could be done in future studies that focus on individual differences in REI self-reflection.

In addition, as SRF is more suitable for evaluation short-term outcomes of training, we cannot make conclusions on training impact on researcher behavior. Future studies could include a more longitudinal aspect to measure training effectiveness and also include the measurement of behavioral aspects.

Based on the results of this research, we provide some recommendations for using a self-reflection tool:

- Collecting self-assessment data and encouraging learners to self-reflect on their learning is beneficial, especially in the context of REI where deliberation and reflection contribute to ethical sensitivity and competence (see also Mustajoki & Mustajoki, 2017);
- Combining facilitator evaluation and learner self-assessment gives a more holistic picture of the learning process (see also Tammeleht et al., 2022);

- Self-assessment and self-reflection should be used repeatedly as this makes the evaluation more accurate (see also Ross, 2006);
- Self-assessment and reflection should be explicitly taught to learners, and they should be encouraged to use it systematically, there should also be a safe environment for using self-assessment (Ross, 2006)
- Self-assessment might be more suitable for evaluating understanding of a specific task or topic and during a shorter time period (see also Ross, 2006);
- Increased accuracy may indicate that the training format is effective, as more expert learners tend to be more accurate with their evaluation.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

Funding

This research is conducted within the BEYOND BAD APPLES project, which has received funding from the EU's Horizon Europe program under grant agreement No 101094714.The data that support the findings of this study are available on request from the corresponding author, AT. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

REFERENCES

ALLEA. (2017). The European code of conduct for research integrity. Retrieved August 5, 2023, from. www.allea.org

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189X11428813

Bakker, A. (2018). *Design research in education: A practical guide for early career researchers*. Routledge.

Barab, S. (2014). Design-based research: A methodological toolkit for engineering change. In *The Cambridge handbook of the learning sciences* (2nd ed., pp. 151–170). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.011

Bebeau, M. J., Rest, J. R., & Yamoor, C. M. (1985). Measuring dental students' ethical sensitivity. *Journal of Dental Education*, *49*(4), 225–235. https://doi.org/10.1002/j.0022-0337.1985.49.4.tb01874.x

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research & Development*, *18*(1), 57–75. https://doi.org/10.1080/0729436990180105

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at University* (3rd ed.). SRHE and Open University Press.

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In *Advances in experimental social psychology* (Vol. 44, pp. 247–296). Academic Press. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences*, *11*(1), 105–121. https://doi.org/10.1207/S15327809JLS1101_4

Finnish Advisory Board on Research Ethics. (2009). *The ethical principles of research with human participants and ethical review in the human sciences in Finland*. Finland: TENK.

González-Betancor, S. M., Bolívar-Cruz, A., & Verano-Tacoronte, D. (2019). Self-assessment accuracy in higher education: The influence of gender and performance of university students. *Active Learning in Higher Education*, *20*(2), 101–114. https://doi.org/10.1177/1469787417735604

Jordan, S. R. (2013). Conceptual clarification and the task of improving research on academic ethics. *Journal of Academic Ethics*, *11*(3), 243–256. https://doi.org/10.1007/s10805-013-9190-y

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

Katsarov, J., Andorno, R., Krom, A., & van den Hoven, M. (2022). Effective strategies for research integrity training—a meta-analysis. *Educational Psychology Review*, *34*(2), 935–955. https://doi.org/10.1007/s10648-021-09630-9

Knowles, M. S. (1975). Adult education: New dimensions. *Educational Leadership*, *33*(2), 85–88.

Löfström, e. (2012). Students' ethical awareness & conceptions of research ethics. *Ethics and Behavior*, *22*(5), 349–361. https://doi.org/10.1080/10508422.2012.679136

Mol, H., & van den Hoven, M. (2022). Is there a relationship between student attitudes and behavior regarding integrity issues? *International Journal for Educational Integrity*, *18*(1), 1–13. https://doi.org/10.1007/s40979-022-00100-5

Mustajoki, H., & Mustajoki, A. (2017). *A new approach to research ethics: Using guided dialogue to strengthen research communities*. Taylor & Francis. https://doi.org/10.4324/9781315545318

Nevgi, A., & Löfström, e. (2015). The development of academics' teacher identity: Enhancing reflection and task perception through a university teacher development programme. *Studies in Educational Evaluation*, *46*, 53–60. https://doi.org/10.1016/j.stueduc.2015.01.003

Rest, J. R. (1986). *Moral development: Advances in research and theory*. Praeger.

Rissanen, M., & Löfström, e. (2014). Students' research ethics competences and the university as a learning environment. *International Journal for Educational Integrity*, *10*(2). https://doi.org/10.21913/IJEI.v10i2.1004

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, *11* (1), 10.

Steele, L. M., Mulhearn, T. J., Medeiros, K. e., Watts, L. L., Connelly, S., & Mumford, M. D. (2016). How do we know what works? A review and critique of current practices in ethics training evaluation. *Accountability in Research*, *23*(6), 319–350. https://doi.org/10.1080/08989621.2016.1186547

Stoesz, B. M., & Yudintseva, A. (2018). Effectiveness of tutorials for promoting educational integrity: A synthesis paper. *International Journal for Educational Integrity*, *14*(1), 1–22. https://doi.org/10.1007/s40979-018-0030-0

Tammeleht, A. (2022). *Facilitating the development of research ethics and integrity competencies through scaffolding and collaborative case-based problem-solving* [Doctoral Dissertation]. University of Helsinki.

Tammeleht, A., Koort, K., Rodríguez-Triana, M. J., & Löfström, e. (2022). Knowledge building process during collaborative research ethics training for researchers. *International Journal of Ethics Education*, *7*(1), 147–170. https://doi.org/10.1007/s40889-021-00138-y

Tammeleht, A., Rodríguez-Triana, M. J., Koort, K., & Löfström, e. (2019). Collaborative case-based learning process in research ethics. *International Journal of Educational Integrity*, *15*(1), 6. https://doi.org/10.1007/s40979-019-0043-3

Teadustava, H. (2017). *Estonian Code of Conduct for Research Integrity*. Retrieved September 5, 2023, from https://www.eetika.ee/sites/default/files/www_ut/hea_teadustava_eng_trukis.pdf

Turner, M. R., Watts, L. L., Steele, L. M., Mulhearn, T. J., Torrence, B. S., Todd, e. M., Mumford, M. D., & Connelly, S. (2018). How did you like this course? The advantages and limitations of reaction criteria in ethics education. *Ethics & Behavior*, *28*(6), 483–496. https://doi.org/10.1080/10508422.2017.1308193

Van den Akker, J. J. H., & Gravemeijer, K., McKenney, S. & Niveen, N. Eds. (2006). *Educational design research*. Oxon. https://doi.org/10.4324/9780203088364

Vesper, J. L., Reeves, T. C., & Herrington, J. (2011). The application of expert review as a formative evaluation strategy within an Educational design research study. http://researchrepository.murdoch.edu.au/id/eprint/11821/(26.06.2023)

Watts, L. L., Medeiros, K. e., Mulhearn, T. J., Steele, L. M., Connelly, S., & Mumford, M. D. (2017). Are ethics training programs improving? A meta-analytic review of past and present ethics instruction in the sciences. *Ethics & Behavior*, *27*(5), 351–384. https://doi.org/10.1080/10508422.2016.1182025

Winstone, N. e., Ajjawi, R., Dirkx, K., & Boud, D. (2022). Measuring what matters: The positioning of students in feedback processes within national student satisfaction surveys. *Studies in Higher Education*, *47*(7), 1524–1536. https://doi.org/10.1080/03075079.2021.1916909

APPENDICES

Appendix I – Feedback form

Please give feedbackto the workshop (tick/specify)

		Which part?/Please specify
1.The tasks were…	easy	
	complicated	
	average	
2. Was any part …	unclear?	
	too simple?	
	Everything was OK.	
3.What was the role of the group? Group-work …	helped	
	made things more complicated	
	disturbed	
	didn't matter	
4. Will participation in the workshop help notice ethical aspects better in the future?		
5. Did the workshop fulfill your expectations?		
6. What would you change? Improve?		
7. What surprised you most? (e.g. was important/new?)		

Appendix II – Initial online SRF

Link to the form: https://en.researchethicscompass.net/enesereflektisoon

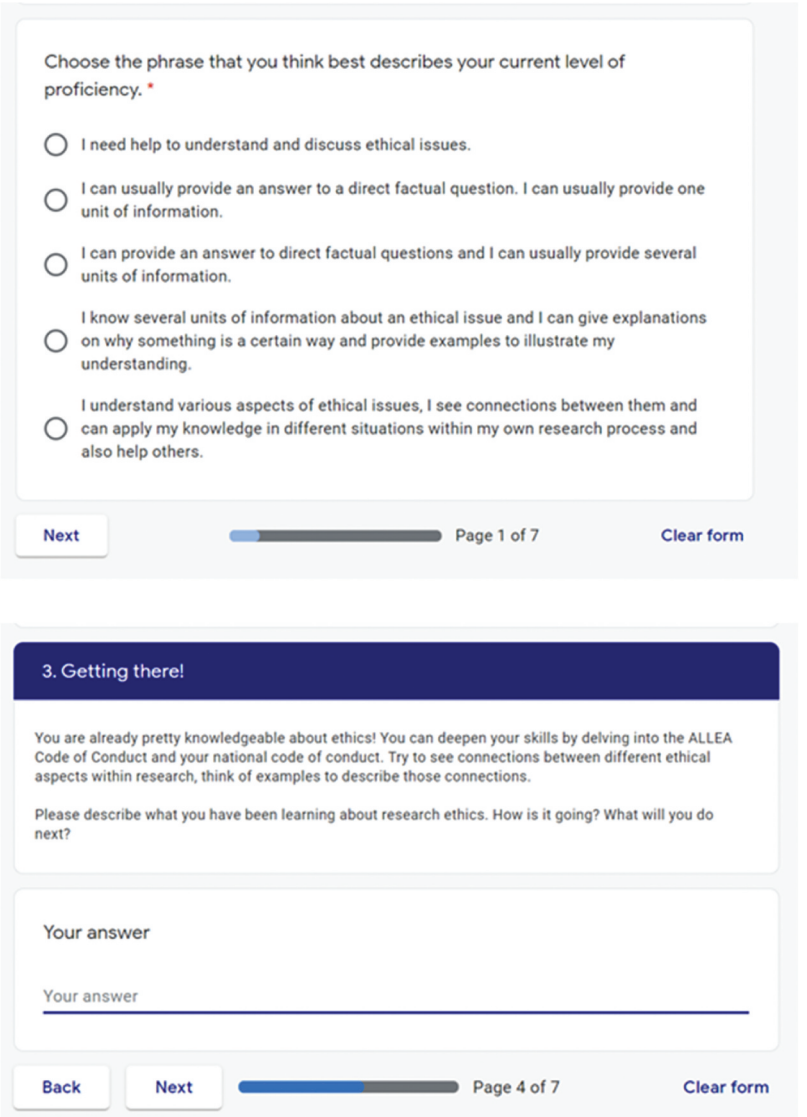

Figure A1. Screenshots of the initial online SRF (Tammeleht, 2022).

Appendix III – Content of the final version of SRF

Occasionally it is important to reflect on your level of understanding. Please think back at your previous activity and evaluate your ethical competence:

- Write down the activity you would like to reflect upon.
- Then choose a statement about your understanding (pick the closest match).
- In addition, write a short explanation to illustrate your choice.
- Receive feedback.

1. Which activity/topic are you reflecting on?
2. Please describe what you have been learning about research ethics in relation to the above mentioned activity/topic. Try to support the statement you chose above. (You can provide examples)
3. Choose the phrase that you think best describes your current level of understanding:
   a. I have difficulties identifying if an issue is ethical in nature, and I find it difficult to begin sorting the issue out.
   b. I can usually provide an answer to a direct factual question. I can usually think of one-two things to say.
   c. I can provide an answer to direct factual questions and I can usually say three or more items related to the topic.
   d. I can think of several items/aspects about an ethical issue and I can give explanations on why something is a certain way and provide examples to illustrate my understanding.
   e. I understand various aspects of ethical issues, I see connections between them and can apply my knowledge in different situations within my own research process and also help others.
4. **Feedback**:
   a. To get on track with research ethics and integrity, you need to do some work Try to make sense why the topics discussed are relevant. Imagine situations in which you feel empathy or in which you feel you are guided by your values. Why are the values important to you? What would a situation, in which opposite or very different values prevail, be like? What would you do? Considering such questions will help you to think about research ethics and integrity. You can improve your ethical competence by consulting relevant codes of conduct, and to discuss with experienced researchers how they think about making right choices in research? What do they say? How do they justify their points of view? Why do you think these questions are important for the research community? What is your role in it?
   b. You are on the way: You know something! You can still deepen your competencies in research ethics and integrity by consulting relevant codes of conduct and discussing with more experienced researchers. Try to grasp more aspects of ethical topics. What else is important or even necessary in order to make the right choices in research? Who all may be concerned about the situation, and would different people perhaps have different views on a situation? What other issues may follow from the current one? Ethical issues in research are often such that there can be related questions, or new questions may follow from the initial issue. Try to map out who all are concerned by a situation, and what kind of different questions they would raise about the situation!
   c. You are already pretty knowledgeable about ethics and integrity! You identify more than just one dimension of ethics or integrity in an issue. Next, try to see connections between different ethical aspects within research, and think of examples to describe those connections. Who are concerned about the situation? Do different interests and points of view lead to different solutions? How does an individual's action influence a group or a community of researchers or even an institution or field? And the other way around: If a group allows for bad research practices to flourish, how does that impact individual researchers and the colleagues they collaborate with?
   You can go over relevant codes of conduct to review principles of good research practice and think about examples of questionable research practice and how to avoid those. What can you do in order to ensure that your research is up to the highest ethical and integrity standards? Share ideas with others in your group or research community!
   d. You know ethics very well, but there is still a little room for improvement! You can consult the relevant codes of conduct again, but also try to generalize your knowledge and try to see beyond your own research. Find ways to apply your knowledge in various contexts and provide help to others. It is also useful to raise questions in your research group or with colleagues in order to identify how the research ethics of others influence people beyond their group, or what kind of practices could be developed to mitigate questionable research practices.
   e. You have good competencies in research ethics and integrity! You can manage your own research and ethics, and help others manage theirs. You may be a research project leader or you have thought about research ethics and integrity a lot, perhaps as a consequence of the type of research you are doing. You probably have the ability to recognize ethical issues in research and to work on those before they turn in to problems. You may wish to think about how to contribute to the wider research community by acting as a role-model. By showing good example, researchers can contribute to more sound research practices and a better research culture.

Appendix IV - Examples of data analysis (from Tammeleht et al., 2019; 2022)

Table A1. Examples of data analysis (Tammeleht et al., 2019; 2022)

SOLO level	Coding description	Example
Extended abstract	The learner goes beyond conceptualising the present ethical issue making steps towards relating the issues to applications beyond the present case. Displaying the ability to theorise, generate, generalise, hypothesise, create, or reflect.	"Solutions could be that there is no filming and they use another data collection method. or provide the same activities for participants in another room – the researcher must not be in the comfort zone, must really give an effort and provide alternatives for participants. Segregation must be avoided at any cost. Researchers should really work hard to see the parents' point of view."
Relational	The learner displays an ability to address the most relevant ethical issues and provide explanations pointing out interrelations and providing examples demonstrating own reasoning.	"Of course, there should be respect for autonomy – everyone should be able to make a choice independently. But on the other hand, someone may still be hurt. How to be just and form the groups in a fair way?! (adding a personal example)
Multistructural	The learner displays an ability to address the most relevant ethical issues and provide explanations pointing out interrelations and providing examples demonstrating own reasoning.	Highlighted text: [. . .] Girls in their early teens [. . .]. Comment: "Vulnerable group of minors; research permits; consent from the girls/parents; ensuring confidentiality."
Unistructural	The learner identifies one relevant aspect displaying some familiarity with relevant concepts but failing to address some more pertinent dimensions of the case.	"All the aspects of the code of conduct (e.g. Hea Teadustava; code of conduct, etc). Yes; from the internet."
Prestructural	The learner fails to identify a relevant ethical perspective or does not approach it in a meaningful way, repeating the words in the case without displaying evidence of own processing.	"Which codes of conduct?"