RESEARCH   **Open Access**

Rethinking success, integrity, and culture in research (part 1) — a multi-actor qualitative study on success in science

Noémie Aubert Bonn[*] and Wim Pinxten

Abstract

Background: Success shapes the lives and careers of scientists. But success in science is difficult to define, let alone to translate in indicators that can be used for assessment. In the past few years, several groups expressed their dissatisfaction with the indicators currently used for assessing researchers. But given the lack of agreement on what should constitute success in science, most propositions remain unanswered. This paper aims to complement our understanding of success in science and to document areas of tension and conflict in research assessments.

Methods: We conducted semi-structured interviews and focus groups with policy makers, funders, institution leaders, editors or publishers, research integrity office members, research integrity community members, laboratory technicians, researchers, research students, and former-researchers who changed career to inquire on the topics of success, integrity, and responsibilities in science. We used the Flemish biomedical landscape as a baseline to be able to grasp the views of interacting and complementary actors in a system setting.

Results: Given the breadth of our results, we divided our findings in a two-paper series, with the current paper focusing on what defines and determines success in science. Respondents depicted success as a multi-factorial, context-dependent, and mutable construct. Success appeared to be an interaction between characteristics from the researcher (Who), research outputs (What), processes (How), and luck. Interviewees noted that current research assessments overvalued outputs but largely ignored the processes deemed essential for research quality and integrity. Interviewees suggested that science needs a diversity of indicators that are transparent, robust, and valid, and that also allow a balanced and diverse view of success; that assessment of scientists should not blindly depend on metrics but also value human input; and that quality should be valued over quantity.

Conclusions: The objective of research assessments may be to encourage good researchers, to benefit society, or simply to advance science. Yet we show that current assessments fall short on each of these objectives. Open and transparent inter-actor dialogue is needed to understand what research assessments aim for and how they can best achieve their objective.

Study Registration: osf.io/33v3m.

Keywords: Research integrity, Research assessment, Pressure to publish, Inter-actor dialogue, Success in science, Misconduct, Questionable research practices, Flanders, Research evaluation

* Correspondence: noemie.aubertbonn@uhasselt.be
Research Group of Healthcare and Ethics, Faculty of Medicine and Life
Sciences, Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium

Background

Excellence is a prominent theme in any funding scheme, university mission, and research policy. The concept of excellence, however, is not self-explanatory. Apart from the fact that excellence is hard to define, it is complicated to translate it into concrete criteria for evaluating whether researchers are successful or not in their pursuit of scientific excellence. Nonetheless, in today's highly competitive setting where talent is plenty and money is tight, determining evaluation and assessment criteria is a necessity.

When researchers are being assessed, it is important that the criteria used for determining success are compatible with our concepts of scientific excellence. However, with poorly defined concepts of excellence [1, 2] and assessment criteria that raise considerable controversy, there is no guarantee that this is actually the case.

The issue has increasingly attracted the attention of influential voices and fora, which resulted in a growing number of statements and documents on the topic, including the Declaration on Research Assessment (DORA [3];), the Leiden Manifesto [4], The Metric Tide [5], and more recently the Hong Kong Principles for Assessing Researchers [6]. In a review of 22 of these documents, Moher and colleagues pointed out that current research assessments are open for improvement, particularly in further addressing the societal value of research, in developing reliable and responsible indicators, in valuing complete, transparent, and accessible reporting of research results as well as reproducibility, and in providing room for intellectual risk taking [7]. As many of the documents mention, however, changing scientific assessment is not straightforward and is likely to face resistance from diverse parties. One of the reasons for this resistance may be the complex inter-actor exchange that governs research and academia. As the European Universities Association (EUA) made clear in a recent report, research institutions, funders, and policy makers must "work together to develop and implement more accurate, transparent and responsible approaches to research evaluations" ([8], p. 13). But although certain actors such as researchers and scientific editors have been highly involved in the debate, other actors have been largely missing from the discussion.

Previous research has mostly focused on opinions of students, researchers, and editors [9]. It was therefore the goal of our research to extend our understanding of success and integrity by also capturing the views of policy makers, funders, institution leaders, research integrity office members, research integrity network members, laboratory technicians, and former researchers who changed career. We present our results in two publications. In the present paper, we discuss how different actors perceive success in science, and in an associated publication [10] we present their view on integrity and

problems in science. Our findings resonate with past efforts by suggesting that, in their current state, research assessments may fuel detrimental research practice and damage the integrity of science.

Methods

For reader's convenience, the methods of our project are described in full both here and in our associated paper [10].

Participants

The present paper reports findings from a series of qualitative interviews and focus groups we conducted with different research actors. This qualitative work was part of the broader project Re-SInC (Re-thinking Success, Integrity, and Culture in science; the initial workplan is available at our preregistration [11]).

In Re-SInC, we captured the views of different research actors on scientific success, problems in science, and responsibilities for integrity. Being aware that the term 'research actor' may be ambiguous, we defined research actors as any person having a role in the setup, funding, execution, organisation, evaluation, and/or publication of research. In other words, we included actors linked to the policing, the funding, the evaluation, the regulation, the publishing, the production (i.e., undertaking the research itself), and the practical work of research, but we did not include sole consumers of science or end users of new technologies.

We used Flanders as a setting, including participants who either participate in, influence, or reflect (directly or indirectly) upon the Flemish research scene. Nevertheless, we will discuss below that our findings are also highly coherent with similar works in different research settings (see for example [12–14] in the UK, [15] in Croatia, [16] in the US, and [17] in Denmark). In most cases (49 out of 56 participants), participants did not know the interviewer before the interviews and focus groups. In selecting participants, we aimed to capture the breadth of the Flemish research scene. Using Flanders as a research setting had the advantage of allowing us to capture perspectives from an entire research system in a feasible setting. The Flemish research scene comprises of five main universities and a number of external research institutes, major funding agencies, a federal research policy department, and one advisory integrity office external to research institutions. We chose to concentrate our research on three of the five universities, and to include partnering European funding and policy organisations as well as international journals and publisher to build a realistic system sample. When participants were affiliated with a university, we focused on the faculty of biomedical sciences. Given the exploratory and qualitative nature of this project, we did not aim for an exhaustive nor a fully representative sample.

Our objective was to shift the focus from the narrow view targeting mainly researchers to a broader view that includes a broad range of research actors. Accordingly, we maximized the diversity of participants in each actor group to ensure that each group encompassed a wide range of potentially different perspectives.

Our main actor categories are PhD students, post-doctoral researchers (PostDoc), faculty researchers (Researchers), laboratory technicians (LT), policy makers and influencer (PMI), funding agencies (FA), research institution leaders (RIL), research integrity office members (RIO), editors and publishers (EP), research integrity network members (RIN), and researchers who changed career (RCC). The composition of each actor group is detailed in Table 1.

It is important to keep in mind that the research world is complex and not organized in distinct actor groups. Consequently, participants could often fit in more than one category, and sometimes felt the need to justify circumstances that would make them fit in the category we selected. Before the interview, we asked participants whether they agreed with the category we assigned them in, and we refined and exemplified the definitions of our actor groups to reflect the participants' distinctions (i.e., further explaining the slight differences between the groups planned in the registration and those used here).

Recruitment

We used several recruitment strategies. For the focus groups with PhD students and researchers, we circulated an email to everyone in the Faculty of Medicine and Life Sciences of the host university and invited them to register on an interest list. We later scheduled a convenient time where most of those who registered were available. We used a similar strategy for the focus group of editors and publishers, but circulated the invitation in a relevant conference. For focus groups with lab technicians and post-doctoral researchers, key players helped us recruit and organize the focus group. Given the diverse methods, and assistance used in organising the focus groups, estimating response rates would thus be challenging.

For interviews, we invited participants directly via email. We sent up to three reminder emails, but did not pursue further if no response was obtained at the third reminder email. All participation was on a voluntary basis. From the individuals invited to participate in an interview, two declined the invitation, one did not feel like a good fit for our project and was therefore not interviewed, and eight did not reply to our invitation. Although invited individuals sometimes referred us to colleagues holding similar roles in the same institution, the remaining invitations led to interviews.

Design and setting

We conducted semi-structured interviews and focus groups, meaning that we asked broad questions in an open manner to target main themes rather than specific answers. All interviews and focus groups were audio recorded and transcribed verbatim. Details about the tools used to guide the interviews and focus groups are available in the tool description below.

To maximise transparency, we provide an extended descriptions of the interviewer and the setting of the interviews in Supplementary file section 1 and a copy of the COnsolidated criteria for REporting Qualitative research checklist (COREQ) in Supplementary file section 2.

Ethics and confidentiality

The project was approved by the Medical Ethics Committee of the Faculty of Medicine and Life Sciences of Hasselt University (protocol number CME2016/679), and all participants provided written consent for participation, for use and publication of anonymized direct quotes, and for dissemination of the findings from this project. A copy of the consent forms is available in the registration of this project [11]. We protected the confidentiality of participants by removing identifiers from quotes included in the text. Nonetheless, Flanders is a small research system and given our actor-specific sample, personal identification within quotes remains a risk despite our efforts. To further protect participants' confidentiality and avoid that identification of individual quotes lead to identification of all quotes from the same participant, we decided not to specify respondents in individual quotes, but to refer only to actor groups.

Following this reasoning, we are unable to share full transcripts, but attempted to be as transparent as possible by providing numerous quotes in the text, in tables, and in appendices.

Tool

To build our focus group guide, we inspired our style and questions from the focus group guide developed by Raymond De Vries, Melissa S. Anderson, and Brian C. Martinson and used in a study funded by the NIH [16]. We obtained a copy of the guide after approval from the original authors, and revised the guide to tailor questions to the topics we wished to target, namely 'success in science' and 'responsibilities for research integrity'. We revised our focus group guide several times before data collection and discussed it with Raymond De Vries — expert in qualitative inquiries and part of the team that built the original guide upon which we inspired ours. We built interview guides based on our revised focus group guide. We adapted specific questions (e.g.,

Table 1 Demographics of participants

Actor group	Abbrev.	Sample description	N, setting, and gender[a]
Researchers	Researchers	Faculty researchers from the Faculty of Medicine and Life Sciences of the host institution.	[■ ■ ▲ ▲]
Post- doctoral Researchers	PostDoc	Post-Doctoral researchers enrolled in the Faculty of Medicine and Life Sciences of the host institution.	[■ ■ ■ ▲ ▲]
PhD students	PhD	PhD students enrolled in the Faculty of Medicine and Life Sciences of the host institution.	[■ ■ ■ ■ ■]
Lab technicians	LT	Laboratory technicians from the Faculty of Medicine and Life Sciences of the host institution.	[■ ■ ■ ■]
Past researchers who changed career	RCC	Although this group was not part of our pre-registration, one RCC asked us whether she could take part in our study after seeing the invitation email. After having a chat with her, we realized that hearing the narrative and perspectives of individuals who did research work but decided to leave academia would deeply enrich our results and inform us on problems which are big enough to drive researchers away from research. Therefore, we invited a few researchers who changed careers (i.e., researchers or research students who decided to leave academia) to participate in interviews. In this group, we selected individuals from each of the three universities included in our project, and ensured to have individuals who left academia during their PhD, after their PhD, after their PostDoc, and during a tenure track. Recruitment of those participants was helped by recommendations from colleagues who were aware of the profiles we were looking for.	■ ■ ■ ▲
Research institution leaders	RIL	We included three Flemish universities in our study. In each institution, we involved several members from the board of directors. These included directors of research, deans, or directors of doctoral schools from the faculties of medicine and life sciences or equivalent.	▲ ▲ ▲ ▲ ▲ ▲ ▲
Research integrity office members	RIO	We included different members from offices in charge of investigating allegations of research integrity and misconduct in three Flemish research institutions and outside research institutions in Flanders (e.g., research integrity officer, policy officers, etc.)	■ ■ ■ ▲
Editors and publishers	EP	We invited both big and small editors and publishers, and were fortunate to be able to involve journals and publishers with a broad range of editorial practices (i.e., open access and subscription based; published in local language and published in English; focusing on reviews and focusing on ground breaking empirical findings). To select the interviewees, we first invited a selection of journals from the top twenty highest Impact Factor for 2017 under the category of 'Medicine, general and internal' in the Journal Citation Reports (Clarivate Analytics), purposively picking different publishing models. In addition, we invited select publishers to take part in our research. After conducting individual interviews with a few agreeing participants from this sub-selection, we organized a small focus group with editors of smaller or differing journals, allowing us to involve a great diversity of editors and publishers.	[■ ■ ▲ ▲] ■ ■ ■ ▲
Funding agencies	FA	We selected national, as well as European funding agencies, making sure to target different funding styles. We made sure to include perspectives from regional public funders, regional private funders, international funders, as well as funders focusing on applied research and funders focusing on fundamental research.	[▲ ▲] ■ ▲ ▲
Policy makers or influencers	PMI	In this group, we included both organisations responsible for setting science policy, and organizations which influenced such policies by serving as informers. Consequently, PMIs do not necessarily write nor decide science policies, but may also be asked to provide data which later influences policy decisions.	■ ▲ ▲ ●
Research integrity network members	RIN	We selected a few actors from the research integrity core experts. These included researchers involved with important European research projects on research integrity as well as one actor involved in writing the European Code of Conduct for Researchers.	■ ▲ ▲
		TOTAL = 56 participants	

[a]Square bullets (■) represent female participants; triangle bullets (▲) represent male participants, and round bullets (●) represent participants with undefined gender ('prefer not to answer'). Bullets displayed in brackets represent participants with whom we conducted as focus groups or joint interviews

responsibilities, evaluation) to each actor group, but preserved the general structure and themes for all interviewees. A general version of the interview and focus group guides are available in Supplementary file parts 3 and 4. More specific group guides can be provided upon request. All guides were constructed around the following four topics:

i) **Success in science**: What makes a researcher successful? Are these characteristics captured in current assessments? What are indicators for success? What do you feel is most satisfying, most rewarding about your career (specifically for researchers and research students)?

ii) **Current problems** (including misconduct and questionable research practices): Do you have experience with research that crossed the lines of good science? How can we draw the line, what are red flags? Why do bad practices happen? Can they happen to anyone?

iii) **Responsibilities towards integrity**: What is your responsibility towards integrity? Where does it end? Who else is responsible? In what ways are other actors responsible?

iv) If you were granted a fairy wish and could **change one thing in how science works**, what would you pick?

It is important to consider that the interview guide was not used mechanically like a fixed questionnaire, but sometimes shortened, expended, or reordered to capture responses, interest, and to respect time constraints. In this manuscript, we mainly report findings from the first (i) and fourth (iv) topics, although we sometimes captured the discussion from other questions when relevant.

Analysis

Recordings were first transcribed verbatim and, where necessary, personal or highly identifiable information was anonymized. We analyzed the transcripts using an inductive thematic analysis with the help of NVivo 12 Software to manage the data. The analysis proceeded in the following order:

v) **Initial inductive coding**: NAB first analyzed two focus groups (i.e., researchers and PhD student) and five interviews (i.e., RIL, RIO, PMI, RCC, and RIN) to have an initial structure of the themes targeted. In this step, she used an inductive thematic analysis [18] while keeping the three main categories — i.e., success, integrity, and responsibilities — as a baseline. Using the inductive method avoided that we limit our analysis to the order and specific questions included in our guide, and allowed us to also identify and note themes that were raised spontaneously or beyond our initial focus.

vi) **Axial coding**: With this first structure, NAB and WP met and took a joint outlook at these initial themes to reorganize them in broader categories and identify relationships between categories. For this step, NAB built figures representing the connections between the main themes, and refined the figures and the codes after the meeting.

vii) **Continued semi-inductive coding**: NAB continued the coding for the remaining transcripts, sometimes coding deductively from the themes already defined in steps 1 and 2, and sometimes inductively adding or refining themes that were missing or imprecise.

viii) **Constant comparison process**: NAB and WP repeated the axial coding and refining steps several times throughout this process, constantly revisiting nodes (i.e., individually coded themes) by re-reading quotes. The nodes and structure were then discussed with RDV to reconsider the general organisations of the nodes. This constant comparison process is common in qualitative analyses, and is commonly used, for example, in the Qualitative Analysis Guide of Leuven (QUAGOL [19];). This repeated comparison led to a substantially solid set of nodes which later guided further coding in a more deductive manner, though we made efforts to remain open to possible new themes in respect of our inductive analysis.

ix) **Lexical optimization**: Finally, after having coded all transcripts, NAB and WP further discussed the choice of words for each node and reorganized the themes to ensure that they were an ideal fit for the data they were describing. NAB and Raymond De Vries met to have a final outlook of the general structure and to reorganise the nodes in clean and natural categories.

Results
Short summary of results

Our investigation of the perspectives of success in science, among which 56 participants spread in eleven different actor groups took part, revealed that the way in which we currently define science and the way in which we assess scientific excellence generates conflicting perspectives within and between actors.

First, we realised that the way in which researchers define their personal successes was not necessarily standard, and that definitions of successes seem to change with different contexts, demands, and career stages. For instance, the desire to make a change in society was particularly strong in early career researchers, while more established researchers also valued simple curiosity, and relational successes.

When involving all different research actors, we were able to build a representation of success which was nuanced and multifactorial. Success appeared to be an interaction between characteristics from the researcher (Who), research outputs (What), processes (How), and luck. Interviewees noted that current research assessments overvalued outputs but largely ignored the processes deemed essential not only for the quality of science, but also for the collegiality and the sense of community that unite scientists. Luck was thought to play a crucial role in success and was often used to explain cases where evaluations of success were considered unfair: bad luck explained the lack of reward for excellent researchers, while good luck explained that regular researchers moved ahead without deserving it more than others.

Interviewees generally agreed that current research assessments did not capture the whole picture of success, but often disagreed on the value of specific indicators

used to attribute success. The relevance of publications, impact factors, science communication, and openness in research assessments raised such disagreements.

Interviewees provided insights on the characteristics they considered essential to any fair and representative assessments. Among those, interviewees suggested that science needs a diversity of indicators that are transparent, robust, and valid, and that also allow a balanced and diverse view of success; that assessment of scientists should not blindly depend on metrics but also value human input; and that quality should be valued over quantity.

Finally, when asked what they would change in science, many respondents targeted the way in which researchers are assessed and rewarded, reiterating that there is an urgent need for fairer distribution of resources and rewards in science.

Complete and detailed results are presented in the subsections below.

Researchers' personal successes

PhD students, post-doctoral researchers, and researchers described a number of factors which made them feel satisfied or which they interpreted as personal success. First, PhD students and post-doctoral researchers strongly supported that **making a change** in practice was something that was central to them.

"I agree with the fact that that feeling that something is done with what you found is crucial for your own feeling. […] I think that's crucial. Even more than the publications or the … yeah … " (PostDoc)

"Yeah it was part of my motivation to give something back to the clinical field by doing research." (PhD student)

For PhD students, realising that their results would remain theoretical or would be too small to make a difference was raised as one of the disappointment they faced in research.

"If I can help people by doing this project, that gives me a lot of satisfaction I think.
[Responding to this participant, another added] *That's true but that was also my first idea when I started, but I have to be honest, my project is so fundamental that I'm almost finishing up, and I don't see anything that will be going to the clinic for years or something. So at that point for me it was a bit disappointing, because... Ok, I wanted to, but I'm so fundamental, basically really molecular stuff, that I don't see it to get really...*

[Later in the discussion, another participant further added] *Yeah I think for me it's the same. Because I'm working on a project that's like this very tiny subset of a subset of [specialised] cells. And then at the beginning you think 'I'm going to change the field with this research', but yeah I don't know." (PhD students).*

Although some researchers also supported that translating their findings in practice was satisfying, they acknowledge that **theoretical knowledge** or simply following their **curiosity** became their "*main drive*", or at least provided its share of satisfaction.

"For me it's good if it goes this direction [i.e., is translated in practice] *but also just creating new knowledge which doesn't really directly impact people, I think is also very very interesting, or I'm also very passionate about that. So it shouldn't always have an implication." (Researcher).*

For researchers, external satisfactions, such as *peer appreciation*, or **fulfilling institutional requirements** were "*also very important*" to personal satisfaction, but as secondary aspects which were not enough for feeling completely satisfied.

"I also have some … still some criteria which I have to do that I also think about those things. But I don't feel bad about it that it's my only drive for some things that it's just publication. On the other hand I also feel that I cannot be satisfied alone by those things." (Researcher)

Finally, post-doctoral researchers added two intriguing dimensions to the concept of success. First, they stated that successes are personal, and that each researcher will likely be successful in different ways. In this sense, personal success was seen to reflect aptitudes and skills in which individuals excel, rather than a universally shared idea.

"[In my group, we don't have strict requirements], and I think it's very beautiful because we have [dozens of] PhD students and they're all — or 99% of them are — successful, but they are so different in being successful. Some are really being successful in the number of publications, some of them are really successful in the network they have with other companies, with other research institutes, some of them are really successful in the perseverance to do something really new and to make it happen, only if there's a small study on 20 patients, but it's so new and they will really make it happen in the hospital.

So, they're so successful on so many different levels and I really like the fact that we don't judge them all in the same way because they can be themselves and be successful in the way that they want to be successful." (PostDoc)

The need for diversity of successes was thus valued, even though it was acknowledged to be a rare feature in research assessments. A second intriguing dimension was raised by a post-doctoral researcher who pointed out that, even within individual researchers, personal conceptions of success may be mutable, likely influenced by career stages, work environments, and expectations of others.

"Participant A: For me I think my idea on what success is is changing a lot of course. When you're a PhD student you just want a breakthrough in your project, that's success, and then by the time you're finishing your PhD you're looking at what ... Is there a possibility for post doc then you realise 'OK they're counting publications, they're doing this' and then you're looking around and then you sometimes get this mixed feeling of someone who you feel was not very creative or did not have to do a lot of work themselves, it was very guided and clear steps, and they have a lot of publications and so they get a post doc position. And then that's sometimes difficult, and you think like 'How does this work here?' [...] then I went into more research coordination and then I was in a [different group] and then it was all the time about metrics. Because the money was divided by metrics, and it was like publications and project funding and ... And then I felt like everything revolved around that. It wasn't important anymore like what projects we're doing as long as it was a project on that funding channel because that counted higher on the metrics and ... So ok, and then you're really like that. And now being here in this setting I'm really seeing the impact of research. Now it's changed again. Now it's really like that kind of research where you can make a difference for an organisation, for patients ... That's the thing that's success. And I think that maybe like you say that in the long run that's what you have to do. But it's kind of the short-term mechanisms, and not always ...
Participant B: Yeah, I think that the definition of success is highly dependent of the institute and the environment you're in like you're mentioning. And if you're constantly told 'This is how we measure success' then ...
Participant A: Yeah, so then you're really guiding yourself to get those key indicators." (PostDocs)

In other words, interviewees revealed that personal success was a mutable variable which appeared to change depending on contexts, demands, and career stages.

Inter-actor views on success

Interviewees mentioned several factors which they believe are essential or useful in becoming a successful researcher. We classified these factors in four main categories: factors visible in the *researchers* themselves (Who), factors from the research *process* (How), factors from the research outputs (What) and, unexpectedly, factors related to *luck*, which was thought to play an important role in success (details on development of these factors is described in the Supplementary file). Figure 1 illustrates the different categories we captured.
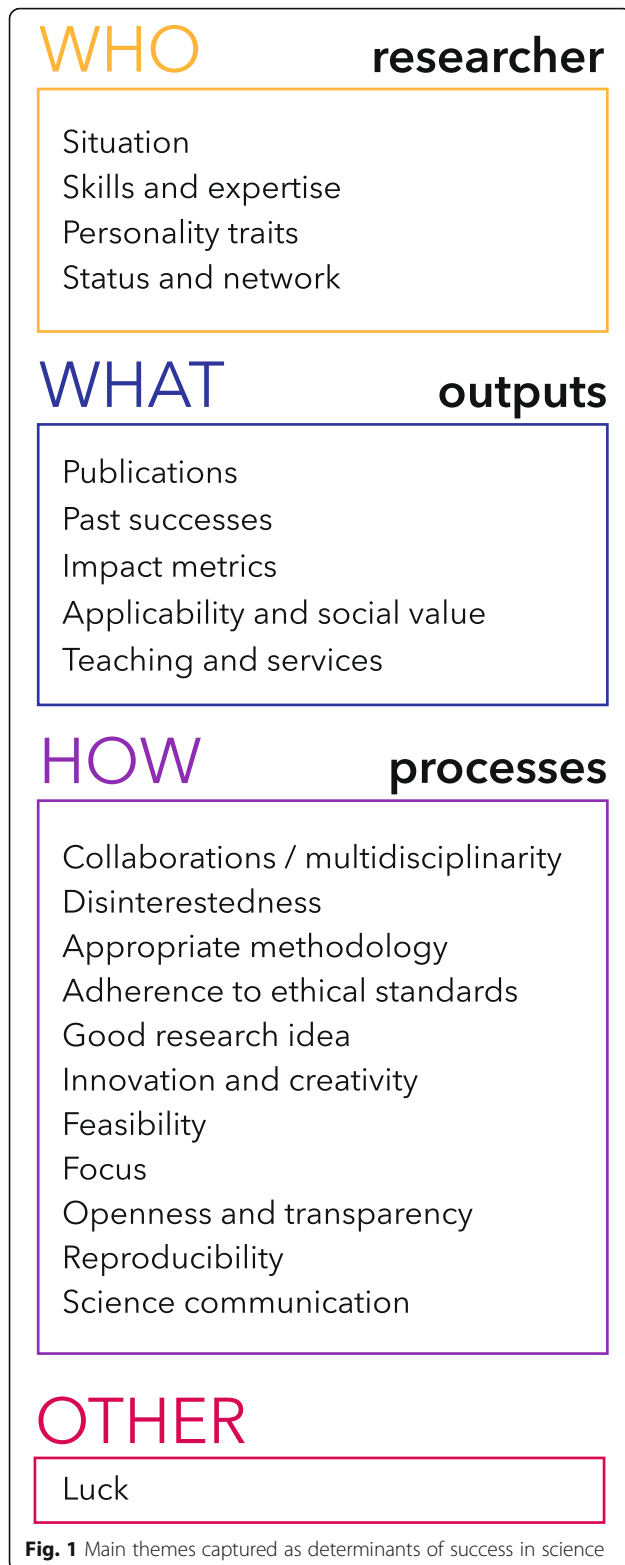
Who
Researcher

Several features related to the *researchers* were considered important in determining and yielding success. While all these individual factors were said to play a role in producing success, they were also described as indicators to look for when selecting researchers for a position, thereby influencing careers and promotions. Among those, participants highlighted *personal traits*, such as ambition, passion, rigorousness, and intelligence, as well as acquired *skills and expertise*, such as business potential, management skills, writing skills, and scientific expertise. Certain respondents also believed that success could be influenced by specific *situation* in which individuals find themselves. In this regard, gender and ethnicity were mentioned as possible obstacles — through pregnancy leaves, family obligations, prejudice, or language inequalities — or advantage for success — through employment quotas. Along the same line, childlessness and celibacy were mentioned as advantages for yielding success since they allowed researchers to devote more time to their work.

Beyond the advantage that extra time and flexibility could provide, they were sometimes considered as conditions to a successful research career. Indeed, some interviewees believed that researchers and research students *should* be able to devote themselves to their career by being mobile and by working beyond regular schedules and conditions.

"I think people have to realize when you do a PhD, it's a stressful thing, you really are going to get the highest degree there is at a university, it doesn't fit between 9 and 5." (RIL)

" ... being passionate about science is almost like being an artist. You live in poverty because you want to pursue your art." (PMI).

WHO   researcher

Situation
Skills and expertise
Personality traits
Status and network

WHAT   outputs

Publications
Past successes
Impact metrics
Applicability and social value
Teaching and services

HOW   processes

Collaborations / multidisciplinarity
Disinterestedness
Appropriate methodology
Adherence to ethical standards
Good research idea
Innovation and creativity
Feasibility
Focus
Openness and transparency
Reproducibility
Science communication

OTHER

Luck

Fig. 1 Main themes captured as determinants of success in science

"That's also what we ask for, excellence for people when they come here. […] Usually those people need to have been abroad for at least six months.

But if it is two years it's better. So these are important factors to create excellence." (RIL)

Many of the researchers who changed career mentioned that the expectation that they should sacrifice family life and private comfort for science played a role in their decision to leave academia. This is explored in the associate paper [10].

Finally, the network and status that researchers bring along with them was also seen as determinant to success. Having an established network and personal recognition from peers was thought to be key to success.

What
Outputs
Indicators which provide information about what researchers have accomplished were univocally considered crucial in determining success. Among those, high academic grades, past success in obtaining funding, publications, and publication metrics (e.g., impact factor, citations, H index) were mentioned as currently being used for determining success, although not all interviewees agreed on the individual value of these determinants. In addition, less traditional products of research were also mentioned, such as the applicability and societal value of the research findings and the researcher's involvement in teaching and services (i.e., mostly referred to as serving on institutional boards, committees, and scientific societies).

How
Processes
Features which indicate 'how researchers work' (i.e., *processes*) were also deemed integral to success, regardless of the output they generate. What truly differentiated *outputs* from *processes* was the perspective that the latter contributed to success regardless of the final, measurable result. On the one hand, some processes were thought to play a part in the success of individual research projects. Collaborations, multidisciplinarity, appropriate methodology, adherence to ethical requirements, good and innovative research ideas, feasibility, and focus were all viewed as pathways to achieve good outputs and related successes. On the other hand, respondents also identified a number of processes which they considered impacted beyond individual projects and were essential to the success of science at large. Openness and transparency, for example, were repeatedly viewed as important aspects of the collegiality which promotes the success of science as a common goal. One interviewee explained that openness was *"very important to help the research enterprise because it's really about facilitating the fact that other people can build upon a research"* (EP). Along the same lines, reproducibility was

considered as the "*most important thing*" (RIL) and as "*a very important element in science*" (PMI). Yet, interviewees noted that reproducibility is "*often lacking*" from research (PMI), and that replication studies are underappreciated in current success assessments (Researcher) or even possibly wasting research money (RIL). Finally, public engagement, mainly in the form of communicating scientific findings to the public, was also mentioned as part of the broader scientific success by building trust in science and by potentially contributing to the quality of research.

Other
Luck
Interviewees also attributed success to *luck*, a feature which transcended outputs, processes, and individuals. In our analysis, we discerned three different definitions of what it meant to be 'lucky' in science. First, researchers could be considered lucky if they worked with distinguished colleagues or in established labs, given that such settings maximized the opportunities for obtaining high end material, publications, and grants. This first meaning brings back the idea of the network that researchers bring with them, and adds an element of arbitrariness to the control that researchers have in building their network. Second, luck was also employed to refer to unexpected evolutions and trends, such as working on a topic which suddenly boomed in visibility and media attention or being "*somewhere at the right moment at the right time*" (FA). In this second signification, luck was perceived as something that one could partially create, or at least grasp and maximize. Finally, luck was sometimes attributed to the output of research results, with positive findings being lucky, and negative findings being unlucky. In this last sense, luck was a factor that was out of researchers' control and independent of their skills. In all three senses, luck was both described as something that had helped mediocre researchers move ahead in their career and as something that had wrecked the success of otherwise talented researchers.

Current assessments of success
Interviewees generally agreed that there was a tension between what formally determines success and what they believed *should* account for success in science.

> "Hm … What the current situation is, or what I think success should be? (laughs)" (EP)

> "I think that you have different views on looking on it. You have the measurable parts, and you have the non-measurable part. And I think that these two are sometimes in contradiction." (RIO)

> "…I started this PhD project because I wanted to have results useful to clinical practice, and I said "I want to do this". And [my supervisors] were already saying for a year "No, no, it's not interesting, no we shouldn't do that." and I said "I want to do this, or my project failed for me." […] Ok, I know it's not going to be so big that it's so interesting for journals, but I think for our clinical field, for Flanders, it's important that we do a study like that. And it was … that was the chapter that people from clinical practice were most interested in too. So… I think when you ask us 'What is success in research', we've got our own points of success, and what we know that's expected from us by the system. So those are two different lists. (laughs)." (PhD)

But despite a general agreement on the need to reintroduce processes in research assessments, participants often disagreed on the value of specific success indicators, particularly on the indicators related to the *What* and the *How*.

Disagreements on outputs
Publications
The emphasis on publication in research assessments raised substantial disagreement. Some respondents considered that relying on publications to measure success was problematic and even damaging for science, while others saw publications as a necessary and representative measure of scientific success. We distinguished three arguments against using publications as the main indicator for assessing research success (Fig. 2).

First, publications were described as a *reductionistic* measure of success. Using publications as the main measure for success ignored "*other very important contributions to the scientific enterprise*" (EP), and could unfairly disadvantage excellent researchers who are simply unsuccessful in publishing their results. Second, publications were also seen as an *arbitrary* measure which does not represent merit, efforts, and quality. Researchers and research students in particular worried that publications often resulted from arranged connections rather than from high scientific value, and argued that publications with high impact factors were not necessarily of high quality. Adding to this, several interviewees supported that publications could be a mere matter of luck. Third, interviewees worried that the increasing dependency of researchers on their publication output (i.e., the publish or perish culture) may introduce *perverse* incentives which might threaten the integrity of research. Indeed, interviewees worried that publication pressure might not only tempt researchers to engage in questionable practices to maximise their publication output, but could also shift the main objective of research projects towards

Should publications serve as a major indicator for success?

Arguments against	Arguments in favour
Reductionistic	**Necessary**
Focusing on publications ignores other essential research behaviours which should be valued in research assessments.	Publications allow sharing and iteration of research knowledge. Assessing publications ensures that they remain a priority for researchers.
Arbitrary	**Representative**
Publications are not representative of the merits and efforts put into work. They often result from biases and sheer luck.	Publications require a huge amount of work. Good publications are thus a good indicator of research achievement and merit.
Perverse	**Measurable**
The emphasis on publications moves the purpose of research from the creation of knowledge to mere publication tactics.	Publications are measurable and quantifyable, so they can provide a reliable and comparable indicator of success in science

Fig. 2 Summarising scheme of arguments for and against using publications as the main indicator of success

'sexy' and 'publishable' topics rather than topics that are '*interesting*' or '*relevant*' to advance science. Illustrative quotes are available in Supplementary file section 5.

Despite these three arguments against the focus on publications for evaluating success, a number of respondents also identified arguments in favour of such focus, sometimes directly opposing the arguments introduced above. First, publications were described as a *necessary* aspect of scientific advancement; emphasising publications in evaluations was thus a way to ensure that researchers keep publications in the forefront of their priorities. Second, some respondents described publications as *representative* indicators of good research and merit. In fact, considering that publications are the endpoint of an extensive and difficult process which could not happen without hard work, some argued that publication outputs helped identify good researchers. Finally, publications were also described by many as the only tangible way to value and *measure* science, which added to the credibility of research assessments.

Impact factor

Similar conflicts were observed when looking at the impact factor. First, the impact factor was acknowledged by many as a being useful for its *measurability*, its *simplicity*, and its *acceptability*. Several interviewees — research institution leaders in particular — perceived the impact factor as a measure of quality of a journal and its review process (see Supplementary file section 6 for associated quotes).

Nonetheless, using the impact factor as a measure of success yielded overwhelmingly negative responses, even among participants who believed it served as an indicator of quality. Most participants mentioned that the impact factor was not adapted to their disciplines, that it was not representative of the impact of individual papers, that is was open to biases and manipulation, and even that is could disrupt science by discouraging research in fields with traditionally lower impact factor (see Supplementary file section 6 for associated quotes).

A few interviewees proposed that direct citations would be more relevant in personal impact assessments, but they also acknowledged that determining the impact of individual articles using direct citations could take years, if not decades in some disciplines. Furthermore, researchers added that their most cited paper was not necessarily the one they considered most important, and that citation counts tended to refer to novelty and timing rather than to quality. Consequently, despite an overwhelming aversion towards using impact factors for

scientific evaluation, concrete alternatives were more difficult to nail down.

Disagreements on processes
Science communication
The importance of science communication also raised conflicting opinions among interviewees. Many supported that sharing science through popular channels such as Twitter, YouTube, Facebook, or Wikipedia should be considered in career evaluations. For example, one respondent in the PMI group noticed that *"Researchers who do a lot of work on Wikipedia are not rewarded for it, but they're doing a lot of good work!"*. The same interviewee however, later warned that appearing in the media was different than actively making the effort to communicate science, *"because then the media decides who is successful"* and *"a lot of researchers will also be successful and you will never hear of it"*. A few researchers mentioned that science communication was essential to maximize the interdisciplinary impact of one's work, and that presenting findings in broad conferences and participating on Twitter could foster this interdisciplinarity. Other respondents even regarded science communication and the ability to simplify and share one's findings with different stakeholders as a core requisite for the quality of research.

> *"I mean we work a lot on, or we try to promote everything which has to do with public engagement and science communication, all these things, but if you're not able to explain to lay people what your research is about … [shakes head meaning it's not a good sign]. I think it's a sort of, how do you call it, a litmus test in a certain way […] Sometimes sort of public engagement … the arguments are sort of normative. You have to do this because you're working with public money and you have to be accountable or … which is ok, but, I really believe it's better than that. It's more important than that. I really believe it's good to discuss with philosophers, with ethicists, with citizens, with patients … For the quality of your research. To be stretched. It's another type of checks and balances than the ones which are done in peer review. It doesn't replace peer review, it's just another level. To look at the relevance, to, yeah … to be confronted with questions you probably haven't ever asked … To be better in communicating which … Better communicating will help you better thinking. I mean I think there is a lot of quality gains." (FA)*

But although some perceived science communication to be an essential component of quality work, others saw it as a component which did not indicate the quality nor the efforts invested in the research. Some researchers even supported that the quality of science might be threatened by the lack of quality control of social media.

> *"I feel there is also a kind of danger in those things, because for example I follow some researchers on Twitter, which have a very… I feel that they're on Twitter all day long I'd say, and everybody follows them… But it's not… the research is not always that good, but because of the fact that everybody is following, this is going to be the new reality, and I start to… yeah… These things, worldwide, have <an> impact on the research impurities, and it's shifting towards…. yeah, it's not controlled, the quality of those things." (Researcher).*

This perspective was echoed by a participant from the RIO group, who admitted having faced substantial resistance from researchers when presenting an action plan meant to promote and value science communication in her institution. This RIO received responses such as *"yeah… that's the one who's always with his head in the newspapers, but is he writing A1s[1]?"*, and concluded that researchers might not be in favour of such a shift. In sum, even though science communication is an important aspect currently put forward in new evaluation processes and policies, researchers do not all agree on its value and its impact on the quality of science.

Openness
Openness also raised diverse thoughts from our interviewees. Although most agreed that open science and transparency were important or even *"necessary for the community of researchers"* (PMI), some doubted that open science would help foster integrity, proposing that it might simply bring a *"different level of cheating"* (RIL). We also understood from researchers and especially from research students that the fear of 'being scooped' was still too vivid for openness to fully happen, at least before publication. PhD students expressed frustration but also helplessness towards their will to be more open, admitting that the risks of losing their data tended to overcome their will to be more open and that opening their work was often discouraged by their supervisors.

> *"Participant A: Yeah we are now trying, or in our group someone is trying to put up a database for all of the data on [our topic]. But then researchers*

---

[1]A1's are a category of publications in research assessments in Flanders. They relate to articles included in Web of Science's Science Citation Index Expanded, Social Science Citation Index and/or Arts and Humanities Citation Index, whose document type is labelled as "Article", "Review", "Letter", "Note" and/or "Proceedings Paper"; or in journals included in the Journal Citation Reports of Web of Science.

would need to hand over their data to make it accessible. And there is a lot of discussion about it, if people would be willing to do that, to hand out your unpublished data... I think it will help the research, and it will help patients, but I don't know if everyone is willing, I don't know if I would be willing to, just put it in..

Interviewer: *Would you all ... would you be willing to put your data in a server?*

Participant B: *I had the question once, but by a supervisor, and we're PhD students so you asked, he said 'No, no we're just going to publish first and then when we did that then we can say here's the data'.[...].*

Participant C: *The problem with research is also it's really a competition in research. I also have it now that I can't present on a congress because there are only three articles published on the subject I'm studying, so the supervisors are scared if I make a poster or I present that other researchers will get interested in the same topic, and then, if they publish first all I'm doing is a waste of time... not exactly waste, but... yeah... so I think in research you really have a lot of competition because some people are focusing on the same subject and the data is not published so they will be first, and they want to be first, and... that's a problem with research. And I think it's also a problem that no one wants to share their unpublished data because they are scared that someone else will go and take the data and will publish first and then, you don't have it anymore.*

Participant D: *Yeah I completely understand the feeling because what we are doing it's also new so it's never been done and my promotor is always so reluctant to let me go and show the data to other people. [...] he is always so scared that other people are going to steal his ideas... Sometimes I do understand, but sometimes I'm also like, I don't really like this kind of environment, it struggles with my personality a lot, I think." (PhD students).*

Beyond open data, issues surrounding open access were also brought up in our interviews. We noticed that PhD students, who are directly affected by the inability to access research articles, strongly supported open access. Some university leaders also criticized the monopole of big publishers, noting that we faced a growing problem where subscriptions may become *"unpayable in the long run"* (RIL). Other university leaders however, stated that they would not advise

their researchers to publish in open access journals, on the one hand for financial reasons, and on the other hand because they perceived the model of open access as biased towards accepting papers regardless of their quality.

"That's another big issue. That's the open access eh? The model of the open access is unfair because the journal makes the profit by publishing because the author has to pay. So I think the review process is probably more biased. [...] I believe. I think that... OK, there is, in my very small field, there is some open access journals, and I feel like whatever review you do they all get published because they get the money. So, is that a good solution? No I don't believe it's a good solution.

Interviewer: *Yeah. So would you not advise to your researchers to.*

Participant: *I don't, no, because we don't have the money (laughs)" (RIL).*

One editor or publisher explained that bad publicity surrounding open access journals may come from the unfortunate reality that the open access model *"opened the door for a number of the so called predatory journals"*. Nonetheless, this interviewee also declared that *"in the end of the day* [journal quality] *depends on the editorial process of the journal and on the editorial criteria of the journal"* rather than on the publication model.

A number of other indicators raised polarized views, such as the need for societal benefit, and the need for focused areas of expertise. In sum, respondents agreed that current research evaluations were sub-optimal. Nonetheless, disagreements on the specific indicators which should and should not be used in attributing success suggest that solutions are far from simple.

Optimizing research assessments

Despite persisting disagreement on the content of good research assessments, several respondents proposed concrete recommendations on the form research assessments should take. Four main characteristics were put forward as essential for fair and representative evaluations (See Table 2 for sample quotes representing the four criteria).

Diversity of indicators

First, many interviewees mentioned that it is essential to use a diversity of indicators to be able to measure different aspects of research. Many respondents worried about the current overreliance on outputs (i.e.,

Table 2 Characteristics for good research assessments

Characteristic	Sample quote	Actor
Diversity of indicators	"With metrics I think there is an important rule to keep in mind is that if you're going to use metrics, you need to use many of them. And you need to really understand what they mean and whether they answer what you're looking for."	EP
	"And maybe also, and I think that the idea of taking other impacts into account can be helpful. I have no one solution, but I think this can be helpful."	FA
	"I think that you have to have different parameters. I think that's important. Not to focus on just one or a few, but have different parameters that focus on different aspects and put these together together with alternatives."	RIO
	"...you need to use [metrics] in combination in terms of other indicators, you need to use what we say 'a basket of metrics', you cannot evaluate people just based on one single metric. I would argue that you need several, and then of course you have different metrics evaluating different things. One thing is valuating excellence. Impact factor is going to be among that one. But then when you evaluate the education part, the capacity of someone to be a good professor, you need different educators also. So that's the first one. You need … It's never in isolation."	EP
Human input	"I don't think you can rely on one or several indicators without human input. You can't make sense of a number on its own."	EP
	"That's also why there is not one penny of research money allocated in the university that is not based on peer-review. Everything is based on peer review. So every proposal submitted is based on peer review."	PMI
	"So you can have indicators, publications of course is a good indicator, but it's an indicator. You should also do, let's say an holistic, what we call an holistic approach, have an holistic approach."	FA
Quality over quantity	"I think they should evaluate again true quality in terms of … And that isn't done easily via metric, one metric, you know. It's actually … You know I would suggest, it's just up the top of my head … That for any appointment, the people say to a researcher 'OK, please choose your best two or three papers in the past five years, and then two additional ones if you want, where there's no time limit. You know, something that you might have published 20 years ago, but is really important. And you submit that with your application. And then people … And you say to us why it's important. And then people need to of course evaluate that. So that would get away to, that would take into account the more longer term strategy of someone, but also … But it needs a qualitative."	EP
Transparent (robust and valid) indicators	"I think that it needs to be transparent, robust, validated, etc. […] Everybody can see the methodology, they can reproduce it if they want. […] As long as you describe really well what the rankings take into account, and why you are first or hundredth … What would be bad is if you rank and you don't tell people on what basis you rank. There you go. You might disagree with the indicator, you might disagree with the ranking, but as long as it's transparent, well validated, robust, etc. There you go."	EP

especially publications and impact factors). Interviewees believed that relying on one or few metrics generated important biases, opened the door to manipulation, and ignored important processes which relied both on different metrics and on other types of evaluations, such as openness, societal impact, or science communication.

Human input
Second, respondents also believed that it was necessary to have human input — in the form of peer review — in the evaluation process to capture what some called a holistic view of success. Peer review was, however, said to also share important weaknesses which must be taken into account. Among those, (i) the potential for conflicting interest (especially worrisome to researchers and students who perceived that funding depended more on status and network than on the quality of the project proposed), (ii) conservatism (an issue that is explored further in the associated paper on problems of science [10]), (iii)

subjectivity, and (iv) costs[2] were mentioned. One research funder proposed that repeating evaluations in different contexts, institutions, and with boards of mixed affiliations could help balance these problems. Another respondent proposed that, to reduce the costs and increase the availability of peer-review, peer-review itself should be rewarded in research assessment.

> "Why shouldn't people be given credit for doing this kind of work? It's really important work, it keeps the whole academic system alive. So I think it's crazy that it's not included as a, you know, a metric, a

---

[2]This aspect was raised when discussing external expert panels for integrity issues, in which one RIO mentioned that "They're paid for leading the report and making the preparations, you have to bring them to your university, you have to put them in a nice hotel, obviously, you have to dine, I mean, the amount of money is just enormous. And then you have... ok what is coming out of this? You have some remarks... […] Yeah I'm a bit critical towards that system".

possible metric or an indicator of being a successful scientist!" (RIN)

Quality over quantity

Third, the importance of evaluating the quality over the quantity was raised many times by different research actors. Many proposed that presenting only a subset of the most relevant work (e.g., three papers most important to the researchers, and why) could help by permitting in depth evaluation rather than reliance on quantity and metrics. Nevertheless, funders and policy makers mentioned that despite criticism from researchers about the over reliance on quantity, peer reviewers — generally researchers themselves — often asked for the full list of publications, the H index, or other quantifiable indicators when evaluating proposals, even when the proposal was purposively adapted to contain only a subset of relevant work. Overcoming this quantifiable culture thus seems to be a must for initiating a change.

Transparent, robust, and valid indicators

Finally, the transparency, robustness (consistency between evaluations) and validity (measuring what is intended) of indicators were also mentioned as a requirement for good evaluation. These last criteria are basic criteria for any reliable metric, yet they are not always met by newly proposed indicators, and the way current indicators are used sometimes compromises the validity of the intended measure (e.g., assessing *quality* of *single* publications using the impact factor, which qualifies average *journal* citations). Added to these four essential characteristics, the importance of being consistent in how evaluations are conducted while considering differences in fields and disciplines were often raised by interviewees.

A wish for change

At the end of our interviews, we ask participants what they would do if they had a 'fairy wish' to changes anything in science. In other words, we ask them to describe one aspect of science they believe needs priority for change. Although not all answers targeted research assessments, the majority of respondents discussed changes relating to research assessments or research funding as their 'fairy wish'. In changing research assessments, the need to value quality over quantity, to reduce output pressure and competition, and to broaden and adapt indicators of success to reflect not only the output, but other aspects of science were mentioned. In changing research funding, participants raised the need for fairer evaluations and distribution methods as well as the wish for long-term funding schemes and baseline research allowances. Supplementary file section 7

illustrates these ideas with a selection of quotes from diverse participants.

Discussion

To advance or even maintain their career, researchers need to be successful. But meanings of success in science are not univocal. Different research actors shared their perspective with us, depicting success as a multifactorial, context-dependent, and mutable construct which is difficult to define. In many cases, our participants' personal views on success conflicted with how success is currently determined by research assessments. We are not the first to showcase this conflict. For instance, our findings are in high agreement with those of the 2014 Nuffield Council on Bioethics report on the research culture in the UK [12]. The report highlights that research assessments and funding structures in place in the UK[3] risk increasing pressures and competition, and argues that current cultures and assessments sometimes fail to encourage elements essential to research quality, such as openness, transparency, and high-risk research. Other recent reports from different organisations further reinforce these perspectives [14, 22, 23]. The current paper adds to this discussion by showing that, even when considering the perspectives of different research actors on the way success is defined in science, research assessments generate a lot of criticism. One recurrent criticism was the fact that current assessments over-rely on research *outputs*, thereby ignoring, if not discouraging, important *processes* that contribute to the quality of research. This issue is central to the current discussions on the topic. For instance, just a few months ago, Jeremy Farrar, director of the Wellcome Trust (UK), stated that the "relentless drive for research excellence has created a culture in modern science that cares exclusively about what is achieved and not about how it is achieved" [24]. Resonating this perspective, The Hong Kong Principles for Assessing Researchers [6] propose that researchers should be assessed on the process of science, including responsible practices (principle 1), transparency (principle 2), and openness (principle 3), and that a diversity of research activities, such as scholarship and outreach should be taken into account (principles 4 and 5). Part of the criticism when assessing outputs also comes from the dominance of inflexible

---

[3]In this case the UK Research Excellence Framework (REF) was criticized for accentuating pressures to publish in high impact factor journals. The UK REF bears considerable similarity with the funding system in place in Flanders. In Flanders, a portion of the federal funding for research institutions is distributed using a special key called the BOF (bijzondere onderzoeksfonds [special research funds]) key [20]. Much like in the UK REF, publications in journals with an impact factors have a considerable weight in the final score for funding distribution between Flemish institutions [21].

and reductionistic metrics, an issue that was also significant in our findings. Echoing such concerns, the Declaration on Research Assessments (DORA; 3) argues against using impact factors for individual evaluations, while the Leiden Manifesto and the Metrics Tide pledge for the development and adoption of better, fairer, and more responsible metrics [4, 5]. In this regard, the issues raised by our interviewees are at the heart of current discussions on research assessments. Yet, our findings also reveal that perspectives on specific indicators remain multi-sided, and that priorities for change are far from univocal. This variety of perspectives may suggest that the core objectives of research assessments are disputed, and that specific actor roles may come into play. For instance, it seems reasonable that research assessments are put in place to ensure that universities *nurture and value good researchers*; that funders and policy makers *maximize the societal value of science*; and that publishers, editors, and researchers *advance knowledge*. To place our findings in perspective, we decided to revisit these three objectives and to question whether current research assessments succeed in fulfilling them.

First, research assessments may aim to *promote and value good researchers*. Our respondents suggest that success in science has an important individual facet. Indeed, success was often attached to personal skills, competencies, and efforts. Research assessments are thus expected to be, at least in part, meant for the fair recognition of researchers' accomplishments. Current research assessments however, were said to fall short on their objective to fairly reward researchers. Interviewees expressed their concern for fairness by blaming (bad) luck for (lack of) success, and by worrying that connections, seniority, and renown could yield unfair advantages which are not related to genuine merit [10]. Valuing good researchers could also mean building capacities and nurturing autonomy in order to create strong and sustainable research units. Accordingly, if a goal for research assessments is to promote excellent researchers, they should also facilitate, support, and sustain the creation of strong research teams by valuing teamwork, diversity, inclusion, and collegiality. Yet, as we describe in the associate paper, interviewees identified important problems in current research climates which inhibit these essential features and foster competition, mutual blame, and mistrust instead [10]. Consequently, current research assessments fall short on their objective to promote and value good researchers.

Second, research assessments may aim to *maximise the benefits that science can bring to society*. Since science is primarily financed through public money, it is often argued that it should profit back (tangibly and intellectually) to society. Following this perspective, research assessments should aim to ensure that scientists involve, communicate, and implement their findings within society. Translational efforts, public engagement, science communication, open access, and feasibility should thus be at the heart of research assessments. Past research has shown that this is not the case [25]. In our findings, we show that science communication generated polarised views among interviewees: some participants considered that it was an intrinsic element of the quality of science while others considered that it was irrelevant or event detrimental to quality. Our findings also suggest that the value of open science and the desire for implementation of research findings tend to diminish with seniority. While this finding may be anecdotal to our sample, it could also suggest that the broad neglect for societal value in current assessments shapes researchers' perspectives of success, encouraging them to prioritize competition and metrics over openness and societal value. Consequently, research assessments may not only fall short on their objective to benefit society, but they may even impact the core culture of acceptance of the public dimensions of research.

Finally, research assessment may aim *for advancing science and knowledge*. Two aspects are then essential to consider here. First, to advance science, we need to ensure that research is conducted with integrity. Assessments should thus encourage the processes which maximise the integrity and the quality of research. We have shown however that openness, reproducibility, rigorousness, and transparency were recurrently mentioned as missing from current research assessments. Certain aspects of current assessments were even thought to discourage integrity and research quality. Many interviewees supported that the lack of consideration for negative results caused tremendous research waste, that competitiveness of assessments compromised transparency, and that the current focus on 'extraordinary findings' discouraged honest reporting and high quality research [10]. Evidently, if research assessments aim to promote the advancement of science, processes which foster integrity and quality must be given due recognition. But even when integrity and quality of research are ensured, advancing science requires continued innovation, creativity, and productivity. According to our findings, this is where most research assessment currently focus. Publications counts and impact metrics are meant to ensure that novel knowledge is created and that the pace of creation is fast. Yet, the overemphasis on outputs and the negligence of processes also risks shifting the focus from 'what is needed to advance the field' to 'what is sexy to publish', or 'what will attract funding'. Current assessment systems were further criticized for their being conservative, for discouraging high-risk research, and for relying on short-term funding schemes (also see 10). Innovative research requires long-

term investments and sufficient freedom for failure. In sum, current research assessments do not even seem optimized to help advancing science and knowledge.

Addressing these key objectives in future discussions on research assessments could help build a common understanding. Ensuring that the discussion on research assessments listens to the perspectives of all research actors and that all parties are transparent and explicit about what they wish to achieve by assessing researchers may be a first step for an open dialogue to enable concrete changes to take place.

Considering the current failures on each of these key objectives, it is clear that research assessments need to change. While it would be impossible — if not detrimental — to impose a single perspective of success in science, it remains important to discuss the core objectives of research assessments and to expose their respective failures. In doing so, the perspectives of all actors must be heard — including the neglected voices of early career researchers and researchers who left academia. Including those voices in our work allowed us to understand that not only do we need to change indicators to consider research processes and quality, but we also need to revisit the format of research assessments and research careers to value team efforts and collaboration instead of individual competition. Finally, research assessments cannot change without the joint effort of everyone involved. Funders, institutions, and journals have obvious roles to play in changing what they consider successful in research. But researchers — including those who built their entire career on the current assessments — also need to have the modesty, openness, and courage to change the cultures that decades of inadequate metrics have solidified.

Study limitations

A few points are important to consider when interpreting our findings. First, given the exploratory and qualitative nature of this project, our sample is neither exhaustive nor fully representative. We chose to ask for personal perspectives rather than official institution or organisation views since we believed it would allow us to capture genuine beliefs and opinions and avoid rote answers. We thus encouraged participants to share their personal thoughts rather than the thoughts that could be attributed to their entire actor groups, institution, or organisation. We consider that these personal beliefs and opinions are crucial in shaping the more general views of organisations, yet we urge our readers to remain careful when making group comparison and generalisations.

Adding to the above concern, it is important to keep in mind that the research world is complex and not organized in distinct actor groups. Participants could often fit in more than one category by endorsing several research roles. We asked all participants whether they agreed with the category we assigned them in, and we refined and exemplified the definitions of our actor groups to reflect the participants' distinctions. Yet, we must consider each actor category not as a closed box with characteristic opinions, but as a continuum which may or may not hold divergent views from other actor groups. Our findings help capture views which may have been overlooked in past research which focused on researchers, but should not be used to discriminate or represent the opinions of entire actor groups.

Finally, it is important to consider that given the richness of the information gathered, certain findings may be displayed with greater importance than others simply based on the authors' personal interests. We were careful to include also the views we disagreed with or found to be of limited interest, yet it is inevitable that some of the selection and interpretation of our findings was influenced by our own perspectives. To maximise transparency on the genuine views of our informers, we supplement our interpretation of the findings with quotes whenever possible, most of which are available in the supplementary file.

Conclusion

The present paper describes the perspectives of different research actors on what defines and determines success in research. In their answers, interviewees raised a number of shortcomings about the approaches currently used for assessing success in science, and these shortcomings lead to important problems in the functioning of science (see the associated paper 10).

Assessing researchers is an issue that has high stakes, not only for individual researchers who wish to continue their career and seek recognition, but also for the future of science. Our findings reiterate that current research assessments need to be revisited, that all research actors must be involved in the discussion, and that the dialogue must be open, inclusive, transparent, and explicit. Acceptability, trust, and joint efforts can only be increased if all actors are involved, understand the other's perspective, and work together to build a solution.

> **Additional file 1.**

changed career; Re-SInC: Rethinking Success, Integrity, and Culture in Science; RIL: Research institution leader(s); RIN: Research integrity network member(s); RIO: Research integrity office member(s); WCRI: World Conference on Research Integrity

Authors' contributions
Authorship contributions: NAB contributed to the Conceptualization, Project administration, Methodology, Resources, Investigation, Data curation, Formal Analysis, Visualization, Validation, Writing – original draft, and Writing – review & editing. WP contributed to the Conceptualization, Funding acquisition, Methodology, Resources, Validation, Supervision, and Writing – review & editing, both intermediate and final versions. All authors read and approved the final manuscript.

Authors' information
NAB is a PhD student supervised by WP at the faculty of Health and Life Sciences in Hasselt University, Belgium. Her research interests are predominantly in research integrity and publication ethics, especially targeting research assessments and the way in which scientific systems influence the integrity and the quality of science.
WP is associate professor at the faculty of Health and Life Sciences in Hasselt University, Belgium. As a medical ethicist with expertise in research ethics, he has increasingly devoted attention to education and research in the field of research integrity, with a specific interest in research culture and the definition of success in science. He is supervisor of NAB.

Availability of data and materials
Data: Given the sensibility and risk of identification of qualitative data, we are unable to share full transcripts. In the manuscript however, we attempted to be as transparent as possible by providing numerous quotes in the text and in tables to exemplify and support our claims.
Materials: Our focus group and interview guides, as well as the consent forms and information sheet we shared with participants are available on the registration of this project [11].

Ethics approval and consent to participate
The project was approved by the Medical Ethics Committee of the Faculty of Medicine and Life Sciences of Hasselt University (protocol number CME2016/ 679), and all participants provided written consent for participation.

Consent for publication
Participants provided written consent for participation and for publication and dissemination of the findings from this project.

Competing interests
NAB has received an award with financial reward from the World Conference on Research Integrity (WCRI) at the 5th WCRI in 2017 for the literature review that lead to this work, and a travel award from the Research Foundation - Flanders (FWO) to present these findings at the 6th World Conference on Research Integrity in Hong Kong. WP has no conflicting interests to declare.

References
1. Moore S, Neylon C, Paul Eve M, Paul O'Donnell D, Pattinson D. "Excellence R Us": university research and the fetishisation of excellence. Palgrave Communications. 2017;3:16105.
2. Hatch A. To fix research assessment, swap slogans for definitions. Nature. 2019;576(9):9.
3. American Society for Cell Biology. San Francisco Declaration on Research Assessment. 2013. Available from: https://sfdora.org/read/.
4. Hicks D, Wouters P, Waltman L, Rijcke SD, Rafols I. The Leiden manifesto for research metrics. Nature News. 2015;520:429–31.
5. Wilsdon J, Liz Allen, Belfiore E, Campbell P, Curry S, Hill S, et al. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. 2015. Available from: https://re.ukri.org/documents/hefce-documents/metric-tide-2015-pdf/.
6. Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V. Dirnagl U. The Hong Kong Principles for assessing researchers: Fostering research integrity. PLoS Biology. 2020;18(7):e3000737. https://doi.org/10.1371/journal.pbio.3000737.
7. Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JPA, Goodman SN. Assessing scientists for hiring, promotion, and tenure. PLoS Biol. 2018;16(3): e2004089.
8. Saenen B, Borell-Damián L. EUA briefing - reflections on university research assessment: key concepts, issues and actors. 2019. Available from: https://eua.eu/resources/publications/825:reflections-on-university-research-assessment-key-concepts,-issues-and-actors.html.
9. Aubert Bonn N, Pinxten W. A decade of empirical research on research integrity: what have we (not) looked at? J Empir Res Hum Res Ethics. 2019; 14(4):338–52.
10. Aubert Bonn N, Pinxten W. Rethinking success, integrity, and culture in research (part 2) - a multi-actor qualitative study on problems of science. Research Integrity and Peer Review in review.
11. Aubert Bonn N, Pinxten W. Rethinking success, integrity, and culture in science (Re-SInC); 2016. https://doi.org/10.17605/osf.io/ap4kn.
12. Bioethics NCo. The culture of scientific research in the UK. 2014. Available from: https://nuffieldbioethics.org/publications/the-culture-of-scientific-research.
13. Wellcome Trust, Shift Learning. What researchers think about the culture they work in. 2020. Available from: https://wellcome.ac.uk/sites/default/files/what-researchers-think-about-the-culture-they-work-in.pdf.
14. Metcalfe J, Wheat K, Munafò M, Parry J. Research integrity: a landscape study. 2020; Available from: https://www.vitae.ac.uk/vitae-publications/reports/research-integrity-a-landscape-study.
15. Buljan I, Barać L, Marušić A. How researchers perceive research misconduct in biomedicine and how they would prevent it: a qualitative study in a small scientific community. Account Res. 2018;25(4):220–38.
16. Anderson MS, Ronning EA, De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. Sci Eng Ethics. 2007;13(4):437–61.
17. Davies SR. An ethics of the system: talking to scientists about research integrity. Sci Eng Ethics. 2019;25(4):1235–53.
18. Elo S, Kyngas H. The qualitative content analysis process. J Adv Nurs. 2008; 62(1):107–15.
19. Dierckx de Casterlé B, Gastmans C, Bryon E, Denier Y. QUAGOL: A guide for qualitative data analysis. Int J Nurs Stud. 2012;49(3):360–71.
20. Decreet betreffende de organisatie en financiering van het wetenschaps- en innovatiebeleid [Decree concerning the organization and financing of science and innovation policy], (2009).
21. Zacharewicz KJT. Research Performance Based Funding Systems: a Comparative Assessment. 2016. Available from: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101043/kj1a27837enn.pdf. Report No.: JRC101043.
22. Saenen B, Morais R, Gaillard V, Borrell-Damián L. Research Assessment in the Transition to Open Science: 2019 EUA Open Science and Access Survey Results 2019. Available from: https://www.eua.eu/downloads/publications/research%20assessment%20in%20the%20transition%20to%20open%20science.pdf.
23. The Wellcome Trust and Shift Learning. What Researchers Think About the Culture They Work In 2020 [updated 15 January. Available from: https://wellcome.ac.uk/reports/what-researchers-think-about-research-culture.

24. Farrar J. Wellcome - News and views 2019. Available from: https://wellcome.
     ac.uk/news/why-we-need-reimagine-how-we-do-research.
25. Alperin JP, Muñoz Nieves C, Schimanski LA, Fischman GE, Niles MT,
     McKiernan EC. How significant are the public dimensions of faculty work in
     review, promotion and tenu0Fre documents? Elife. 2019;8:e42254.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in
published maps and institutional affiliations.