

## VISION STATEMENT FOR EVALUATING RESEARCH AT GHENT UNIVERSITY

(approved by the Board of Governors on 8 November 2016)

### CONTEXT

**Indicator-driven methods to evaluate research are particularly prevalent at Ghent University.** Since it is easier to translate the quantity of research into indicators rather than quality, indicator-driven evaluations tend to target the quantity more than the quality of research.

This quantity-driven focus can be attributed to **intrinsic motivation**, on the one hand – Ghent University’s ambition to clear its backlog of research visibility two decades ago – and **extrinsic motivation**, on the other – a funding mechanism by the Flemish Government which, since 2003, has mainly stimulated quantity and – only to a very limited extent – quality. The focus on quantitative indicators has served the University well. It allowed Ghent University to grow in the early 21st century from a modest higher-education establishment with regional focus and a few pockets of high-quality research, into a fully-fledged, research-driven higher-education establishment, recognized worldwide for its wide range of high-level research in which governments and private partners are happy to invest their research funds.

In many ways, the quantity-driven system is an *internal driver* for improvement. Moreover, the indicator-driven methods are increasingly better at passing a balanced, quality judgement, such as, for example, on the basis of citation analyses.<sup>1</sup> In several applications at Ghent University, however, **an excessive focus on quantitative indicators compromises the quality of the research.** As a consequence this creates risks on a number of fronts: for the integrity of the research process, for the chances of science finding a place in society and even for the personal ambition of researchers.

---

<sup>1</sup> In this paper, we draw a distinction between an indicator-driven assessment, on the one hand, and an assessment by peer-review experts, on the other. Many indicators capture the *quantity* of research in one mark (e.g. the number of publications, the number of doctorates); more and more indicators also approximate the *quality or the impact* of the research (e.g. citation impact) – even though these quality and impact indicators are more often than not controversial. We do not argue, therefore, that an indicator-driven assessment cannot take quality into consideration. We do warn, however, that quality-driven indicators at best have a *proxy* value and so cannot be used as a substitute for a personal assessment by experts or stakeholders.

In addition, the *extrinsic drivers and context* have changed. First of all, the **funding mechanism of the Flemish Government** has been adapted to the extent that – at least for the exact, applied and biomedical sciences – **a sustainable investment in quality is rewarded much more than the quantity of the achievements**.

Second, thanks to globalization, **international standards** have become more important than ever to acquire funding (e.g. EU funds), for the international career development of researchers and for the appeal of Ghent University as a research partner, employer, place of learning – even if these international standards are at odds with the Flemish funding model.

Third, while the practice of research evaluation is under scrutiny **internationally**, ever starker warnings are sent out regarding the undesired side effects of a quantity-driven model. Pioneering initiatives in this are the San Francisco Declaration on Research Assessment (DORA, 2012)<sup>2</sup> and the Leiden Manifesto (2015).<sup>3</sup>

Fourth, we point out that the research world has developed such a strong evaluation culture that the **risk of over-evaluation** is very real. Every investor, employer, manager or supervisor wants the guarantee that the investments made into research will pay off – but the way in which this is checked separately often leads to an accumulation of evaluation processes for the same individuals involved.

## A BETTER VISION, A BETTER IMPLEMENTATION

Substituting a quantity-driven/indicator-driven assessment for a model that is more focused on quality-based assessment/peer-review assessment is not that simple. First of all, there are different criteria and processes *within* each of these methods. Sometimes, one indicator is more reliable than the other, or one working method may be better suited than the other.

**Moreover, both approaches have their advantages and disadvantages.** In a polarized representation, indicator-driven methods are efficient and transparent but lack depth and ignore the complexity and diversity of the research, while peer-review-driven methods tend to be less efficient and transparent but do have an eye for depth, complexity and diversity. **In an evaluation context, both approaches are, therefore, best used in a complementary capacity.**

Depending on whether the level of aggregation of the evaluation is high, the indicator component will be more prominent, while the peer-review component often prevails in

---

<sup>2</sup> <http://www.ascb.org/dora/> The most significant recommendation in DORA is the elimination of the Journal Impact Factor in all evaluations for funding, appointment or promotion and the appreciation of the content of publications, rather than the academic journal in which your paper is published. In addition, it calls for a more efficient use of the digital scope. In each of its specific recommendations, it is the transparency in the evaluation process that takes centre stage.

<sup>3</sup> <http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351> This manifesto, drawn up by experts in bibliometrics, examines the many pitfalls that can arise in the evaluation process because readily available indicators can easily obstruct a proper assessment. Very concrete do's and don'ts offer a guideline to allow bibliometrics and peer review to reinforce each other.

personalized evaluations. The balance, however, is not set in stone. Depending on the objective of the evaluation – which is, in turn, closely linked to the level of aggregation – the ideal balance between appropriate indicators and appropriate evaluation methods shifts from one to the other.

To complete the nuance, we draw attention to the fact that the opposites "quantity >< quality" do not exactly match the opposites "indicator-driven >< peer-review-driven", since indicators are increasingly better at measuring quality aspects, and more and more peer-review evaluators retrieve their information from indicators. Both the expertise and the practice of research evaluations evolve with the available technologies and insights.

Ghent University thus needs a conceptual approach that helps (1) strike **the right balance between indicator-driven and peer-review-driven assessment methods for each objective**; (2) **guarantee** that each of these methods are **properly applied**; and (3) **build sufficient flexibility into the system** so that we can respond to methodological innovation in indicators, to general changes in the research context and to discipline-specific developments in a timely manner.

This vision statement has been drawn up on the Research Council's initiative as a basis for evaluating research at Ghent University - whether it is about the recruitment of researchers, distributing funds from the university's Special Research Fund (BOF), a large-scale evaluation of a research discipline or a decision on the promotion of a professorial staff member.

This statement is to be followed up by a more concrete implementation of each of these principles in the case of, for example, the recruitment of professors, evaluations of research projects or promotion in the academic career.

## EIGHT BUILDING BLOCKS FOR A QUALITY EVALUATION OF RESEARCH

Ghent University's view of research evaluation is based on the following eight building blocks:

**1. THE CHOICE OF AN APPROPRIATE EVALUATION METHOD FOR RESEARCH IS IN LINE WITH THE OBJECTIVE OF THE EVALUATION.** Since the choice of the most appropriate evaluation method depends on the objective of the evaluation, the objective of each evaluation should be explicitly determined for each type of evaluation.

In the following list, the eight main targets for evaluation are listed in order of prevalence of qualitative tools (such as a peer-review report of the academic reputation of a publisher) and quantitative instruments (such as publication numbers, average citation impact or number of opinion pieces in national newspapers). There will only be a limited number of evaluation methods in which only quality or only quantitative criteria are used, but **the more personal the evaluation, and the more process-related, the more the evaluation will be based on an**

**assessment by experts/peers from the portfolio of the researcher, the research team or the establishment.**<sup>4</sup>

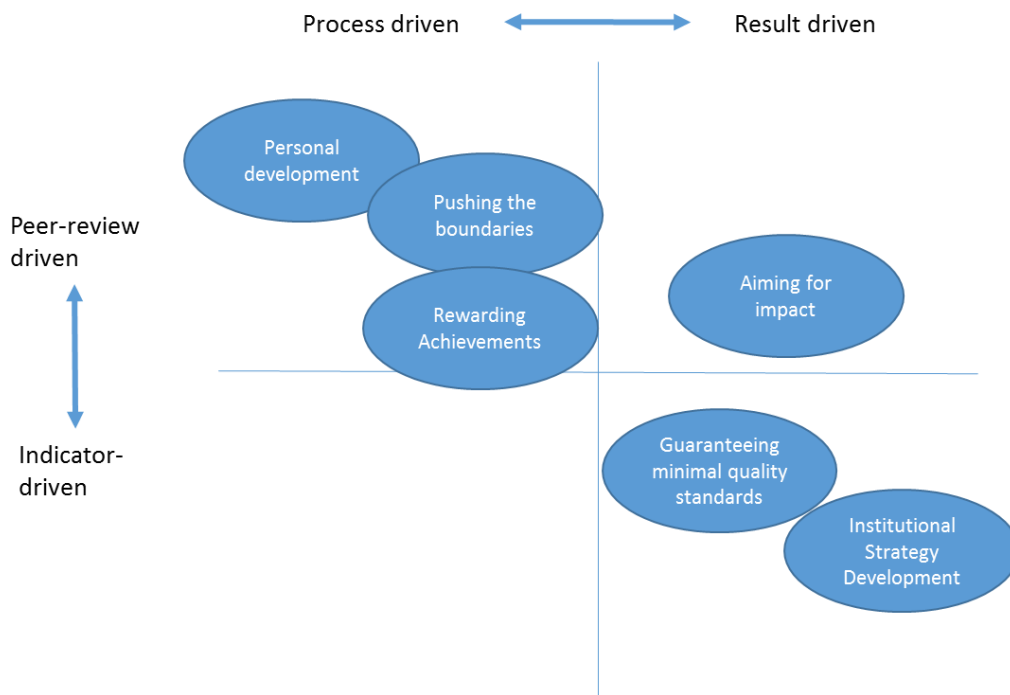
Criteria used at a **broad level of aggregation** for a particular purpose cannot simply be translated into **finer levels of aggregation** that pursue other goals.

- **PERSONAL DEVELOPMENT:** The evaluation aims to support the researchers' personal development, to help them find their place in their field. The evaluation is future-oriented and potential-driven; not sanctioning.
- **REWARDING ACHIEVEMENTS:** promotions and prizes are awarded on the basis of an ex-post evaluation that rewards previous research achievements. The expectation of this reward is often seen as an indirect incentive to influence some form of research behaviour - even more so if there is a realistic chance to actually obtain this reward.
- **PUSHING THE BOUNDARIES:** The science system strives for excellence in the form of ground-breaking research. Therefore, excellent achievements from the past are often regarded as a predictor for outstanding achievements in the future, but potential assessment also takes into account issues of scale, management skills, vision and motivation.
- **AIMING FOR IMPACT:** Research (results) is (are) assessed in terms of the potential to make a difference in the practice or in the work of colleagues (academic impact), the potential for innovation and application (economic impact), or with a view to explaining/changing/ improving social processes (societal impact). External stakeholders and internal experts offer complementary visions.
- **GUARANTEEING MINIMUM QUALITY STANDARDS:** The evaluation aims to detect substandard quality and remove it from the ecosystem of research. This form of quality control monitors the 'lower limit' of research quality and sets minimum requirements. This fits in with the social responsibility for spending public funds.
- **INSTITUTIONAL STRATEGY DEVELOPMENT:** Research is evaluated with a view to the strategic positioning of the institution itself: to identify in-house talents, strengths and potential.

In many - but not all - evaluation forms, we see a **correlation** between a process-related evaluation and qualitative tools, on the one hand, and between a results-oriented evaluation and quantitative indicators, on the other:

---

<sup>4</sup> "Peer review" is a personalized, expertise-driven assessment by experienced fellow-academics. This may be an internal expert (e.g. a supervisor of a doctoral student within the university) or an external expert (e.g. an anonymous reviewer who assesses a scientific article for a magazine). When in this document, only the last group of external reviewers is meant, the term "external" is added in each case.



The **DISTRIBUTION of FUNDING** has deliberately not been included in the above list as an 'objective', even though research funding is distributed according to research achievements and research proposals. The distribution of funds, however, is not an end in itself, but a means to achieve other, more important objectives, not least giving researchers the opportunity to develop and/or pushing the frontiers of science. So within the practice of distributing research funds, a distinction should be drawn between different evaluation criteria, each tailored to the main objective of the funding instrument.

**2. THE EVALUATION TAKES INTO ACCOUNT THE INTENDED IMPACT OF THE RESEARCH; STRICTLY ACADEMIC, ECONOMIC, SOCIETAL,** or a combination of these. Even though there is scope for different types of research at a university, 'impact' forms an inherent part of excellence, and this applies to all forms of research.

In the evaluation of research, we consider 'impact' as an intrinsic part of excellence. Any kind of research has to be based on excellent methods and sources. Yet, one research project will rather go for the curiosity-driven, ground-breaking aspect of science, the other will focus more on innovative applications with research-transfer value, while yet other research initiatives mainly seek to change societal processes or applications. Quality research is research that has an impact on science (the field itself or larger), the economy or society.

For the scientific or economic impact, many indicators have already been developed that roughly indicate what the impact is on science or economics. These include, for example, the average citation impact, highly-cited papers, or the article influence score on the one hand, and patents, spin-offs or employment figures on the other.

The instruments for measuring societal impact, however, are less well developed. Partly for that reason, this role of science has long been neglected. Societal impact will, therefore, be assessed on the basis of a combination of processes (peer review, storytelling and alternative

measuring instruments), by various actors (peers, stakeholders and experts) and with due regard for the broad range of impact-related aspects.

Ghent University also monitors the balance between fundamental and applied (societal + economic) research. While scope for ground-breaking research must be guaranteed, the relevance for society deserves further consideration.

**3. THE EVALUATION TAKES INTO ACCOUNT THE DIVERSITY BETWEEN DISCIPLINES.** The publication and citation culture, as well as the research-transfer potential of research vary from field to field; the most appropriate evaluation method must, therefore, be developed in consultation with the experts in the field.

In some fields, there is a relatively **strong consensus** on the appropriate indicators for assessing the quality of the research, and these indicators are also readily available. The biomedical and exact sciences spring to mind, where articles published in high-impact journals and the indicators derived from this are regarded as indicators of excellence. Unfortunately, this consensus does not apply to all areas and niches.

There is also diversity in approaches within one and the same discipline: one expert may be attracted to **risky fundamental research**, while the ambition of the other may lie with the **impact of research in society**.

Top journals are not always open to **specialized niches** within dominant disciplines. Moreover, **risky, ground-breaking** or **interdisciplinary** research does not usually find its way to these publication channels without considerable delay.

In other disciplines, especially the applied and social sciences as well as the humanities, other publishing outlets also play a key role alongside Web of Science journals, like, for example, conference proceedings or books. That means there is **less consensus** on indicator-driven quality criteria based on citation analysis.

Finally, research which is **excellent** but only **locally relevant** may not be disadvantaged by the fact that it is difficult to publish it in English-language journals with a global audience.

The best way to appreciate this diversity will vary according to the level of aggregation. At the individual level, the one-size-fits-all approach is to be avoided at all costs.

**4. FOR EACH CHOSEN EVALUATION METHOD, THE SIMPLICITY OF THE PROCEDURE IS WEIGHED UP AGAINST THE COMPLEXITY OF THE RESEARCH.** No evaluation of the quality of research may be based on one indicator only.

Ghent University has invested in a **proper information system** of research output (biblio, IWETO/FRIS) for many years and is currently extending this to other research-related areas (Gismo). As soon as publications and activities are properly registered in the information system, the **administrative burden on researchers** in the context of an evaluation (promotion, project application, ...) should be reduced to a minimum.

The paperwork must also be cut down by **only requesting information that will actually be used**. The person responsible for the evaluation should, therefore, carefully consider the information that is absolutely essential to be able to apply the criteria that are used in the evaluation.

**No matter how tempting the simplicity of this information system is, it should not stand in the way of a nuanced assessment by peer-review experts**, not least in the case of individual evaluations and in cases where the main objective is furthering the researcher's personal development or stimulating ground-breaking research. Indicators can at most offer experts a starting point for a thorough assessment.

The value and usefulness of many indicators are also often the subject of discussion. Sometimes, they offer nothing but false precision (the marginal difference between impact factors after the decimal point, for example, has no value), while at other times, they can be manipulated.

A **combination of multiple criteria** is hence necessary to appreciate the complexity of the research and the research system. They can never be reduced to one indicator - not even a single composite indicator.

**5. THE EVALUATION CRITERIA ARE DRAWN UP AND COMMUNICATED TO ALL STAKEHOLDERS IN ADVANCE.** These criteria must be clear and transparent to all parties involved in the evaluation of research (committee members, authors of project proposals, applicants, ...).

A thorough consideration of the appropriate evaluation method and criteria is done **in advance in consultation with all stakeholders**. Experts in the field and experts in research quality assurance play an important role in this.

All those involved in this evaluation (for example, external members of an assessment committee) have taken note of the applied evaluation method and criteria.

The evaluation process **must then abide by these criteria**. As soon as a procedure is up and running – e.g. a recruitment procedure, a selection process after a call for projects, a promotion cycle – no new selection criteria can be added which were not known at the outset.

This is not to say that criteria cannot be adjusted over time. However, within a specific evaluation cycle or procedure, the rules cannot be modified ad hoc.

**6. THERE ARE SUFFICIENT EXPERTS ON THE EVALUATION COMMITTEE WHO ARE IN A POSITION TO ADEQUATELY ASSESS THE QUALITY OF THE RESEARCH.** Those who are evaluated must be confident that they are assessed objectively.

When peer-review experts are involved in the evaluation of research (e.g. doctoral committee, selection committee, promotion committee), they must meet several conditions:

- ❖ Committees must be **balanced** in their composition, both in terms of gender and in terms of discipline-specific experience.

- ❖ Depending on the objective of the evaluation, the **involvement of external experts** may be greater or smaller.
- ❖ Depending on the objective of the application, the degree of expertise can vary. **Generic experts** can have a keen eye for research management or interdisciplinary opportunities; **niche experts** are necessary in order to adequately assess risky research or research at a high level of excellence.
- ❖ Members of the evaluation committee should be able to take **sufficient distance** from the person(s) evaluated. In the event that their independence is not guaranteed or in the event of a potential conflict of interest, a solution is sought that is fair to all concerned.

Indicators support an evaluation committee in their work and contribute to the quality of the peer review. They cannot, however, replace an assessment by experts. It is, therefore, the **experts who bear full responsibility** for the evaluation they conduct, and not the criteria.

## 7. THE ABOVE PRINCIPLES ARE IMPLEMENTED BY MEANS OF A SMART CHOICE OF EVALUATION INDICATORS AND BY ADOPTING A HOLISTIC APPROACH TO PEER REVIEW.

The application of the above principles will affect the choice of the indicators used. In addition, by applying these principles to peer-review processes (where indicators and a personal, but objective, assessment by experts often go hand in hand), the evaluation forms will not just be **output-oriented** but also **process-oriented**.

While for a number of frequently used indicators, Ghent University adopts the following positions, new developments in the evaluation of scientific research may necessitate regular updates or revisions:

- ❖ **Impact Factors:** Even though impact factors can be easily manipulated, they are still commonly used on an international scale to distinguish prestigious from less reputable journals. We cannot, therefore, ignore their existence within a research policy. Young researchers should become familiar with the (changing) importance of impact factors in their field, and develop a critical view of the practice of editors and reviewers and the manipulability of this indicator.  
However, in evaluation practice, impact factors are *not suitable* for cross-disciplinary comparisons or for evaluations of the research at individual level. Rankings of journals based on impact factors (e.g. the top 5% or 10% journals with the highest impact factor) can, with some caution, be used when they are the only viable alternative to other, more complex evaluation forms.  
For that reason, Ghent University cannot endorse the San Francisco Declaration on Research Assessment. This decision, however, does not stand in the way of an overall appreciation of the principles in this statement.
- ❖ There are better alternatives available for impact factors to try to summarize quantity and quality in indicators, such as the **Normalized Citation Impact**. This measures the academic impact of one publication or a group of publications, taking into account the citation intensity of the discipline. Additionally, bibliometricians recommend using a combination of indicators in any assessment (e.g. the Article Influence Score, H-Index,



Scimago Journal Rank, Relative Citation Rate) instead of basing an assessment on a single type indicator.

- ❖ Excellent work is also published in journals with a lower, or no, impact factor. Especially in the case of individual evaluations, bibliometric **assessment criteria at article level** are more reliable than those at journal level.
- ❖ **Book Publications in the Humanities and Social Sciences** at Ghent University deserve the same appreciation they receive internationally. Researchers should be encouraged to publish their work at a reputable academic publisher that sets the bar high in terms of quality content and peer review, editorial involvement, design and international distribution. In assessment contexts, high-standard monographs should, therefore, be taken into account.
- ❖ **The Flemish Academic Bibliography (Het Vlaams Academisch Bibliografisch Bestand - VABB)** is a selection of scientific peer reviewed publications, but does not distinguish between the quality of publications. Given the approval process via a Flemish committee, the VABB is not ideally placed to assess the publications of academics working outside Flanders. In evaluations or allocation keys on a sufficiently broad level of aggregation, however, publications with a VABB label may be useful, particularly when alternative options are unavailable.
- ❖ **Co-publications** resulting from academic cooperation are, in some cases and, depending on the submitted research topic and research objective, more innovative. Consequently, in these cases they generate a greater impact than academic publications by one single author. Quality cooperation can in those cases be assessed by, for example, gauging the author's specific contribution in the evaluation method.
- ❖ **Quality is more important than quantity.** In each of the evaluation processes at Ghent University, it should be considered whether, for example, a brief description of the academic, economic and societal impact of the results indicated in a limited set of publications would not better serve the quality of the assessment process rather than a listing of publications with their possible quality indicators (ranking impact factor, average citation impact, etc.)

Not all evaluations should be output-driven. Depending on the purpose of the evaluation, as set out in the first principle, **the process of the research** will play a greater role, quite apart from the output. A quality research process includes the following components:

- ❖ **Ethics and integrity:** Academic integrity should be encouraged at every stage of the research process, from the conception of the research idea to the moment the results and data are published. Ethical aspects should be integrated in all types of research that affect humans and animals. Moreover, the way in which researchers behave within the research context, e.g. in collaboration with colleagues, is essential in quality research. Questionable Research Practices do occur in the academic world in every guise. We cannot simply assume that Codes of Ethics and integrity principles are respected in each and every instance. Not only those who take responsibility for research projects, but also (each member of) a research team and those who evaluate this work, should monitor the quality of this process. When assessing candidates in the recruitment process, when considering the quality of project proposals or when setting out a strategic vision development of research, great care must be taken to guarantee the quality of the vision, the tools and methods, as well as the internal processes that monitor integrity and ethics of research.

- ❖ **Data management:** The academic process requires transparency, not least in respect of the (substantial) data that underlie academic analysis. Sound data management is a basic requirement for this and provides additional guarantees for a flawless methodology, for sharing and reusing data by other researchers in an Open Science context and for the accountability of a researcher's own academic integrity.
- ❖ **Management and Leadership:** Research is increasingly the result of teamwork. However, even when the research is largely the result of the idea of one individual, as is the case in many human sciences, sound leadership by the head of the research group and good management of a research project are crucial elements to allow all the researchers involved to excel in their work. Since these research-driven structures do not always coincide with education-based structures, such as departments, it is not always straightforward to bring about cooperation and coherent research strategies. In addition, each researcher also bears full responsibility for the management of their own set of tasks. A good evaluation process should take into account the potential for leadership and management and/or earlier experiences of leadership and management, as well as the appropriate organizational level.

## **8. ANY COMMITTEE OR POLICY MEASURE EVALUATING RESEARCH, MAKES A BEST EFFORT COMMITMENT TO TRANSLATE THE ABOVE PRINCIPLES INTO PRACTICE.**

Evaluations in general, and indicators in particular, have a bearing on the system of which they form a part, as they invite researchers to change their behaviour. This can lead to "**gaming**" and "**goal displacement**": namely, that the indicator becomes a target in itself, and that the objective for which the indicator was a measuring instrument, becomes secondary.

A **combination of indicators** is one way of avoiding, and of anticipating, this. In addition, a **meta-evaluation** of the effect of the indicators used should be conducted on a regular basis. Only in this way can the effectiveness of the research system remain guaranteed.

Indicator-based incentives or procedures which no longer reach their goal or cause undesirable side effects to the system, should be adjusted.

At regular intervals, Ghent University's Research Council takes the initiative to test the effectiveness of several research evaluation processes at Ghent University, both within its own operations and elsewhere at Ghent University, and to make recommendations for improvement.