

INSTITUTE
OF ECONOMICS



Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

**How should evaluation be? Is a good
evaluation of research also just? Towards
the implementation of good evaluation**

Cinzia Daraio ^a
Alessio Vaccari ^b

^a Department of Computer, Control and Management Engineering "Antonio Ruberti" (DIAG),
Sapienza University of Rome, Italy.

^b Department of Philosophy, Sapienza University of Rome, Italy.

2021/39

November 2021

ISSN(OBJECTIVE) 2284-0400

How should evaluation be? Is a good evaluation of research also just? Towards the implementation of good evaluation

Cinzia Daraio¹ and Alessio Vaccari²

¹ *Department of Computer, Control and Management Engineering “Antonio Ruberti” (DIAG), Sapienza University of Rome, Rome (Italy), E-mail: daraio@diag.uniroma1.it,*

² *Department of Philosophy, Sapienza University of Rome, Rome (Italy), E-mail: alessio.vaccari@uniroma1.it*

Abstract

In this paper we answer the question of how evaluation should be by proposing a good evaluation of research practices. A good evaluation of research practices, intended as social practices à la MacIntyre, should take into account the stable motivations and the traits of the characters (i.e. the *virtues*) of researchers. We also show that a good evaluation is also *just*, beyond the sense of *fairness*, as working on good research practices implies keep into account a *broader sense of justice*. After that, we propose the development of a knowledge base for the assessment of “good” evaluations of research practices to implement a questionnaire for the assessment of researchers’ virtues. Although the latter is a challenging task, the use of ontologies and taxonomic knowledge, and the reasoning algorithms that can make inferences on the basis of such knowledge represents a way for testing the consistency of the information reported in the questionnaire and to analyze in a correct and coherent way the data gathered through it.

Keywords: research evaluation, research practices, virtue ethics, good evaluation, fair evaluation, just evaluation, ontology-based modelling, questionnaire

1. Introduction

We live in an evaluation society (Dahler-Larsen, 2011): evaluations, quantifications and assessments are everywhere and analyse, scrutinize and monitor all kind of aspects of scholarly activities. In addition, in recent decades, the rapid changes taking place in the production, communication and evaluation of research have been signs of an ongoing transformation. Largely, we are facing a transition from a traditional evaluation model, based on indicators (e.g. number of publications and citations) to a modern evaluation, characterized by a multiplicity of distinct, complementary and new dimensions including the so called *altmetrics*. This situation is led by the development and increasing availability of data and statistical and computerized techniques for their treatment, including among others the recent advancements in artificial intelligence and machine learning.

Generally, moral responsibility is a typical human’s ability to discern, choose and operate in accordance with principles deemed to be of universal value (see Talbert 2019, for an articulated discussion on moral responsibility). Recently, *responsible research and innovation* have received increasing attention. A recent literature review (Burget et al., 2017), building on 235 articles analyzed, identifies four conceptual dimensions of responsible research and innovation that are: inclusion, anticipation, responsiveness and reflexivity; with in addition the 2 emerging dimensions of sustainability and care. Wilsdon et al. (2015) drawing on discussions over responsible research and innovation, propose the notion of *responsible metrics* as

“a way of framing appropriate uses of quantitative indicators in the governance, management and assessment of research. [...] Responsible metrics can be understood in terms of a number of dimensions: (i) *Robustness*: basing metrics on the best possible data in terms of accuracy and scope; (ii) *Humility*: recognising that quantitative evaluation should support – but not supplant – qualitative, expert assessment; (iii) *Transparency*: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results; (iv) *Diversity*: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system; (v) *Reflexivity*: recognising and anticipating the systemic and potential effects of indicators, and updating them in response (Wilsdon et al. 2015, 134-135)”.

Responsible research assessment according to Curry et al. (2020) “is an umbrella term for approaches to assessment which incentivize, reflect and reward the plural characteristics of high-quality research, in support of diverse and inclusive research cultures.” They list the following fifteen movements (including Wilsdon et al. 2015 at the point iii) that have influenced current debates on responsible research assessment:

- (i) DORA: The San Francisco Declaration on Research Assessment (<https://sfdora.org/>);
- (ii) The Leiden Manifesto for research metrics (<http://www.leidenmanifesto.org/>);
- (iii) The Metric Tide (<https://re.ukri.org/sector-guidance/publications/metric-tide/>);
The Independent Review of the Role of Metrics in Research Assessment Management;
- (iv) Science in Transition (<https://scienceintransition.nl/en/>);
- (v) Hong Kong Principles for Assessing Researchers (<https://www.wcrif.org/guidance/hong-kong-principles>);
- (vi) HuMetricsHSS (Humane Metrics Initiative, <https://humetricshss.org/>);
- (vii) INORMS Research Evaluation Working Group (<https://inorms.net/activities/research-evaluation-working-group/>);
- (viii) EC Open Science Policy Platform and Next Generation Metrics (<https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-policy-platform>);
- (ix) Science Granting Councils Initiative (<https://sgciafrica.org/en-za>);
- (x) Global Young Academy Working Group on Scientific Excellence (<https://globalyoungacademy.net/activities/optimising-assessment-promoting-excellence/>);
- (xi) Helsinki Initiative on Multilingualism in Scholarly Communication (<https://www.helsinki-initiative.org/en>);
- (xii) FOLEC: Latin American Forum on Research Assessment (<https://www.clacso.org/en/folec/>);
- (xiii) Science Europe Position Statement on Research Assessment Processes (<https://www.scienceeurope.org/>);
- (xiv) European University Association (EUA) Roadmap on Research Assessment in the Transition to Open Science (<https://eua.eu/resources/publications/316:eua-roadmap-on-research-assessment-in-the-transition-to-open-science.html>);
- (xv) Wellcome Trust’s campaign to Reimagine Research (<https://wellcome.org/what-we-do/our-work/research-culture>).

In this context, it becomes increasingly important to “evaluate evaluations” (Larson and Berliner, 1983) and discuss about how evaluation should be. Interesting questions on the Science of the 21st Century were posed by Sweeney (2021) during his keynote address at ISSI2021: “Are we starting from the right place? Do we need to look at where we’re going, not where we’re coming from? Do we understand enough about what Society needs/wants, and how our People and Research Culture will deliver it?”

Contrary to responsible evaluations and metrics that abound in the literature, in the current debate, looking for “good” evaluations is not a recurrent topic. Although there is a proliferation of increasingly sophisticated quantitative methods to evaluate research, there is still a lack of clarity on how to understand and operationalize the notion of “good” evaluation of research practices. In a previous recent work (Daraio and Vaccari, 2020) we recognize researchers’ motivations and their specific traits of character (i.e. researchers’ *virtues*) as important factors that must be considered to make a “good” evaluation. Good evaluation takes into account the constitutive elements of *research practices*, intended as *social practices*, according to the MacIntyre view (MacIntyre, 1985; Murdoch, 1998). Following this line, good evaluation must consider researcher’s virtues in the realization of the “internal goods” of the social practice they are involved in. Virtues allow the achievement of the goods of the practice and are appropriate to the different roles that researchers play according to their skills and experiences. In a good research practice, in fact, there are three main kinds of researchers involved: *leaders*, *good researchers* and *honest researchers* (Daraio and Vaccari, 2020). The main result of this recent study is the preliminary version of a questionnaire to assess the researcher’s virtues, to be able to complement current research evaluations with additional elements referring to scholar’s motivations and personal traits of character currently excluded by bibliometric indicators in use.

In this paper we maintain that evaluation must be good and investigate the connection between a *good* evaluation and a *just* evaluation. We show that by adopting the notion of good evaluation proposed in Daraio and Vaccari (2020) automatically lead us to a *just* evaluation that encompasses the *fairness* but goes also beyond it, relying also on a broader *general concept of justice*. Moreover, we propose the development of ontologies and taxonomic knowledge as a way for pursuing the implementation of the good evaluation through a questionnaire on researchers’ virtues.

The paper unfolds as follows. The next section illustrates the main aim and contribution of this paper. Section 3 describes the methods and materials of the paper, arguing that a good evaluation implies a just evaluation. Section 4 presents the main challenges and first steps towards the implementation of good evaluation describing an ontology-based modelling of research virtues. Section 5 summarizes and concludes the paper.

2. Aim and contribution

In this work, we propose to adopt a framework for the evaluation of research based on a good evaluation of research practices. We think that considering the good evaluation could be useful for *concretely declining*, in real research practices, the principles and statements postulated by the existing literature on responsible evaluation which otherwise would remain vague and indefinite.

We build on the notion of good evaluation introduced in our previous research (Daraio and Vaccari, 2020) and extend the discussion on it deepening two important points: i) the virtues of researchers that are the cornerstone of the good evaluation, and ii) showing that a good evaluation is also morally just.

Moreover, we propose an advancement towards the implementation of good evaluation in practice. It is an ontology-based modelling approach to represent the domain of researchers’ virtues in order to prepare a questionnaire to be administered to researchers. This approach will allow us 1) to check the consistency and the coherence of the questionnaire content and structure before its application (or use) and 2) to correctly and coherently interpret the data gathered through the questionnaire.

3. Methods and materials

The methods applied in this paper are i) philosophical argumentation and ii) ontology-based modelling for designing a questionnaire on researchers' virtues. By applying philosophical argumentation we deepen the basis of good evaluation discussing the virtues of researchers (see Section 3.1) and show that a good evaluation of research practice is also morally just (see Section 3.2). The main ideas of an ontology-based modelling of researchers' virtues are described in Section 3.3, while Section 4 illustrates the first steps towards the implementation of good evaluation.

3.1 The cornerstone of the good evaluation: The virtues of researchers

Let's start with a general characterization of the nature of virtue. Virtues are dispositions to believe, to feel emotions and act in certain ways that are activated when we perceive some relevant characteristics in the world. A courageous person, for example, will face danger firmly when she believes that someone or something to which she attaches a value is at risk of being harmed. Similarly, a charitable person when she believes that someone is suffering will tend to sympathize with that suffering and, consequently, take care of it. Furthermore, an accurate person is one who does not accept any belief-shaped thing that enters his or her mind, but takes special care in forming his or her beliefs. Actions of these type will not be singular but will be stable and predictable: whenever she perceives certain relevant characteristics of a situation, the virtuous agent will tend to give the appropriate response to those characteristics. The possession of virtue moreover seems to be something that admits degrees of competence along a spectrum that goes from knowledge about what to do to a more complex one about why to do a certain thing. This element depends on two factors that concern our practice of attributing virtuous traits:

- 1) we sometimes attribute a virtue even when those who perform the corresponding action are not able to give an articulated justification of what they have done and/or their actions are not cross situationally consistent.
- 2) we believe that those who give an articulated justification of their virtuous actions, and tend to manifest the virtuous trait in a coherent plurality of situations, have a greater knowledge of virtue than those who do not. Unlike the former, these agents tend to know how to distinguish circumstances that require a virtuous response from similar ones that do not.

An entirely brave person, for example, will tend to discriminate between the real danger from a mere apparent threat and to avoid facing danger just for the sheer pleasure of the adrenaline that follows. It will also be able to recognize situations that require a courageous response and to be motivated accordingly in a variety of situations: not only those involving physical confrontation, but also those involving, for example, the defense of an unpopular idea or pursue a complex line of research never explored before. In a similar way, an entirely accurate person takes into account that research is both time- and resource-consuming and is able to put different investment policies in place depending on circumstances. These agents seem to possess greater knowledge of the virtue of those who simply act virtuously in a limited number of cases and lack the ability to formulate a justification for what they do. They are able to articulate the reasons in favour of virtuous behaviour in a plurality of contexts and use them to justify their conduct to themselves and others. For those who possess this knowledge virtue will not be a mere disposition to act, but a disposition to act infused by a reflective ability that involves the mastery of concepts. As with other character traits, finally, virtues are *profound qualities* that reveal what kind of person is the one who possesses them. The attribution of character traits is in fact a common practice that we use as much to get a general idea of someone

we do not know well as, if we know him better, to predict what he will say or do or to explain why she has made certain choices in the past. Moving from the spectator's point of view to that of the agent, there are facts that are typically explained (retrospectively) by referring to motives that depend on general principles of our conduct that survive that individual case. Although not all of these principles are something we necessarily approve of, some of them are aspects of our character that are valued by ourselves and others and that we strive to maintain, creating opportunities to test them and reflect on how we express them in our behaviour. In these cases, virtue becomes a crucial element of our narrative and plays a normative role in our future choices.

In our research we are going to use two different types of virtues: the *intellectual* or *epistemic virtues* and the *virtues of character*. The former involve dispositions to exercise a set of capacities that relate to how we acquire our beliefs and communicate them to others. The latter, on the other hand, concern functional ways in which we enter into relationships with ourselves and interact with others by promoting our own and their good. The former are developed primarily through teaching by specialized staff (professors; research directors; etc.). The latter, on the other hand, require having entered into a non-emulative educational relationship with figures who have the role of educators in the community (parents, teachers, etc.).

Below we provide an open-ended list of the virtues that enable the constitution of a good research practice and constitute something that should be taken into account by those making comparative judgments about the relative value of different research practices in addition to existing and currently used individual bibliometric indicators. For an extensive discussion on these individual bibliometric indicators, see Schubert and Schubert (2019) and Wildgaard (2019). Schubert and Schubert (2019) offer an overview on h-index related indicators while Wildgaard (2019) presents a summary of existing author-level indicators of research production.

The virtues that enable the constitution of a good research practice include:

a) Intellectual virtues:

Accuracy: it is the disposition that consists in the care with which the individual researchers collect data that will constitute the pool of information shared in the research practice. Since collecting and evaluating information is time and resource consuming, the accurate researcher is one who implements several "policies of investigation" that are appropriate to the research circumstances (Williams, 2002).

Sincerity, Honesty: it is the disposition to tell the truth to others and, when this does not happen, it is the capacity to indicate good reasons why this did not happen where good refers to the fact that these reasons have a constitutive reference to the interests of other people (McIntyre 1985, Williams 2002).

Creativity: is the ability, which finds expression both in our social interactions with others and in the products of our research, to produce something that not only has value but is characterized by the elements of novelty and the capacity to arouse surprise in others (Swanton 2003, pp. 162, 165).

b) Virtues of character useful to oneself:

Humility: it is the ability to accept the authority of the standards related to the rules that define the practice. I have to recognize that other participants know rules and know how to apply them better than I do. I have to be willing to learn from these people and accept their criticism (MacIntyre 1985, p. 193).

Pride: it is evaluative attitudes towards ourselves (Ardal, 1966; Cohon, 2008; Taylor, 2015). Unlike other emotions, which simply motivate us to pursue or avoid objects, this traits of character fix our attention on persons, casting a positive or negative light on them. If I am proud of my child's success at school, my pride does not fix my attention on the 'merits of my child,' and still less on 'me in the role of father,' but on the whole of myself. As Cohon has rightly said, "when I feel pride, I am proud of something in particular [its cause] ... But the attitude of pride is a pleasure or satisfaction not in that particular accomplishment or possession, but in myself in my entirety" (Cohon, 2008, p. 166). We believe that the pride associated with one's own achievements in research and the consequent approval of one's peers or superiors is a fundamental spring that drives researchers to perform at best in their area of research (Tangney, 1999). Pride is then a synonym of self-respect, dignity, honour, self-esteem and self-worth.

Patience is the ability to curb one's own urge to complete a research in order to obtain as soon as possible the gratification of a positive result. To be able to wait and to be guided by a cautious skepticism that prompts us to control accurately the different steps of our investigation.

Prudence: is the capacity to sacrifice the satisfaction of less important pleasures closer in time than the satisfaction of more distant but more important pleasures. Where the degree of importance is defined with respect to the long-term objectives that characterize our lives (Parfit 1984).

Resilience: together with pride, this ability is indispensable to move forward in the research. It allows us to leave behind failures (paper rejected, unfunded projects, etc.) and to focus on future projects (Hormann, 2018).

c) *Virtues of character useful to others*

Courage is the capacity to risk damage or danger to oneself when individuals, values, goals that are crucial to the existence of the practice are at stake. Courage is therefore a way of showing that our attachment to these elements of the practice is genuine (MacIntyre, 1985, p. 192).

Empathy, Benevolence: in line with extensive literature, by this term we mean the human ability to feel the emotions and feelings of other people through a vicarious feeling that is similar to that of the person with whom we sympathize. We do not believe, however, that empathy in itself is a virtuous capacity in research practices. Since empathy is an instrument for reading the other's mind, it can also be used to manipulate other researchers in malicious ways. Empathy must be cultivated in such a way that it is rooted in the benevolent tendencies of human beings (Batson 2017, p. 2). In this way, empathy can allow the creation of a climate of *trust* between those who work within research institutions. Indeed, mutual trust is an indispensable component in these practices given the fundamental fact of the asymmetry of power that characterizes those interactions (Baier, 1991).

Integrity: is the willingness to behave in such a way that our actions are the outcome of our deepest values and commitments and that we tend to refuse making them hostages to imposed obligations or duties that we do not endorse on reflection.

Justice: following Aristotle (Aristotle 2014, Book V), we distinguish two senses of justice: *general justice* and justice as a particular virtue. The former indicates the ability to understand and apply the good rules that guarantee the good functioning and survival of a particular institution and research practice. This capacity typically includes creative knowledge: applying a rule means being able to apply it correctly to new cases that were not initially foreseen when the rule was formulated, or to complex cases that seem to involve several conflicting rules. Possessing the virtue of general justice means being able to make this creative use of rules out of a reflection on the aims of research institutions or practices, i.e. on their internal and external goods.

The second sense of justice is about encompassing our sense of justice as *fairness*. This aspect of the virtue of justice concerns the individual's attitude not to demand more from the institution in which one works than what is due to us in view of the role one occupies in the research practice. Following the taxonomy we have introduced in our research (leader, good researcher and honest researcher), this means that a *good* researcher should not expect to receive what a *leader* receives just as an *honest* researcher cannot expect to receive what is due to a *good* researcher and a *leader* respectively.

Practical wisdom: it is a kind of *super-virtue* essential to make each virtue effective. In line with Aristotle, we believe that this rational capacity enables the virtuous agent to acknowledge and respond properly to the items in the field of the research practice, choosing the appropriate means for their own ends (McDowell, 1979). Moreover, it also allows the different virtues within an individual's character to operate and develop harmoniously with each other.

3.2 A good evaluation is also morally just

In order to show that a good evaluation is also morally just, three different conceptions of justice will be examined: 1) the deontological conception; 2) the utilitarian conception; 3) the neo-aristotelian conception.

The basic idea of the *deontological conception* of justice is that human relationships must be fair. Justice according to Rawls (1971) is therefore a set of principles that constitute constraints on each individual's pursuit of his or her individual goals. In this sense, justice is therefore a set of rights and opportunities that ensure that each person pursues his or her individual aims while respecting the same pursuit for other members of the community (Rawls (1971, p. 50). According to this conception, the just society is composed of rational, autonomous, equal, and independent individuals who treat each other with respect and who are governed by principles that protect this type of mutual relationship.

The *consequentialist perspective*, on the other hand, considers justice from the point of view of the overall consequences of actions. The most influential version of consequentialism, i.e. *utilitarianism*, has two components: a theory of good and a theory of just action. The first holds that pleasure is the only property that has value. The conception of right holds that the right action is the one that maximises the pleasure or satisfaction of the preferences of all individuals affected by that action. According to utilitarianism, the achievement of equity is compatible with situations where human beings are still in a state of great suffering. The goal of justice should be to promote the development of each individual's capacities, and thus happiness, and not equity in itself. According to Sen (2010), for example, a fair distribution of resources may not be sufficient to achieve this goal. A person with a walking disability might need more wealth or a higher share of state services to be able to move around as a person without a disability. According to utilitarianism, the achievement of equity is compatible with

a situation in which human beings are in a state of great suffering. The task of justice should be to promote the development of each individual's capacities, and thus overall happiness, and not equity in itself. Moreover, as Singer (2003) has pointed out, increasing inequality could lead to a significant improvement in the conditions of those who are worse off. This outcome, for people living in conditions that are just above the absolute minimum necessary to live can make a considerable difference (on this point, see Donatelli, 2015). According to the utilitarian conception, therefore, a just society is not defined primarily by the existence of a certain type of relationship between individuals, but by the fact that people are more or less happy.

According to a richer and more sophisticated utilitarian version, however, originally formulated by Mill (see Mill, *On Liberty*, in Mill and Ryan, 1997), justice is to be understood in a broader way that includes the protection of the enduring interests of human beings as 'progressive beings'. In this sense, justice is closely related to the rights of freedom, not only the negative ones, which protect the individual sphere from state interference, but also those rights that concern instead the promotion of material conditions so that individuals can actually freely choose their own life projects. Mill thus links justice to respect for the so-called perfect rights of the *Natural law* tradition, that is, to that class of rights with respect to which each individual possesses a "valid claim on society to protect him in the possession of it, either by the force of law, or by that of education and opinion."

Finally, the third, which emerges within the communitarian tradition and has been famously advocated by Sandel (2010) draws directly on the Aristotelian view of justice. Like Mill's, this conception presents a broad notion of justice that is not confined to the realm of the distribution or redistribution of goods in society, but instead inquires more broadly about what kinds of goods should be promoted by the different institutions on which society is founded. Unlike some versions of the Kantian-style deontological conception, which tend to be neutral on ethically sensitive issues, it promotes an ongoing public debate about different conceptions of the *good life* that can result in the adoption and enactment of ethically oriented laws. In relation to what most directly affects this paper, this conception holds that the identification of the multiple goods that should be distributed among citizens must be based on a shared conception of the functions of social institutions on which society is based. Identifying which goods education or public health should promote presupposes a conception of the function of hospitals, schools and universities. A conception that, in turn, is also connected to a view regarding what virtues are respected and honored in those institutions or practices.

Sandel's neo-Aristotelian conception is a fruitful approach to responding to distributional problems that arise within research practices and their comparative evaluation. Following Aristotle, we distinguish between the virtue of *general justice* and the virtue of *justice*. Both are qualities of the character of researchers.

Let us dwell on the first. This indicates not only the ability to abide by the general rules that enable research practices to stand and flourish, but also the ability to identify behaviors that are appropriate to those rules in the presence of circumstances that either were not anticipated by those rules or are too complex to be addressed by simply applying those rules mechanically. The researcher who possesses the general justice is able to identify appropriate actions that extend the content of the rules governing the proper functioning of a practice to new cases. This is possible because the virtuous researcher is able to go back and reflect on the constitutive ends of each research practice or institution, i.e., its internal and external goods, and based on that reflection indicate the appropriate actions to solve the concrete issue under discussion.

Justice as a special virtue has a different but equally important role. Still following Aristotle we can identify this virtue with the ability to counteract our desire to receive more than our *fair share*. Respect for this virtue, underlies the very structure of work in good research practice which is organized between leader, good and honest researchers. This distinction is

maintained by assuming that each individual accepts what is due to him or her in his or her role in research practice. And that is why this second virtue of justice is so important.

The inclusion of the virtue of justice, in the two senses indicated, among the virtues of researchers makes it possible to identify a new potential characteristic in good research evaluation. We have argued elsewhere (see Daraio and Vaccari, 2020) that good research evaluation is one that takes into account the virtues of researchers, understood as their ability to excel in the different spheres of activity that characterize the internal and external goods of research practices. As we have shown, these virtues must also include the virtue of justice. The content of this virtue depends on the ends, the attainment of which constitutes the good of the practice. This then means that a good evaluation is also a just evaluation, since an evaluation that takes into account the virtues of researchers cannot fail to take into account i.e., measure, the degree to which the virtue of justice is present in the components of research practice.

3.3 Designing an ontology for the modelling of researchers' virtues

Formally, an ontology in Description Logics is a knowledge base. It is a couple (pair) $O = \langle TBox, ABox \rangle$, where TBox is the Terminological Box that represents the *intensional* level of the knowledge or the conceptual model of the portion of the reality of interest expressed in a formal way; and ABox is the Assertion Box that represents the *extensional* level of the knowledge or the concrete model of the portion of the reality expressed by means of assertions (instances).

The use of ontology-based modelling in our context allows us to implement *cognitive interviewing methodology* to address the challenges outlined in the previous section. Cognitive interviewing is a psychologically oriented method for empirically studying the ways in which individuals mentally process and respond to survey questionnaires. Cognitive interviews can be conducted for the general purpose of enhancing the understanding of how respondents carry out the task of answering survey questions. However, the technique is more commonly conducted in an applied sense, for the purpose of pretesting questions and determining how they should be modified, prior to survey fielding, to make them more understandable or otherwise easier to answer. The notion that survey questions require thought on the part of respondents is not new and has long been a central premise of questionnaire design. However, cognitive interviewing formalizes this process and it has become an interdisciplinary field (for an overview, see Willis, 2004 and Miller et al., 2014).

An ontology-based semantic modelling approach offers several advantages, including:

- i. a conceptual specification of the domain of interest, in terms of knowledge structures;
- ii. the mapping of such knowledge structures to concrete data (the answers of the questionnaire);
- iii. the reasoning over the abstract representation of the domain prior to the data collection;
- iv. a flexible conceptual system that can be easily updated;
- v. an open conceptual system that can be used as a common language for the research community.

The languages for representing ontologies and taxonomic knowledge, and the reasoning algorithms that can make inferences on the basis of such knowledge have been addressed by a large body of research in Artificial Intelligence and Knowledge Representation. Their formal characterisation is nowadays based on Description Logics (Baader et al., 2003), which provides

a syntax for concept and role expressions and formal semantics to interpret them in a set theoretic framework. Concepts (i.e. classes) model sets of individuals and roles model binary relations. The representation of ontological knowledge is thus achieved by defining concepts and the properties (relations) that link them to other concepts in the domain of interest. The concepts are arranged in a hierarchical structure based on the subsumption relation (i.e. set containment). Along with the formal language, systems that allow to model and use ontologies are accompanied by various forms of syntactic representation, including graphical models. Protegé (Gašević et al., 2009) is a standard tool that builds its success, among other things, on the capability to handle multiple syntactic representations that allow the user to model the domain of interest using the most convenient notation, while grounding it to a well understood formal counterpart. Another key feature of Protegé is the decoupling of the representation from the reasoning tool that is adopted to make inferences. Protegé, for all the reasons explained above, will be used for the development of the ontology for the assessment of researchers' virtues.

4. Towards the implementation of good evaluation

4.1 Challenges and first steps

The assessment of researchers' virtues is challenged by several pitfalls and problems. The first issue is related to the question of the measurement of virtues. Virtue, it has been argued, seems to consist of a special sensitivity that escapes empirical measurement (Murdoch, 1998). Recently, however, some scholars have tried to undermine this pessimist assumption by giving hope to those of us who seek to develop a model of evaluation of research including also a component based on the virtues. Snow (2014) has recently launched a promising line of research based on the elements of psychology that characterize the virtues. In its perspective *virtue* is composed of the following three elements:

- 1) *intelligence*, which highlights the fact that virtue proceeds from a set of cognitive and emotional mental states that enable us to be sensitive to some morally relevant features of the situations in which, really or imaginatively, we find ourselves (Snow, 2014, pp. 4-5). See also Snow (2010 and 2012);
- 2) *dispositionality*, refers to the fact that this state is a trait of the personality of the agent and is not an occasional element of his psychology;
- 3) *behaviour*, i.e. virtue typically manifests itself in the actions and other behavioural responses of the virtuous person (Snow, 2010, pp. 4-5).

Snow argues that each of these characteristics of virtue can be measured and she outlines a model that consists of three measurement criteria. First, the agent's performance must be taken into account, i.e. the presence of the virtue in question must be verified from the agent's ability to repeatedly perform a given behavioural pattern in the different situations that constitute, so to speak, the field of action of a specific virtue. Secondly, Snow believes it is crucial to take into account the reports that agents make of their emotional and cognitive life during the performance of actions that they consider virtuous. To facilitate this task, Snow believes it is desirable that, on the model of some US colleges, research institutions make available to their participants' special apps that can be downloaded on any electronic device, allowing them to collect the results of the self-observations of agents. Gathering the products of introspection, in addition to offering a useful material to those who are called to assess the presence of virtues in others, also allow agents to take into account the health of their virtues and measure any flexing or, on the contrary, increases in their readiness and effectiveness in

responding to the pressures the world exerts on them. Finally, Snow argues that it is important to connect these data with those that impartial observers, in the figure of external evaluators, can collect in the course of annual surveys covering both the outputs of the research and the way in which the researcher dwells in different spheres of social interaction with other participants in the practice.

A further problem to be addressed is which questions to introduce in the questionnaires. These must be sufficiently diversified to allow the evaluators to answer not only the dry question about whether or not there is a virtue, but also to determine the *quantum* of it. Snow suggested four levels to be introduced in the questionnaires.

- I. The first verifies the presence in the agent of receptivity to the *stimulus* that typically activates virtue.
- II. The second examines its ability to recognize the virtue appropriate to the given circumstance.
- III. The third verifies the most complex ability to generate a virtuous response.
- IV. The fourth, finally, measures the ability of the agent to generate the virtuous response in a cross-cutting way to a plurality of situations.

Following the four levels questions introduced by Snow, it is possible to measure on a scale from 0 (minimum) to 4 (maximum) the researcher's mastery of virtue. This is done over a spectrum ranging from (1) the ability to understand the importance of the problem to which virtue constitutes an answer, to (2) the ability to recognize the virtue in question, to (3) the ability to express virtue occasionally, to (4) the ability to manifest it in all situations that constitute the scope of that virtue.

On top of these problems, the traditional problems related to the development of questionnaire and the collection of the necessary information through questionnaire and interview arise (Hochschild, 2009; Rabionet, 2011; Kvale, 2008 and Wolcott, 2008).

In this paper we suggest to exploit the advantages of an ontology-based modelling that we will illustrate after the next section, for overcoming the aforementioned problems.

4.2. Semantic modelling of the virtues of researchers

The starting point for the semantic modelling of the domain in exam are the virtues of researchers. A first attempt to develop a questionnaire for the evaluation of virtues in research practices has been done by Daraio and Vaccari (2020). Our Table 1 below, elaborating further Table 1 of Daraio and Vaccari (2020, p. 1067-1068), proposes some examples of questions to consider in evaluating the virtues of researchers.

Table 1. Examples of questions to include in the Questionnaire on Researchers' Virtues

a) Intellectual Virtues	Questions
<i>Accuracy</i>	<ul style="list-style-type: none"> - Do you thoroughly collect all the pieces of information that constitutes the body of knowledge on which the practice revolves? - Do you evaluate this disposition as instrumental or intrinsic? - Do you have a tendency to share data, results, methods, ideas, techniques, and tools used in the practice?
<i>Sincerity, Honesty</i>	<ul style="list-style-type: none"> - Do you think there are circumstances in which your colleagues can be manipulated? - Are you inclined to admit publicly when you make mistakes?
<i>Creativity</i>	<ul style="list-style-type: none"> - Are you able to explore and follow your own line of research? - Do you have the tendency to question your own ordinary experience and look with suspicion at what is the result of habit?
b) Virtues of character useful to oneself	Questions
<i>Humility</i>	<ul style="list-style-type: none"> - Are you incline to recognize that other researchers (participants to the practice) know rules and know how to apply them better than you do? - Are you to be willing to learn from these people and accept their criticism?
<i>Pride</i>	<ul style="list-style-type: none"> - Do you have the ability to feel fulfilment for academic success through demonstrating competence according to social standards and to draw strength from one's achievements? - Do you think you have a stable awareness of your own value that is not shaken by the successes of others? Do you have the capacity to enjoy and congratulate other people's accomplishments?
<i>Patience</i>	<ul style="list-style-type: none"> - Are you able to curb the rush to hastily complete a search to achieve the gratification that comes with a <i>prima facie</i> positive result? - Are you willing to be guided by a cautious scepticism that prompts you to control accurately the different steps of our investigation?
<i>Prudence</i>	<ul style="list-style-type: none"> - What do you think about people who tends to restraint of actions, inclinations, and impulses that are likely to upset others? - Do you tend to respect the fundamentals of the research practice within which one works?
<i>Resilience</i>	<ul style="list-style-type: none"> - Would you describe yourself as a person who leaves behind failures (paper rejected, unfunded projects, etc.) and to focus on future projects?

c) <i>Virtues of character useful to others</i>	Questions
<i>Courage</i>	<ul style="list-style-type: none"> - Are you willing to risk damage or danger to oneself when individuals, values, goals that are crucial to the practice are at stake? - Do you have a propensity to apply for highly competitive grants?
<i>Empathy, Benevolence</i>	<ul style="list-style-type: none"> - How do you feel about people who are sensitive to the suffering of their colleagues caused by failures or exclusions? - How do you feel about people who seeks feedback to improve interactions with others? - Do you think that preservation and promotion of the welfare of people with whom one is in frequent personal contact is important? - How would you describe collegial engagement?
<i>Integrity</i>	<ul style="list-style-type: none"> - Do you think that to hold beliefs that are consistent with actions has always a positive value?
<i>Justice</i>	<ul style="list-style-type: none"> -Do you tend to think that the rules defining good research practice are outdated? Do you find them adequate to deal with the new issues posed by more advanced research? Do you accept the principle that you receive what is due to your role in research practice? Do you tend to feel under-determined when this happens?
<i>Practical wisdom</i>	<ul style="list-style-type: none"> - Would you describe yourself as someone who tends to choose the most effective means to achieve their ends? if not, why? - Have you ever been in a situation of conflict between things that have identical value to you? If so, what considerations do you turn to in an attempt to resolve the conflict?

Note: The content of this table is an elaboration of the content of Table 1 by Daraio and Vaccari (2020, pp. 1067-1068).

4.3 Related works

While there is a rich literature and several approaches for extracting information through ontologies, for a review see Wimalasuriya and Dou (2010), the literature on ontological modelling in support of questionnaires development is scant. Notable exceptions include Sherimon et al. (2013 and 2014) where an ontology-based model for gathering the patient medical history based on dynamic questionnaire ontology is developed. The model is implemented and explained for diabetes domain by using Protegé. Another interesting contribution is Borodin and Zavyalova (2016), in which the authors focused on the problem of semantic representation of questionnaires. They constructed the generic ontological model of questionnaire, which provides a possibility of question structure description including complex questions with a set of answers of different kind and question order, including skipping and branching.

Surveys in fact, may be conducted to gather information through a printed questionnaire, over the telephone, by mail, in person or on the web, etc. and the structure of survey questionnaires and feedback if described in ontological terms provides an opportunity not only to structure the survey data but also to analyse the responses.

We contribute to enrich the limited existing literature, showing the potential of ontology based modelling for challenging and intriguing topics as the “good” evaluation of research practices, that relates the development of a questionnaire for the assessment of the motivations and the stable threats of characters (or virtues) of researchers.

5. Concluding remarks

In this contribution we have elaborated a model of good research evaluation. The model argues that good evaluation must take into account the three characteristics that constitute good research practice. These are the virtues of the researchers, the internal and external goods of the practice and the division of the components of the research practice into *leader*, *good* and *honest* researchers. In this paper we also argue that good evaluation of a research practice is also just evaluation. This depends on the fact that good evaluation will also have to take into account the virtues of justice of the researchers. We have argued that the virtue of justice is not reducible to fairness and includes two capacities, that of being able to creatively and extensively interpret the rules that underlie a research practice and the ability not to ask for a greater share of goods than we are entitled to by the role we play in the practice. Based on the conceptual tools used by normative ethics, in particular from the perspective known as ethics of virtues, we have started to develop a questionnaire capable of revealing the presence and quantity of virtues in individual researchers. We believe that the use of ontology-based modelling in this context might enable us to further implement a cognitive interviewing methodology which may help us to address the many challenges of this research field.

Acknowledgments

The financial support of the Sapienza University of Rome (through the Sapienza Awards no. PH11715C8239C105 and no. RM11916B8853C925) is gratefully acknowledged. This paper is a substantially extended version of the paper by Daraio and Vaccari (2021) presented at the 18th International Conference on Scientometrics & Informetrics (ISSI2021), held virtually in Leuven on 12–15 July 2021. In this extended version, we added the discussion on how evaluation should be, the explanation that a good evaluation implies a just evaluation that goes beyond the concept of fairness and include a broader sense of justice and extended the part on justice in the implementation section about the questionnaire.

References

- Ardal, P. S. (1966). *Passion and Value in Hume's Treatise*. Edinburgh: Edinburgh University Press.
- Aristotle (2000, revised edition 2014). *Nicomachean Ethics*, R. Crisp (Ed.), Cambridge, Cambridge University Press.
- Baader, F., Calvanese, D., McGuinness, D., Patel-Schneider, P., Nardi, D. (2003) (Eds.). *The description logic handbook: Theory, implementation and applications*. Cambridge university press.
- Baier, A. (1991). *A Progress of Sentiments: Reflections on Hume's Treatise*. Harvard: Harvard University Press.
- Batson C D (2017) The Empathy-Altruism Hypothesis: What and So What? In Emma M. Seppälä, Emiliana Simon-Thomas, Stephanie L. Brown, Monica C. Worline, C. Daryl Cameron, and James R. Doty (eds.), *The Oxford Handbook of Compassion Science*, Oxford, Oxford University Press.

- Borodin, A. V., & Zavyalova, Y. V. (2016) An ontology-based semantic design of the survey questionnaires. In 2016 19th Conference of Open Innovations Association (FRUCT) (pp. 10-15). IEEE.
- Burget, M., Bardone, E., Pedaste, M. (2017). Definitions and conceptual dimensions of responsible research and innovation: A literature review. *Science and engineering ethics*, 23(1), 1-19.
- Cohon, R. (2008). *Hume's Morality: Feeling and Fabrication*. Oxford: Oxford University Press.
- Curry, S., de Rijcke, S., Hatch, A. et al. (2020). The changing role of funders in responsible research assessment: progress, obstacles and the way ahead. Working Paper. Research on Research Institute (RoRI) <https://doi.org/10.6084/m9.figshare.13227914.v1>.
- Dahler-Larsen P. (2011), *The Evaluation Society*, Stanford Univ. Press, Stanford.
- Daraio C., Vaccari A. (2021), An Ontology-Based Semantic Design for Good Evaluations of Research Practices, 18th International Conference on Scientometrics & Informetrics (ISSI2021), 12–15 July 2021 Proceedings, edited by W. Glänzel, S. Heeffer, P.-S. Chi and R. Rousseau, ISBN 9789080328228, July 2021, pp. 305-316.
- Daraio C. Vaccari A. (2020) Using normative ethics for building a good evaluation of research practices: towards the assessment of researcher's virtues, *Scientometrics*, 125, 1053-1075.
- Donatelli, P. (2015), La giustizia fra deontologismo e consequenzialismo, *Parolechiave* 1 (2015), 83-96.
- Gašević D.; Djurić D.; Devedžić V. (2009). *Model Driven Engineering and Ontology Development* (2nd ed.). Springer.
- Hochschild, J. L. (2009) Conducting intensive interviews and elite interviews. In Workshop on interdisciplinary standards for systematic qualitative research. National Science Foundation.
- Hormann, S. (2018). Exploring Resilience: in the Face of Trauma. *Humanistic Management Journal*, 3(1), 91–104.
- Kvale, S. (2008) *Doing interviews*. Sage.
- Larson, R. C., Berliner, L. (1983), On evaluating evaluations. *Policy Sciences*, 16(2), 147-163.
- MacIntyre, A. (1981 first ed., 1985) *After Virtue*, London, Duckworth.
- Murdoch, I. (1998) *Existentialists and Mystics: Writings on Philosophy and Literature*. Allen Lane/the Penguin Press.
- McDowell, J. (1979). Virtue and Reason. *The Monist*, 62(3), 331–350.
- Mill, J. S., and Ryan A. (1997). *Mill: Texts, Commentaries*. New York; London: Norton, Print. Norton Critical Edition.
- Miller, K., Chepp, V., Willson, S., & Padilla, J. L. (Eds.). (2014). *Cognitive interviewing methodology*. John Wiley & Sons.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Rabionet, S. E. (2011) How I Learned to Design and Conduct Semi-Structured Interviews: An Ongoing and Continuous Journey. *Qualitative Report*, 16(2), 563-566.
- Rawls, A. (1971). *Theory of Justice*. Original ed. Cambridge, Mass. Web.
- Sandel, M. J. (2010). *Justice: What's the Right Thing to Do?* 1st Pbk. ed. New York.
- Sen, A. (2010). *The Idea of Justice*. London: Penguin.
- Sherimon P.C., Vinu P.V., Krishnan R., Takroni Y., (2013) Developing a Survey Questionnaire Ontology for the Decision Support System in the Domain of Hypertension, IEEE South East Conference, April 4-7, Florida, U.S.
- Sherimon, P. C., Vinu, P. V., Krishnan, R., Takroni, Y., AlKaabi, Y., & AlFars, Y. (2014) Adaptive questionnaire ontology in gathering patient medical history in diabetes domain. In Proceedings of the first international conference on advanced data and information engineering (DaEng-2013) (pp. 453-460). Springer, Singapore.
- Singer, P. (2003). *One World. L'etica della globalizzazione*, Einaudi, Torino.

- Schubert, A., & Schubert, G. (2019). All along the h-index-related literature: A guided tour. In *Springer Handbook of Science and Technology Indicators* (pp. 301-334). Springer, Cham.
- Snow N. (2014) Virtue intelligence, unpublished paper presented for the conference “Can Virtue Be Measured?” held by the Jubilee Centre for Character & Value, 9th – 11th January 2014, <https://www.jubileecentre.ac.uk/485/conferences/can-virtue-be-measured>.
- Snow, N. (2010) *Virtue as Social Intelligence: An Empirically Grounded Theory*, New York, Routledge.
- Snow N. (2012) Notes Toward an Empirical Psychology of Virtue: Exploring the Personality Scaffolding of Virtue, in *Aristotelian Ethics in Contemporary Perspective*, ed. by J. Peters, New York, Routledge, pp. 130-144.
- Swanton, C. (2003). *Virtue Ethics: A Pluralistic View*. Oxford: Clarendon Press.
- Sweeney D. (2021). The Science of 21st Century Science, Keynote at ISSI2021 Conference, 12 July 2021.
- Talbert, M. (2019), “Moral Responsibility”, The Stanford Encyclopedia of Philosophy (Winter 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/>, last accessed on 17 June 2021.
- Tangney, J. P. (1999). The self-conscious emotions: Shame, guilt, embarrassment and pride (pp. 541–568). In Tim Dalgleish & M. J. Powers (eds.), *Handbook of Cognition and Emotion*, New York, Wiley.
- Taylor, J. (2015). *Reflecting Subjects: Passion, Sympathy, and Society in Hume’s Philosophy*. Oxford: Oxford University Press.
- Wildgaard, L. (2019). An overview of author-level indicators of research performance. In *Springer Handbook of Science and Technology Indicators*, 361-396.
- Williams, B. (2002). *Truth and Truthfulness: An Essay in Genealogy*. Princeton: Princeton University Press.
- Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage publications.
- Wilsdon, J., et al. (2015). The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. DOI: 10.13140/RG.2.1.4929.1363.
- Wimalasuriya, D. C., Dou, D. (2010) Ontology-based information extraction: An introduction and a survey of current approaches, *Journal of Information Science*, Volume: 36 issue: 3, page(s): 306-323.
- Wolcott, H. F. (2008) *Writing up qualitative research*. Sage Publications.