D3.4a Report on developing a Quality Checklist for RCR education (prototype tool)

Deliverable Code	D3.4_20200116_V1.0
Distribution Level	Public
Responsible Partner	UU
Authors	André Krom & Mariëtte van den Hoven
Checked by	Orsolya Varga (UD) Date: 01-14-2020
Approved by Project Coordinator	Mariëtte van den Hoven (UU) Date: 01-16-2020

This project received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 824586

Contents

Introduction

The Horizon 2020 Integrity project aims to empower students[1] and early career researchers through education[2] for Responsible Coduct of Research (RCR). It does so in an evidence-based way, and by using a scaffolded approach. This means that education for RCR will be tailored 1) to their educational level and discipline, and 2) to the specific needs that student groups may have to enable them to responsibly navigate issues of research integrity – current and new (i.e. in a way that is 'future proof').[3]

To these aims this report (D3.4) presents a benchmark for assessing the quality and effectiveness of teaching practices in RCR training. We present a prototype quality checklist that can be used to help determine how to design and develop innovative tools for RCR education in WP 4.

To develop the quality checklist and the prototype tool, we described in D3.3 what strategy would be used to arrive at a qualtiy checklist. Having composed a database tailored to the aims of task 3.3 as the primary source for developing criteria for a quality checklist, we came across a publication that seemed particularly promising for developing the quality checklist. The publication in question is the work by Mulhearn et al. (2017), who have developed and validated what they call a Predictive Modeling Tool (PMT). We observed that this tool focuses mainly on formal education, while the H2020 Integrity project also includes informal and nonformal educational settings and include a wide variety of learning aims that can be used in RCR education. Moreover, we wanted to gain insight in what determines the quality of RCR education for several disciplines and study phases, thus to be able to tailor educational innovations in WP 4 to the needs of the target groups students in our project.

---

[1] 'Empowerment' will be operationalised in WP 4 of the Integrity-project. Generally, this will include competences of individual students/researchers, as well as institutional capacities that may be required to maintain or promote Responsible Conduct of Research.
[2] 'Education' is taken in a broad sense, covering formal education (e.g. courses), informal education (e.g. mentoring) as well as non-formal education (e.g. via science museums).
[3] The notion of 'future proof' competences and capacities will be fleshed-out in WP 4 of the Integrity-project.

This report specifically presents a first version of a quality checklist and prototype tool for assessing education for RCR. Moreover, 'assessing' should be primarily understood here as peforming quality checks on basic aspects of RCR education, based on the results and considerations contained in the tool. We can expect the tool to be further developed when results from other tasks and work packages of the Integrity-project become available, some of which will yield their final outcomes that are relevant for the checklist and the prototype tool after this deliverable is due. We can also expect that WP 4, where we will bring together the information from WP2 and WP3 into a more comprehensive view on teaching RCR to empower students will also contribute to the further fine-tuning of the tool. Lastly, we will seek for other media to present the tool in an even more user-friendly manner, as Excel prototypes also have limitations.

The quality checklist and tool for assessing education for RCR will also be described in a scientific article. That allows us to make the most of the work done in other work packages as well, specifically, the co-creative exercises with intended audiences of the educational tools that will be developed in WP4-6. We will use the draft quality criteria of this report (D3.4) to perform quality checks during those co-creative exercises, which – we expect – will also help to further refine the quality criteria.

Outline

This report consists of two parts. First we describe the methodology that was used to develop a prototype tool of a checklist for performing quality checks on RCR education (part I). Then, we will introduce the prototype tool of the Quality Checklist and provide guidance on how it can be used. (part II). The tool itself, including all main lessons and quality criteria is on the H2020 Integrity website. Finally, in Appendix 1 we have included a brief comparison to a comparable product that had been developed for designing, assessing and improving RCR education: the Predictive Modeling Tool (PMT). Here, reasons are given for developing the prototype tool as a complement to the PMT.

Part I - Methodology for developing the Quality Checklist

This section describes the methodology used to develop the prototype tool of the Quality Checklist. In all fairness, this was not a straightforward process. Indeed, it was an iterative process with new insights leading to a number of adjustments. This included *prioritising* a focus on formal education for the first version of the Quality Checklist (prototype tool), reasons for which will be provided below.

Literature search

As part of the same work package a systematic review of the literature on RCR was planned, focusing on the period 1990 – 2018 (task 3.1). In addition, a questionnaire on RCR teaching practices was organized in a range of European countries (task 3.2).[4] The development of the quality checklist and prototype tool presented here (task 3.3) is based in part on some of the preliminary results of task 3.1. To start with, the literature review in task 3.1 yielded the conclusion that there have been a range of meta reviews on the efficacy of RCR education. Most of the meta reviews used for developing the prototype tool were suggested based on the literature review in task 3.1. The list was completed during an additional literature search, updating the previous one. Besides yielding an extra meta review, this search also brought to our attention a tool that seemed quite promising for the purposes of developing a quality checklist, at least at first. The tool in question is the Predictive Modeling Tool (PMT), that was developed and validated by Mulhearn et al. (2017).

Workshops

To determine whether the PMT could be useful for the purposes of developing a quality checklist, and if so to what extent, two workshops were organized. An internal workshop at Utrecht University (Workshop 1, 28 May 2019), and a workshop with WP3 partners at the University of Zürich (Workshop 2, 15 July 2019). In brief, workshop 1 yielded the conclusion that it was worthwhile to further explore the added value of the PMT for the purposes of task 3.3 (quality checklist). Workshop 2, in turn,

[4] See D3.2 (mapping results of current teaching practices).

maintained that conclusion, but also yielded a range of questions and issues concerning the PMT. Both in preparation for workshop 2 and afterwards the corresponding author of the work on the Predictive Modeling Tool was contacted to collect additional material on the PMT and to pose some of the main questions and issues concerning the PMT that came up during our second workshop.[5]

Interim-plan: develop the quality checklist by adding to the Predictive Modeling Tool

For some time, our plan was to develop the quality checklist by a) specifying what kind of input and information was needed for the quality checklist, b) examining to what extent the PMT could provide this information, and c) to complement any missing information by performing a limited and focused additional literature search. An extra literature search of limited scope was indeed performed to find information on additional target audiences (specifically, high school students), additional disciplines (other than, for instance medical sciences, social science and science & engineering) and on additional forms of education (specifically mentoring). Parallel to the additional literature search, and some initial coding exercises, ideas about what would be a solid base for developing the quality checklist began to shift, more and more away from the Predictive Modeling Tool. Based in part on the ambition to develop a tool that could be used by those involved in developing and assessing RCR in a broad range of disciplines, it was decided to at least include the main lessons of all meta reviews in the quality checklist. The Predictive Modelling Tool is primarily based on meta reviews [8] and [9], which focus on a limited range of academic disciplines. In the final analysis, it was decided to develop the quality checklist more or less independent from the PMT[6], and to include the results of 11 meta reviews in our prototype tool. A separate manuscript is being prepared that will offer a systematic comparison between the Predictive Modeling Tool and the prototype tool of the Quality Checklist.

[5] We are very grateful to prof. Mumford (University of Oklahoma) for his exceptionally rapid and friendly responses to a range of requests.

[6] See Appendix 1 for a brief discussion of the Predictive Modeling Tool, and the main reasons for assigning a (much) less prominent role to the Predictive Modeling Tool in developing the Quality Checklist.

Structure and visualisation

The structure and visualisation of the Quality Checklist was developed in the following way. First, based on earlier work on quality of education, our WP3 project partners from the University of Zürich suggested to categorize the material in the Quality Checklist by using the basic distinction between INPUT (what is put into education), PROCESS (the education itself), and OUTPUT (focusing on assessment of education). Based on the review of the literature in task 3.1, our partners from Zürich also made a range of suggestions for what to include in the Quality Checklist.

Second, we organized an internal workshop at the University of Utrecht (25 November 2019) on how we could (further) visualise the quality checklist. We considered visualisation to be important since the literature on efficacy of RCR education is so extensive, and this presents a real challenge when it comes to how to provide those involved in RCR education with relevant (and if needed in-depth) information in an accessible way. Central ideas coming out of this workshop where a) that we should use as little text as possible, b) that we would use colour codes to indicate effect sizes (shades of green for positive effects, red for negative effects), and c) that information should be provided in a layered way, leaving it up to the user to determine whether s/he would like to view more detailed information on how to improve RCR education. We considered that it would be worthwhile to try and develop an *interactive* tool, but at this stage it was kept open to what extent we would be able to actually produce an interactive version of the prototype tool. The workshop also yielded the basic structure of the Quality Checklist prototype tool.

Introducing interactive elements to maintain accessibility

During the development of the prototype tool, a tension arose between trying to keep the quality checklist as simple as possible, on the one hand, and providing enough information that allows users to adequately contextualize the results, on the other hand. The latter point essentially revolves around a strong conditionality of conclusions about what could promote the efficacy of RCR education. For instance that what may work in one context, does not automatically work in another context. We consider it to be important to keep this conditionality on the radar when developing and/or adapting RCR education. That meant, however that quite a lot of more detailed information had to be included in the tool, increasing the risk that the tool would become inaccessible or at the very least that using the tool would become overdemanding. The solution was sought in a combination of introducing layers

(most importantly distinguishing between "main lessons" and "details"), and introducing interactive elements. The interactive prototype tool of the Quality Checklist was the result of a gradual and iterative process aimed at maintaining accessibility while also keeping the conditionality of the main lessons on the radar. The choice to do this in Excel at first was primarily motivated by this being the program that the project team was most familiar with. As indicated we will seek for other media to present the tool in an even more user-friendly manner, as Excel prototypes also have limitations.

Prioritising formal education

Meta reviews provide the most systematic and rigorous reflections on the efficacy of RCR education. They serve as a filter in that they only include studies that meet a range of high standards for examining the efficacy of RCR education. One such requirement is that studies should provide information on the effect sizes found, and how they were calculated. [8] This is the main reason for why we focused on meta reviews in developing the Quality Checklist, and, as a consequence also to focus on *formal* education, as these meta reviews mainly discuss formal educational trainings. As a result, a choice had to be made with regard to what the focus of the first version of the prototype tool of the Quality Checklist would be. As indicated in D3.3 the initial plan was to include informal education as well, that is the impact of mentoring practices.

We decided to prioritise formal education for the first version of the prototype tool. This was done for pragmatic reasons. First, the decision to develop the quality checklist from the results of 11 meta reviews – and not just the meta reviews [8] and [9] which underly the Predictive Modeling tool - meant a huge increase regarding the amount of information that had to be included in the prototype. It was impossible, practically, to also include additional insights on the impact of mentoring practices within the time-frame set for this deliverable.[7]

Including informal education at a later stage

Importantly, the choice to prioritise formal education in the current version of the Quality Checklist does not pose problems for subsequent tasks. Indeed, we think that subsequent tasks in WP 4, in

[7] This is supported by the fact that the deliverable is available two weeks later than originally planned.

which formal RCR education will be developed and tested, benefit most from the decision to initially prioritize formal education in the Quality Checklist. In brief: the choice results in a knowledge base that is much more solid than it would have been had we tried to include informal education as well at this stage.[8] As the first version of the Quality Checklist is a prototype tool, that will be developed further during the project, we see ample opportunity to also include the results from our literature search on e.g., mentoring practices in the Quality Checklist. Specifically, this can and will be done in combination with preparations for tasks 5.3 and 5.4, that focus on mentoring.

Building on other tasks and work packages

Preliminary results of prior tasks and work packages have informed the development of the current version of the Quality Checklist. As indicated, a selection of meta reviews and suggestions for quality criteria followed from the literature review in task 3.1. In addition, results from a questionnaire that was developed in task 3.2 clearly show that RCR education is much used in some disciplines than in others[9], supporting one of the main quality criteria included in the Quality Checklist, namely that one size does not fit all, and that we should always keep in mind that the main lessons on the efficacy of formal RCR education are based on studies in a limited range of academic disciplines. Conclusions on what works, therefore, cannot easily be generalized. Finally, interviews that were part of qualitative research for WP 2 confirm that there are strong differences between disciplines in terms of knowledge (of RCR, and what it requires) and needs. This supports a key consideration on the Quality Checklist, namely that RCR education should take the specific needs of participants into account.

In further developing the prototype tool, we will increase its empirical robustness even more by including important insights that will follow from subsequent tasks and work packages. Specifically, as indicated in D3.3, the Quality Checklist will not only be used to perform quality checks on courses and tools that will be developed as part of the H2020 Integrity project, we expect that applying the quality criteria will also yield important information for further refining the Quality Checklist itself.

---

[8] As this would have meant that substantially less time would have been available for describing main lessons on how the quality and efficacy of formal education can be checked and improved.
[9] See D3.2 (mapping results of current teaching practices).

Selection of what to include in the Quality Checklist

Studies on the efficacy of RCR education contain a rich variety of information, with some findings being more relevant to the Quality Checklist than others. While most findings from all 11 meta reviews have been included in prototype tool, a very small number of findings has been left out. For example, information on correlations between efficacy of RCR education and "publication outlet", e.g., whether a study had been published in the social sciences or in ethics,[10] was omitted. Such findings, we thought, are too remote to qualify as a main lesson on what can contribute to the efficacy of specific educational activities.[11]

A note on categorisation

In some cases, we have chosen to categorize key topics related to the efficacy of RCR education slightly different than it was done in the underlying meta reviews. One example is trainer characteristics. While meta reviews commonly discuss for instance the impact of trainer expertise as part of "general instructional characteristics" (e.g., [8]), we have grouped characteristics related to trainers under "Input" and "conditions". This does not affect what main lessons are drawn, it just means that these lessons are to be found at a different location.

Towards a view on assessment

The H2020 Integrity project takes Responsible Conduct of Research as referring to an overarching perspective on research conduct/practices. Empowerment of students is an essential element of that perspective. The project focuses on examining what can contribute so that researchers know how to act in a responsible manner in conducting research.

[10] For information on effect sizes related to "publication outlet" see meta review [1].
[11] In case we considered omitting a finding from a meta review (again, this was done only sporadically), we checked whether the authors considered those findings as a main finding. If not, we considered this to be a confirmation that the finding was indeed too remote from qualifying as a main lesson on what can contribute to the efficacy of do specific educational activities.

This section presents a preliminary view on the assessment of Responsible Conduct of Research (RCR) education. Its preliminary status derives in large part from how the field of RCR is currently conceptualized: unclear and not always consistent. Depending on what we take RCR to entail, we might arrive at different conclusions concerning the extent to which RCR education has shown to be effective. For example, if we take RCR to be a matter of conforming to the law and to rules and regulations of professional practices only, then RCR education contributing to improvements in those areas would support a conclusion that educational activities have been effective to a certain extent. If, however, we think that Responsible Conduct of Research also involves taking on board insights from research ethics, and that different normative theories might specify in different ways what values that are central to RCR require, then this could yield different conclusions on whether or not educational activities have been effective.[12] There is thus a need for conceptual clarification and consistency in our approach to RCR and RCR education. There is at least no unilateral view on RCR and RCR education deriving from the literature and, as a consequence, neither on when exactly RCR education is effective in the relevant respects.

That there is a need for conceptual clarification in the field of RCR is not new. Recent examples addressing the issue include Tauginienė et al. (2018) and Shaw (2019). Both publications make suggestions for clarification and a unified terminology. However, they do so in ways that are at odds with each other on certain key topics. For instance whereas Tauginienė et al. (2018) tie the term (academic and research) 'fraud' inherently to a form of *deception*, Shaw (2019) counts doing irrelevant research as a form of fraud as well.' In developing a Quality Checklist for the assessment of RCR education, we are dependent on the how the field is currently conceptualized, but cannot solve every conceptual unclarity and inconsistency. Perhaps that is also not necessary. Indeed, we think that for the purposes of this report, a preliminary view on assessment suffices, on the condition that when performing quality checks on specific educational activities, we remain aware that conclusions about the efficacy of RCR education are at least partly dependent on what we take RCR to entail. That being said, we think that (further) conceptual clarification is important, also for the broader purposes of the

[12] For instance, Steneck (2006) argues against research ethics having a role in research integrity, other than posing questions.

H2020 Integrity project. This will be offered in WP 4, in which we will also develop a unified account of RCR as a perspective on empowerment via RCR education.

The relation between quality and effectiveness

The aim to develop a *quality* checklist for the assessment of RCR education derives from the fact that if we are able to pinpoint the specific characteristics of RCR and how it is currently trained to researchers, we must also be able to distinguish the good and bad conditions for its success, for its quality. A core element of our view on assessment of RCR education, then, is that it is and should be about *quality* of education.

Surely, the extent to which education is effective in promoting certain learning aims, contributes to the quality of that education. Quality of education, however, is not determined *solely* by how effective that education is in promoting specific learning aims. For instance, we take the position that learning aims (should) come first, and that assessment should follow. This is an example of a quality criterion that is surely related to efficacy, but not solely determined by it.[13] Depending on what is thought to be important to teach specific trainees, quality may require choosing an approach that is known to have smaller effect sizes. For instance, meta reviews indicate that RCR education has larger effects sizes if the majority of trainees is male[14] and if participants are older.[15] But surely, this does not mean that we should only offer RCR education to this demographic. Indeed, it is important to offer it to younger female trainees as well. It seems reasonable, therefore, to consider RCR education to be of lower quality if it does not promote specific learning aims in all relevant target groups. Even if this means that larger effect sizes could be achieved by (only) targeting others.

There are further reasons for why quality of education should not be reduced to how effective that education is, are the following. First, what is considered important from a quality perspective might not always be easily measurable (including whether 'improvements' in that respect have been achieved).

[13] Cf. NAS (2017: 199); meta review [1], [2], [5], [8], [11].
[14] Meta review [8] and [9].
[15] Meta review [1], [2], [3], [5], [8], [9].

Second, factors external to education for RCR might impact how effective education for RCR can be. Suppose that impact is negative. Would it then be fair to conclude that the educational activities are or must be of poor quality? We are inclined to think that this would not be fair.

Along these lines, we think that assessing the quality of RCR education would ideally be a combination of (1) assessing the extent to which educational activities contribute to specific learning aims for specific target groups; and (b) meeting certain quality criteria that are less directly connected to the efficacy of education. Regarding efficacy of RCR education, it is clear that one size does not fit all.[16] This reinforces the idea that learning aims should come first, and that assessment should follow. Key is what we think specific participants should learn, not that we might achieve larger effect sizes by targeting others.[17] This, in turn, also means that "more effective" and "less effective" do not necessarily equal "good" and "bad".[18] Moreover, in determining what participants should learn RCR education should take the specific needs of participants into account.[19] That, we think is an essential requirement for RCR education to be able to empower concrete individuals.

[16] Cf. meta review [3], [4], [6] – [11].
[17] Cf. meta review [8].
[18] Cf. meta review [9].
[19] Cf. meta review [5].

Part II - Guideline to the Quality Checklist prototype tool – What is it and how can it be used?

The Quality Checklist was developed as an Excel-file.[20] The aim was to make an interactive tool that contains important insights from the scientific literature on what contributes to the efficacy of RCR education, and that can be used to perform basic quality checks on RCR education.

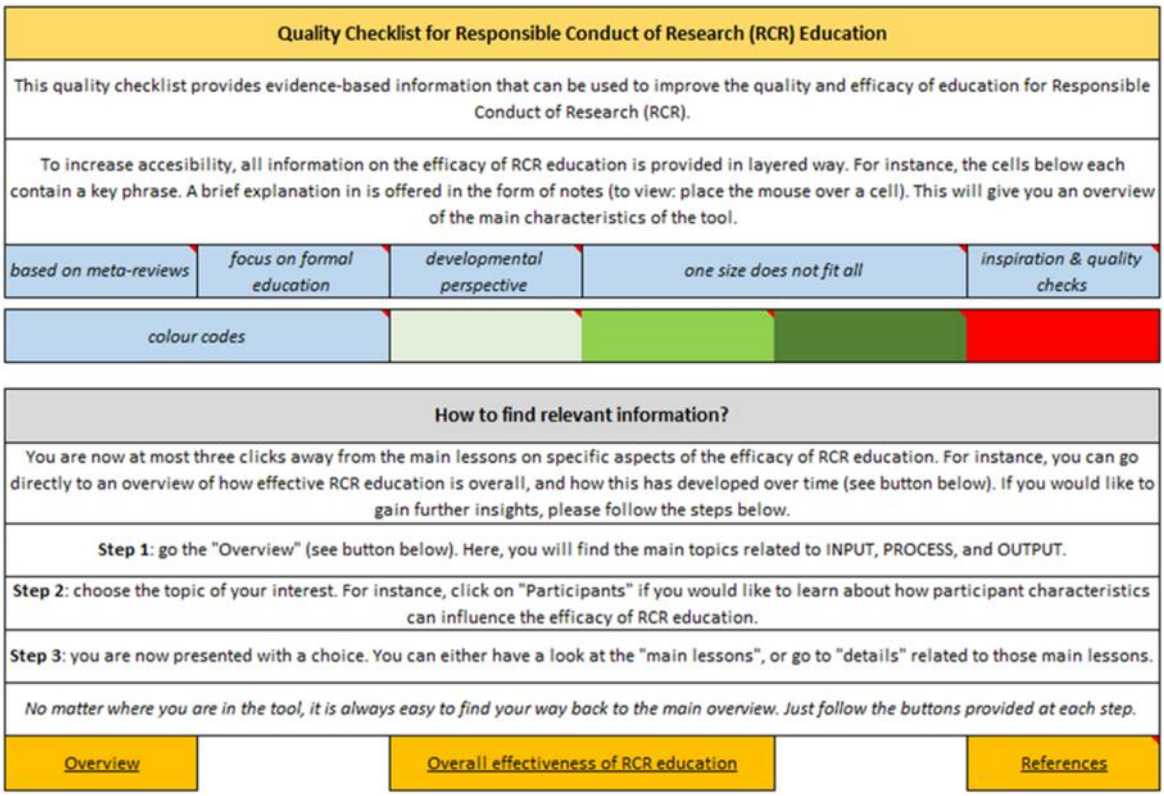
This Section describes the main characteristics of the Quality Checklist (prototype tool), how the tool can be used, the type of information it contains, and a general indication how to interpret the main results by way of an example. A selection of screen shots from the tool will be used as illustration. The tool itself is available on the H2020 Integrity website: http://h2020integrity.eu/about-us/documentation/.

"Intro"

Opening the prototype tool of the Quality Checklist, users are taken to the "Intro" page. Here, users can find basic information about the tool: its main aim and key characteristics, how to find relevant information, and how quality is related to the efficacy of RCR education. We will first discuss key

[20] The H2020 Integrity-consortium will explore ways to further develop the tool.

characteristics of the tool, and how users can find relevant information. These topics are situated on the left-hand side of the Intro page (see screenshot below).



Key characteristics of the tool

Key characteristics of the tool are that it is based on meta reviews on the efficacy of RCR education, that it focuses on formal education (i.e. on courses), and that it takes a developmental perspective. The latter entails that, as far as possible, main lessons on the efficacy of RCR education are connected to the educational or career stage of participants. In line with the broader H2020 Integrity project, the educational/career stages of central interest are high school, undergraduate students and early career researchers (graduate students and PhD candidates).

For each characteristic of the tool a brief explanation is available in a note, that users can read by hovering the mouse over a box. For instance, the box “based on meta reviews” contains a note explaining that 11 meta reviews were used as input for the tool, and that numbers [1] to [11] in the

tool each refer to a specific meta review. The list of meta reviews can be accessed via the "References" button (bottom-right). The availability of a note is signalled by a small red triangle in the top right of a cell.

The "Intro" also contains several pointers related to how the insights and conclusions from the meta reviews can be interpreted. To start with, a key lesson is that *one size does not fit all*. This means that what is the most effective approach for RCR education, may differ depending on, among other things, what are the central aims of that course, and who are the intended audiences. The tool allows users to examine what is known about effective approaches for different target groups.

The main added value of the quality checklist is that it allows users to perform basic quality checks on important aspects of RCR education. The tool can serve as a source of inspiration but cannot offer guarantees. It is a heuristic device, more than anything else. This is because studies on the efficacy of RCR education allow for conclusions about *correlations* between specific approaches and how effective that education is. They do not allow for conclusions about cause and effect.[21] Still, the guidance that the prototype tool can offer, provides anyone involved in developing and/or assessing RCR education with important insights on how different approaches may improve its efficacy.

To improve accessibility the quality checklist uses colour codes. Green is used to signal positive effects related to certain characteristics of RCR education, red to signal negative effects. In line with the scientific literature, we use three shades of green to indicate different sizes of positive effects. *Light green* for a small positive effect, *dark green* for a large positive effect, and *a shade of green in-between* for a medium-sized positive effect.[22] Finally, when presenting the main results on a certain topic, for instance the impact that participant characteristics may have on the efficacy of RCR education, we use

[21] In order not to overcomplicate the discussion, from here on we will speak of 'effects' and 'impact' instead of correlations. Please note that in each case conclusions are about correlations, not cause and effect.
[22] Commonly, effect sizes are expressed in terms of Cohen's d. Effects are considered to be small when d = .20; medium when d = .50; and large when d = .80.

the colour grey to indicate what is the most effective. This is because the exact effect sizes may differ, depending on for instance the scientific discipline.

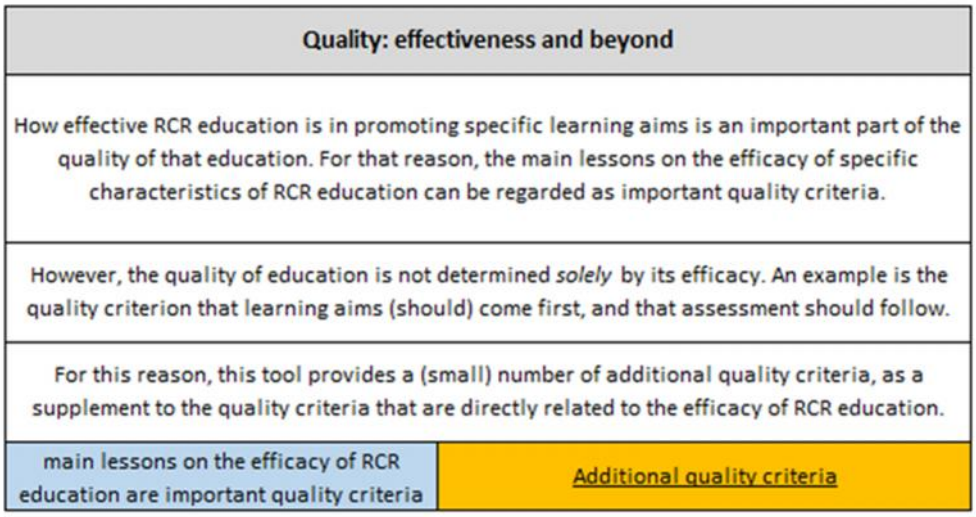
How to find relevant information?

The "Intro" page also describes how users can get to the information they are interested in. Information on the overall efficacy of RCR education, and how this has developed over time is available in one mouse click by going to "Overall effectiveness of RCR education". Main lessons on each topic are available three mouse clicks at most.

- Step 1 is to go to the "Overview" page, where key topics related to the efficacy of RCR education are grouped under "INPUT", "PROCESS", and "OUTPUT" (more on this below).
- Step 2 is to choose a topic of your interest. For every topic, this will take you to a basic menu with two options. Users can either go directly to the "Main lessons" on how this topic is related to the efficacy of RCR education or have a look at the "Details" underlying/ supporting the main lessons on that topic.
- Step 3 is making that choice.
- Finally, it is explained that each page of the tool contains shortcuts to the "Intro" and "Overview" pages, so users can move around quickly and easily.

Two types of quality criteria: effectiveness and beyond

The right-hand side of the "Intro" page describes how quality is related to the efficacy of RCR education (see screenshot below). Here, it is explained that quality of RCR education is determined in part by how effective that education is in promoting specific learning aims. Main lessons on how certain characteristics of RCR education may improve its efficacy can therefore be regarded as important quality criteria for RCR education. As indicated in the section on methodology (see Part I), quality of education is not determined *solely* by how effective that education is in promoting specific learning aims. In addition to main lessons on the efficacy of RCR education, the checklist offers a (limited) number of quality criteria quality that are less directly related to how effective RCR education is in promoting specific learning aims.

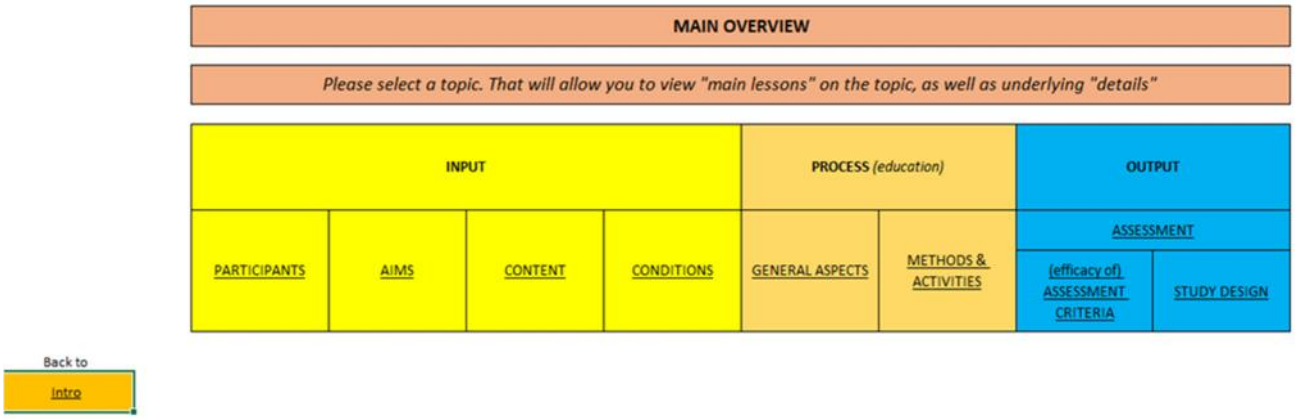


Readers interested in the collection of quality criteria (I.e. those related to efficacy, and the additional quality criteria) are advised use the prototype tool on the H2020 Integrity website: http://h2020integrity.eu/about-us/documentation/

This concludes our discussion of the "Intro" page of the prototype tool. We will now move on to the "Overview" page.

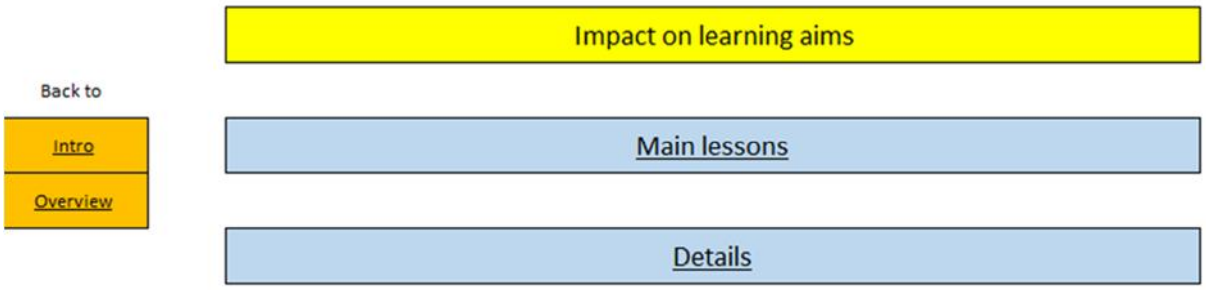
"Overview"

The "Overview" page (see screenshot below) can be regarded as the central station of the tool. It contains an overview of key topics related to the efficacy of RCR education. These topics are grouped as being primarily related to INPUT (what is put into education), to PROCESS (the education itself), and OUTPUT (focusing on assessment of education). From this point on, users are two mouse clicks away from either the main lessons on how a topic is related to the efficacy of RCR education, or to more detailed information on which those main lessons are based.

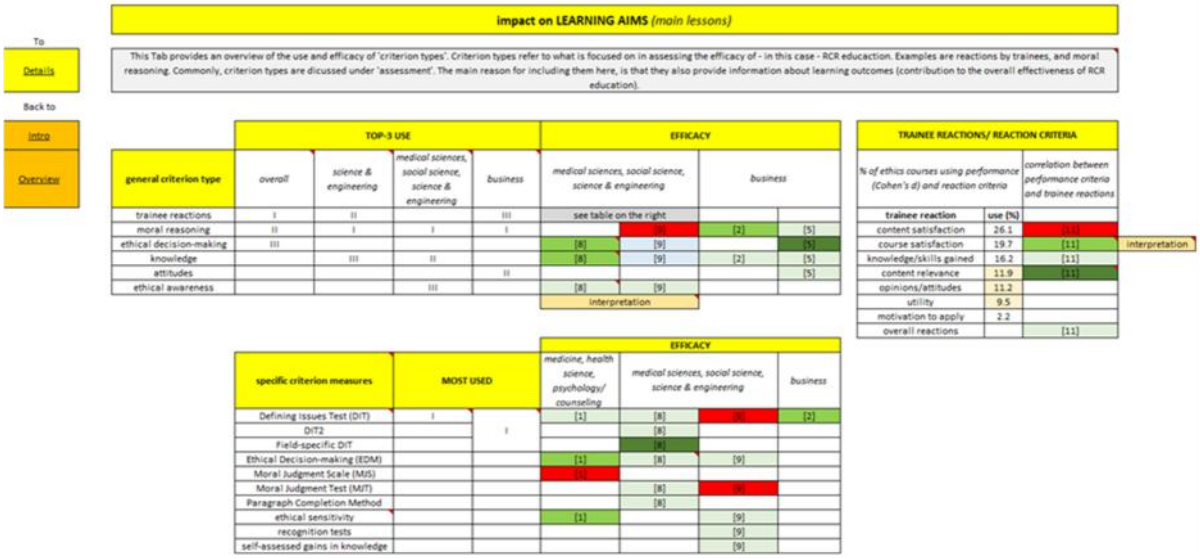


An example: "Impact on learning aims"

By way of illustration we will now show what the page with the main lessons related to learning aims looks like, and how the results should be interpreted. Users can get there by first clicking on "AIMS" in the main "Overview". This will take you to the menu that you see below.



When choosing "Main lessons here", the following page will appear (see screenshot below). Notice that even when focusing on main lessons some pages may still contain quite a lot of information. We will now explain how the page on the impact on learning aims can be read.



impact on LEARNING AIMS *(main lessons)*

To Details

Back to Intro Overview

This Tab provides an overview of the use and efficacy of 'criterion types'. Criterion types refer to what is focused on in assessing the efficacy of - in this case - RCR educaction. Examples are reactions by trainees, and moral reasoning. Commonly, criterion types are dicussed under 'assessment'. The main reason for including them here, is that they also provide information about learning outcomes (contribution to the overall effectiveness of RCR education).

general criterion type	TOP-3 USE: overall	science & engineering	medical sciences, social science, science & engineering	business	EFFICACY: medical sciences, social science, science & engineering		business	
trainee reactions	I	II		III	see table on the right			
moral reasoning	II	I	I	I		[9]	[2]	[5]
ethical decision-making	III				[8]	[9]		[5]
knowledge		III	II		[8]	[9]	[2]	[5]
attitudes				II				[5]
ethical awareness			III		[8]	[9]		
					Interpretation			

TRAINEE REACTIONS/ REACTION CRITERIA: % of ethics courses using performance (Cohen's d) and reaction criteria		correlation between performance criteria and trainee reactions
trainee reaction	use (%)	
content satisfaction	26.1	[11]
course satisfaction	19.7	[11]
knowledge/skills gained	16.2	[11]
content relevance	11.9	[11]
opinions/attitudes	11.2	
utility	9.5	
motivation to apply	2.2	
overall reactions		[11]

interpretation

specific criterion measures	MOST USED		EFFICACY: medicine, health science, psychology/ counseling	medical sciences, social science, science & engineering	business
Defining Issues Test (DIT)	I	I	[1]	[8]	[2]
DIT2				[8]	
Field-specific DIT				[8]	
Ethical Decision-making (EDM)			[1]	[8]	[9]
Moral Judgment Scale (MJS)			[1]		
Moral Judgment Test (MJT)				[8]	
Paragraph Completion Method				[8]	
ethical sensitivity			[1]		[9]
recognition tests					[9]
self-assessed gains in knowledge					[9]

To what extent RCR educational activities help promote specific learning aims, is of key importance. However, just like the studies that they are based on meta reviews do not offer a direct answer to that question. The closest that studies on the efficacy of RCR education come to this is by examining to what extent different so-called "criterion types" contribute to the effectiveness of that education. Criterion types refer to what is focused on in assessing the efficacy of – in this case – RCR education. Trainee reactions and moral reasoning are two examples. Criterion types are commonly discussed under "assessment" in the scientific literature. We chose to include this information under learning aims as well, given that the results on criterion types also provide information about learning aims and outcomes, namely their contribution to the overall effectiveness of that education. This is explained to users in a general remark at the top of the page.

The remainder of the page "Impact on learning aims (main lessons)" indicates which criterion types are used most, and to what extent each criterion type contributes to the overall effectiveness of RCR education.

Most used criteria

To limit the amount of information the tool focuses on the top-3 use of criterion types that are used the most. As the screenshot above shows, this may differ between scientific disciplines. For instance, generally *trainee reactions* are the most-used way to evaluate RCR education, followed by moral reasoning and ethical decision-making (see screenshot, top left). In science & engineering, however,

moral reasoning is the most-used criterion to evaluate RCR education, followed by trainee reactions. And the third most-used criterion type in science & engineering is *knowledge*. The top 3 may change again if we look at different (combinations of) disciplines.

Tools that are used to assess the efficacy of RCR education are mentioned under "specific criterion measures" (bottom left). The screenshot shows that (variations of) the Defining Issues Test (DIT) are among the most used specific criterion types. The DIT is a measure of moral development and/or moral reasoning. This corresponds to moral reasoning being the number 1 most used general criterion type in science & engineering; in business ethics education; and if we look at a collection of disciplines (such as medical sciences, social science, and science & engineering).

Most effective criteria

The main lessons also contain information about which criterion types are the most effective. 'Most effective' here means which criterion type contributes the most to the overall effectiveness of RCR education. Again, this may differ per academic discipline. Looking at general criterion types first (top central), it seems that ethical decision-making and knowledge are more effective criterion types than ethical awareness. And knowledge, in turn, seems to be more effective than ethical decision-making. The latter conclusion does not become immediately clear by looking at the colour codes. For both knowledge and decision-making have the same colour, corresponding to a medium large positive effect. In principle, however, the general effect sizes 'small,' 'medium,' and 'large' each leaves room for much variation. In this case Cohen's *d* = .78 for knowledge, and .51 for ethical decision-making. To not overburden users with too much details, specific info on exact effect sizes are omitted, with a few exceptions. One such exception is a difference as large as in the example above, with ethical decision-making *barely* yielding a medium effect size, and knowledge *almost* yielding a large effect size according to meta review [8]. In cases such as these, information about exact effect sizes is included in a comment, leaving it up to the user whether s/he would like to get more detailed information.

Regarding which criterion type is the most effective, a complicating factor should be mentioned. As the screenshot shows, conclusions about which criterion types are the most effective may differ between meta reviews, even if they focus on the same disciplines. For instance, whereas knowledge and ethical decision-making have a medium positive effect according to meta review [8], they have a small positive

effect according to meta review [9]. Both meta reviews focus on the same specific selection of academic disciplines: medical sciences, social science, and science & engineering.[23] An important difference between the two meta reviews is that whereas [8] examines characteristics of RCR education in isolation, [9] modelled multiple characteristics *simultaneously*. Looking at different characteristics simultaneously, meta review [9] found a (small) negative effect for moral reasoning. That is an interesting finding, given that the screenshot also makes clear that moral reasoning is the most used criterion in science & engineering, in business ethics education, and if we look at a combination of disciplines such as medical sciences, social science, and science & engineering. At the same time, however, the authors of meta review [9] urge caution with regard to interpreting their results on criterion types, given that 'they contradict prior meta-analytic work and other well-established notions in the ethics instruction and training literatures.'

Conclusions on most used and most effective criteria

The prototype tool allows users to compare criteria that are used most to criteria that are most effective. For instance, as the screenshot shows, ethical decision-making, a criterion with a medium positive effect in science & engineering according to meta review [8] is not in the top-3 of most used criteria in that domain. On the other hand, the criterion with the largest positive effect in that domain, knowledge, takes second place in the top-3 of most used criterion types in that domain. The results also show that, as far as specific criterion measures go, different versions of Defining Issues Test have rather diverging effects, with a (small) negative effect for the standard DIT according to meta review [9] and a large positive effect for a field-specific DIT according to meta review [8].

Examples like these were an important impetus to include in the current prototype tool: a) not just main lessons but detailed information on which those main lessons are based as well; and b) occasional notes on how specific conclusions should be interpreted. That way, the status of general conclusions remains visible, which to us seems important if they are input for decisions on key aspects of educational activities.

[23] Indeed, the meta reviews are based on the same underlying studies.

Trainee reactions

Lastly, the 'main lessons on the impact on learning aims' also contains a section on the use and efficacy of trainee reactions in the assessment of RCR education (top-right). Using trainee reactions for the assessment of RCR education means asking participants what they think about specific aspects of a given course. Meta review [11] focuses on trainee reactions and examines if specific type of trainee reactions show a positive correlation with performance criteria (Cohen's *d*). Interestingly, the screenshot shows that the most used trainee reaction to assess RCR education – content satisfaction – shows a negative correlation with performance criteria. This means that if an evaluation shows that students are satisfied with the content of a course, this is not a sign of that course's effectiveness. In other words, content satisfaction is not a good (indirect) measure for the efficacy of RCR education. This is different for 'course satisfaction,' 'knowledge/skills gained,' and 'content relevance,' trainee reactions that are among the most used reaction measures. According to meta review [11] these show a medium, weak and strong correlation with performance criteria. In other words, these specific reaction measures could be useful tools in assessing the efficacy of RCR education, albeit in an indirect way.

This concludes our introduction of the prototype tool of the Quality Checklist. Main lessons and details for all other key topics can be found via the "Overview" page. Readers who are interested in using the tool to perform quality checks on basic aspects of RCR education, are kindly referred to the Horizon 2020 project website: http://h2020integrity.eu/about-us/documentation/.

Literature

Meta reviews[24]

[1]] Alison L. Antes, Stephen T. Murphy, Ethan P. Waples, Michael D. Mumford, Ryan P. Brown, Shane Connelly & Lynn D. Devenport (2009). 'A Meta-Analysis of Ethics Instruction Effectiveness in the Sciences,' Ethics & Behavior, 19:5, 379-402, DOI: 10.1080/10508420903035380

[2] Waples, E.P., Antes, A.L., Murphy, S.T. et al. (2009). 'A Meta-Analytic Investigation of Business Ethics Instruction,' J Bus Ethics 87: 133. https://doi-org.proxy.library.uu.nl/10.1007/s10551-008-9875-0

[3] Antes AL, Wang X, Mumford MD, Brown RP, Connelly S, Devenport LD. (2010). 'Evaluating the effects that existing instruction on responsible conduct of research has on ethical decision making,' Acad Med Mar;85(3):519-26. doi: 10.1097/ACM.0b013e3181cd1cc5

[4] Logan M. Steele, Tyler J. Mulhearn, Kelsey E. Medeiros, Logan L. Watts, Shane Connelly & Michael D. Mumford (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, DOI: 10.1080/08989621.2016.1186547

[5] Medeiros, K.E., Watts, L.L., Mulhearn, T.J. et al. (2017). 'What is Working, What is Not, and What We Need to Know: a Meta-Analytic Review of Business Ethics Instruction,' J Acad Ethics 15: 245. https://doi-org.proxy.library.uu.nl/10.1007/s10805-017-9281-2

[6] Tyler J. Mulhearn, Logan L. Watts, Brett S. Torrence, E. Michelle Todd, Megan R. Turner, Shane Connelly & Michael D. Mumford (2017). 'Cross-Field Comparison of Ethics Education: Golden Rules and Particulars,' Accountability in Research, 24:4, 211-224, DOI: 10.1080/08989621.2016.1274974

[24] The numbers [1] to [11] correspond to the numbers of the meta reviews in the Quality Checklist prototype tool.

[7] Brett S. Torrence, Logan L. Watts, Tyler J. Mulhearn, Megan R. Turner, E. Michelle Todd, Michael D. Mumford & Shane Connelly (2017). 'Curricular Approaches in Research Ethics Education: Reflecting on More and Less Effective Practices in Instructional Content,' Accountability in Research, 24:5, 269-296, DOI: 10.1080/08989621.2016.1276452

[8] Logan L. Watts, Kelsey E. Medeiros, Tyler J. Mulhearn, Logan M. Steele, Shane Connelly & Michael D. Mumford (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, DOI: 10.1080/10508422.2016.1182025

[9] Logan L. Watts, Tyler J. Mulhearn, Kelsey E. Medeiros, Logan M. Steele, Shane Connelly & Michael D. Mumford (2017). 'Modeling the Instructional Effectiveness of Responsible Conduct of Research Education: A Meta-Analytic Path-Analysis,' Ethics & Behavior, 27:8, 632-650, DOI: 10.1080/10508422.2016.1247354

[10] E. Michelle Todd, Brett S. Torrence, Logan L. Watts, Tyler J. Mulhearn, Shane Connelly & Michael D. Mumford (2017). 'Effective Practices in the Delivery of Research Ethics Education: A Qualitative Review of Instructional Methods,' Accountability in Research, 24:5, 297-321, DOI: 10.1080/08989621.2017.1301210

[11] Megan R. Turner, Logan L. Watts, Logan M. Steele, Tyler J. Mulhearn, Brett S. Torrence, E. Michelle Todd, Michael D. Mumford & Shane Connelly (2018). 'How Did You Like This Course? The Advantages and Limitations of Reaction Criteria in Ethics Education,' Ethics & Behavior, 28:6, 483-496, DOI: 10.1080/10508422.2017.1308193

Other

National Academies of Sciences, Engineering, and Medicine (2017). Fostering Integrity in Research. Washington, DC: The National Academies Press. https://doi.org/10.17226/21896.

Shaw, D. (2019). 'The Quest for Clarity in Research Integrity: A Conceptual Schema,' Sci Eng Ethics 25: 1085-1093, doi: 10.1007/s11948-018-0052-2

Tauginienė, L, Gaižauskaitė, I, Glendinning, I, Kravjar, J, Ojsteršek, M, Ribeiro, L, Odiņeca, T, Marino, F, Cosentino, M, Sivasubramaniam, S. (2018). Glossary for Academic Integrity. ENAI Report 3G [online], http://www.academicintegrity.eu/wp/wp-content/uploads/2018/10/Glossary_revised_final.pdf

Steneck, N. H. (2006). 'Fostering integrity in research: Definitions, current knowledge, and future directions,' Sci Eng Ethics, 12(1), 53–74, DOI: 10.1007/pl00022268

Tyler J. Mulhearn, Logan L. Watts, E. Michelle Todd, Kelsey E. Medeiros, Shane Connelly & Michael D. Mumford (2017). 'Validation and Use of a Predictive Modeling Tool: Employing Scientific Findings to Improve Responsible Conduct of Research Education,' Accountability in Research, 24:4, 195-210, DOI: 10.1080/08989621.2016.1274886

Predictive Modeling Tool and supporting documents

Excel file: Ethics Predictive Modeling Tool_Final Version.xlsx

Walkthrough example: Ethics Predictive Modeling Tool_Final Version.xlsx

Appendix 1 – A brief comparison between the Predictive Modeling Tool and the Quality Checklist

The design for the prototype tool for the quality checklist was inspired by work by Mulhearn et al. (2017) who have developed and validated a "Predictive Modeling Tool" for RCR education (PMT from here on). The aims of the tool are ambitious, namely that it can be used to design, assess and improve education for the Responsible Conduct of Research. Developing the PMT was part of a larger project, involving and building on carefully organized meta studies on the efficacy of education for RCR (Watts et al. 2017a; Watts et al. 2017b) which resulted, among other things in a list of 77 variables or performance indicators for RCR education. These performance indicators are the core of the PMT. In brief, the PMT allows users to enter the characteristics of a RCR course. Employing 77 performance indicators the tool provides an overall score for the course, as well as more detailed scores for different performance categories. In addition, it provides recommendations on how the effectiveness of that course can be improved. According to the authors, the effectiveness of a course can often be greatly improved by making only a small number of changes. (Mulhearn et al., 2017)

One of the strentghs of the Pedictive Modeling Tool is that users can simply answer a range of questions on characteristics of a RCR course (existing or planned), and receive constructive feedback on what could be changed to make that course more effective. At first we thought to build the quality checklist out of the PMT, or that the PMT already presented the type of quality checklist that we were looking for, but on second thought, we came to the conclusion that it has some important disadvantages for the purpose of our H2020 project. Among the main disadvantages are 1) that the PMT does not ask about learning aims. True, the PMT does ask users to specify which specific criterion measures are used to assess the effectiveness of the course under examination, such as knowledge,

ethical decision-making and perceptions of self. (Mulhearn et al., 2017; Predictive Modeling Tool[25]) And specific criterion measures can help to gain insights into how the overall effectiveness of a RCR course can be broken down into how different criterion types (knowledge, moral reasoning, et cetera) contribute to the overall effectiveness of a course. Still, the use of specific criterion measures, by itself, does not necessarily provide information on which learning aims are central to that course. For it is possible that the course has learning aims that are not assessed. And it is possible that a course is assessed by using criterion types that do not match the learning aims that are central to the course under examination. Hence, use of the PMT provides a challenge for meeting a central quality criterion, namely that goals (should) come first, and that assessment should follow.

A second disadvantage of the Predictive Modeling Tool is that by providing scores and recommendations for any course the characteristics of which are filled-out in the PMT, it hides from view that it is based on studies that show what works in specific disciplines, and that what works may differ significantly between disciplines. True, all publications related to the development of the Predictive Modeling Tool, highlight their conclusions being based on studies from a selection of disciplines as a limitation, and a cause for caution with regard to interpreting the results, and particularly with regard to generalizing the results to other domains. (e.g., [8], [9], Mulhearn et al., 2017) Still, users of the PMT will receive a score for how effective their course and recommendations for improving their course which are based on a comparison to effective courses from a specific selection of academic disciplines, even if the course that is assessed by using the PMT is from a different academic field. Depending on the circumstances, users may receive recommendations that do not improve the efficacy of their course. In theory, it is possible that recommendations that have positive effects in one academic field, has a negative impact in another academic field. For instance, whereas meta review [9] found a small negative effect for moral decision-making being used as a criterion type in medical sciences, social science and science & engineering, meta reviews [2] and [5] found a (small to medium) positive effect of this criterion type being used in business ethics education.

[25] Here "Predictive Modeling Tool" refers to an Excel-based tool. See references. We thank prof. Mumford (University of Oklahoma) for kindly sending us the actual Excel-based tool, and accompanying documentation.

While the PMT may have advantages that the prototype tool for the Quality Checklist does not have, the Quality Checklist does not have the disadvantages of the PMT mentioned above.[26] Indeed, the Quality Checklist promotes that learning aims are put first, and that assessment follows. It does so by highlighting that quality is not determined solely by efficacy. If learning aims come first, then what might be needed is focusing on specific trainees, and focusing on promoting specific knowledge, skills, and attitudes (KSA), even if focusing on other trainees and other KSA's would yield larger effect sizes. More importantly, the Quality Checklist keeps front and center that the main lessons on the efficacy of RCR education are connected to specific studies, in specific disciplines, et cetera. For that very reason the Quality Checklist is explicitly positioned as a heuristic tool; as providing ingredients for quality checks on basic aspects of RCR education; as an inspiration for anyone interested in improving the quality of their RCR education. For a tool that is meant to serve and inspire those involved in RCR education in a broad range of academic disciplines, we think it is important to keep the conditionality of the main lessons front and center, and to provide users with easy acces to details (and variation) underlying those main lessons.

[26] We are preparing a manuscript in which we will offer a systematic comparison between the Predictive Modeling Tool and the prototype tool of the Quality Checklist.