

# ACUMEN

academic careers understood through measurement and norms

FP7 Grant Agreement | 266632

Deliverable No and Title | D1.11 The impact of evaluations on academic careers

Dissemination Level | PU (public)

Work Package | WP1 – Evaluation Impact

Version | 1.0

Release Date | 30 November 2013

Author(s) | Frank van der Most

Project Website | <http://research-acumen.eu/>

European Commission  
7th Framework Programme  
SP4 - Capacities  
Science in Society 2010  
**Grant Agreement: 266632**



# **The impact of evaluations on academic careers**

An study of academics in four disciplines, four countries and three cohorts

ACUMEN Deliverable D1.11

Frank van der Most<sup>1</sup>

e-Humanities Group and DANS, KNAW

30 November 2013

---

<sup>1</sup> frank.vandermost@inter.nl.net , Telephone: +31 6 1428 8768, [www.frankvandermost.net](http://www.frankvandermost.net)  
The e-Humanities Group, P.O.Box 94264, 1090 GG Amsterdam, The Netherlands

# Table of Contents

List of tables.....	iv
Acknowledgements.....	vi
1 Introduction.....	7
1.1 Impact, invisible work and informal evaluations.....	7
1.2 Evaluation regimes and career regimes.....	9
1.3 Structure of the report.....	12
2 Research design.....	13
2.1 Measuring the impact of evaluations in academics' daily life.....	13
2.2 Outline of the research design.....	14
2.3 Selection of cases.....	14
2.4 Organizing and doing the interviews.....	18
2.5 Processing data with a purpose-built database and text-analysis software.....	20
3 An indicator of the impact of evaluations.....	23
3.1 Introduction.....	23
3.2 Academic Individuals.....	23
3.3 Deans, HR managers and HoDs.....	34
3.4 Comparison and conclusions.....	41
4 Invisible work.....	43
4.1 Introduction: A taxonomy of invisible work.....	43
4.2 Method details.....	44
4.3 What did the three most influential evaluations overlook?.....	45
4.4 What is important in academic work and careers that is not evaluated, but should be?.....	46
4.5 Conclusions.....	54
5 Informal evaluations.....	61

5.1 Introduction: evaluations that remain invisible.....	61
5.2 Method details.....	61
5.3 Identifying informal evaluations.....	62
5.4 Informal job applications.....	63
5.5 Invitations to .....	65
5.6 Career and life choices.....	66
5.7 Oral feedback.....	67
5.8 Conclusion: Portfolio use and the dark side of informal evaluations.....	69
6 Conclusions.....	73
6.1 Impact and preparing for contingency.....	73
6.2 Improving the evaluation process and criteria, suggestions for GEPs and the ACUMEN Portfolio.	74
6.3 Opportunities for ACUMEN portfolio in informal evaluation .....	75
7 References.....	77
8 Abbreviations.....	79
Appendix 1: Interview protocol for researcher interviews.....	81
Appendix 2: Interview protocol for deans and managers.....	93
Appendix 3: Information sheet and consent form.....	105
Appendix 4: Time-line format.....	111
Appendix 5: Event codes.....	113



# List of tables

Table 1: Selection of countries.....	15
Table 2: Overview of interviews and distribution of gender.....	17
Table 3: Gender distribution over cohorts.....	18
Table 4: Three most important developments mentioned by 37 AIs, by gender.....	24
Table 5: Three most important developments mentioned by 10 AIs in the early career cohort, by gender.....	25
Table 6: Event code details for groups 'Jobs' and 'Other' of most important developments.....	25
Table 7: Breakdown of job types for the event code 'Acquiring/starting of a job' for most important developments mentioned by AIs.....	26
Table 8: Importance of 110 MIDs for 37 AIs, by gender (cut off at 4).....	27
Table 9: Relations between importance and event-code groups of 110 MIDs for 37 AIs (cut off at 4).....	28
Table 10: Three most influential evaluations mentioned by 37 AIs, by gender.....	29
Table 11: Event code details for group 'Grant applications' of most influential evaluations.....	30
Table 12: Breakdown of 'Grant application - successful' in Table 11.....	30
Table 13: Distribution of relations between MIDs and MIEvals for 33 AIs.....	32
Table 14: Relations between MIDs and MIEvals for 33 AIs.....	33
Table 15: Three most important developments mentioned by 10 DHs, by gender.....	35
Table 16: Three most important developments mentioned by 10 DHs, detailed by event code and gender.....	36
Table 17: Importance of 30 MIDs for 10 DHs, by gender (cut off at 2).....	37
Table 18: Relations between importance and event-code groups of 30 MIDs for 10 DHs (cut off at 2).....	38
Table 19: Three most influential evaluations mentioned by 10 DHs, detailed by event code and gender.....	39
Table 20: Distribution of relations between MIDs and MIEvals for six DHs, by gender.....	40
Table 21: Relations between MIDs and MIEvals for six DHs.....	40
Table 22: Overview of reported work that is invisible to evaluations.....	56
Table 23: Why evaluate invisible work and how - overview of answers.....	59
Table 24: Three most influential evaluations mentioned by 37 AIs, by gender, by formal/informal status.....	62

# Acknowledgements

During the preparation of this project, Grit Laudel and Paul Benneworth (Center of Higher Education Policy Studies, University of Twente), Gerben Moerman (University of Amsterdam), Pleun van Arensbergen, Inge van der Weiden, Rosalie Belder and Edwin Horlings (Science System Assessment department, Rathenau Institute) provided valuable feedback, suggestions and discussions during various meetings.

Mereie de Jong of Vertaalbureau Medea did a wonderful job on the transcription of a great number of interviews, for which I am very thankful.

A question on the LinkedIn group Kwalon<sup>2</sup> resulted in stimulating and helpful feedback and suggestions, mostly on the practical side of big interview projects.

I am grateful for the useful feedback from members of the e-Humanities group, my project partners within ACUMEN and the ACUMEN's reviewers. Sally Wyatt (e-Humanities group, Royal Netherlands Academy of Sciences and Arts), Andrea Scharnhorst (e-Humanities group and DANS, Royal Netherlands Academy of Sciences and Arts) and Lorna Wildgaard (Royal School of Library and Information Science, University of Copenhagen) gave valuable feedback on drafts of this report, for which I am indebted. All remaining mistakes are on my account.

Last but not least, I thank Sally Wyatt and Andrea Scharnhorst for their continuous feedback, enjoyable and stimulating discussions, and rock-solid support.

---

<sup>2</sup> Kwalon is the Dutch platform for qualitative research



# 1 Introduction

## 1.1 Impact, invisible work and informal evaluations

The ACUMEN project (2010) aims to develop guidelines to improve evaluation practices and to design a portfolio tool that allows individual academics to present themselves in a more encompassing way. Both the good evaluation guidelines and the portfolio tool address tasks and work that academics nowadays do that may be under-valued by existing evaluations.

In order to focus the development of the portfolio tool and the guidelines, a clear understanding of the impact of evaluations on the careers and daily work at the individual level is necessary. Which types of evaluations play important roles and how do they work in practice? This report addresses that question as part of ACUMEN's Work Package two (WP2). Identification of important forms of evaluation can focus the guidelines and the design of the portfolio tool, and details of how evaluations impact careers and daily work may lead to detailed suggestions.

This report covers important work for ACUMEN but at the same time is also quite orthogonal to the other WPs. A number of them focus on Dos and Don'ts to inform individual bibliometric evaluation (See also Glänzel & Wouters, 15 July 2013). Those studies have been informed by the inspection of quantifiable traces and web-presence of academics. The work reported in this deliverable is the only instance in the project, where the voices of academics can be heard, when they report about their experiences with different kinds of evaluations.

To our knowledge<sup>3</sup> only a few studies have investigated how particular forms of evaluation have impacts on individual academics. Luukkonen & Stähle (1993 p. 32-33) report that scientists who were 'negatively' evaluated during nation-wide field evaluations, claimed that they had received no new ideas from the evaluation. Those 'positively' evaluated claimed the contrary. In addition they claimed to have learned what others had been doing or they had acquired an overview of their respective fields or their position in relation to colleagues. In general, the interviewees assessed their evaluation's impact on their research orientation as small. In addition, four out of ten scientists appreciated the preparations for the evaluations as useful because it forced them to reflect on their research, its strong and weak points, the significance of their findings and their future plans.

A few studies report personal consequences of field evaluations, ranking systems and performance-based reward systems. Individuals were fired (Luukkonen & Stähle, 1993, p. 44-46; Thakur, 2007, p. 89) or labeled as incompetent (Luukkonen & Stähle, 1993, p. 44-46). Anxiety, depression, loss of motivation and psychological problems were reported by interviewees about colleagues (Luukkonen & Stähle, 1993, p. 46). Ter Bogt & Scapens (2009) also report increased levels of stress due to performance based reward systems implemented in departments in the Netherlands and the UK.

Clearly, effects of particular evaluations on individuals were found, but these studies focused on particular types of evaluation whereas this report focuses on the daily work and careers of individuals and asks how evaluations of different sorts and kinds have an impact.

One immediately understands that given the variety of individual careers, the differences between academic practices due to Internet and globalization, and the persistency of disciplinary differences in norms and values make an interview-based study in a three years period a great challenge. This is why this report details the careful strategy and the consideration around the design of the interviews and the sampling process. This study sought to find a balance between a large parade of individual accounts, the achievement of some comparability, and the identification of generic and persistent elements in those various accounts. This is why coding was used, a database was build, and some of the results are presented in a formal way. As the description of the interview planning reveals, at the end the rather optimistic number of 100 interviews proved to be not feasible. Still, by using up-front techniques in

---

<sup>3</sup> This paragraph and the next are copied and adapted from Van der Most (2010).

digital methods, the project delivered not only a description of interviews of a total amount of 2500 pages, but coding and extractions using innovative techniques for mid-scale interview sampling.

In addition to the main question of the impact of evaluations in the daily work and careers of academics, we focus on two topics: invisible work and informal evaluations, which will be introduced in the following two sub-sections.

### **Invisible work**

Tasks and works that are not addressed in evaluations are in fact unappreciated. After all, appreciation is a result of a process of focusing attention to the thing that is appreciated and subsequently or concurrently evaluating it. The shape and form of official evaluations is, like all artifacts, an outcome of social processes of construction, full of, among other things, negotiations, translations and contingency. Thus whether a task or work is addressed in an evaluation is also a result of these processes. Some work is highlighted through evaluations and other work is made invisible. (Star & Straus, 1999)

ACUMEN is particularly interested in this invisible work. Starting with the observation that "Currently, there is a discrepancy between the criteria used in performance assessment and the broader social and economic function of scientific and scholarly research" (ACUMEN, 2010 A p. 5), ACUMEN identifies five aspects: 1) evaluation criteria are dominated by mono-disciplinary measures that reflect academic quality and societal relevance in a limited way. 2) public science has grown in size and complexity which has put pressure on existing evaluations and quality control. 3) researchers have turned to the World Wide Web to create knowledge and communicate, which evaluation systems have not sufficiently addressed. 4) there is a lack of viable bibliometric and scientometric indicators that address the individual level. 5) public science still hinders women more than men to develop as researchers (ACUMEN, 2010 B p. 4-5) Two of these address invisible work directly.

First, mono-disciplinary criteria focusing mostly on publications in international peer reviewed journals and limited bibliometric indicators dominate evaluation practices. This leaves other dimensions of quality unaddressed, such as societal relevance and contributions to teaching.

Second, the World Wide Web opened up possibilities of new research methods, modes of working and new ways of sharing knowledge, discussing research with other sectors of society and addressing societal problems. Participating in on-line discussions, editing blogs, writing blog-posts, tweeting, on-line video presentations are regarded unimportant by most evaluations.

Although ACUMEN identified some invisible work and potential activities that remain under-evaluated, this report aims to check this, identify other invisible work, find out why academics deem this work important and how they feel it should be evaluated.

### **Informal evaluations**

Within ACUMEN's context, the flip-side of asking which work remains invisible in evaluations, is, asking which evaluations remain invisible in academic work. Double-blind peer review of papers, grant applications and job applications are highly visible evaluations in academic life. The PhD thesis defense, habilitation, and applications for promotion are more visible or less depending on the country. Annual appraisal talks are usually considered a nuisance. These and other forms of evaluation become visible when they are formal or formalized, which differs from country to country or location to location. In addition their visibility is also constructed through the amount of prestige that is attributed to them.

Out of curiosity, we kept an eye out for informal evaluations: evaluations that have no official status, but happen in daily practice. A researcher may listen to a colleagues' presentation, be positively impressed and perhaps later ask him/her to participate in a project application. Heads of departments may be assessing their group members to see if they could be eligible for promotion or have the necessary qualities to apply for a career grant. Which forms of informal evaluation exist in academic practice and how do they have an impact? These are relevant questions for ACUMEN because they may inspire new

uses of the portfolio tool. Moreover, formal and informal forms may turn out to be mixed, which may have consequences for good evaluation practices.

## 1.2 Evaluation regimes and career regimes

Before moving on to the problem of how to measure the impact of evaluations on academic careers, at least two concepts need to be outlined: 'evaluations' and 'careers'. Both have a broad meaning. Because the project takes an internationally comparative approach, it needs a common concept broad enough to talk about the national and local shaping of careers and evaluations.

### Evaluation and evaluation regime

In this study, a wide notion of evaluation is being used: anything or any activity occurring in an academic's career or daily work that has an aspect of evaluation to it would be included. Although one can identify a number of forms of evaluations, as indicated in the section 'Informal evaluations' (p. 8), it will be hard to define the concept in a generic way so that it would encompass all known forms of evaluation. Moreover, the study does not a priori determine which evaluations would be taken into account and which not. Since, it also set out to identify invisible evaluations, it would be wise not to exclude any in advance. Thus, this study deliberately and explicitly left the task of identifying evaluations to its interviewees.

Research<sup>4</sup> has had its internal peer review practices in different shapes as practiced in scientific academies and journals for over a century and a half or more (Shapin & Shaffer, 1985). In the course of the second half of the 20th century, the rise of research councils and other funding organizations added similar peer review practices to university life, amounting to a set of evaluation practices along a project's life cycle (Latour & Woolgar, 1979). Somewhat later, ministries, research councils and other actors became interested in new forms of evaluation of research, basically on a more aggregated level than individual projects, papers or researchers. As of the late 1960s, they became interested in the impacts of research on innovation (Kostoff, 1994). About a decade later, evaluation of institutes and nation-wide disciplinary efforts also became target for evaluation (Ståhle, 1987; Van der Meulen, 1992) as did the impact of technology on society through technology assessment. In the course of the 1980s, bibliometric and scientometric methods for the study of science came to the attention of some governments as a base for research funding (Philips & Turney, 1988; Van der Meulen, 1992). During the 1990s also the functioning of funding programs and research councils was evaluated. In the same decade, methods for societal evaluation of research were being developed (Spaapen, Dijstelbloem, & Wamelink, 2007) but not practiced on large scale. However, in countries like the Netherlands, our interviews indicated that the topic is gaining attention. Finally, during the 1990s, world-wide and national scope rankings of universities came to the fore (Thakur, 2007; Usher & Savino, 2006, January No4 1983). For educational programs such rankings had been developed before that, but overall university rankings, also taking research into account were developed.

The developments that are listed here, did not occur in all countries at the same time and not all types of evaluations have been practiced in all countries. Moreover and depending on the type of evaluation, they were not all systematically covering all research: not all disciplines, nor all levels of aggregation, ranging from individual to group to discipline. Still, in some countries, the resulting 'patchwork' has become rather dense and/or systematic in some dimensions as testified by Whitley (2008) who speaks of strong and weak 'research evaluation systems'.

Notice that other forms of evaluation, such as job applications, annual appraisal talks, promotion applications, PhD defense and habilitation, i.e. evaluations at the individual level, are still missing from

---

<sup>4</sup> This overview was copied and adapted from Van der Most (2010).

this short literature overview. To our knowledge they have not been studied as evaluation practices. But they do play important if not essential roles in the work and careers of academics.

Instead of systems, I prefer to speak of 'evaluation regimes' because the notion of system suggest that things are deliberately designed as one logical, consistent whole. The notion of regime used here is loosely based on that of 'technological regimes'. "The notion of regimes refers to complexes of (more or less) shared rules on how to act and to interact, and to a system of corresponding interdependencies between (groups of) actors." (Deuten, 2003 p. 31)

Similarly, the evaluation regime is the combined result of laws, rules, and common practices concerning evaluations<sup>5</sup>. It includes forms of evaluation at all levels of aggregation, ranging from the international level to the individual<sup>6</sup>. The regime concept can apply to geographical or institutional areas of varying size, for example a nation, a university or even a school or faculty within a university. Lastly, regimes change over time because laws, rules and practices change over time and changes are not necessarily synchronous between regimes.

### Career and career regime

A suitable concept to grasp academic careers in the context of this research has to be highly flexible. Firstly, because we are interested in the impact of evaluations on actual careers, which means that we want to put the experiences of individuals central.

Secondly, we want to be open to a multitude of forms of evaluations, so the career concept should not limit that. In addition, we are searching for biases in evaluation practices. Some of these biases or sources of bias can easily be identified, but some may only become visible when actual careers and experiences of researchers are tracked.

Thirdly, besides actual careers, also the career-models as outlined by laws and rules are subject to change. These models usually follow a career-ladder metaphor, for example from PhD student, to post-doc, to assistant professor, to associate professor to full professor. The higher up the ladder, the more freedom the academic has, the more income he/she receives and the more certain his/her position is.

However, these systems undergo change. Enders (2001)<sup>7</sup> mentions a loss of status for the academic profession, tighter resources for public higher education, a rise of managerial power to the cost of academic guild powers, and a blame for too little attention to societal needs (p. 2). A concrete example: quite a number of European countries once guaranteed life-long employment for all university staff. Roughly, in the course of the 1990s these guarantees were replaced by 'permanent' or 'indefinite' employment, meaning that one could lose one's job if financial problems arose for one's university. In practice, universities would still try not to fire their permanent staff, but at present in some countries, even full-professors may fear for their jobs. So, although the career-ladder model still exists, the ladder as a whole is 'dropping', so to say. Grimm perspectives do not see a ladder at all anymore: researchers are mainly in a continuous flow from one temporary job to another, hanging on while waiting for a chance to apply for permanent positions that are becoming increasingly more scarce.

Fourthly, mobile researchers who travel across borders to their new jobs, may also experience changes because they move from one career-model to another. The differences may be small but also may not always be to their benefit: to them, they move 'down'.

Neither science policy studies nor higher education studies has a dominant concept for the study of academic careers<sup>8</sup>. However a number of approaches have surfaced. The following list is not complete, but discusses a number of potentially interesting approaches.

---

<sup>5</sup> Clearly, there are corresponding interdependencies also present in evaluation regimes, but this study does not go into that dimension of Deuten's description.

<sup>6</sup> For a similar concept, see Dahler-Larsen (2011) who introduced the notion of 'evaluation machines'. His notion addresses society and organizations as unit of analysis. We want to address also the individual level.

<sup>7</sup> (Felt, 2009 p. 27) : "At the time this book is being written, quite a substantive patchwork of reordering and reforms have taken place in many countries recently, or are still happening. In many cases, this concretely means the State backing away from its role as central direct financier of research"

<sup>8</sup> Personal communication Grit Laudel (26 August 2011). Also confirmed by Paul Wouters during the ACUMEN web-meeting (4 October 2011)

### *Life cycles and research trails*

The central idea of life-cycle and research-trail approaches is that careers develop over time. Verbree et al. (2011) conceptualize 'life cycle' as "The life cycle of research leaders refers to years of experience." (p. 8) and distinguish three phases in life cycle of 'research leaders' (p. 11). Others use the notion to study productivity (Aizenman & Kletzer, 2008; Levin & Stephan, 1991; Rauber & Ursprung, 2008) or citation (Goodwin & Sauer, 1995) development in the lives of academics.

Instead of counting production or citations, one can also look at the content of the work being developed in the course of time. Chubin & Connolly (1982) coined the phrase 'research trails' as "representing a sequence of work by an individual or a small team of researchers" (p. 295) and as "distinguished by some continuity of focus - be it methodological-theoretical and/or problem oriented - in published research" (p. 295). Recently Gläser and Laudel (2009) described a procedure to visualize research trails based on publications and citations. See also Horlings & Gurney (2012) who claim a "novel method for mapping a corpus of publications" (p. 1145) of an individual or group by grouping the corpus' papers into groups of problem areas. Hellsten (2007, p. 481) used a 'bar code' graph to show the field-mobility of individual researchers.

Although the life cycle and research trails take the perspective of the individual academic and provide interesting perspectives on academic careers, they do not focus on particular events such as evaluations.

### *Careers in overlapping contexts*

Hermanowicz (2007) offers an outline for a sociology of scientific careers. He sets out "to show how careers can be studied sociologically with a person-oriented emphasis – an approach I call 'careers in context'" (p. 626). He is interested in how "occupational self-identities emerge" (p. 629), wants to ask people "What do you think about what you do?" (p. 628) and study their careers in context. Hermanowicz operationalizes the latter by emphasizing time and place. Time and place are not the most distinctive dimensions, be it still important ones, for scientific careers.

Laudel & Gläser (Laudel & Gläser, 2008) fill the notion of context with a combination of the traditional career-ladder concept from organizational studies and a career line in terms of research agenda suggested by the scientific community. They apply this to a problem in the current day transition from apprentice to colleague.

Similarly, Kaulisch & Enders (2005) propose a framework of three overlapping institutional contexts, comprising of 1) the international science system/academic discipline 2) national settings and cultural contexts and 3) organizational context or local context.

Careers do take shape in contexts, but as pointed out above, these change over time and differ between places. When one is interested in the impact of evaluations on careers, then one may not need to identify relevant dimensions of these contexts in advance. It would make sense to analyze how 'impact-full' evaluations in someone's career relate to certain dimensions of context. However, these attempts may fail when this context is as dynamic as Enders (2001) suggests.

For the purposes of this study, the most logical would be to take the dimensions of 'career' and 'evaluation' as a starting point. Evaluation regimes have been introduced above. A career regime would be the combined result of laws, rules, and common practices concerning evaluations.

Observing that a career takes shape in a multi-dimensional context does not conceptually grasp what would constitute a career and how to research the central question of this project.

### *Critical career points and critical life-events*

Bazely (2003) aims to arrive at a definition of 'early career' in research, but uses no theory to study careers. Based on interviews and surveys in six disciplines in Australia, she identifies five points that academics pass in their career paths of becoming an established researcher: 1) completion of high-level research training; 2) obtaining an academic position; 3) balancing teaching and research activities; 4) maintain a research profile under increasing pressure of other tasks; 5) achieving the status of established researcher.

Baldwin & Blackburn (1981) use the 'critical life-event instrument' to study career progress and argue for a developmental perspective on academic careers in which the academic proceeds through several developmental stages. The stages ranged from "Assistant professors in the first three years of full-time college teaching" to "Full professors within five years of formal retirement" (p. 601)<sup>9</sup>. They authors claim

---

<sup>9</sup> Unfortunately, it is not clear how the authors arrived at these stages.



that the critical life-event instrument is based on Havighurst et al. (1979), but Havighurst et al. (1979) do not introduce it<sup>10</sup>.

Conceptually, these approaches (see also Orvis, 1996) are of use because they put the individuals' perspectives central. Independent of what career models say, individuals have their own views of what have been important events in their careers. Thus, a career-concept that puts the individual's perspective central should be open to whatever the individual finds important. Arguably, such a concept can not define much more than the fact that there are changes and developments in a career.

### *Careers as strings of consecutive career changes and multi-dimensional mobility*

We propose to conceptualize careers as strings of career steps and work changes which take place in and move through one or more dynamic social contexts. These social contexts can be analyzed along multiple analytical dimensions. Two important dimensions are the career regime and the evaluation regime, whereas regimes are the combined result of laws, rules and common practices regarding a chosen dimension.

Such a definition allows us to take an actor-central perspective, investigate particular events - such as evaluations - within careers, and analyze these along multiple dimensions of our interests. The definition also takes into account that individuals' careers move through different spaces (social contexts) and that the properties of these spaces may change in the course of time.

The analyst can choose dimensions that are of his/her interest. For example, academic output, citations, research trails, gender, age/aging, affiliation and so forth. However, taking the actor's perspective may point out other dimensions that are relevant for the chosen dimension of interest.

## 1.3 Structure of the report

Having introduced the main question and concepts that drive this report, the following chapters will first introduce the research design and then present the results for impact, invisible work and informal evaluations in separate chapters. The 'Research design' chapter will introduce a way to measure the impact of evaluations on the work and careers of individual academics with the use of an indicator.

Chapter 3 will present the data in tabular form. That is, it will present a lot of numbers. The idea is not to give a statistical analysis but to provide overviews of what happened in close to 50 interviews. In addition, because we test our proposed indicator on a limited number, the tabular form is also used in order to see if upscaling to statistically relevant studies could work. Chapters 4 and 5 will present categories and counts of similar answers to questions, but in these chapters, the stress will be on quotations from the interviews.

The chapters on invisible work and informal evaluations will have sections that go into details of the research method for those two subjects. Each of the three results chapters will have its own conclusions and Chapter 6 will then wrap up the main conclusions from these chapters.

---

<sup>10</sup> The study is very interesting though: the authors found very interesting differences between careers of 'High publishers' and 'Low publishers' and between the activities these groups performed after their retirement. For example, low publishers not only started publishing later, but also stopped publishing sooner than high publishers. Low publishers made careers within the organization/institutes in which they worked, High publishers in disciplinary organizations.

## 2 Research design

### 2.1 Measuring the impact of evaluations in academics' daily life

The research project reported here studies the impact of evaluations on the careers and daily work of individual academics. It addresses both a quantitative and a qualitative dimension of this topic: how big is the impact and what is the nature of this impact?

Research projects investigating these questions run into a common problem of social research. Since one can not do experiments with 'the social', in this case peoples' careers, it will be hard to isolate the impact of one factor. Academics, like many other people, have contingent and dynamic lives in which one can identify multiple dimensions. Many events determine the course of one day. Often, many events determine even just one other event. This is depicted in Figure 1 for the impact of an evaluation on a change in someone's career or daily work.

Because different events of different nature play a role, it will be virtually impossible to assess the relative weight of each of them. It would require careful consideration and intimate knowledge of the actual events, which could only be done through an ethnographic approach.

One would thus first need to answer the question about the nature of the impact before addressing the size of the impact. Ethnographic studies are time consuming and many would be required before one can answer the questions in general or on an aggregated level.

A second best solution would be not to compare the role of the evaluation in relation to other events leading to a career change or a change in daily work, but to see which evaluations academics deem influential in their work/career and if or how these evaluations relate to important developments in careers and daily work. In other words: instead of contrasting evaluations with other causal events leading to an important change in career/work, one can contrast salient evaluations with other salient events.

Such an approach can be implemented through interviews rather than ethnographic research. It would take the perspective of the individual researcher, concurrent with ACUMEN's point of departure. It also allows some insight in the nature of the impact of evaluations on someone's career/daily work, and one can devise a simple indicator for impact based on individual experiences.

If the three out of the three most influential evaluations refer one-to-one to the three most important events in someone's work/career, then one can say that evaluations have a high impact. If only one of the three most influential evaluations relates to one out of three most important developments, then evaluations had a low impact. Medium impact is the label in case two evaluations relate to two developments<sup>11</sup>. If none of the evaluations relate to any of the three most important developments then one can say that evaluations have no significant impact. In this project, we tested how many out of three most influential evaluations refer to three most important developments, but instead of the top three one could also investigate higher or lower numbers.

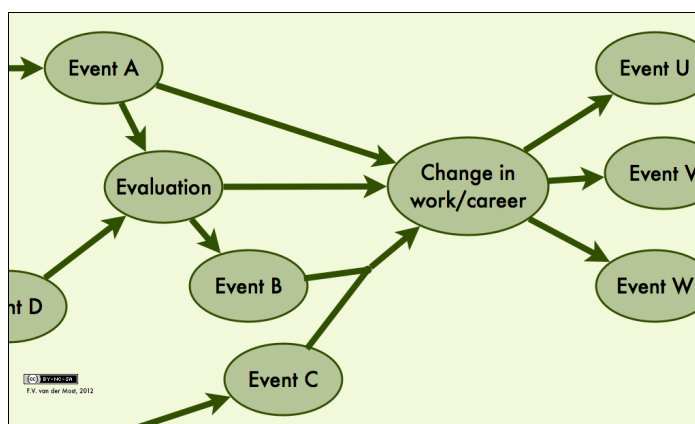


Figure 1: Identifying the impact of an evaluation

<sup>11</sup> Of course, two or three evaluations could relate to same development, but it is difficult to decide how to label this.

## 2.2 Outline of the research design

To construct the indicator set out in the previous section, we need a set of interviews in which the basic questions are : 1) what are the three most important developments in your career? and 2) what are the three most important evaluations in your career?

Of each development and evaluation we want to know:

- why does the interviewee consider it important, since we can not define the importance in advance
- how things happened in practice in order to get a good grasp of the structure of events and to grasp the nature of the evaluation and the development
- how the developments and evaluations relate to the interviewee's career.

For the analysis we need to know roughly what the interviewee's CV looks like and for the ACUMEN portfolio and good evaluation practices we need a characterization of the evaluations involved.

ACUMEN intends to cover all EU and associated countries, all disciplines and all academics. An interview project indeed covering all of these would be too big for ACUMEN's budget, but a certain diversity needs to be achieved. This is also good for explorative purposes.

Finally, because careers intersect with and go hand in hand with university development, we are also interested in the organizational perspective on the role of evaluations in academic careers. To what extent does the perspective of academic individuals align with that of their employers?

## 2.3 Selection of cases

Below we will first describe the dimensions for the selection and the choices made in these dimensions. Next, we will explain how the selection process was set up.

### *Disciplines*

The ACUMEN project selected four disciplines to focus and harmonize its work across all Work Packages. Basically, it selected one discipline within each of the three main disciplinary groups of natural sciences, social sciences and humanities and added engineering. The choice fell on astronomy and astrophysics (AA), public health (PH), philosophy including the history and philosophy of science (P+) and environmental engineering (EE) respectively. PH and EE are considered to be relatively new interdisciplinary fields or hybrid fields where new practices are likely to occur.

For the purposes of this project, two of these disciplines had to be expanded somewhat in order to find enough universities that housed them. Environmental Engineering was expanded to also include Environmental Science and Public Health to include Health Sciences. In terms of institution building, these EE and PH are not full-fledged disciplines.

### *Countries*

Because within a country, career paths are shaped in large part through laws and national regulations, it makes sense to find countries with contrasting career regimes in order to broaden the potential scope of answers from interviewees. Roughly speaking, there are three regimes:

- 1) the chair regime in which researchers have temporary assistant positions for a decade or two and then jumps to professor and attains a permanent or guaranteed position. In such regimes, a professor chairs a research group which besides the chair consists of his/her assistants. The chair determines the direction of research, the distribution of resources and the hiring of personnel. This regime can be found in Germany, Switzerland and Austria
- 2) The lecturer regime in which a researcher enters into a permanent position early in his/her career and then moves up on the ladder. In this regime a department has a head, but the control over the research agenda and the distribution of resources is more equally shared. Typically, one would find multiple

professors and other senior staff with comparable academic freedom in one department. This regime exists in for example the UK and Australia. It also existed in the Netherlands until about the end of the 1990s.

3) The tenure track regime in which a junior researchers are given some time, for example five years, to work their way up to become acknowledged in their field and are then rewarded with a permanent position, perhaps as professor or associate professor. This regime exists in the United States, is not common in Europe but some universities are experimenting with it<sup>12</sup>.

The differences between actual career regimes are often not this sharp, although some countries do have rather 'pure' versions. The chair regime exists in Germany and other German-speaking countries. The lecturer regime exists in the United Kingdom and Scandinavian countries. Mixed versions with stress on the lecturer regime can be found in the Netherlands.

Evaluation regimes are to some extent nationally shaped, be it to a lesser extent than career regimes and depending on the type of evaluation. The Netherlands for example have nationally organized evaluations of disciplines but such a system does not exist in many other countries. In Sweden, some universities organize university-wide self evaluations. When research-funding originates from only one or a few sources such as research councils or ministries, then the evaluation of project applications can be highly standardized.

Having said this, perhaps the most determining evaluation systems are those related to university block-funding. On this dimension also two regimes stand out with regard to research<sup>13</sup>. One in which block-funding for research is mostly or fully determined on the basis of an evaluation of research and one in which this is not the case or only for a limited percentage of the overall budget. The UK's Research Assessment Exercise<sup>14</sup> (RAE), introduced in the mid 1980s, is a famous example of the former. To our knowledge, within Europe, Poland is one of the few other countries (if not the only other country) that also goes in this category, albeit the evaluation system is strictly quantitative whereas the RAE is of a qualitative nature.

		Career regime	
		Chair regime	Lecturer regime
Evaluation regime (for university block- funding for research)	Evaluation based	Poland	United Kingdom
	Not evaluation based	Germany, Austria, Switzerland, Italy, Spain	Netherlands, Sweden, Norway, Finland, Denmark

Table 1: Selection of countries

Summing up: career regimes can be split up in two categories and evaluation regimes related to university block-funding too. Thus we arrive at a two by two matrix (see Table 1) of countries in which we expect to find differences between interviewees' answers. On the career-dimension we expect this

<sup>12</sup> We learned through the interviews that this is happening in the Netherlands.

<sup>13</sup> When it comes to education, universities in most European countries are for a substantial percentage based on student numbers and/or output of graduated students.

<sup>14</sup> The name 'Research Assessment Exercise' is neither the first nor the last name of the system, but perhaps still the most famous name. Currently it is called the Research Excellence Framework.

because career regimes outline the shapes of possible careers. On the evaluation-dimension we expect to see differences because universities' salary-commitments to their employees are enormous and when universities' budgets are dependent on evaluations then this is likely to influence hiring and firing decisions.

### *Selection of universities and research groups*

Universities were selected randomly while aiming to keep travel costs low. In order to protect the interviewees' anonymity neither the names of the universities nor their cities will be disclosed. The number of universities per country ranges from two to 4.

For each of the four disciplines, we identified one department or research group and selected all interviewees from that unit. Taking into account that many candidate interviewees would decline to participate we looked for units with at least 10 individuals.

### *Academic individuals - cohorts*

The selection of candidates also tried to capture diversity in order to increase the explorative value of the current study. In the selection of academic individuals (AI), two aspects played a role: seniority and gender.

We measure seniority in terms of 'number of years since acquiring the doctor title' rather than considering the job title. In other words we opted for cohort diversity rather than diversity in career-speed. The reason has to do with the changes in career and evaluation regimes that most countries have seen in the past two to three decades.

Riley (1998) points out a pressing aspect of life course history and career studies: if at one point in time one studies a group of people of different ages, one should be aware that although they live at the same time, they are from different 'cohorts'. They were born in different periods and they have lived through different chains of events. There are two fallacies connected to this in life course research:

- 'life course fallacy': "erroneously interpreting cross-sectional age differences as if they referred to the process of aging." (p. 42)
- 'fallacy of cohort-centrism': "erroneously assuming that members of all cohorts will grow older in the same fashion as members of our own cohort" (p. 43)

Obviously, these fallacies equally apply to career studies. Considering the waves of changes that have gone through the governance of academia in recent decades, the present study should take these warnings seriously. It means that people from one cohort may be more homogeneous in their answers than people with the same job-title.

We defined the following cohorts:

- early career (EC) researchers: people who have worked five to 10 years in academia since their PhD
  - mid career (MC) researchers: people who have worked 11 to 25 years in academia since their PhD
  - senior career (SC) researchers: people who have worked 26 years or more in academia since their PhD
- Within each country within each discipline we aimed to find one individual for each of the three cohorts. In some cases, and in particular in public health, the measurement does not work very well because individuals may not have a PhD or got it only relatively late in their career.

We tried to have a rough representation of gender distribution per discipline, but since we aimed to find all individuals for one discipline in one research group or department, this was not always possible. Also the low number of interviewees per group limited the precision of representation.

### *Individuals - deans, human resources managers and department heads*

For each discipline within each country/university we aimed to interview at least one dean, human resources manager or head of department. This group will be identified as DH (Dean or Human resources).

### Overview of cases

In total we aimed to do 64 interviews. Four disciplines, four countries, and three cohorts plus one DH per discipline per country.

We contacted in total 135 individuals, of whom eventually 48 agreed to an interview. When we ran out of candidates within specified cohorts we would target other group members outside but as close as possible to the targeted group.

Unfortunately we were unable to achieve the goal of 64 interviews. In case of a few research groups eventually all group members were contacted or we ran out of time to contact alternative candidates. Besides that, two groups stood out when it comes to non-participation. Firstly, German deans were hard to reach, so in some cases we could not even get a response and in others they were simply declining. We had few alternatives since the university websites were not so informative that we could identify human resources managers. Secondly, many Polish scholars did not respond to e-mail nor picked up the telephone.

In total we did 48 interviews. We interviewed 38 (13 women and 25 men) academic individuals about their personal experiences and 10 with DHs. In the DH group all women are human resources managers whereas all men are deans or heads of department (HoDs). Table 2 provides an overview of the interviews and the gender distributions. Table 3 provides the totals of gender distribution over the cohorts.

Although we interviewed less people than intended, they more than suffice for our purposes. We have substantial and similar amounts of data from AIs in both genders, all four disciplines, all three cohorts and three out of four countries.

	Astronomy & Astrophysics				Public Health				Philosophy				Environmental Engineering				Total	
Country	EC	MC	SC	DH	EC	MC	SC	DH	EC	MC	SC	DH	EC	MC	SC	DH	AIs	DHs
UK	F	M	M	M	M	M <sup>p</sup>	F	F	-	F	M <sup>sp</sup>	F	M	M	M	M	3F + 8M	2F + 2M
DE	F	M	-	-	F	M	F	-	-	M	M	-	M <sup>p</sup>	M	F	M	4F + 6M	1M
NL	M	M	M	-	F	F	F	M	M	-	F	F <sup>d</sup>	M	M <sup>s</sup>	M	M	4F + 7M	1F + 2M
PL	-	M	-	M	-	-	F <sup>w</sup>	-	M	F	M <sup>g</sup>	M	-	M <sup>g</sup>	-	-	2F + 4M	2M
Total	2F + 7M				7F + 3M				3F + 6M				1F + 9M				13F + 25M	3F + 7M

s = summary instead of transcription

d = interview in Dutch

w = written response instead of interview

g = interview in German

p = phone interview

Table 2: Overview of interviews and distribution of gender

	Gender		Total
	F	M	
Early career	4	6	10
Early career or mid career	0	2	2
Mid career	7	10	15
Senior	2	7	11
<b>Total</b>	<b>13</b>	<b>25</b>	<b>38</b>

Table 3: Gender distribution over cohorts

## 2.4 Organizing and doing the interviews

For each country, the process of organizing, doing and processing the interviews went roughly according to the following scheme:

- a two week period of searching and identifying suitable universities and research groups based on university websites
- a two month period of contacting potential candidates and planning and preparing the interviews and travels
- a two week period of interviews for the UK and Poland. In Germany the interviews were divided over two periods and in the Netherlands interviews were spread out over many months.

The United Kingdom: 14 to 25 May 2012

Germany: 25 June to 4 July, 29 September

The Netherlands: 29 November 2012 to 20 March 2013

Poland: 19 to 26 February 2013

- The interviews were transcribed and coded during two periods: nine interviews during the summer of 2012 and the remaining interviews from April to August 2012.

### *Recruiting participants*

Candidate interviewees were initially contacted via e-mail. They received a standard invitation for participation which included an information sheet. (See Appendix 3) If they had not responded after two weeks, they received an e-mail reminder and one week later we would try to phone them. Closer to the planned trip, the waiting times were decreased. If a candidate declined we would contact someone else in the same cohort in the same unit. If there were no such candidates left over, we would contact individuals as close as possible to that cohort.

If a candidate accepted, we scheduled a meeting and unless a CV with sufficient information would be available on-line we asked them to send their CV for our preparation. We prepared time-lines on paper which contained space for information about jobs, residence, family situation, diplomas, age, gender and nationality and brought these to the interviews (see Appendix 4 for the time-line format).

### *Interview setting*

Almost all interviews were done face-to-face usually in the office of the interviewee. A few interviews were done in a bar or restaurant or in the office of the interviewer, and a few could not be done during

the planned trips. Instead they were done over the phone or Skype (See Table 2). All interviews were conducted by the author of this report. He made notes in answer forms<sup>15</sup> during the interview.

Almost all interviews were done in English. A few interviewees asked to do the interview in Dutch or German (See Table 2). One interviewee gave written responses in English. She wanted to do the interview but felt we needed a translator. Because that was difficult to arrange and costly, we decided on written responses, which were translated by a colleague or acquaintance of the interviewee.

The average recorded interview-length was one hour and 45 minutes. Two interviews lasted half an hour and three interviews lasted more than two hours.

### *Consent from interviewees*

At the very beginning of the interview, we asked interviewees for their approval to record the interview and after the introduction to the project and any questions from the interviewees, we asked them to sign the consent form. For full details see the form in Appendix 3. By signing it, they agreed to participation in the project under certain conditions. Among these was the condition that data provided by them would be presented anonymously in reports and publications. Upon their request, they could see and edit the transcript in any way. One interviewee asked for the transcript, but made no changes. Thirdly, at any time after signing the form, interviewees could and still can withdraw, which will be taken into account for all publications and reports following the withdrawal. At the time of writing, no interviewee has withdrawn.

In addition to these conditions, interviewees could indicate whether they agreed or disagreed to : 1) recording and digital processing of the interview; 2) circulation of the transcript among three named partners of ACUMEN's KNAW team; 3) circulation of the transcript among all ACUMEN project partners. One interviewee objected to recording and digital processing. Instead of a transcript we made a summary. We did not count his answers in the overviews. Six interviewees did not agree to circulation to ACUMEN project partners. Two of these also did not agree to circulation within the KNAW team. The data from these interviews is taken into account in this report and the report may contain anonymous quotes from these interviewees since they did agree to that. From the interviewee who sent in written responses we did not obtain a signed consent form<sup>16</sup>. We did use the data in this report.

### *Interview protocol*

The interviews were structured through two protocols: one for the academic individuals (AIs; Appendix 1) and one for the deans, HoDs and HR managers (DHs; Appendix 2). The protocol for the AIs was tested on two researchers in the interviewer's network.

In short, the protocol for the AIs has four parts. The first part consists of a check of the CV as represented in the time-line and asking for additional data that could not be found in the CV. This was meant not only to collect the data but also as a way to refresh the interviewees' memories about their careers. Secondly, the protocol asks for the three most important developments in the interviewee's career and work, for details about each of these, and why these are important to the interviewee. Thirdly, the protocol asks to first list and then ten most influential evaluations in the interviewee's career. After that it asks for details about the top three. Fourthly, the protocol ends with a number of questions about the interviewee's opinions about evaluations in general.

The protocol for the DHs is different but similarly structured. The first part asks about the career and evaluation regimes in the interviewee's faculty. Secondly, it asks for the three most important developments in an academic career, for details about these, and why these are important for the faculty. Thirdly, the protocol asks which three evaluations have the biggest effect on a researcher's career. Finally, this protocol also ends with a number of questions about evaluations in general.

After the first interview trip, we adapted the protocol<sup>17</sup>. Initially, we used a particular meaning for the notion of 'influential evaluation'. It was meant to refer to an evaluation whose outcome had a big influence on a subsequent change in someone's work or career. The change itself would not necessarily

---

<sup>15</sup> The empty forms are not added as appendix, but can be made available on requests.

<sup>16</sup> The interviewee had incredibly little time to spare, it took her a long time to respond and to get the answers translated, so we felt we could not bother her any more.

<sup>17</sup> Other minor changes were also made in the course of time.



have to be big or important. However, it took too much time to convey this nuance and most interviewees did not understand the difference. They translated this into 'important evaluation': an evaluation that was important to them for whatever reason. We changed the question into 'influential evaluation' in the common-sense meaning of relating to a big change in daily work or career. However many interviewees also translated this into 'important evaluation'.

The protocol also included a basic characterization of all evaluations that were being performed with questions such as 'what was evaluated?', 'Who evaluated?', 'What were the elements of the evaluation?', 'Did the evaluators and evaluated know each other's names?'

It turned out that the protocol often took far more time than the planned one hour and a half. In those cases, especially the second part (most important developments) took more time than envisaged. In other interviews, interviewees needed a lot of time to come up with ten evaluations, or they elaborated on all ten. It was difficult to find a balance between on the one hand wanting to know the details of an event, which requires a kind of 'story telling' from the interviewee and on the other hand, restricting interviewees once they started telling. Unfortunately, often, not all questions could be asked.

We aimed to at least establish lists of three most important developments and the three most influential evaluations. When there was little time to discuss the evaluations, in part of the cases the fact that some of the evaluations were discussed as part of the three most important developments, compensated for the loss.

## 2.5 Processing data with a purpose-built database and text-analysis software

This report sheds light on qualitative issues about the impact of evaluations on careers, on the kinds of informal evaluations and invisible work. As such it complements the quantitative studies within (and outside) ACUMEN and gives qualitative insight in what is often reduced to numbers. Having said that, 48 interviews of an hour and a half produce around 2500 pages of transcription. 38 Interviews with academic individuals about three developments in their lives and three influential evaluations created produced

around 220 small stories. So, we had a lot of data to process for a qualitative study

For analysis we not only needed easy access to the transcriptions and summaries of the interviews, but also see the interviewees' answers and stories in the historical context of the career events that they were addressing, and in connection to properties (age, cohort, gender, discipline) of the interviewees. Advanced software packages for text analysis such as NVIVO and TAMSAlyzer, however, are not well equipped to handle structured data. They allow for the addition of data columns to coded fragments, but these need to be filled on row-by-row or fragment-by-fragment basis. A combined use of a relational database for the structured data on the one hand and text analysis software on the other seemed effective.

In order to deal with the amount of data and its complexity, we developed a database (in FileMaker Pro v. 11) for the structured data and the transcripts, plus a 'synchronization' routine to exchange data and transcripts with text analysis software (TAMSAlyzer). The remainder of this section will briefly describe the transcription process and the coding process and structure. During the transcription and coding, we developed a workflow document<sup>18</sup> to describe in detail how we produced the transcriptions, imported them into the data base, added other data into the data, exchanged data between database and TAMSAlyzer, and coded the transcriptions. This way we could establish a consistent treatment of the data over time.

### *Transcriptions*

We outsourced the transcription of 28 interviews to Mereie de Jong of Vertaalbureau Medea. The interviewer transcribed the remaining ones. All interview recordings were transcribed verbatim using

---

<sup>18</sup> The document is not added as appendix, but is available upon request.

Transcriva software. Small utterances like 'Yes', 'Hmmhmm' etcetera to indicate that one understands what the interview partner is saying are mostly left out.

## Coding

We coded the transcripts and summaries in two steps. First, we imported the transcriptions into the database and coded at paragraph level to construct a question and topic based entry to the transcriptions: the interviewee's group (AI or DH), which question was asked, which was actually answered<sup>19</sup>, which main topic was discussed and roughly which period the spoken text was about. The main topic was a code for Most Important Development number one to three, Most Influential Evaluation number one to three, or one the other evaluations in the list of 10. Elsewhere in the database, the actual events were stored, linked to the interviewee and linked to the main topic code for that interviewee.

The table in the database that stored these events also stored other events such as jobs, diplomas and titles from the interviewee's CV/time-line. Often the main topics discussed were CV events or related events, but not necessarily so. Each event was described with the following information: the period of the event, a short text description, two codes to categorize the event, the interviewee's affiliation during the event and the country in which the event took place.

The two codes categorizing the events are an 'event code' and an 'event code group'. The groups and event codes are listed in Appendix

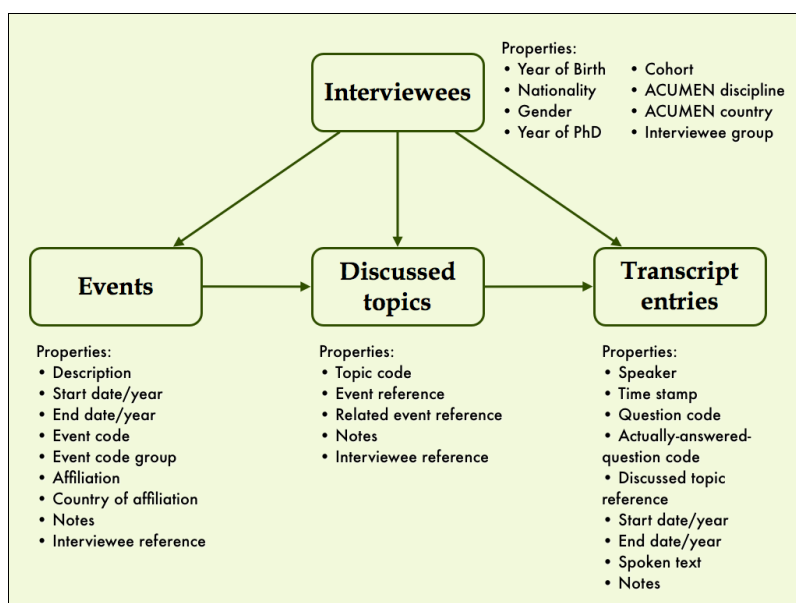


Figure 2: Basic database structure

After the first step, we transferred the transcript to the text analysis

software to code the content of the transcript in more detail. We did not have a pre-defined code list, but used an open coding process: we developed it during the coding process based on our initial areas of interest and the events, ideas and values that the interviewee mentioned or discussed.

The codes could be structured through the use of the greater-than character to indicate further detailing. For example like this:

```

"importance>moving_up_the_academic_ladder
importance>moving_up_the_academic_ladder>not_that_important
importance>moving_up_the_academic_ladder>tenure_track
importance>moving_up_the_academic_ladder>the_PhD_as_the_start
importance>moving_up_the_academic_ladder>to_professor_level
importance>moving_up_the_academic_ladder>to_professor_level>not_that_important".
  
```

TAMSAAnalyzer allows to add descriptions of the codes, but we made no use of that. Firstly, because in the course of the coding the meaning might change, and secondly because codes themselves and the tree-structure already allows to express enough meaning for the purpose of exploring the transcriptions and rough analysis.

All coding was done by one person, who was also the interviewer. After all transcripts were coded the list was about 280 items long. It would go too far to describe the list, but a number of things should be pointed out. Firstly and not surprisingly, four main groups of top-level-codes developed: 'career\_regime', 'evaluation regime', 'importance' and 'micro\_stories'.

<sup>19</sup> Often, interviewees answered a different question more than the one asked. However, often the answer still may have meaning for the question asked.

Secondly, these categories are partly overlapping, i.e. the same sub-codes can exist in both top-level codes. Especially, overlap exists between `career_regime` codes and `evaluation_regime` codes because evaluation of work is part of the `career_regime`. For example when the interview deals with evaluation criteria for certain jobs or titles. Also a lot of overlap between 'importance' and 'micro\_stories' exists, probably because people tend to talk about what they find important and while talking about what they find important they convey information about what happened.

Thirdly, the codes in general refer to something positive: something that did happen or something the interviewee evaluated positively. When the interviewee talks about something negative or something that was not important or happening, then the code would be qualified with a sub-code. For example:

```
'micro_stories>mobility>international_mobility'  
and  
'micro_stories>mobility>international_mobility>non_mobility'
```

After coding the transcripts, the transcripts with the codes were re-imported into the database. This results in system that has a high level of interrogation and exploration possibilities. Because interviewee properties, the events discussed (plus their properties) and the interviewee's CV are all connected to the transcript-entries of that interviewee, calculations and queries can be made to select sub-sets of the transcripts. The transcripts plus additional data can then be exported to the text-analysis software for further searching, selecting and analyzing of the transcripts' contents.

## 3 An indicator of the impact of evaluations

### 3.1 Introduction

Direct measuring of the impact of evaluations on the daily work or careers of individual academics is virtually impossible to measure, simply because an event (the impact) is a result of many other events. Hence, a second best solution is to use some sort of indicator, as discussed in the previous chapter (Section 2.1 in particular).

We separated the results for Academic Individuals (AIs) and Deans, Human resources managers and department heads (DHs). Not only do the two groups represent different perspectives, viz. of evaluated and evaluators, but we also asked them slightly different questions. We asked the AIs what were the three most important developments in your work and career and what were the three most influential. The DHs we asked what are the three most important developments in the careers of their staff members and the three most influential evaluations the staff members can encounter. Whereas we asked the AIs what was important about the developments for them, we asked the DHs what why the three most important developments were important to the faculty or department because we are interested in the organizational perspective.

Thus, Section 3.2 will present the results for AIs and Section 3.3 the result for DHs. Each of the two sections will first discuss the answers to the question about important developments, then the answers to the question about the most influential evaluations and thirdly the relation between the two sets of answers. Each of the two sections has its own conclusions and Section 3.4 will compare the answers of the two groups.

### 3.2 Academic Individuals

#### Most important developments

To the question 'What were the three most important developments in your career?', 37 interviewees mentioned 110 events, which we coded and then grouped. Table 4 gives an overview of how these events are distributed over the event code groups.

Event code group	Gender		Total
	F	M	
Job	12	25	37
Other	10	16	26
Changes in context – organization or work related	3	7	10
Diplomas	2	7	9
Doing a PhD (except diploma)	2	4	6
Changes in context – non-work related or family related	3	2	5
Changes in main tasks	1	4	5
Grant applications	2	3	5
Evaluations of teaching	1		1
Feedback	1		1
Grants		1	1
Job applications		1	1
Informal evaluations	1		1
Paper and book reviews		1	1
Titles		1	1
<b>Total</b>	<b>38</b>	<b>72</b>	<b>110</b>

Table 4: Three most important developments mentioned by 37 AIs, by gender

Two things stand out. Firstly, quite a number of event code groups refer to evaluations of sorts: Diplomas, grant applications, informal evaluations, evaluation of teaching, job applications, paper and book reviews and titles. Together they account for one sixth of the answers.

Secondly, about a third of the reported most important developments have to do with jobs, whereas two thirds have to do with things that happen within jobs. Considering their size, the event code groups 'Jobs' and 'Other' are worth unpacking. Table 6 shows the number of answers with event codes in those two groups.

We broke the table down in disciplinary groups and country groups, but found no striking differences in the distribution of the answers over the code groups nor the gender categories. A breakdown over cohorts showed no striking differences of the totals of answers over the code groups. However, in the early career cohort, with answers from four women and six men, answers from women were more oriented towards other event code groups than 'jobs' (See Table 5).

Event code group	Gender		Total
	F	M	
Job	2	9	11
Other	3	5	8
Changes in context – organization or work related	1	1	2
Diplomas		2	2
Grant applications	1	1	2
Changes in context – non-work related or family related	1		1
Changes in main tasks	1		1
Doing a PhD (except diploma)	1		1
Evaluations of teaching	1		1
Feedback	1		1
<b>Total</b>	<b>12</b>	<b>18</b>	<b>30</b>

Table 5: Three most important developments mentioned by 10 AIs in the early career cohort, by gender

Event code group		Gender		Total
		F	M	
Job	Acquiring/starting of a job	9	19	28
	Getting tenure / a permanent position	2	4	6
	Getting a tenure track position	1		1
	Non-academic		1	1
	Earning an income		1	1
Other	Choice for academia	3	4	7
	Project	3	3	6
	Visiting scholar	1	3	4
	Leaving science	1	1	2
	Founding of own company	1	1	2
	Developing oneself in a broad way, inside and outside academia	1	1	2
	Visibility		1	1
	Change in research orientation		1	1
	Reputation		1	1
<b>Total</b>		<b>22</b>	<b>41</b>	<b>63</b>

Table 6: Event code details for groups 'Jobs' and 'Other' of most important developments

In Table 6 we notice that the group 'Jobs' mostly refers to acquisition or starting of a particular job. Please notice that the codes reflect the answer to the initial question, not the reasons why the answer was important. For example, the event code 'Getting tenure / a permanent position' suggests that people who gave a differently coded answer did not care for a tenure or permanent position. This is not necessarily so - as we will see below.

For the event code 'Acquiring/starting of a job' the type of jobs is mixed: it ranges from starting on a PhD position<sup>20</sup> to professor level jobs. See Table 7. We checked how the numbers would break down over countries and cohorts. One thing stood out: six out of nine answers in the 'Post doctoral worker' code came from the early and mid career cohorts from UK interviewees (4 men and two women)

Job types	Gender		Total
	F	M	
Post doctoral worker	2	7	9
Assistant prof. level	1	2	3
Scientific assistant	1	2	3
Professor level (job)	2	1	3
Associate prof. level	1	1	2
Non-academic		2	2
Unknown (job)	2		2
PhD Worker		2	2
Visiting scholar		1	1
Student assistant		1	1
<b>Total</b>	<b>9</b>	<b>19</b>	<b>28</b>

Table 7: Breakdown of job types for the event code 'Acquiring/starting of a job' for most important developments mentioned by AIs

A final remark about Table 4: Contrary to what one might expect expectations, 'Change in research orientation' was mentioned only once as one of the three most important developments in someone's career.

To give a rough impression of why interviewees found these developments important, Table 8 provides the codes for 'importance' (See Section p. 21). That it is, it counts the occurrence of each code in the transcription for each MID. If an interviewee mentions for example 'income' for all three MIDs, then income is counted three times. It does not tell how often an important thing was mentioned within the discussion of one MID, nor how much time or words the interviewee used to address the thing that was important. To save space, the table only shows those items that were counted at least four times. Obviously, most interviewees would mention more than one thing that was important about a development for them.

To give an example, including the coding

" {importance>academic\_standing}well on the one hand, as I said, it a very prestigious thing, ... , so it was ... important for my further career.{/importance>academic\_standing}  
 {importance>personal\_development}{importance>having\_the\_necessary\_skills\_or\_qualifications}  
 ehm, but it was also very important for my individual development, because there I, well, I mean ...  
 I was the director of the group I had to make plans about how the group should work and what we would do. and I guess I acquired many skills that I didn't have before ... I mean, for instance, the ...  
 supervision of the PhD students, but not only supervisions, but the thing that I had to take care about, the organizational aspects and so on. and that wasn't always very easy. so we had very tensed times. and I had to work on it. ... {/importance>personal\_development}  
 {/importance>having\_the\_necessary\_skills\_or\_qualifications}" (Interviewee 1, 00:34:37.27)

<sup>20</sup> In some countries PhD students are in fact employed.

Importance code	Gender		Total
	F	M	
• recognition	4	11	15
• developing_one_s_own_research_agenda	7	6	13
• enjoying_one_s_work	2	10	12
• doing_research	4	8	12
• personal_development	7	4	11
• societal_impact	6	5	11
• academic_standing	3	7	10
• long_term_stability	3	7	10
• income	3	7	10
• contacts_or_network	3	6	9
• having_the_necessary_skills_or_qualifications	3	6	9
• moving_up_the_academic_ladder	3	5	8
• personal_achievement	1	7	8
• long_term_stability>income	3	5	8
• academia_or_not_question	4	4	8
• move_from_non_academia_to_academia	2	4	6
• self_esteem	3	3	6
• family	3	3	6
• mobility>non_mobility		6	6
• a_stimulating_environment	2	4	6
• self_confidence	5		5
• engagement	2	3	5
• resources>infrastructure	1	4	5
• having_an_academic_career	1	4	5
• long_term_stability>time	1	4	5
• time	2	2	4
• family>children		4	4
• getting_things_published		4	4
• family>combining_two_careers>both_academic	3	1	4
• having_a_life	3	1	4
• contacts_or_network>a_supportive_supervisor_for_further_career	2	2	4
• acquisition_of_funding	1	3	4
• change_in_topic	1	3	4
<b>Total</b>	<b>88</b>	<b>153</b>	<b>241</b>

Table 8: Importance of 110 MIDs for 37 AIs, by gender (cut off at 4)



Importance code	Job	Other	Diplomas	Grant applications	Doing a PhD	Changes in context - non-workrelated or family related	Changes in context - organization or work related	Changes in main tasks	Feedback	Grants	Evaluations of teaching	Other informal evaluations	Paper and book reviews	Titles	Job applications	Total
• recognition	5	2	3	1	1		1	1					1			15
• developing_one_s_own_research_agenda	4	4	1	3										1		13
• enjoying_one_s_work	4	3	1		1	1	1	1								12
• doing_research	6	2			1	2		1								12
• personal_development	2	4	1	1			1		1		1					11
• societal_impact	1	6		1		2		1								11
• academic_standing	4	2	1	1		2										10
• long_term_stability	7	3														10
• income	2	2	1		3	1	1									10
• contacts_or_network	3	5	1													9
• having_the_necessary_skills_or_qualifications	1	2	2	2							1				1	9
• moving_up_the_academic_ladder	3	2			1		1	1								8
• personal_achievement	2	3	1		1					1						8
• long_term_stability>income	6	2														8
• academia_or_not_question	2	3						1	1			1				8
• move_from_non_academia_to_academia	3	2			1											6
• self_esteem		3	1		1			1								6
• family	3	1			1	1										6
• mobility>non_mobility	2	1	1				1							1		6
• a_stimulating_environment	2	2					1	1								6
• self_confidence		3							1			1				5
• engagement	1	3							1							5
• resources>infrastructure	1	2	1							1						5
• having_an_academic_career	3	2														5
• long_term_stability>time	3			2												5
time				1		1		2								4
• family>children	2	1					1									4
• getting_things_published	2	1											1			4
• family>combining_two_careers>both_academic	4															4
• having_a_life	1	2				1										4
• contacts_or_network>a_supportive_supervisor_for_further_career	2	1			1											4
• acquisition_of_funding		1	1	1			1									4
• change_in_topic	1	1					1			1						4
<b>Total</b>	<b>82</b>	<b>71</b>	<b>16</b>	<b>13</b>	<b>12</b>	<b>11</b>	<b>10</b>	<b>10</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>241</b>

Table 9: Relations between importance and event-code groups of 110 MID's for 37 AIs (cut off at 4)

## Most influential evaluations

37 Interviewees in the group of Academic Individuals (AIs) gave 111 answers to the question what are the three most influential evaluations in your career? We coded these answers and grouped the codes as shown in Table 10. Below the table we will detail four out of the five biggest groups. The group 'Informal evaluations' will be discussed in Chapter 5.

Event code group	Gender		Total
	F	M	
Job applications	9	17	26
Grant applications	5	14	19
Diplomas	3	12	15
Job-related evaluations	7	6	13
Feedback	5	6	11
Other	1	6	7
Paper and book reviews	2	5	7
Evaluations leading to an invitation to something	4	2	6
Other informal evaluations	2	3	5
Evaluations of teaching	1	1	2
<b>Total</b>	<b>39</b>	<b>72</b>	<b>111</b>

Table 10: Three most influential evaluations mentioned by 37 AIs, by gender

### *Job applications*

In the group of job applications, we distinguished between successful and unsuccessful applications and applications irrespective of the success. Twenty-two answers out of 26 referred to successful applications. For two answers the success did not matter and two answers referred to unsuccessful applications. The 22 answers referring to successful applications included applications for full professor jobs (4), associate professor jobs (2), assistant professor jobs (5), post doctoral jobs (6), scientific assistant jobs (2), department chair (1) and one job in industry<sup>21</sup>.

Job applications were mentioned most often during interviews in the UK (13), not at all in Polish interviews and six or seven times in the German and Dutch interviews. The relatively high score in the UK is consistent with the common knowledge that UK researchers are more mobile than in other countries.

### *Grant applications*

Grant applications and diplomas are relatively more often mentioned by women than by men, taking the number of men and women interviewees into account. The answers referring to grant applications are mostly about successful applications (Table 11). Considering the low success rates of grant applications, one may expect these to be mentioned more often.

Grant applications were more often mentioned during interviews with AA and EE researchers (five and seven answers) and less often during interviews with P+ and PH researchers (three and four). Senior researchers mentioned them less often (four answers) than early career (6) and mid career researchers (9).

The successful grant applications referred to different types of grants. (Table 12).

<sup>21</sup> The remaining answers referred to unspecified jobs

Event codes for group of 'Grant applications'	Gender		Total
	F	M	
Grant application - successful	1	10	11
Grant applications - irrespective of success	1	2	3
Grant applications - unsuccessful	1	1	2
Grant applications - successful	1		1
Feedback - form - on grant applications		1	1
Grant application - success unknown	1		1
<b>Total</b>	<b>5</b>	<b>14</b>	<b>19</b>

Table 11: Event code details for group 'Grant applications' of most influential evaluations

Types of grant for event code 'Grant application - successful'	Gender		Total
	F	M	
Career grant (national)		2	2
Project grant (national)		3	3
PhD scholarship grant (international)		1	1
PhD scholarship grant (national)		2	2
Project grant (international)	1		1
Visiting scholarship grant (international)		2	2
<b>Total</b>	<b>1</b>	<b>10</b>	<b>11</b>

Table 12: Breakdown of 'Grant application - successful' in Table 11

### *Diplomas*

The 15 Diplomas mentioned in Table 10 refer to seven doctoral diplomas, three habilitations, two MA/MSc diplomas and two other diplomas. Half of these were mentioned by mid-career researchers. It seems that academics in Poland (7) and Germany (5) consider the diploma exams more important than those in the UK (1) and the Netherlands (2). Polish researchers seem to stand out even more, since we interviewed six AIs there, compared to 10 or 11 in other countries. However, Polish and German researchers also found it harder to identify evaluations simply because they have had comparatively few experiences, so it is more likely that these very obvious types of evaluation end high in their list. We found no striking differences between disciplines

### *Job-related evaluations*

Job-related evaluations are evaluations occurring as part of regular employment, so excluding job applications. Job-related evaluations (13 answers) were mostly applications for promotion (5), annual appraisal talks (4), and evaluations related to reorganizations (3). Job-related evaluations were mostly reported during Dutch interviews (8 answers) and within those eight, there seems to be a gender difference: four female interviewees mentioned five job-related evaluations whereas seven male interviewees mentioned three.

Early (3 answers) and mid career researchers (3) seemed less impressed by job-related evaluations than senior researchers (7 answers). We found no striking differences between disciplines.

### *Negative evaluations / unsuccessful applications*

Unsuccessful evaluations do not stand out, neither in grant applications nor job applications. Eight interviewees mentioned 18 negative evaluations at any time in the interview<sup>22</sup>. Two interviewees in

<sup>22</sup> Which means, these were mentioned in such a way that they were coded as a most important development, an evaluation related to a most important development, or as an influential evaluation.

particular (both from the UK, one in EE and one in P+) were quite aware of the role of negative evaluations. They mentioned three and six negative evaluations respectively.

Out of these 18 negative evaluations mentioned, four answers from four interviewees (two in the UK, one in Germany and one in the Netherlands) were in the top threes of most influential evaluations. Two interviewees referred to one or more unsuccessful job applications and two referred to multiple grant applications.

Please note that although the interview protocols mention questions that stimulate interviewee to think also about unsuccessful evaluations, in practice the interviewer almost always forgot them. So, interviewees mentioning unsuccessful evaluations, did so spontaneously.

## **Relations between important developments and influential evaluations**

Table 4 gives an overview of the three most important developments that interviewees mentioned and Table 10 the three most influential evaluations. As set out in section 2.1 we propose an indicator for the impact of evaluations as the number of relations between the two sets of three answers. In other words: to which extent do the three most influential evaluations refer to the same events as the three most important developments. The more that is the case, the more reason we have to conclude that evaluations have a high impact on someone's work or career.

In terms of numbers, we found the following relations. 33 out of 37 interviewees<sup>23</sup> mentioned three most important developments (MIDs) and three most influential evaluations (MIEvals). The distribution of the relations for these 33 interviewees is shown in Table 13.

We checked break-downs over disciplines, countries and cohorts but found no remarkable differences in the distribution.

The table makes clear that in most cases one or more MIDs are related to one or more MIEvals. Although in principle one MID could also relate to more than one MIEvals and vice versa, this does not occur often.

---

<sup>23</sup> One interviewee mentioned in total only two MIDs, two interviewees mentioned three MIDs but either four or only one MIEval

Number of MIDs with related MIEVals	Number of MIEVals with related MIDs	Gender		Total
		F	M	
3	3	1	1	<b>2</b>
3	2	0	0	<b>0</b>
3	1	0	0	<b>0</b>
2	3	0	1	<b>1</b>
2	2	2	8	<b>10</b>
2	1	0	0	<b>0</b>
1	3	0	0	<b>0</b>
1	2	0	1	<b>1</b>
1	1	7	6	<b>13</b>
0	0	2	4	<b>6</b>
<b>Total</b>		<b>12</b>	<b>21</b>	<b>33</b>

Table 13: Distribution of relations between MIDs and MIEVals for 33 AIs

The distribution shows a discrimination between different pairs of  $x$  MIDs and  $y$  MIEVals. However, besides that, it is not very telling. The question is also what kind of important developments are related to what kind of influential evaluations? For example acquiring a new job is often mentioned as an important development (Table 6) and job applications are frequently mentioned as influential evaluations (Table 10). To which extent do these job applications refer to these jobs?

The short answer: only to a limited extent. Table 14 gives a count of the relations at the level of event-code groups for the 33 AIs of Table 13<sup>24</sup>. It shows that indeed in some cases, influential evaluations are related to likely important developments. The most striking example is that in 11 cases where an interviewee mentioned a 'Job' type of MID, he/she also mentioned the job application as an MIEval. Another striking example is that in six cases where an interviewee mentioned a diploma as an MID, the interviewee also mentioned the exam leading to the diploma as an MIEval. And the same for three answers regarding grant applications.

Although the examples may be striking within the context of the table, when compared to the lists of MIDs and MIEVals, the result is not impressive. In total 37 MIDs were about jobs and 26 MIEVals were about job applications, yet in only 11 were the reported applications about the reported jobs. The same goes for diplomas: diplomas were mentioned nine times as a MIDs and 15 times as a MIEval, but were linked in only six cases. Interviewees mentioned grant applications five times as MIDs and 19 times as MIEVals, whereas in three cases, they were linked.

<sup>24</sup> We also generated the table for 37 interviewees, which showed the same pattern.

MIDs event code group	MIEvals event code group									Total
	Job applications	Diplomas	Grant applications	Other	Job-related evaluations	Feedback	Evaluations leading to an invitation to something	Other informal evaluations	Paper and book reviews	
Job	11		1	1	2					15
Other				2	1	2	1	1	1	8
Diplomas		6								6
Doing a PhD	1	1	1	2			1			6
Grant applications			3							3
Changes in context – organization or work related	1				1					2
Changes in context – non-work related or family related								1		1
Feedback						1				1
Grants			1							1
Job applications	1									1
<b>Total</b>	<b>14</b>	<b>7</b>	<b>6</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>44</b>

Table 14: Relations between MIDs and MIEvals for 33 AIs

## Section conclusions

Our results are more or less in line with the findings of Bazely (2003) and Baldwin & Blackburn (1981) the only two studies that also aimed to establish important career-developments from the individuals' perspective (See section , p. 11). Baldwin & Blackburn (1981) focus on the roles of assistant professor and higher up the academic ladder, probably as a result of their research question which focuses on the staff already present at the universities: "How can colleges and universities most effectively capitalize on the potential of their currently employed, experienced faculty?" (p. 598).

Compared to the five steps Bazely (2003) identified<sup>25</sup>, we found the additional step of the choice for academia. To many this is a step at the start of their working life, but to some it comes later on, or at a couple of moments in their career when they first decide to leave and later return back to the university. Besides this difference, all steps that Bazely identified can be recognized in our findings and vice versa.

One overall difference is that Bazely presents these steps as five "common and critical points" (p. 261) for all researchers, even after pointing out how much diversity exists between careers. We did not systematically find this. An important reason for this difference may be that we asked for three important developments while leaving it to the interviewees to explain what made the developments important, whereas Bazely puts her questions in the perspective of striving to become an established researcher. That is, not all our interviewees may be striving (or have been striving) to become an established researcher. Moreover, even if Bazely's five career points are critical to all researchers at this point, not all our interviewees may perceive them as equally salient or important because they may have had no problems passing these points: it happened, as it were, 'naturally' to them.

<sup>25</sup> 1) completion of high-level research training; 2) obtaining an academic position; 3) balancing teaching and research activities; 4) maintain a research profile under increasing pressure of other tasks; 5) achieving the status of established researcher.

In Section 2.1 we set out a design for an indicator of the impact of evaluation on academics' daily lives and careers. In short, it measures how many of the three most influential evaluations (MIEs) refer to how many of the three most important developments (MIDs) in someone's work and career.

We found that a few interviewees showed three MIDs related to three MIEvals, and a few showed no related MIDs and MIEvals. The in-between categories of two MIDs to two MIEvals, and one MID to one MIEval occur in similar numbers of interviews. This means that we found a reasonable distribution of interviewees over the possible values for overlap. So, it seems that the indicator works, at least for the group of 33 interviewees. The number of interviewees is not large enough for statistical analysis, but the results encourage upscaling of the test via a survey-based approach.

Secondly, with this indicator, we can say that in the perception of the researchers evaluations do have an impact on their career and daily life.

Thirdly, this relation is however not straightforward. In quite a number of cases, interviewees mentioned important developments which normally include an evaluation in the process, but these evaluations were in themselves not deemed important. Perhaps more puzzling is the fact that some evaluations were deemed influential or important, but the thing gained through it was not.

Another thing we can learn is that important developments in someone's career not only may-or-may-not be influenced by those types of evaluations that we or DHs would expect to be important (such as job applications for jobs and grant applications for grants) but also by other types of evaluations, such as informal evaluations (Chapter 5).

### 3.3 Deans, HR managers and HoDs

This section will report and discuss the basic answers that deans, human resources managers and heads of departments (in short DHs) gave to similar questions that we asked the researchers. The main difference in the questions was that we asked for the organizational perspective.

Please note that the interviews did not cover all four countries in similar ways - see Table 2, p. 17. In Germany we spoke with only one dean. In Poland we interviewed two DHs, one of which did not answer the main questions<sup>26</sup>. The numbers presented in this report are meant to give an impression of what we found through the interviews. In this chapter breakdowns of tables over countries and disciplines make limited sense because for Poland and Germany we have little data compared to the UK and the Netherlands.

#### Most important developments

To the question what are the most important developments in the careers and work for members of the scientific staff, the DHs answered as summarized in Table 15 and further detailed in Table 16.

---

<sup>26</sup> The interviewee used most time to tell about the current developments around the introduction of a new evaluation system of habilitation applications. He misunderstood the question or simply did not know what to answer.

Event code group	Gender		Total
	F	M	
Job	5	6	11
Changes in context – organization or work related	1	4	5
Grant applications	1	3	4
Other	1	2	3
Job-related evaluations	1		1
Other official or formal evaluations		1	1
Titles		1	1
<b>Total</b>	<b>9</b>	<b>17</b>	<b>26</b>

*Table 15: Three most important developments mentioned by 10 DHs, by gender*



Event code group	Event code	Gender		Total
		F	M	
Job	Getting tenure / a permanent position	1	3	4
	Promotion	3		3
	Professor level (job)		2	2
	Acquiring/ starting of a job	1		1
	Associate prof. level		1	1
<b>Total</b>		<b>5</b>	<b>6</b>	<b>11</b>
Changes in context – organization or work related	Internationalization		2	2
	Local reorganization	1	1	2
	More collaborative work is needed		1	1
<b>Total</b>		<b>1</b>	<b>4</b>	<b>5</b>
Grant applications	Grant applications – successful	1	2	3
	Grant application – successful		1	1
<b>Total</b>		<b>1</b>	<b>3</b>	<b>4</b>
Other	Other		1	1
	Publish (as an activity)		1	1
	Valorization of research	1		1
<b>Total</b>		<b>1</b>	<b>2</b>	<b>3</b>
Job-related evaluations	Annual (or otherwise regular) rewards or bonuses	<b>1</b>		<b>1</b>
Other official or formal evaluations	National or international rankings		<b>1</b>	<b>1</b>
Titles	Professor level (title)		<b>1</b>	<b>1</b>
<b>Overall total</b>		<b>9</b>	<b>17</b>	<b>26</b>

Table 16: Three most important developments mentioned by 10 DHs, detailed by event code and gender

The DHs seem to agree more between them than the AIs about what are important developments. In their top three, actually entering the academic organizations as an employee does not occur. We do not think that this means entering the academic organizations is not on their agenda. Our finding simply means that DHs do not find it that important that it reaches the top three of most important developments. Perhaps it would come at a fourth or other lower place.

The DHs' answers are about what happens after that and here the important steps are: getting tenure or a permanent position, get promoted, i.e. move up on the hierarchical ladder, and finally, arrive at a professor position or title<sup>27</sup>. Next to that they have a number of general concerns which have to do with

<sup>27</sup> In the Germany, the Netherlands and Poland, professor jobs are separate functions, which are scarce and for which one has to apply. In the UK someone may acquire a professor position by applying for a position or by making promotion.

recent policy issues in science, such as valorization, ranking, publishing pressure, and last but not least: acquiring funding.

The answers in the 'Job' event-code group were equally divided between the Netherlands and the UK, which is not surprising in view of the distribution of interviewees. The distribution over disciplines shows that the two interviewees in PH gave five answers in the 'Job' event-code group and only one answer outside that. 'Grant applications' and 'Changes in context' were only mentioned in EE and P+.

Table 17 provides an overview of what the DHs felt is important for the institute about the three MIDs of researchers. (See Section p. 21). That is, the table counts the occurrence of each code in the transcription for each MID. If an interviewee mentions for example 'attracting\_highly\_qualified\_staff' for all three MIDs, then income is counted three times. It does not tell how often an important thing was mentioned within the discussion of one MID, nor how much time or words the interviewee used to address the thing that was important. To save space, the table only shows those items that were counted at least two times.

Importance code	Gender		Total
	F	M	
• academic_standing>of_the_institute	3	8	11
• acquisition_of_funding	2	7	9
• attracting_highly_qualified_staff	3	2	5
• long_term_institute_research_agenda	1	3	4
• teaching		3	3
• institute_research_profile		3	3
• academic_standing	1	2	3
• developing_one_s_own_research_agenda	2	1	3
• contacts_or_network	1	1	2
• moving_up_the_academic_ladder>to_professor_level		2	2
• attracting_students	1	1	2
<b>Total</b>	<b>14</b>	<b>33</b>	<b>47</b>

Table 17: Importance of 30 MIDs for 10 DHs, by gender (cut off at 2)

The items of importance are not clearly linked to particular event codes or even-code groups, as Table 18 shows. Put differently, an individual getting for example a tenured position is important for a group, faculty or university for different reasons.

Importance code	Event code group						Total
	Changes in context - organization or work related	Grant applications	Job	Other	Other official or formal evaluations	Titles	
• academic_standing>of_the_institute		3	6		1	1	11
• acquisition_of_funding	3	3	2	1			9
• attracting_highly_qualified_staff	1		3		1		5
• long_term_institute_research_agenda	2	1	1				4
• teaching			3				3
• institute_research_profile	2	1					3
• academic_standing		1	2				3
• developing_one_s_own_research_age nda	1		1	1			3
• contacts_or_network	1			1			2
• moving_up_the_academic_ladder>to _professor_level			2				2
• attracting_students		1			1		2
<b>Total</b>	<b>10</b>	<b>10</b>	<b>20</b>	<b>3</b>	<b>3</b>	<b>1</b>	<b>47</b>

Table 18: Relations between importance and event-code groups of 30 MID's for 10 DHs (cut off at 2)

## Most influential evaluations

Table 19 gives the overview and details to the DHs' answers to the question what are the three most influential evaluations for academic staff?

Event code group	Event code	Gender		
		F	M	Total
Job-related evaluations	Annual (or otherwise regular) appraisal talks	2	2	4
	Promotion application (independent of level)	2	2	4
	Tenure track intermediate evaluations		2	2
	Application for sabbatical	1		1
	Application for tenure track as a promotion		1	1
<b>Total</b>		<b>5</b>	<b>7</b>	<b>12</b>
Job applications	Job application – irrespective of success		1	1
	Job application – irrespective of success – for a professor position		1	1
	Job application – successful		1	1
<b>Total</b>			<b>3</b>	<b>3</b>
Other	Reputation	1	1	2
	Other (multiple)		1	1
<b>Total</b>		<b>1</b>	<b>2</b>	<b>3</b>
Other official or formal evaluations	National evaluation scheme	<b>1</b>	<b>1</b>	<b>2</b>
Grant applications	Grant applications – successful	<b>1</b>		<b>1</b>
Other informal evaluations	Other (informal evaluation)		<b>1</b>	<b>1</b>
Paper and book reviews	Paper/book reviews – multiple – successful	<b>1</b>		<b>1</b>
<b>Overall total</b>		<b>9</b>	<b>14</b>	<b>23</b>

Table 19: Three most influential evaluations mentioned by 10 DHs, detailed by event code and gender

Consistent with the most important developments, DHs seem not that concerned about the evaluations necessary to enter the university, but more with moving up on the hierarchical ladder.

During the interviews the interviewer accepted the answer of 'Reputation', however, on second thoughts reputation is not an evaluation. It is the result of evaluations or a criterion for evaluation. In any case, two interviewees, both in the UK, appeared to see it as a form of evaluation.

## Relations between important developments and influential evaluations

Similar to Section 3.2, we will see here to which extent the influential evaluations do or do not relate to the important developments that the DHs mentioned. Out of 10 DHs, six mentioned three MIDs and three MIEvals<sup>28</sup>. Table 20 gives the distribution of the number of relations between MIDs and MIEVals for

<sup>28</sup> Two DHs mentioned one MID, one DH mentioned three MIDs but two MIEvals and one mentioned three MIDs but no MIEvals.

these six DHs. We provide the table below, but since six interviewees is very little, the table is not very informative, not even for explorational purposes.

Number of MIDs with related MIEVals	Number of MIEVals with related MIDs	Gender		Total
		F	M	
3	3	0	1	1
3	2	0	0	0
3	1	0	0	0
2	3	0	0	0
2	2	0	1	1
2	1	0	0	0
1	3	0	0	0
1	2	0	0	0
1	1	2	0	2
0	0	1	1	2
<b>Total</b>		<b>3</b>	<b>3</b>	<b>6</b>

Table 20: Distribution of relations between MIDs and MIEVals for six DHs, by gender

Table 21 shows the relations between the MID event-code groups and the MIEVals event-code groups. Again, the table is not very telling. However, it does show that although the relations seem big, compared to Tables 15 and 19, in many cases the answers are not related: the DHs mentioned 10 MIDs in the 'Job' event-code group and 14 MIEVals in the event-code groups 'Job applications' and 'Job-related evaluations'<sup>29</sup>.

MIDs event code group	MIEs event code group		
	Job-related evaluations (not applications)	Job applications	Total
Job	5	2	7

Table 21: Relations between MIDs and MIEVals for six DHs

## Section conclusions

The DHs interviewed for this report are quite consistent in their views on what academic careers mean for their organization and the role of evaluations in these careers. This mostly applies to the UK and the Netherlands, since we interviewed only one or two DHs in Germany and Poland. Hiring permanent staff is seen as a long-term investment and commitment, which is highly related to the future research agenda of the organization. The next concern is that permanent staff moves up in the hierarchy as it is taken as a sign of quality improvement of research and teaching. The pinnacle of this is the status of professor, which not only is a public signpost for quality but also for the public image of the university in general.

<sup>29</sup> Tables 15 and 19 report results for all 10 DHs, but for the six DHs of Table 20, the numbers are 10 MIDs in the 'Job' event-code group and 11 MIEVals in the event-code groups 'Job applications' and 'Job-related evaluations'

Entering the university system through education and initial post-doctoral positions are of less importance than a permanent position and subsequent steps.

### 3.4 Comparison and conclusions

This section will focus on the comparison of the answers of the AIs and DHs. For conclusions focussing on these two groups of interviewees, we refer to p. 33 and p. 40 respectively.

Academic individuals (AIs) on the one hand and deans, department heads and human resources managers (DHs) on the other, have partially different views on what the three most important developments in academic careers are. We can identify two types of developments: 1) changes in career, actual work or organizational or family context that occur in a short period of time, such as a new job, a reorganization, an important project or a promotion; and 2) those developments that are more gradual and take place over longer periods such as changes in student numbers, societal demands, funding or governance regimes of universities. Only a limited number of interviewees mentioned developments in the second group.

In the first group, we found the following salient agreements and differences.

Firstly, AIs often mentioned events that have to do with their entry into academia: positively answering the question whether they want to start a career in a university (or return to academia), an event that marks this start, either the PhD project or first university-job after that or the first permanent position. To DHs such developments are outside their top-three: for the organization it is less important.

For the organization it is more important whether individuals become part of the permanent staff or not. Since it is much harder to fire permanent staff than to discontinue hiring of temporary staff, hiring permanent staff constitutes a long-term bet on the future. It also constitutes a long-term financial commitment for which funds will need to be found.

To AIs, achieving a permanent position is also an important development, but for different reasons. It means financial stability, although this is not as important for everybody. Although not documented in this chapter, some interviewees reported giving up a permanent position in order to be able to follow their partner or find an institute or city where their and their partners' careers could be combined. To an AI, a permanent position also means recognition of their academic quality and an increase or perhaps a start of academic freedom.

To AIs, recognition, increasing academic freedom and increasing possibilities to develop a research agenda and/or a research group are important reasons to strive for promotion. Often, the higher one gets, the better the chances or possibilities to apply for funding. To some AIs, the professor level is the ultimate goal, but some interviewees also see downsides to that final level (more management tasks and 'political' hassle within the organization).

DHs also see promotion and reaching the final level as important developments, but again, for different reasons. To them, promotions are a quality indicator: the more seniors in the department the better it looks within scientific circles. The more professors are hired, the better the academic standing and the better the university becomes visible also outside academia.

We compared how answers were distributed over several dimensions: country, location, gender, cohort and discipline. These distributions should be considered with care because the total number of interviews is too small for statistical analysis. We found some differences, as reported in the previous sections. In general, we more often found differences between cohorts and between countries, than between disciplines. The gender dimension takes an in-between position.

The fact that differences between countries showed more often than differences between disciplines, indicates that when it comes to career and evaluation regimes, and in the perspective of individual academics, national rules and regulations have a bigger influence on careers and daily work than the (international) disciplinary codes of conduct and criteria. This could mean that in international

competitions, the competitors' main career-country<sup>30</sup> plays a big role since they have been accustomed to national levels of productivity and excellence. However, national competitions often do include some comparison to international standards too, for example by using foreign reviewers or asking reviewers to compare to international standards.

As expected (See p. 16) we found differences between cohorts (See also Must, Otsus, Denton, Mustajoki, & Tiirik, 2012, p. 35). This means that it makes sense to take cohorts into account in evaluation practices. One interviewee pointed out that criteria changed in the course of her career, which made it difficult to comply to the newly introduced criteria.

"When I look at my career and what has happened at the universities during these years is that when I started and when I look at the generation before me, the criteria that they had to ... comply to, were very different than the criteria that were put on me. And the criteria that I had been confronted with, have changed over time ... It's like playing a game and during the game they are changing the rules." (Interviewee 2, 00:54:45.86)

One way to take differences in cohorts into account is to consider whether and how existing cohorts can live up to new criteria that are being considered in order to prevent loss of earlier investment. This would be a concern during the development of evaluation procedures and criteria. During the actual evaluation, differences between cohorts can be taken into account by looking at a researcher's development over time, rather than the present state. For example, instead of looking at 'time-less' quantitative indicators, time-series or indicators that show a trend should be used.

In some respects we found less differences than expected. We selected interviewees such that per country and per discipline, interviewees would belong to the same group or department. We wanted to test if local (i.e. within the organization) career and evaluation regimes may have a particular influence on the answers. However, we found no trace of such influence. DHs sometimes referred to or showed elaborate systems of criteria for promotion or job applications for higher functions, but these were never mentioned by AIs. A possible explanation could be that even though these systems do influence the outcome of those evaluations, the interviewees can not notice it. Often they only know the overall outcome: their application succeeded or failed.

We did find differences between countries, but not always in the way we expected. For example, we choose the UK because of the decades-long practice of the Research Assessment Exercise. However, UK AIs did not perceive that as a major influential evaluation in their work or career, even though they are obviously aware of its existence and occasionally referred to it. One explanation could be that most of these interviewees have managed to have a career, i.e. they have been successful enough in publishing their work, so they could not see the selective effect of the RAE. Another explanation is that not all contributions of all staff members' are actually included in the RAE.

In Poland, all contributions of all staff members are counted on an annual basis and the continuation of their job may depend on the results<sup>31</sup>, but here too, interviewees did not mention the evaluations as an influential one for their work or career. When asked why he did not mention these he explained that it is part of business as usual and since the criteria are very clear, he knows exactly what to do to pass. "I don't know if this system will change. But now it is very predictable you know what you can get from this system, so this is not evaluation that I eh that I'm worried about." (Interviewee 3, 01:32:04.04)

---

<sup>30</sup> I.e. the country where competitors developed the main part of their respective careers.

<sup>31</sup> This means that in case of too low scores, the Dean may decide not to renew a temporary contract, or fire those on permanent contracts in case of budget cuts.

# 4 Invisible work

## 4.1 Introduction: A taxonomy of invisible work

The ACUMEN project addresses output, skills and forms of influence that are overlooked by current evaluation practices. We propose to conceptually address these activities as 'invisible work' (Star & Strauss, 1999).

Star & Strauss (1999 15) point out that work is not inherently visible or invisible. Rather, it is made visible or invisible, mostly through defining what counts as work and what not. They identify (p. 15-16) three forms of invisible work: one form is to make the worker or employee invisible. This is related to power differences between employer and employee. (p. 15)

We note that this situation is not that common in science, at least not when it comes to scientific staff, because in academic organizations the employer (vice-chancellor, dean, department heads) are often a scientist or representative of science. In some countries, such as the Netherlands, a trend is visible in which such scientist in leading positions are replaced by non-scientists. Deans may be recruited from consultancy companies and board members may be from industry or government. Perhaps in the future this trend will lead to a rise of making the employee invisible, but at present it hardly not the case. Scientists in general and professors in particular are highly visible figures in universities and elsewhere in society.

A second form of invisible work can occur when the worker is very visible, but the work itself is made invisible or dismissed to the background as unimportant. Star and Strauss (1999) give the example of nurses who are very visible in hospitals and other care-taking institutions, but whose actual work in-full is mostly hidden from sight and relegated to a secondary stage of mere 'care'. This may apply to academics. They are highly visible, but what it entails to do research or to teach remains hidden from sight.

Lastly, both worker and the work may be made invisible. Here, Star and Strauss (1999) identify two cases. In one case, the products of the work are visible as commodities, bought in a location far remote from the location and circumstances of production. In the other case "Formal and quantitative indicators of work are abstracted away from the work setting, and become the basis for resource allocation and decision-making." (p. 15) The latter case<sup>32</sup> is at the center of ACUMEN's focus, in particular where individuals' efforts are measured by indicators such production, citation and derived measures such as the H-index.

Star & Strauss (1999 15) applied the ideas summarized above to the field of computer supported cooperative work. Antonijevic et al. (2013, p. 61) elaborated a taxonomy of a particular form of invisible work, viz. affective labour.

Affective labour "refers to activities that create, sustain, and/or modify behaviors and judgments" (p. 61) It could be formal or informal, and it could take place between academics or between them and other societal actors. Antonijevic et al. (2013) identify three categories of affective labor: care work, articulation work and persuasion work. Figure 3 provides an overview of the taxonomy

- 1) Care work is "work done in looking after our colleagues, our tools, and our outputs" (p. 66)
- 2) Articulation work concerns<sup>33</sup> "labor practices that support the articulation and the coordination of distributed work" (p. 71). This is a most time consuming type of affective labour, which includes not only formal and informal work, but also ad hoc and unplanned work, corridor-talk (and the loss of it when not co-located)
- 3) Persuasion work: "the persuasion work that is part and parcel of scholarly practice" (p. 76) with three sub-categories

---

<sup>32</sup> The first case, arguably could also apply to science, when we think of on-line learning and of-the-shelf consumption of scientific texts.

<sup>33</sup> The notion builds on Strauss (1985) and Star & Strauss (1999)



- 3a) credibility work : academics' labor aimed at persuading others such as peer reviewers, colleagues, research councils, the general public etceteras, but also themselves (p. 76)
- 3b) reputation work : "aimed at demonstrating that a scholar is capable of producing analysis that meets those criteria", i.e. accepted criteria of validity and reliability (p. 76-77)
- 3c) position work : "scholarly activities related to achieving, confirming, and preserving a certain status or position in an academic community" (p. 77)

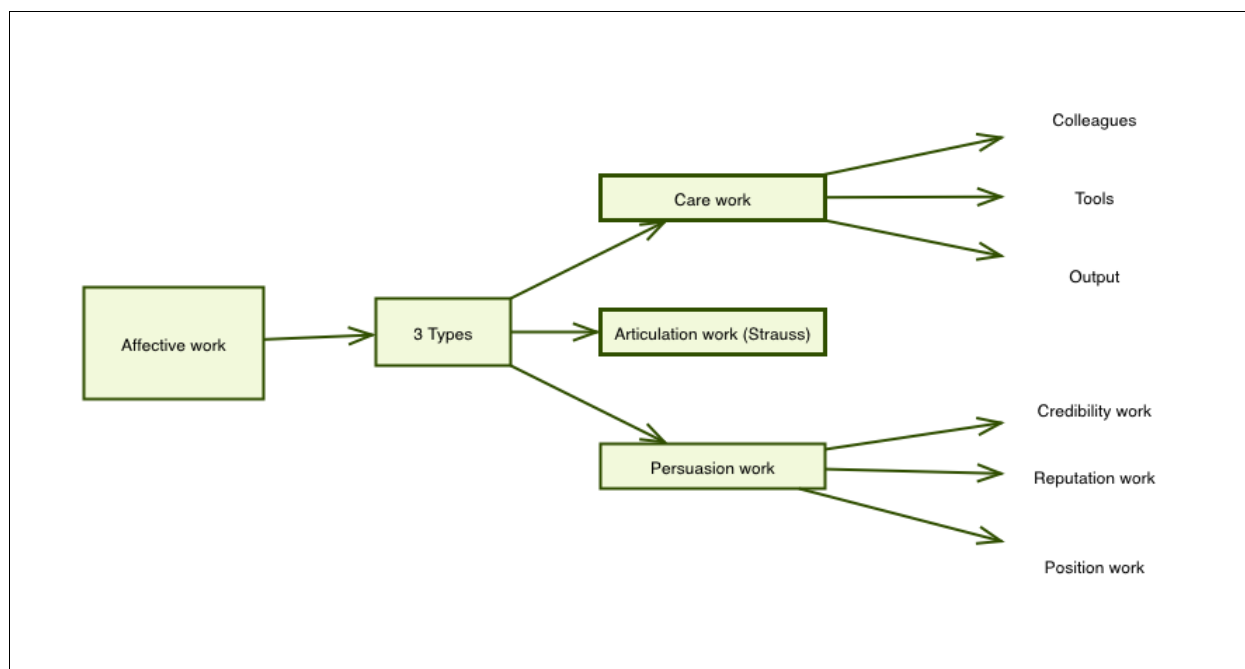


Figure 3: Affective labour and its sub-categories

Similar to Star & Strauss (1999 15), Antonijevic et al (2013) used this frame to study collaborative work with computer support. Since science in part is a collaborative enterprise on many levels, these frameworks offer a good starting point for our analysis. However, not all collaboration is mediated by computers and depending on the discipline and the exact work, academic work can be solitary work. Thus, the categories outlined above, do not cover all forms of invisible work or affective work.

## 4.2 Method details

In Section 2.4, we set out our research method and Appendices 1 and 2 provide the interview protocols. To find invisible work of any kind, we used two means. First, when discussing the three most influential evaluations with AIs, we asked what these evaluations did not evaluate, but should have evaluated<sup>34</sup>. So here we took the evaluations as the starting point and asked what was wrong with them. We did not get many answers, so we describe and discuss them in section 4.3.

Secondly, during the rounding-off questions, we took work and career development as the starting point. We asked AIs what is important for their work in general that is not evaluated in any way, but should be? We also asked why it is important and how it should be evaluated?<sup>35</sup> We asked DHs similar questions referring to the work of researchers rather than their own<sup>36</sup>.

<sup>34</sup> This is question 3.8 in the protocol.

<sup>35</sup> Question 4.

<sup>36</sup> Question 4.

### 4.3 What did the three most influential evaluations overlook?

When discussing the three most influential evaluations with interviewees in the AI group, they generally could not answer the question what the evaluation did not address but should have. Nothing came to their minds. Although we hoped for a bigger response, it can be explained by pointing out that most interviewees mentioned successful evaluations (see p. 30) and one is usually not inclined to criticize them. Or perhaps one would but it may not be among the things one remembers many years after the event.

One interviewee who had had experience with working in industry (AA), noted a general lack in evaluations done at universities: they focus mostly on doing science and how they can help one to do it better. They do not pay much attention to how one is doing personally, whereas in good companies it's the other way around. They reason that if you feel better, you work better.

Of those four interviewees who did mention a negative evaluation in their top three of most influential evaluation, two gave an answer to the question<sup>37</sup>. One interviewee in the Netherlands migrated from Germany some years ago, after he had acquired a PhD and habilitation. He mentioned as a general criticism that evaluations of job applications, promotion applications and career-grant applications do not take into account that the applicant may be from abroad. For the interviewee this meant that he graduated a bit late compared to Dutch standards and the fact that he had his habilitation makes no impression in the Netherlands.

To his estimate, he should have been hired as an UHD (associate professor) instead of as an UD (assistant professor). He aims to move up the ladder, but to get a promotion, he needs to be a successful in the acquisition of grants, in particular a career grant. This counts for 90% of one's application success. However, still according to the interviewee, the major career grants in the Netherlands (known as Veni, Vidi, Vici) also do not take foreign career paths into account.

His experience was a big disappointment for him. "... you are evaluated but then the basis for the evaluation is sort of, yeah, it's, it's what you wrote. It's fit to a traditional career, and if you don't fit this exactly, it's sort of you fall between the cracks or something ... which is in Europe where we all believed it, ... the Bologna process and everything is transparent and we work together and blablabla" (Interviewee 4, 00:49:28.98)

Another interviewee mentioned that the focus on high publication numbers does not take into account that researchers can be in more or less favorable groups. Working in a small group he observed that in bigger groups "where lots of people published and put each other on papers as co-authors. So there was obvious strategy in groups for putting each other on co-authorships" (Interviewee 6 ; 01:07:47.81)

Secondly, getting his research data from remote countries instead of a laboratory also reduced his production speed: "It took me half a year to [travel to a remote place] and and get my samples. Which was great to do but of course is different than someone who hits the [unclear] running the first day he or she arrives in the laboratory." (01:08:32.44)

The publication pressure that the interviewee referred to existed in another country where he did his PhD. He noticed that the evaluation of his Marie Curie application took a much broader perspective and found him good enough.

Two other interviewees who replied to the question, pointed out that an evaluation lacked a broad view. One interviewee in the Netherlands mentioned that a good committee for judging job applications would not narrowly look for matches between the applicants' characteristics and formal criteria but would wonder "When do you have that click?" (Interviewee 2, 01:13:57.45)

The second interviewee, who worked in PH and was interviewed in Poland, criticized both her MA defense and PhD defense for not giving her the opportunity to present herself in a broader sense and present her work to the committee. Instead she was only given the opportunity to answer questions. She also pointed out that the criteria for a job application were geared towards applicants with an MD who could publish based on medical cases, whereas she had a college background and thus had little opportunity to do so<sup>38</sup>.

<sup>37</sup> In case of the other two the interviewer did not ask the question.

<sup>38</sup> This was the interviewee who replied in written form. Unfortunately, the text was not very clear on this point.

Two more interviewees (NL EE, UK EE) pointed out that although they had no problem with the evaluations discussed, the evaluation itself may not be the determining factor. Other factors such as internal politics within the institute may play a role. One of the two interviewees later added that the argument of 'fitting in' may help one's case even though one may not fit other criteria. He also saw this as one of those other factors since there is a lot of politics behind 'fitting in' (Interviewee 6, 01:11:07.58).

Instead of answering what was lacking in the evaluation, two interviewees answered what was stressed too much.

One German interviewee was quite sure that she was asked as a reviewer of EU project applications in her field because she is a German woman and both women and (at that time) Germans were underrepresented in the review committees. She did not like this form of positive discrimination. Similarly, a Dutch woman explained about a job application

" ... I would hate to think, but perhaps it did play a role that my being a woman eh was taken into account, which I would not have liked. I'm not quite sure. I'm reasonably sure that they ... they urged me to apply for associate professorship, which at that time, I was not quite enthusiastic about anymore. ... I'm the only woman on the, on the staff here at [my group]. And they are a bit ashamed about it. So they wanted to show perhaps [laughingly] they were nice to women. But I wouldn't have liked that if that were the case. I want to be evaluated on my, other things that matter, and not on gender, which is quite irrelevant." (Interviewee 10, 01:01:45.85)

She also did not see the point of such positive discrimination, since it will backfire later on:

" Yes, they do, they do a bit of positive discrimination. And I'm, I'm not in favor of it, because eh, if it would have been the case, I didn't know, that I would have been, or other women would have been positively discriminated, they would be discriminated negatively afterwards. People would say: it's not your work, it's just, yes, it's an easy job for you. I haven't hear that by the way, ever." (Interviewee 10, 01:01:45.85)

### 4.4 What is important in academic work and careers that is not evaluated, but should be?

This question asks the interviewees to identify everything they feel is important but made completely invisible in all evaluations. We also asked why they feel the things they mention are important and how it could be evaluated?

Contrary to the answers discussed in the previous section, the answers mostly were about 'work' that had to be done. We summarized the answers and grouped them according to the categories set out in section 4.1. However, we also needed to introduce additional categories.

A number of answers were not taken up in the following overview. Two AIs and two DHs answered that nothing was not evaluated but should be. Apparently they were content with the existing evaluation regime. In one interview the question was not asked, in three more interviews the interviewee provided an answer to a different question (for example, what is evaluated too much?). Five interviewees gave answers that do not refer to activities or traits that refer to themselves and thus fall outside our search for invisible work of academics<sup>39</sup>.

### Caring for colleagues and students

Three interviewees mentioned caring for colleagues and students (see also the interviewee mentioned in 4.3). In all cases the interviewees found the activity important because it enabled them to perform a task,

<sup>39</sup> These answers were: In formal evaluations they don't get the answers they need, because people do not fit; The functioning of my leaders is not evaluated; One always wants more good words for things that go well. At present, one only hears when things go wrong; Besides teaching, nothing is evaluated at all; Qualities rather than achievements

i.e. teaching, doing research or providing service. One woman (UK PH) pointed out that it simply takes time and sometimes a lot of time.

She also noted "They may be inequity thing, in ehm who is doing them. This is a stereotype but I think women more often take on those kind of pastoral roles, supporting students who are failing." (Interviewee 11, 01:09:25.15) Perhaps she is right, but in this research, three out of four interviewees who mentioned care for colleagues and students were men.

Two of the three interviewees saw no way to evaluate care work. The third (UK PH) mentioned time keeping to evaluate the care for students in combination with other quality surveys. However, she added that she had done some time-keeping when she worked in another university, but did not notice that anything was done with that.

## Caring for tools

An early career researcher and a senior researcher from the same group in AA happened to work on the same international project, which aimed to develop a complex measuring instrument. They both mentioned the development work they did on the instrument. The junior researcher mentioned the software development that she did and the management and reporting work that she later took up. The senior researcher mostly referred to management work. An important aspect are the distributed international dimensions, but these are not noticed in his university's appraisals.

Both complained that the university's annual assessments had no eye for this work but only for output in terms of published papers. The junior researcher:

"So only who write paper, produces paper, and by paper I mean refereed publications, is considered. And the others are doing a different job. But, but as far as I know, in the case of [my institute] for instance, the people that are doing other jobs, technical jobs, are the ones that bring the money in. So, and allow the others to do the research. But then they don't get any credit for that." (Interviewee 12, 01:40:04.98)

As the senior researcher explained it:

"I do have a reasonable scientific output. Yeah, but I've got other members in the team who've got far less of scientific output. Simply because of the pressure of the work. It is simply not fair to judge them on their scientific output. Because they've got other commitments other obligations. And they're at least as important for the [field] in general. Yeah, to get a project like this working properly, is in the end beneficial to everyone." (Interviewee 13, 01:15:29.37)

A third interviewee, in the field of EE, mentioned his work as research officer and laboratory manager as important work that is not being evaluated. Here too, his university was only interested in his publications.

When asked why the activities are important the senior researcher mentioned that he has great responsibilities which are not acknowledged. The biggest problem he noticed in the university's annual appraisals is that the people doing these appraisals do not have similar experience with the management of large scale development projects with big time horizons. The research councils do better in this respect: they have committee members with similar experience, be it in different fields

The early career researcher mentioned the importance of the instrument to the field, the time invested and the skills required. Her reporting is actually evaluated in some way through their processing after she files them, however, we note that this does not reach the annual appraisals. Her management work could be evaluated by looking at how many people one is managing and how happy they are with your work.

The EE interviewee also mentioned that it takes time and that one contributes to the work of the organization.

".. ultimately we all work for our employer, ... and if it's strives to be the best in the world at what it does. ... [it] doesn't ... sort of take into consideration that you might be fulfilling a role, which is directly influencing that excellence. influencing because running the lab is looking after the labs. I'm providing it. a [unclear] which is being used by many many people, e than my researchers use. ehm,

and that is in its own way, directly feeding in to every publication that comes out and every piece of piece of research work that's done. and that's just not measured, [unclear] use, so that [is] somewhat frustrating ... ehm ... but it's this black-and-white system that the [university] has chosen to operate." (Interviewee 14, 01:36:20.74)

To improve on the evaluation system which the interviewee reported to focus on number of publications, number of PhD students supervised and amount of grant-money acquired, he proposed a weighted system based on how much time one spends on a task and how well it is done.

All three cases suffer from the fact that the evaluation and reward systems they were part of, do not recognize 'mixed roles' where academic skills and knowledge are needed to do work that would otherwise be categorized as 'support'. Perhaps it is no coincidence that the three interviewees all worked in the UK, where the research assessment exercises have put a lot of emphasis on teaching and research output.

### **Caring for output**

None of the interviewees mentioned care for output as something that was important but not evaluated. Also none of the other questions lead them to implicitly mention this type of affective labour as under-evaluated or under-appreciated.

### **Caring for the organization**

Ten AIs and four DHs gave 18 answers that fall into the category of Care work for the organization. It refers to work and skills that is done or needed to keep the department, faculty or university running. Teaching, committee work and administrative work, and 'good citizenship' were reported subcategories here. They will be described below

This type of work was relatively more often mentioned by women: five out of these 10 AIs were women (compare to the totals of 13 women 25 men), and two out four DHs (compare to totals of three women and seven men in the group of DHs). The distribution of these interviewees over countries and disciplines is more or less even. Eight out of 10 AIs were mid career researchers, the remaining two were early career researchers.

### **Teaching**

Teaching was mentioned six times. Sometimes, the interviewee mentioned the quality of teaching, or student projects in particular. Most of them mentioned the time spent on teaching as the most important reason why it should be evaluated. One mentioned the importance of teaching for the university.

Perhaps it is not so much the evaluation that is lacking, because all the other interviewees did not mention teaching, and many university actually do organize teaching evaluations in some form or another. However, researchers seem to feel a lack of feed back after their efforts have been evaluated.

One interviewee (NL EE) noted that he spend a lot of time, talking and advising students of all levels about their thesis or a project, but got no reward for it at all. The least these students could do is to mention him in the acknowledgements of their theses, he suggested. As things are, however, he is loosing interest: "more and more I don't care, I say: I don't have time. sorry. Cause I know what will happen at the end. You will spend hell amount of time and then they will not remember you". (Interviewee 8, 01:41:05.69)

One other interviewee (AA NL) also mentioned some form of acknowledgement as the major problem.

"if I sit down and write a paper, the return is almost guaranteed, while if I sit down and write a good lecture eh, then, there are warm words, saying that it is nice, but eh if it doesn't really pay back in terms of recognition or something." (Interviewee 4, 01:26:20.77)

Also when you do a bad lecture, it may be noticed by the 'Opleidingscommissie' (Course committee) but it will have no real consequences, according to this interviewee.

Finally, one Polish interviewee (EE) argued for an increase of the points given for teaching. In Poland, academics regularly are being assessed for their work in research, teaching and academic organization. It is a quantitative assessment with weighed scores for output and activities in each of these three tasks. He proposed to increase the weight of teaching activities in this system.

## **Committee work and administrative work**

Five AIs and two DHs mentioned committee work and/or administrative work in general. Those who mentioned administrative work did not specify clearly what it entails, but it's likely that committee work would be included. One interviewee (P+ DE) mentioned his role as director of studies, two interviewees (PH, UK and NL) mentioned their work in the 'ethics committees', one interviewee her work in the equal-opportunities committee (AA DE). Lastly, one interview mentioned all administrative work with a long-term horizon, such as establishing a new chair or reorganization of the curriculum.

Like with teaching, the most often mentioned reason why this work should be evaluated is that it costs time, or as one interviewee put it: it takes time away from things that are evaluated. One interviewee mentioned that the task of director of study takes also takes a lot of energy and skills. Finally, one interviewee saw a need for evaluation of the medical ethics committee because the committee or some of its members may function badly.

On the 'how' question, interviewees were not very outspoken and differed. For some, it was more a matter of being compensated for it through a relief of other jobs, an assistant, or a temporary raise. When it comes to department management, one could check the overall department's functioning. In case of the work of some committees (medical ethics or appointment committees) privacy issues may prevent 'outsiders' looking into the committees' work. One interviewee reported that the medical ethics committees are evaluating each other through mutual working visits.

## **Good citizenship**

One DH (UK) mentioned 'good citizenship': staff members volunteering for administrative work. Although difficult to assess, his university declines applications for promotion when the applicant has clearly shown selfish behavior. Another DH (DE) noticed the phenomenon that not all staff members contribute equally to what we call care work for the organization. Some volunteer easily, others do not. A third DH (NL) felt that PhD students who are often requested to do extra tasks or services (such as help organizing a conference or take on extra teaching) or who voluntarily take on more than they need to should be compensated for that with extension of their contract<sup>40</sup>, which would require that those extra activities are being registered. She noticed that some types of work done by staff members is poorly registered, such as the coaching of junior staff members.

## **Caring for the discipline**

Academic researchers function in at least two professional worlds: the one of the employer and the one of their discipline (See p. 11). This apparently causes some work to be made invisible, at least to and by the employer. Three interviewees mentioned paper reviews and one referred to conferences. One added

---

<sup>40</sup> In the Netherlands, PhD students are employed

reviewing of project proposals and one added editorial work for a journal. DHs did not mention these activities.

Two interviewees mentioned time as a reason why it should be evaluated. However, they also see it as a necessary part of doing science. As one interviewee explained

" It's a contribution to the broader academic good if you like. Academia couldn't go on if if people didn't agree to peer review. Didn't take the sort of unpaid extra roles on in terms of editing, working with journal editors. " (Interviewee 11, 01:06:52.66)

The interviewee who mentioned discussing each other's work at conferences was objecting to a Polish practice of merely attending and presenting conferences to gather points for the appraisals. She noted at one conference there was no discussion at all, "but what [is the] point for me? [laughingly] I want to know something new, I want to learn something, I want to discuss my project." (Interviewee 15, 01:31:10.84)

Interviewees did not offer views on how this could be evaluated, either because the question was not asked or because they did not know how it could be done in job applications. One interviewee however pointed out that a problem is that the reviews are anonymous, but that one's reviews are being evaluated:

" [W]e do get evaluated, but in a different way. Just by the whole community. Well anyone who has seen your referees reports will form a judgement on ... based on that. So you are evaluated, but just in a kind of general way" (Interviewee 16, 01:23:54.91)

An interviewee (who himself did not mention reviewing as something that is not evaluated), backed this up with a story about how he was asked to become an editor. He was quite sure that the former editor had seen his reviews and asked him to become a successor based on that. He also felt evaluated as an editor by his reviewers: if he would send them bad work, they would evaluate him on that.

### Caring for oneself

One Polish AI, one in the UK and a UK DH all mentioned activities that have to do with self-development, such as self-education, self-evaluation and skills development. The three responses were quite different from each other, but seem salient enough to install a category of invisible affective labour. Some of these things are arranged in some countries or universities through the annual appraisal talks. Apparently, they are not addressed in the Polish interviewee's university, and affiliated non-paid staff and other groups of researchers may be exempted from appraisal talks.

The UK-based AI pointed out that self-evaluation not only identifies weaknesses in skills or output, but also makes one realize how successful one is. "you see that eh you're not wasting your time by putting more efforts in that you can create new methodology, that you can create new, master new areas, new, open new research fields." (Interviewee 17, 01:12:47.36)

The DH pointed out that although staff members have access to training facilities for personal development and career development, usually the same people show up, whereas most staff does not show up at all. Still, she argued, it is very good for the individual because people rarely stay at the same post. But it is also beneficial for the organization, because people might return in more senior positions.

### Articulation work

11 answers from six interviewees in the AIs group (including four from the UK) and three in the DHs group (from Germany, the Netherlands and the United Kingdom) fell in the category of articulation work.

Management and leadership on research projects were mentioned five times by AIs and one time by a DH<sup>41</sup>. Sometimes they would qualify their answer. One AI remarked that project management is not

---

<sup>41</sup> Management and administrative work, when referred to in general is discussed under Caring for the organization

evaluated unless one overspends, and one mentioned that they were evaluated but could be evaluated more. One remarked that the university completely overlooked the international dimension of his management work.

Other interviewees pointed at other forms of articulation work. One AI mentioned dealing with difficulties. One AI and one DH mentioned 'dealing with unreasonable people' and social competence. Another AI and another DH mentioned collaboration. Interestingly, the latter two worked in different disciplines and different universities, but both were interviewed in the Netherlands and both pointed out that the demand to collaborate conflicts with the demand for competition. The DH put it as a question: "it is a rat race and there is a pikorde [pecking order]. But how can you work together in such a way that your success can help the success of others?" (Interviewee 18, 01:25:39.44)

The AI contextualized the difficult combination of demands and confronted it with the policy demand for societal relevance.

" what I also miss and maybe it cannot be in an evaluation, ... it's more the ehm collaboration between. Because we are doing research and we're doing research for the society. With money from society and most of the time we are just in [competition] with each other. Instead of working together to get new knowledge. Eh, and ehm, that's makes the the whole situation now ehm, how it works in the universities and what the government ehm is ehm stimulating. [It] makes it more difficult for us to work as a group and to, and not only here at the group, but also with other universities" (Interviewee 19, 01:19:37.27)

## **Persuasion work: credibility**

Seven AIs mentioned a form of credibility work. Their answers fell apart in two groups: being an intellectual and outreach to society.

### **Being an intellectual**

One early career philosopher in the Netherlands and one senior researcher in Poland phrased it in different ways but mentioned being an intellectual. One phrased it as "what is important that people have an ability to as we say eh to open eh a bit of a new horizon. eh to be ehm flexible in their mind ... sometimes I miss the kind of eh ability to see things in a wider context ... a good scientist should be able to put things into perspective" He put this in the perspective of grant applications where it would be handy if reviewers indeed do have this flexibility in their thinking.

The Polish interviewee phrased it like this:

" Und in der Bewertung was ist für mich wichtig aber nicht bewertet wird, ist eine allgemeine, wie soll das sagen, Humankultur, Humanbildung. Die ich nicht nur als ich Bücher lese, wissenschaftlich arbeite, aber auch dann wenn ich ins Kino geh, ins Theater gehe, wenn ich ein interessantes Gespräch mit meinen bekannten oder andere Leuten eh führen kann." (Interviewee 20, 01:07:57.82)

" And in the evaluation, what is important for me but not evaluated is a general, how shall I put it, humanistic attitude or education. Not just when I read a book, doing scientific work, but also when I go to the movies or to the theater, or when I am having an interesting conversation with acquaintances or other people." (our translation)

In the humanities, such a general, broad perspective is necessary, but the Polish appraisals do not take it into account, the interviewee argued.

Although both interviewees saw a need to evaluate the intellectual aspect, they both claimed that it is very hard to evaluate, or at least did not know how to.



## **Societal outreach**

The second form of credibility work that interviewees reported as not evaluated, concerns outreach to society. Five AIs mentioned this: three women and two men, three in PH, one in P+ and one in EE; two in the Netherlands, one in Poland and two in the UK.

One UK interviewee, working in public health:

"Some of the important work I'm doing ... is more implementing and changing eh local systems that will only get into low impact journals. But it may have a bigger impact on health policy and practice on the ground in those settings. So ... evaluating actual change in practice. Improvements in health is, is almost impossible to do well so we tend to focus on what's easy to evaluate, which is the journal, the citation rates." (Interviewee 11, 01:14:27.67)

A Dutch interviewee, also pointed out that it is not easy to translate academic findings into Dutch clinical practice. She noted that nobody is interested in giving companies or physicians any support to implement findings into practice. However, a colleague in the same faculty does observe a change in policies and evaluation practices. She is now asked to list everything related to societal impact, not just implementation, also for example media appearances.

Besides the fact that the work takes time (mentioned twice) one interviewee explained that applied research is her main interest and for another it brought fun to his research.

As to how societal impact could be evaluated, two interviewees mentioned outreach publications, one added indications of international exposure, such as being invited by international institutions like UNESCO for a project. In the other interviews the question was not asked or not answered, or the interviewee did not know how to.

## **Persuasion work: reputation**

No interviewee mentioned something that we could categorize as reputation work.

## **Persuasion work: position**

Two interviewees mentioned position work, or rather the result of that as something that should be evaluated whereas it is currently not.

## **Reputation in the evaluation of funding applications**

One AI (NL EE) answered that evaluations of FP7 grant applications hardly take someone's reputation into account, but look at what the applicant is doing in the present. He argues it should be taken into account because someone's past success is a good indicator of his future success in projects. The interviewee had worked in industry for 15 years, and there it works the other way around. There, if one asks for big amounts, they do take one's reputation into account. Or, it may happen that one received collaboration proposals from people one has never heard of who found you because of your reputation. Moreover, still following the interviewee, in science, once one is a successful applicant, people from other groups want to cooperate.

In funding applications, one can simply take past success into account. "what would be very interesting is at the end of the project, eh, to know from the, the person that, that eh decided to grant the project, if he thinks that what was claimed eh, at the beginning, is, is actually realized, or, or that at least the impact is realized. and maybe that the project has gone a little bit different way, be-, because it's science. you never know what's happening. ... and we don't do that." (Interviewee 21, 01:51:39.05) Whereas "... industry does that very very clearly. Not in a quantitative way, but eh, captains of industry

meet with each other and they tell each other about ... successful ... things that they, that they do."  
(Interviewee 21, 01:52:32.38)

## Reputation databases

Later in the interview, when answering another question, the interviewee mentioned how "the scientific community silently evaluates" the depositing of certain data and the quality thereof. Examples of this are databases for stellar data, software algorithms or crystal structures. In the latter case "just putting the data, just, just putting the, the long axis, the the second axis, the third axis, their angles. that is considered a scientific publication. With no words, hè. Just [laughs], just the, the data about your compound."  
(Interviewee 21, 02:05:41.83) Names of individuals and groups are linked to the entries and they are being seen as scientific achievements that end up in CVs.

The interviewee added that such use never was the purpose of these databases.

"but it's, it's not ex-, it is not explicit. This, the people who set up those databases set them up with, simply with the idea that the data would be available. Not with the, the, the philosophy that the persons that would entered ... would get higher academic status. It's just a result." (Interviewee 21, 02:06:44.84)

We can add that this is exactly what happened to the Science Citation Index, which was devised to connect scientific articles, but is now used to indicate someone's reputation. Typically, the interviewee also spoke of "the most esteemed eh databases of algorithms" (Interviewee 21, 02:07:28.98) which suggests that similar to scientific journals, there is a ranking of algorithm databases.

Back to the interview: when we suggested to start using these databases in more formal ways in evaluations, start building indicators to rank databases and entries and so forth, he agreed it would technically be possible, but when asked whether we should do it responded :

"it is dangerous. I think now there is already a little bit of eh eh irritation. If somebody thinks that he has a new compound and the reviewer thinks he has not, eh, there is a eh, some error in the measurements or whatever. Eh, that irritation would grow more, I think if you're, if you are aware that this is a important step in your scientific career." (Interviewee 21, 02:11:04.42)

He continued to argue that it is not so important because it is not like a journal publication which people in the end will forget.

"I think now it's eh, people sort of know it, that it's important, but it's not so. ... It's not like a publication, hè.

In a publication, the journal may eh, may be forgotten in the end or it will, may end. Nobody may read it. Nobody reads things that are 40 years old. hè? eh, but eh, if your name is in these kind of databases, it is there forever. ... it's even, it's even worse than a patent, hè? A patent eh, is something for 20 years or so and then people will forget about it, but this is really, yeah. That compound carries your name and it's something of nature, of course." (Interviewee 21, 02:11:49.37)

Although, he seemed to argue that entries in the crystal database are not that important for someone's reputation, we would not be surprised if material scientists would argue the opposite. That would increase rather than reduce the risk that the interviewee is warning for.

## Individuals' reputation

On dean (PH NL) also mentioned authority, which he saw as a professor's quality which is the basis to influence research, teaching and the general public. To the interviewee, authority and academic leadership are strongly related. He repeatedly stated that it is very hard to measure these qualities and agreed that it takes some tacit knowledge on the part of the evaluators to establish someone's authority. Researchers assess authority when they watch someone giving a presentation or answering questions.

Sometimes it is very clear when someone does not have authority. A nice thing about evaluation committees is that they can detect authority. Other than that, still following the interviewee, the only way to measure it, is to ask for references.

### Creativity

We do not propose creativity as a category of affective labor, but mention it as such because it does not translate easily into one of the outlined categories (see section 4.1).

Two department heads (EE UK, P+ PL) listed creativity as something that is important but not evaluated. Whereas the UK interviewee mentioned creativity as another thing that is difficult to evaluate (like leadership), the PL interviewee elaborated more on it: "probably the most important is creativity and eh ... ability to to think in a new way ... to search to new problems, or new perspectives ... but it's very difficult and I don't think that institutions, they promote this kind of thinking. " (Interviewee 22, 01:26:21.37) He argued that it is an important part of philosophy and probably the most difficult. Many important developments in philosophy came from outside the field.

Both departments heads agreed that it is difficult to measure and evaluate creativity. The UK department head observed that it is easier to spot the absence than it is to tell which out of three candidates is more creative. The Polish department head agreed that it is difficult to evaluate but added that creativity as it finds its way in innovative research proposals is usually received negatively because evaluators have difficulties to see what the applicant wants to do.

## 4.5 Conclusions

### Invisible and overly visible aspects of personal properties and backgrounds

We received little response to the question: what did the discussed evaluations not take into account but should have? (Section 4.3) The responses we did get, did not refer to particular activities, skills or output, but to personal properties or circumstances in the contexts of job and grant applications: gender, general well-being, nationally or disciplinary uncommon career paths or research histories, and overall 'fit' of an applicant in the work environment. When we asked 'What is not evaluated at all but should be?' (Section 4.4), most of the answers did refer to activities. A notable exception was 'creativity', which was mentioned twice and which also could not be categorized within affective labor<sup>42</sup>. Although it is not difficult to see that it is related to affective labor, it is not labor itself. It is a quality of a person or an act.

Except for the gender dimension, interviewees complained about evaluations' lack of consideration for these properties and circumstances. A general solution would be to more thoroughly discuss the use or non-use of these properties and criteria and then take them up into formal criteria or guidelines. There are however practical limits to this. For example, it will be very hard to take foreign career paths into account. There are many and it requires a huge amount of knowledge of such paths on the part of all committees. Secondly, things such as the 'fit' of a job applicant in the future research environment are difficult to objectify.

Another solution would be to accept the fact that committees for grants, jobs and promotions need some discretionary space to diverge from formal criteria. This will possibly solve the problem of 'fit', it broadens or opens up an avenue to bring in the 'political' arguments which applicants consider inappropriate. See also the conclusions of the next chapter (p. 69), which deal with these 'political' arguments.

---

<sup>42</sup> Some other answers also referred to qualities such as management skills, social skills, or leadership, but those can easily be related to work.

In terms of our taxonomy of invisible work, the answers elude our categories of invisible work and affective labor because they do not refer to work or output nor to the person as a whole. Rather, evaluations render certain aspects of a person invisible. These aspects are very real and relevant to these persons but apparently not to the organizations doing the evaluations.

Interestingly and further complicating the issue, for some aspects, such as gender and nationality, it is the reverse: the evaluations render the aspect too visible, whereas the evaluated would prefer them to be invisible. In some countries, such as the United States, it is common not to include information about age and gender in order to reduce discrimination.

## **Under-evaluated work**

Interviewees had no problems identifying important work or things that are important for their work that are not evaluated. Table 22 gives an overview, ordered by categories of affective labor.

We had to add categories of affective labor to those proposed by Antonijevic et al. (2013). Their categorization was designed for work in collaboratories (computer supported collaborative work), which obviously does not cover all academic work. To be able to cover all the answers we added three sub-categories to care work: caring for the organization, for the discipline and for one-self. Assuming that the latter two speak for themselves, caring for the organization differs from articulation work in that articulation work refers to a specific project/collaboration, whereas caring for the organization refers the organization in general.

We checked how the answers were distributed over gender, disciplines, cohorts and countries. The table makes clear that women identified more invisible work than men. However, within the group of AIs, interviewees in PH produced over 20 answers whereas interviewees in other disciplines produced around 10. In PH we interviewed more women than men, contrary to the other three disciplines. Thus the high amount of answers related to women could also be related to the discipline.

Furthermore, we found that about half of the answers came from UK interviews. It could mean that there is simply more overlooked in UK evaluations, but it could also mean that in the UK, evaluation practices are further developed and consequently, academics are more engaged and scrutinizing them.

Lastly, mid career individuals produced comparatively more answers (27) than one would expect based on their share in the total, and senior academics slightly less (10).

Interviewee group	Invisible-work category	Answer	Gender		
			F	M	Total
AI	1a Care work > colleagues	Collegiality / cooperative attitude	1	3	4
	1b Care work > tools	Technical/methodological support	1	2	3
	1d Care work > organization	Administrative work	4	1	5
		Management skills	1	1	2
		Teaching	2	4	6
	1e Care work > self	Other	1	1	2
	1f Care work > discipline	Administrative work	1		1
		Refereeing work	3	2	5
	2 Articulation work	Collegiality / cooperative attitude	1	1	2
		Leadership skills	1		1
		Management skills	2	1	3
		Other		1	1
	3a Persuasion work > credibility work	Being an academic/intellectual		2	2
		Societal impact or relevance	3	2	5
	3c Persuasion work > position work	Other		1	1
		[Question not answered]	1	2	3
		[Question not asked]		1	1
		Nothing		2	2
		Other	3	1	4
		Total		25	28
DH	1d Care work > organization	Administrative work	1	1	2
		Good citizenship	2	1	3
	1e Care work > self	Management skills	1		1
	2 Articulation work	Collegiality / cooperative attitude		1	1
		Good citizenship		1	1
		Leadership skills		1	1
	3c Persuasion work > position work	Other		1	1
	98 Other	Creativity		1	1
	99 –	Nothing		2	2
		Other		1	1
	Total		4	10	14
Overall total		29	38	67	

Table 22: Overview of reported work that is invisible to evaluations

In some categories both AIs and DHs are represented, whereas in others only one of the two groups is. Both groups mentioned committee work and administrative work (Caring for the organization), Caring for oneself, Articulation work and Position work. DHs did not mention Caring for colleagues and students, Caring for tools, Caring for the discipline, nor Credibility work. Unfortunately, we can not tell

why they did not give answers in these categories. Did they overlook them, did they think these things are evaluated or did they not deem them that important to be in need of evaluation? Two DHs did however mention creativity, which was not mentioned by the AIs. Here too, we do not know why none of the AIs mention this.

This chapter lists a number of activities and aspects, that can be taken up in the design of the ACUMEN Portfolio. Some already were part of the design, others were added.

In view of the ACUMEN goals and ideas, conspicuously absent from the list are on-line activities such as blogging, twittering, participating in on-line discussions, preparing and up-loading on-line resources as web-lectures, slides and the like. An obvious reason for this could be that most interviewees do not have a substantial on-line presence in this sense, other than staff pages at their universities. A quick scan revealed that one interviewee maintained a project website, two had personal websites, four had somewhat expanded webpages. Another reason could be that they see these activities as part of work that is or is not evaluated.

## Importance of invisible work

Table 23 gives an overview of the reasons why the different kinds of invisible work are important. Three groups can be distinguished. One is that the activity takes resources such as time, energy and skills. Secondly, the activity is a sine qua non of research itself, research projects, the functioning of a department or faculty, the functioning of the discipline or of the individual. In these cases the work could be considered support work, but of such quality that it can not be done by support staff but requires people with academic roles. In many cases the problem for the individuals is that the high-level support work limits possibilities to produce papers or other output that evaluations take into account.

Thirdly, an important group of reasons refers to maintenance of quality: the quality of the courses taught, the research done, the functioning of the organization, or the functioning of the world outside science (in case of societal outreach).

Lastly, a group of reasons refer to things like influence and creativity, things that 'make science tick'. The interviewees seem to argue that these are different from high-level support work, but things that make science possible in the first place.

The most often mentioned reason why particular work should be evaluated is that it takes time. When it comes to peer review of project applications, our ACUMEN colleagues also found that those respondents in their survey that decline requests to review, do so because in large part because their institutes' time allocation policies prevented them<sup>43</sup> (Must et al., 2012, p. 30).

The argument that an activity costs time and thus needs to be evaluated goes a long way. If something takes up a lot of time, it can be assumed that employers want their staff to do it. In that case, the interviewees rightly feel it should be evaluated so that it becomes visible.

However, the argument that certain work takes time has a few problems. Perhaps a task takes a lot of time, because someone is not good at it or very inefficient. Secondly, evaluators may consider that some of the work mentioned is an aspect of tasks or work that actually is evaluated. It could then be that this aspect takes an unexpected amount of time and as such stands out as under-evaluated.

Finally, as with all work, one may not receive any appreciation of it through evaluation, it may still very well be worth one's while. Two of the reasons why respondents do participate in application review is that it gives an overview of their field and/or that it helps them to assure the quality of their field<sup>44</sup> (Must et al., 2012, p. 29).

---

<sup>43</sup> Other important reasons were that no payment was offered, there was a conflict of interest, or the respondents find the work boring.

<sup>44</sup> Other important reasons are : feeling a general obligation to the field and helping/educating fellow researchers.

Invisible work category	Why should it be evaluated?	How to evaluate it?
Caring for colleagues and students	<ul style="list-style-type: none"> <li>- it enables accomplishment of another task</li> <li>- takes time</li> </ul>	<ul style="list-style-type: none"> <li>- don't know</li> <li>- time keeping and surveys</li> </ul>
Caring for tools	<ul style="list-style-type: none"> <li>- general importance for the field</li> <li>- enables department colleagues to produce</li> <li>- enables organization to function</li> <li>- brings in money</li> <li>- takes time</li> </ul>	<ul style="list-style-type: none"> <li>- survey among colleagues</li> <li>- weighted system based on time spent</li> <li>- recognizing mixed roles (academic and support)</li> </ul>
Caring for output	-	-
Caring for the organization: teaching	<ul style="list-style-type: none"> <li>- takes time</li> <li>- guarding quality of education</li> <li>- status of the university</li> <li>- lack of feedback</li> <li>- lack of acknowledgement</li> <li>- lack of consequences</li> </ul>	<ul style="list-style-type: none"> <li>- student surveys</li> <li>- acknowledgement</li> </ul>
Caring for the organization: committee work and administrative work	<ul style="list-style-type: none"> <li>- takes time</li> <li>- takes skill and energy</li> <li>- quality of committee work</li> </ul>	<ul style="list-style-type: none"> <li>- compensate for time</li> <li>- check the functioning of the department</li> <li>- mutual evaluation of 'closed'<sup>45</sup> committees</li> </ul>
Caring for the organization: good citizenship	<ul style="list-style-type: none"> <li>- general functioning of the organization</li> </ul>	<ul style="list-style-type: none"> <li>- compensation</li> <li>- listing or registering</li> </ul>
Caring for the discipline	<ul style="list-style-type: none"> <li>- takes time</li> <li>- necessary part of science</li> </ul>	<ul style="list-style-type: none"> <li>- the community evaluates</li> </ul>
Caring for oneself	<ul style="list-style-type: none"> <li>- personal development</li> <li>- self esteem</li> </ul>	<ul style="list-style-type: none"> <li>- annual appraisals</li> </ul>
Articulation work	<ul style="list-style-type: none"> <li>- project progress</li> </ul>	<ul style="list-style-type: none"> <li>- problem that collaboration does not sit well with competition</li> </ul>
Persuasion work: credibility > being an intellectual	<ul style="list-style-type: none"> <li>- general necessity to be open and have a flexible mind</li> </ul>	<ul style="list-style-type: none"> <li>- difficult to evaluate</li> </ul>
Persuasion work: credibility > societal outreach	<ul style="list-style-type: none"> <li>- improves practices in the real world</li> <li>- takes time</li> <li>- is fun to do</li> </ul>	<ul style="list-style-type: none"> <li>- list everything</li> <li>- indications of exposure outside science</li> </ul>
Persuasion work: reputation	<ul style="list-style-type: none"> <li>- past success is an indicator of future success<sup>46</sup></li> </ul>	<ul style="list-style-type: none"> <li>- ex-post evaluation of granted projects</li> </ul>
Persuasion work: position	<ul style="list-style-type: none"> <li>- draws attention and proposals to collaborate</li> <li>- basis of having influence on science, education and society</li> </ul>	<ul style="list-style-type: none"> <li>- difficult to evaluate</li> <li>- mentioning in databases (better not use this criterion)</li> </ul>

<sup>45</sup> Committees whose tasks do not allow non-members to have insight. For example committees on research ethics.

<sup>46</sup> Do notice that Must et al. (2012) in another sub-project of ACUMEN in fact warn for the Matthew effect, which in part may be caused by this idea.

Invisible work category	Why should it be evaluated?	How to evaluate it?
Creativity	- innovation in science	- difficult to evaluate - absence can be spotted

Table 23: *Why evaluate invisible work and how - overview of answers*

## How to evaluate invisible work?

Table 23 summarizes the ways that the invisible work could be evaluated according to the interviewees. The proposed methods were quite straight forward: time registration, registration of activities, surveys among colleagues and students.

In some cases interviewees admitted that a particular activity is difficult to evaluate. The more abstract the activity is, the more often interviewees admitted this.

The fact that an activity takes time was most often mentioned as reason why the activity should be evaluated. Accordingly, time registration or simply registration of the activity was often mentioned.

In some cases the problem was not that the activity could not be evaluated, but that the interviewee received no feedback, no acknowledgement and/or no compensation in any way. It means that evaluations themselves not necessarily make an activity visible. Evaluation is a necessary condition but not sufficient.

Besides the registration of activities and some forms of acknowledgement (teacher of the year), these answers are difficult to translate to the ACUMEN Portfolio. Or put differently, innovative ways to evaluate invisible work is likely to find a grateful audience.





# 5 Informal evaluations

## 5.1 Introduction: evaluations that remain invisible

Within ACUMEN's context, the flip-side of asking which work remains invisible in evaluations, is, asking which evaluations remain invisible in academic work. Double-blind peer review of papers, grant applications and job applications are highly visible evaluations in academic life. The PhD thesis defense, habilitation, and applications for promotion are more visible or less depending on the country. Annual appraisal talks are usually considered a nuisance. These and other forms of evaluation become visible when they are formal or formalized, which differs from country to country or location to location. In addition their visibility is also constructed through the amount of prestige that is attributed to them.

Out of curiosity, we kept an eye out for informal evaluations: evaluations that have no official status, but happen in daily practice. A researcher may listen to a colleagues' presentation, be positively impressed and perhaps later ask him/her to participate in a project application. Heads of departments may be assessing their group members to see if they could be eligible for promotion or have the necessary qualities to apply for a career grant. Which forms of informal evaluation exist in academic practice and how do they have an impact? These are relevant questions for ACUMEN because they may inspire new uses of the portfolio tool. Moreover, formal and informal forms may turn out to be mixed, which may have consequences for good evaluation practices.

In the first couple of interviews we actively looked for informal evaluations, and when pointed out, interviewees agreed that they were informal and worth pointing out. However, we soon also noted that interviewees also pointed them out on their own initiative. In the end we encountered many more than expected: about one fifth of the reported top threes of most influential evaluations are informal evaluations. Although almost all interviewees reported positively about informal evaluations, they do have a dark side.

The next Section will briefly describe how we went about identifying informal evaluations, Section 5.3 will list them and the subsequent sections will go into the details of the different categories that we distinguished.

## 5.2 Method details

To identify informal evaluations, we used two main means: the event codes and event-code groups and the detailed coding of the interview transcriptions (see the subsection Coding, p. 21). In addition we did full-text searches to find references to particular details.

During the coding of the events that interviewees reported, we paid attention to the events' formal/informal character and registered in the event code when an evaluation had an informal character or when the status was unclear or unknown. For example "Feedback - informal - from colleagues - in general" for a clearly informal form of feedback, or "Feedback - formal/informal unknown - from 'client' organizations" where it was unclear.

In principle, event-codes were grouped per evaluation category. When an informal evaluation event did not have a clear event-code group, it was grouped in 'Other informal evaluations'. When in the course of the coding, it turned out that an evaluation category could be determined, a new event-code group plus event codes were made and these events moved into that group. This is how the event-code groups of "Feedback" and "Evaluations leading to an invitation to something" came into existence.

During the open coding process that was used to code the transcriptions resulted in a list of about 280 codes. From this list we selected codes that we knew indicated informal evaluations, or expected that they might cover informal evaluations. Based on this list we made a selection of interview fragments which we then browsed.

### 5.3 Identifying informal evaluations

Table 10 (p. 29) shows the counts per event-code group of the answers to the question what are the three most influential evaluations. It includes the group of 'Informal evaluations' which ranks as third most often mentioned, after job applications and grants. Further down the table, we find 'Evaluations leading to an invitation to something' and 'Feedback', which both contain mostly event codes for informal evaluations. Next to these, other event-code groups may contain event codes, referring to an informal form of the group's form of evaluation, for example, informal job applications (See Appendix 5).

In order to give an impression of the role of informal evaluations we took Table 10 and separated out all the events with an informal event-code. Table 24 displays the result.

Event code group	Formal/unknown			Informal			Total
	Gender		Total	Gender		Total	
	F	M		F	M		
Job applications	8	16	24	1	1	2	26
Grant applications	5	14	19				19
Diplomas	3	12	15				15
Job-related evaluations	7	6	13				13
Feedback				5	6	11	11
Other	1	6	7				7
Paper and book reviews	2	4	6		1	1	7
Evaluations leading to an invitation to something	2		2	2	2	4	6
Other informal evaluations				2	3	5	5
Evaluations of teaching	1	1	2				2
Total	29	59	88	10	13	23	111

Table 24: Three most influential evaluations mentioned by 37 AIs, by gender, by formal/informal status

Of the 111 answers, one fifth refers to informal evaluations, which makes the group the second biggest, still smaller than job applications but bigger than grant applications.

Another remarkable trait is that women mentioned informal evaluations relatively more often than men when compared to the total number of men and women interviewees. This is the case for the top three most influential evaluations, but not for the top ten of influential evaluations that we asked for before asking for the top three. In the top, men mention informal evaluations slightly more often<sup>47</sup>.

Although, more answers came from women than from men, those men that gave those answers gave more of those answers per person. So, in Table 24, eight women produced the total of 10 answers referring to informal evaluations, whereas eight men produced 13 answers<sup>48</sup>. Remarkably, there were two men (one in PH and one in AA), whose top three of most influential evaluations consist of informal evaluations exclusively.

<sup>47</sup> In the top ten of influential evaluations, 15 answers referring to informal evaluations came from women, and 35 came from men.

<sup>48</sup> In the top ten of influential evaluations, eight women produced 15 answers and 13 men produced 35 answers.

There seems to be a relation with disciplines: eight answers came from interviewees in AA and 11 in PH, whereas EE and P+ each were the source for two answers. Also, senior academics mentioned informal evaluations far less often (1 time) than early career researchers (9 times) and mid career researchers (13 times). The distribution over the four countries follows the distribution of all interviewees over the countries.

In the 23 answers of DHs, informal evaluations occurred only once, which could indicate a big difference in perspective. During the interviews both DHs and AIs were equally invited or encouraged to think of evaluations in a broad sense, so we do not think the result is an artifact of the interview protocol. DHs clearly are aware that informal evaluations exist because some did mention particular ones during the interviews. Some reported that HoDs would make a kind of pre-assessment of group members' chances to apply for promotion.

Considering the relative size of this group of answers within the AIs, it is worth investigating what informal evaluations are about and what kind of roles they play in an academic's career. Informal evaluations take a number of different shapes. In the remainder of this section, we will discuss the most frequently mentioned kinds: job applications, invitations, career choices and life choices, and feedback.

## **5.4 Informal job applications**

### **Informal job interviews**

The two informal job applications that were listed as influential were both instances of job applications where a lot of evaluation work was done before the actual application. In one case, the interviewee already worked at a research group when a vacancy occurred and his professor suggested him to apply. He did so and he got the job. In the other case, the interviewee was looking for her first position after her PhD. She applied for a job and found out that she had met and talked to one of the two people deciding who to hire some time before when this person was visiting her institute.

" Actually, the two people deciding in that occasion were [A] and [B]. ... And I had met [A] before, during my PhD. Well, he was working on something that was well, related to something that I was doing during the PhD, the type of research I was doing during the PhD. And eh, yeah, he seemed to find interest in what I was doing, so I think he had already a good impression of me and I think he let the choice to [A]." (Interviewee 12, 00:50:55.19)

The contact with [B] was not the actual application interview. There was one with [A] as part of her application.

The first interviewee explained why he was not surprised that he got the job.

" ... , he obviously felt that I was, you know, the right sort of person for that role. not just probably academically, but probably the right sort of person who would, who fitted in okay. because when you're applying, when, I I know myself when I've appointed researchers. it's not, when you're advertising a role and you a-, and you're appointing somebody, you need to make sure that they will fit in.

... I think you see the, you'll see there's, if you talk to other people and ask similar questions, many people you'll probably find, ... , chances are their prof or their supervisor will, if they are the right person, in the right place, because funds aren't always available. if there's somebody available who can fulfill a role, ehm, and you're comfortable with them in the group, you'll try and keep them. because it saves an awful lot of ... problems. if you're then recruiting a new person and they, for whatever reason, they may interview fine, but they don't always fit in." (Interviewee 14, 00:44:36.72)

## Fitting in

And indeed, other interviewees had similar experiences, in particular 'fitting in a group' as colleague or collaborator was mentioned more often.

When asked whether a job application was mostly a disciplinary evaluation (testing for disciplinary knowledge) or functional (functioning as an employee), two interviewees (DE AA, DE EE) mentioned fitting in. As one of them phrased it

" Actually it was the start-up, ... , I was the first PhD or assistant for a new professor. And there was not a team I had to fit in ... actually he chooses whether he thought we could work together. So it was more a function, or it was more a yeah, it was the minimum condition was that that the topic fits and then the decision was made on whether I as a person fit into this team. " (Interviewee 9, 00:55:14.81)

To make sure, none of the interviewees made the point that it was only about fitting in a group, as also this dean pointed out about job applications in his faculty when asked whether the evaluations were mostly disciplinary or functional.

" It's both. It is both. Because in some case one could imagine an individual operating as an individual, but in most cases you have to fit in with the department and get on with your colleagues and motivate students and so on. So you need both. I mean in very rare cases, again mathematics is the obvious answer, where there are people just literally work on their own. And it's probably also true within the arts. But eh mostly, you have to interact with your colleagues, and so that's one of the criteria." (Interviewee 23, 01:03:48.75) (AA, UK)

Finally, one interviewee (EE UK) pointed out a problem with the 'fitting in' criterion

" A job application, always from the institutional side, happens within a broader context, which of course as an applicant you don't know, or, and so. Even, you might choose not for one with the largest number of job applications because it also depends on how the applicants would fit in the larger. And you could see that as positively to make sure that he or she fits in, or negative way because there is a lot of politics behind. That that people want someone their own, more compatible with their own profile." (Interviewee 6, 01:11:07.58) (UK)

When an applicant fits into a group, he/she will be pleased if that was a criterion that helped him/her secure the position. However, a potential downside is that 'fitting in' can be a way of smuggling more subjective feelings and organizational 'politics' into the recruitment process.

## 5.5 Invitations to ...

Researchers occasionally or regularly are personally invited to participate in a workshop, write a paper for a special issues, deliver a key-note speech and the like. The invitee was somehow selected and one can assume that some form of evaluation or an entire string of evaluations had taken place before the invitation arrived. Our interviewees also reported a range of such invitations based on informal evaluations. Three forms stuck out: invitations to start on a PhD, to participate in a project and to apply for a job.

### Invitations to start on a PhD

One AIs in Germany and one in the Netherlands mentioned that their MA thesis supervisor invited them to start on a project or apply for a job leading to their PhD<sup>49</sup> and marked this as the most influential evaluation. Two more interviewees in Germany mentioned the event too, but for them it was the fourth most influential evaluation. In all four cases and two cases from two additional interviewees, this evaluation was part of one of the three most important developments, typically (the start of) doing the PhD. In short, although this type of informal evaluation seems rather influential for four interviewees and very influential for three.

The interviewees are equally divided over the three cohorts, but five out of six were interviewed in Germany. In half of the cases, the interviewee was suggested to apply and did so successfully, in the other half of the supervisor offered a job or applied for external money to be able to hire the interviewee.

### Invitations to lead a project

Two interviewees in the UK, both in PH mentioned an informal evaluation leading to an invitation to participate in a project or become project leader. They ranked these evaluations as one of the three most influential evaluations in their careers. One of them mentioned another such case and one more interviewee, also in the UK, mentioned three such cases, but none of these ended up in the top three.

In all these cases, a colleague whom had had earlier work experience with the interviewee was looking for someone to participate, lead or take over the lead in a particular project. As one interviewee put it:

"I mean, what happened was [the colleague] knew me already. He had seen me working. He was positively impressed with that and I was one of the very few people that was available in this kind of field. Yeah, so. ... there was not even an interview or so. [laughs]" (Interviewee 13, 00:38:00.50)

And another

"Yeah I was asked, I was asked by the head of department if I was willing to take on a piece of work. And it was a fairly big piece of work. That I think, she'd been asked to do, but could [not] manage herself. So, I managed as a fairly junior research fellow to take on write a eh significant project, if not huge, that was originally offered to a professor." (Interviewee 11, 00:57:15.21)

### Invitations to apply

Although these evaluations were not mentioned as influential evaluations themselves, at least 12 interviewees (8 AIs and four DHs - not counting the interviewees mentioned above whose MA supervisor

---

<sup>49</sup> In Germany, the Netherlands and Poland this means a job offer as a paid student or assistant, in other countries it means an invitation to continue as a student.

suggested them to apply) at some point mention that 'someone suggested to apply' for a job or a grant in at least 18 different occasions. In most cases the interviewees mentioned it almost en-passant.

"[my supervisor] told me and said: there's a position. You should give eh an, how do you call it? ... I should apply. Yes. And I did. And then [laughs] and eh. It was without any complications, so."  
(Interviewee 24, 00:34:31.22)

Often, neither the interviewee nor the interviewer realized that such suggestions or invitations are the result of an evaluation which is often of an informal nature.

In a number of cases the one who suggests is a supervisor or HoD, which brings us to the accounts of the four deans and human resources managers, three from the UK and one from the Netherlands. As discussed in Section 3.3, promotion of temporary to permanent staff and of permanent staff to a higher level on the career ladder, are important to faculties and universities. They are regulated through elaborate formal promotion and tenure track procedures. Part of these is that the staff needs to apply and HoDs play an important role in checking that people who seem eligible for a 'step up' actually do apply. Thus, HoDs do a preliminary and informal evaluation of their staff members' qualities, comparing them to the promotion criteria.

When it comes to recruitment, existing staff plays a role through identifying likely individuals for a specific function. As one HR manager explains, sometimes it is hard to find people for senior functions.

" But sometimes, those in the know will say, don't advertise yet because the environment isn't such you know. And then they say, I think they do a bit smoothing, and a bit of, you know, oh we've got this job coming up, will you apply? I think there's a lot of that that goes on and then we advertise and people apply." (Interviewee 25, 00:38:50.34)

It is not her favorite way of recruiting:

" Well, if you talk about being, you know, equal and diverse, sometimes, I think that sort of activities isn't great, but it goes on. It does happen" (Interviewee 25, 00:39:50.15)

In fact, sometimes principal investigators are very particular about who they want to hire, as she later in the interview pointed out.

" [W]e do have occasions where people have gone to conferences, met somebody, and thinks, oh, they'd be good. And then come back and say, I want to recruit that person. And ehm. So obviously from human resources, the answer would be, well, you know, you have to advertise it. But I don't want to advertise it, I want that person. So then [unclear] Then [unclear] there might be better people, as you rightly say. So there's always a battle. But I think that goes back to the fact that principal investigators are very focussed on their research. They want their job done. More than properly. They're very protective, though driven." (Interviewee 25, 00:42:44.17)

Summarizing, informal evaluations find their way in recruitment and promotion even though these are highly controlled and formalized in terms of process and criteria. In the cases described the informal evaluation of potential applicants takes place before the applicants have sent in their letters or even know about an application opportunity.

## 5.6 Career and life choices

In Table 24, the group of 'Other informal evaluations' contains four evaluations, mentioned by three interviewees that refer to life choices.

One interviewee (DE AA), who argued that whatever happened to him now as an academic was a result of accumulation of successive developments and evaluations starting a long time ago, went back to his childhood and reported his parents' choice to send him to high school which allowed him to start on his studies later on. The evaluation here, was their assessment of their son's capabilities.

Another interviewee (PL EE) mentioned his choice to become an engineer.

"Wie ich mich erinnere, ich wollte damals, ich wollte damals [anonymized] studieren. Ja. Das ist natürlich sehr schön. Aber ehrlich gesagt, das ist nicht so gute fach. Also, ich habe damals entschieden. Ich will ein Ingenieur sein. Und das war, das war, alle der Faktor Ich will der Ingenieur Ich will Richtung technische Sachen gehen. Das war sehr wichtig." (Interviewee 26, 01:20:36.46)

"As I remember, back then I wanted, back then I want to study [anonymized]. Yes. That is of course very nice. But, honestly speaking, it is not such a good profession. So I decided at that time. I want to be an engineer. And that was, that were all the factors. I want to go in the direction of technical things. That was very important." (Our translation)

This choice was very much influenced by the fact that his father, his uncle and his cousins were all engineers, so one may wonder to what extent it was an independent evaluation of options, but the interviewee felt it was an important evaluation anyway.

The third interviewee<sup>50</sup> had experienced a near-death situation a number of years before the interview, which lead him to realize how important life choices are. So, as one of the most influential evaluations he listed his evaluation of life choices in general. It meant that he decided that a career in science with a permanent position, working 60 hours or more per week and traveling all the time was not what he wanted. Also, he noticed how some of his colleagues got a burn-out, and became more and more unhappy.

In this category of career and life choices, one would also expect those evaluations that lead people to choose to enter an academic career, step out or back in. Table 6 (p. 25) lists eight answers coded as 'Choice for academia' as one of three most important developments. Except in a few cases, neither the interviewee nor the interviewer realized that an informal evaluation of career options and/or life options often are behind such a choice.

## 5.7 Oral feedback

8 Interviewees reported feedback 11 times as a top-three most influential evaluation, making it the biggest group of informal evaluations. One interviewee mentioned three feedback events, one mentioned two, the rest mentioned exactly one. They are five women and three men. Three interviewees were interviewed in the Netherlands, two in Germany, two in Poland and one in the UK. Four Interviewees work in PH (2 in Germany and two in the Netherlands), two in AA and two in P+ (both in Poland). Three interviewees are mid career researchers and five are mid career researchers, so this was clearly not an important form of evaluation for senior researchers.

8 Interviewees mentioned feedback 14 times as an influential evaluation, but did not rank these in their respective top threes. Among these eight is one interviewee who listed five feedback events. This was the same as the one mentioned earlier who had three such events in his top-three of most influential evaluations. Clearly, he, more than other interviewees felt feedback was the most important form of evaluation.

The overlap between the two groups is four interviewees. The distribution of the second group over the four countries is similar as is the distribution over the disciplines, but otherwise the composition is markedly different: here two interviewees are women and six are men, and this group includes three senior scientists, one mid career and four early career scientists.

When we focus on the group that reported feedback in their top-three, two interviewees' answers referred to feedback in general from colleagues.

"eh, ehm, what was really eh helpful for me when I started here at eh, I eh back, I always eh, got feedback for what I was eh doing. it was [unclear] are very challenging, all the time [unclear] after I finished university and starting here. is very different [laughingly] from studying here. so ehm, eh, it was very helpful to eh, to have the possibility to like reflect all things I'm thinking about, I'm writing, I'm doing, eh, in our group here, with our professor. that was eh, I think fo-, for the part of

<sup>50</sup> The interviewee was very cautious about keeping his anonymity, so we do not provide any details about him.



education as a researcher, that was eh, ja, one of the most important parts for eh for, for the beginning." (Interviewee 27, 00:12:57.61)

This interviewee (DE PH) found feedback from her colleagues thus important that she also listed it as a top-three most-important development in her career. This first project was a test-case for her to see if research was a viable career option for her.

Whereas it was important support for this interviewee, another mentioned feedback from collaborating colleagues: feedback in general and in relation to recognition, not just positive but perhaps also negative:

"It's more like, not evaluation, but recognition that people recognize you as as, and tell you: oh wonderful what you are doing. Or how terrible is it." (Interviewee 2, 00:59:27.60)

At the same time, feedback improves academic work

"I think that recognition for me is always about when you collaborate and you are able to immediately respond to each other and in that response there is already an evaluation of what is going on. And that is ehm an evaluation that improves what you do. Ja" (Interviewee 2, 01:07:02.30)

Besides these two interviewees who referred to feedback from colleagues in general, other interviewees mentioned more specific feedback: feedback during particular events, from particular persons or on particular points.

The interviewee who mentioned informal feedback three times in his top three of most important evaluations, seemed to refer in all three cases to feedback revolving around his role and position in the organizations where he was working. The interviewee, looking to learn new skills and develop different sides of himself, had moved out of academia after his PhD, worked at two companies and then returned to university.

When he moved to industry, he landed at a small company that had never taken in someone like the interviewee with good analytical skills but quite different from the consultants that the company worked with. This meant that he had many discussions and talks with the company's director about his role. Similarly, in the second company where he worked, he was hired because of his network in science and the fact that he could pass as a scientist which made the company's contacts with scientific institutes easier. This role however also required discussions with his director. Although he stopped working for both companies, he still has regular contact with the two directors.

When the interviewee moved back to university, he had discussions with a professor and the research director about how he could contribute

"I had several discussions with ehm, with either the professor at the university or the director of the institute. about ehm, ... about, you know, my role at the university. I think that the, let's say the discussions I had with the director of the institute, ... , we talked about you know, just brainstorming about eh if I would come to work for them, what would I do? and how would I fit in within all the projects? and w-, eh exactly what were they looking for and what was I looking for?" (Interviewee 5, 00:40:47.90)

Coming from industry, he brought in valuable management skills and his network, which he would use to assist the professor. In addition he would be supervising MA students and PhD students, but he would not be publishing a lot. However, a good publication list was a criterion to provide him with a permanent position as assistant or associate professor, which is what both he and the professor wanted.

Another interviewee (PL P+) mentioned feedback that she receives during conferences. It also revolved around the direction of her work, but instead of functional as with the previous interviewee, the feedback is about intellectual development.

"it will be okay, because there are sometimes a lot of different experience in that area. but at every time it is an evaluation, and it [is] very important to me because this gives me a perspective to see where I'm exactly going in my scientific life. And I, I take it very seriously. If you need me to evaluate of importance of this evaluations, evaluations of the conference will be ... very high [in her ranking of evaluations]" (Interviewee 3, 01:28:28.02)

To some interviewees it is very important who provides the feedback. This person needs to be a trusted or highly regarded person.

"yeah, and eh the feedback, eh he was very important for me because I admired him and his work and I think he's eh very very clear and ehm analytical ehm analytically thinking, and ehm this feedback I got from him is amongst the turning points." (Interviewee 7, 01:00:54.31)

An interviewee in the Netherlands has a kind of mutually-coaching relationship with a colleague whom she chose carefully: "oh, eh maybe ehm, because you, you pick your own eh person, it's, it's a, you already know that she is a little bit like you in the same way you are doing your work." (Interviewee 19, 01:07:38.88) She is aware of the downside of choosing someone similar to herself, and also realizes that she cannot choose someone who is too close to her field.

Interviewee "yeah. we are really individuals here. when you are at the university, [unclear], we have our own group and we are doing our own projects. and it's also a lot of ... competition. and that's, that's not really good for ehm, to create such a ehm, environment where you can" (Interviewee 19, 01:08:44.20)

FvdM "ask anybody"

Interviewee "no. because it's your, it's your colleague, but also your eh"

FvdM "yeah, your competition"

Interviewee "yeah ... and the higher the position, the the more difficult it eh, unfortunately it is. ... yeah, she's doing really another kind of research. so she's not in my field." (Interviewee 19, 01:09:34.12)

Two interviewees mentioned particular feedback on output. One mentioned a point a colleague from a different sub-discipline made on many of his draft texts: "well, em, I appreciate your creative ideas very much and eh I understood. It's new and it's it's completely correct, but the way you eh sell it is [overly complex]" (Interviewee 7, 01:04:52.63) The colleague used a word-pun with the interviewee's name, which can not be quoted but which effectively made him remember the critique.

The other of the two interviewees reported praise on a presentation as one of the three most important evaluations in her career. Shortly after she was hired on an internal collaborative project she presented her ideas about her work to an important partner-group. Their positive feedback was an important reassurance, both to her and her supervisor.

Summarizing, feedback constitutes an important category of informal evaluations to early career and mid career scientists, but not to senior scientists. It may concern feedback on academic production, such as conference papers and presentations, which the academic uses to improve the quality of his/her work. However, in many reported cases, feedback was about other things as well: one's role in an organization, long-term strategy, confirmation on an approach and recognition of one's work. More often than not, it was important who exactly provides the feedback: not just a random colleague, but a specifically chosen and trusted or well regarded person.

## 5.8 Conclusion: Portfolio use and the dark side of informal evaluations

Although interviewees reported quite a number of informal evaluations, a few forms remained almost unobserved, i.e. invisible. These are probation evaluation, reference letters, and evaluation of reviewers (by editors) and editors (by reviewers), each of which have been mentioned by exactly one interviewee. Perhaps other interviewees did consider these evaluations but deemed them not important enough to mention them.

Probation evaluation is a formal evaluation, but may have remained invisible because those who did not pass may have stepped out or forced to step out of academia altogether, which means that we have not reached them as interview candidates.

Applicants know that reference letters exist, but not many have seen what their former supervisors, bosses and colleagues wrote. Nor have they experienced how their evaluators treated these letters. In fact reference letters in job applications may undergo a double evaluation, both of which remain invisible. One evaluation assesses the value of the writer, i.e. the reference, and the other assess the application/applicant in light of the letter. Again, most applicants only know in case of successful job-

applications that the letter was no hindrance to their success. Interviewees hardly mentioned unsuccessful applications, whereas these could be very important for the course of their career. Thus, reference letters remain not only literally invisible to the applicant, they may be invisible because the applicants can not see what the effect of the letter on the application was.

Finally, reviewers and editors are subject to evaluation too, and their work strongly influences publication failure and success, something that careers nowadays are highly dependent of. However, we interviewed mostly people who have managed to stay in academia, at least until we interviewed them.

So, these evaluations have in common, that our interviewees have passed them successfully, which probably made their effect on the interviewees' careers fairly invisible. Perhaps they were taken for granted and thus deemed less important than others.

Having outlined this limitation, we will now turn to the evaluations that were reported. In the diversity of informal evaluations as presented above, a number of family characteristics can be identified.

- Informal evaluations often consider an individual or his/her work 'from nearby': the evaluated and the evaluator know each other well. The evaluation is far remote from double blind peer review.
- The evaluation is based on an integral understanding of the evaluated as scientist and/or as a close colleague or friend.
- The relation can be reciprocal: close colleagues give feedback on each other's work; one may tip the other for an application opportunity at some time and vice versa at another.
- Follow-up action is left entirely to the evaluated.
- Informal evaluations occur on a daily or on-going basis.

An important question for the ACUMEN project is what informal evaluations and their characteristics mean for the Portfolio design and use, and how can they be related to good evaluation practice.

### Portfolio use

One potential use of the portfolio lies in the feedback that scientists receive about their long-term plans and career development. Anybody advising on such things could use the portfolio, be it a friend, close colleague, HoD or the portfolio's owner him/her self.

The Portfolio provides overview to identify strengths, weaknesses, salient qualities and the development of these. This could lead to strategic advice on which fields of research to go into, what kind of jobs to pursue, which journals to target for new publications, and so forth.

Portfolios can have multiple uses for informal evaluations. They can be used to identify candidates for a certain job, project proposals, special issues and the like. Often, this is what people normally do 'in their heads', but a portfolio may help. Sometimes one meets someone at a conference and wants to know more about this person. Another use lies in the phase after the identification of a candidate. The Portfolio can be used to forward information about the candidate to committees, individuals or groups that take final decisions.

An issue to address is how the Portfolio's owner provides access rights to others. Technically, this problem has been partly solved by websites and social media such as LinkedIn, Facebook and Google: the Portfolio's owner gives access rights to certain individuals, groups and/or everybody. Besides the rights to see the Portfolio and parts thereof, also forwarding rights should be addressed. It will be difficult to implement these because as soon someone has access, data can be printed or converted into a PDF file and forwarded.

On the other hand, the forwarding problem is not that big. Firstly, forwarding also happens with CVs and this seems to work fairly well. Secondly, scientists' careers are public to a large extent anyhow, at least those working in public universities and research institutes, as targeted by ACUMEN. And thirdly, in a number of situations, but not all, the informal evaluator already has a good knowledge of the evaluated.

## The dark side of informal evaluations

Some informal evaluations play a role in formal evaluations, or in processes that involve formal evaluations. For example, a principal investigator evaluates a colleague to be an interesting candidate for a job and then suggests the colleague to apply. Evaluating if someone 'fits in the group' as part of a job application, is another example. The HoD checks whether group members could apply for promotion and suggests those who pass to do so. The professor who wants to recruit an outstanding PhD student for a post-doc position. She wants to hire him but her HR department forces her to open a position, formulate criteria, advertise publicly and consider all applicants.

The interviewees reported evaluations in which they were the ones singled out positively: they got the job or the promotion. Perhaps, they rightfully did so. Perhaps the informal evaluations they were subject to, were as good as any and perhaps the interviewees were indeed the best candidates available. However, we have to be aware that behind them are those who applied too but did not get the job or promotion: those who wonder why they did not get it whereas on paper they appear equally good. What is informal evaluation to one, may be 'bias', 'departmental politics', 'animosity', 'time wasting' or 'window dressing' to another.

A call for pre-determined, measurable, scientific criteria for success and rules of conduct may help but will not solve the problem. The professor who wants to recruit a particular person can come up with a job-description and criteria that will single out her preferred candidate. Secondly, philosopher Martha Nussbaum points out that it is impossible to design a rule-based system to guide decisions and do the right thing in all situations (Nussbaum, 1986). Thirdly, such criteria and rules have not been developed, even after decades. Fourthly, a problem is that what actually counts as a scientifically relevant criterion and what not, is a matter of dispute (Gieryn, 1983).

Considering that we can not develop the ultimate set of criteria and rules of conduct for effective and fair evaluations in science, and considering that informal evaluations will mix in with formal ones, then what are the consequences of introducing a Portfolio and GEPs?

First, let us consider decision processes (recruiting people, promoting staff, and project funding of research) that involve formal evaluations. They can be conceptualized as a three-staged process. In the first stage, actors establish the official criteria for success. These criteria can be of variable quality, i.e. measurable or not, vague or concrete, 'scientific' or not. They are then published with the call for applications. In the second stage, applications are collected and then evaluated against the criteria. A number of applicants may not pass those criteria that are measurable and will be rejected. These rejections can easily be justified. During the third stage, actors need to decide about the remaining applications when they outnumber the granting possibilities. It is very likely that this will happen<sup>51</sup>. During this stage, additional criteria may be formulated and vague criteria may be further specified. In a way it is a continuation of the first stage, but possibly with a different set of actors and thus a different set of views and arguments.

In all three stages, informal evaluations, departmental or organizational politics and bias may occur, but once measurable criteria are published, actors can refer to them to build their case in favor or against applications or decisions. During the third stage, actors will be limited by those criteria that were measurable and were used to reject applications during the second stage.

Via the Portfolio, ACUMEN argues to take a broader picture of the evaluated into account. This has three possible effects. First, introduction and use of the Portfolio increases attention to more, measurable qualities of the evaluated individuals. The more successful that is, the better criteria can be formulated during the first stage of decision processes. Through an iterative processes, the Portfolio stimulates articulation of measurable criteria.

Secondly, Portfolios may also support the design of a set of criteria to single out preferred individuals, during the first stage. This happens for example in the case of the hypothetical professor who wants to hire a promising graduate for a postdoctoral position. Clearly, this would be a non-desirable use of Portfolios. If however, this professor has no access to Portfolios of potential competitors, the attempt may not be that successful.

Thirdly, the portfolio steers away from criteria-based decision making, and moves towards case-by-case comparison because individuals are invited to present themselves as broad as possible. This means

---

<sup>51</sup> Hiring or funding parties cannot fully specify measurable criteria (and assigning relative weights) that will single out the exact amount of 'winners' that they seek. When the count of those passing the criteria is smaller than the demand, then none of the applicants will feel mis-treated, but the hiring party will need to start searching again with less specific criteria in order to 'throw a bigger net'.

that applicants, through their Portfolios, implicitly invite evaluators to consider a broader range of criteria or a broader range of 'evidence' for certain criteria. Such invitations could find a willing audience during the third stage of the process, when the evaluators have used-up the initial set of published criteria and still need to reject a number of the remaining applications.

It still begs the question how to compare things that cannot be compared, but the Portfolio allows the evaluated to propose their criteria for success.

## 6 Conclusions

In this research we set out to understand the impact of evaluations on the daily work and careers of individual academics. Not only did we want to know how big the impact is, but also of what nature: which types of evaluations play important roles and how do they work in practice? In particular, we addressed two sub-topics. The first was invisible work, the work that academics need to do but that is not addressed in any evaluations and thus may remain invisible to evaluators. The second sub-topic concerns informal evaluations. Informal evaluations do not stand out between the many formal and often-discussed forms of evaluations. They are invisible as evaluations, but as we found out, quite influential.

### 6.1 Impact and preparing for contingency

Because it is virtually impossible to isolate, the effect of an evaluation on actual work or a change in someone's career from the effect of other events that also have an impact on that work or change, we devised an indicator. In short, it is based on a count of the relation(s) between the three most important developments in someone's work and career on the one side, and the three most influential evaluations on the other. Interviewees were asked to identify these six events, why they were important and what happened in practice (See Section 2.1).

According to this indicator, most of the interviewed researchers do not perceive the impact of evaluations in general to be very high, but still fairly influential (Section 3.2). We asked for personal experiences and views, so we received subjective stories. Still, the interviewees may be right, also from a hypothetical external/objective point of view, simply because of the high contingency in careers and lives in general.

In addition, the indicator works with hindsight, which means that evaluations and developments that may have seemed important when they happened, were not deemed that important in retrospect. This means that the impact of a particular evaluation on a career remains to be seen. Even in cases of successful 'high stakes' evaluations (such as major grants, permanent positions, professorships), only hindsight will tell how influential the evaluation was on an individual's career.

On the evaluators' side, a similar argument can be made, based on the interviews, in particular for the evaluations of applications for permanent positions (job applications, promotion application and tenure track applications) and the applications for professorships (either as new job or as promotion). After the decision has been made, it remains to be seen whether the successful candidate will become/remain productive or successful. The candidate may change his/her research interests, but the organization and environment may also change. The top-ranked applications may only seem the best within a short time horizon.

For the good evaluation practices this means that a kind of 'contingency margin' could be taken into account. For example, criteria should not be defined too narrowly, or some applications that were rejected in the course of the process may need a second look/opinion. Such procedures may cause friction with other criteria that apply to evaluation processes, such as setting clear and specific evaluation criteria in advance, and treating all applicants equally. Also, in practice, such contingency margins do not reduce the amount of evaluation work.

The fact that the importance of an evaluation for an individual can only be determined with hindsight does not mean that evaluations should not be taken lightly. Those evaluations that were reported as influential were predominantly job applications and grant applications, with a shared third place for diploma evaluations, informal evaluations and job-related evaluations (such as for promotion

applications or annual appraisals) The question then is how can we, based on our findings, improve on these evaluations with the help of the ACUMEN Portfolio and Good Evaluation Practices (GEPs)?<sup>52</sup>

## **6.2 Improving the evaluation process and criteria, suggestions for GEPs and the ACUMEN Portfolio**

We identified a number of ways to improve evaluation processes. They are listed below and address both the ACUMEN Portfolio and GEPs. Lastly, we identified an additional opening for the use of the portfolio in informal parts of formal evaluations, which we will present in the next section.

### **Partial mismatch between employers' and applicants' views**

A partial mismatch exists between the important developments and influential evaluations reported by individual academics and those reported by HR managers. (Section 3.4) Where their answers agree in terms of identification of these developments and evaluations, they disagree about the reasons why they are important. A complicating factor is that a successful application can be important for multiple reasons.

All this may mean that evaluation criteria and procedures, which are set up by employers, may not take the perspective of the evaluated into account. Arguably, they should, because what looks like an indication or evidence of a certain quality to the evaluators, may mean something different to the evaluated. Thus, this indicator or evidence may be a bad predictor of future behavior.

For example, an applicant may make sure to have an impressive publication list because he/she knows it is a criterion, not because he/she has an intrinsic motivation to publish. Once he/she achieves a permanent position, he/she may be hard to motivate to continue publishing at the same level. Another example concerns couples, where one partner may consider a permanent worthwhile because it assures the successful applicant of long-term income and status. Some people give up such a position in order to be able to find a location where also the career of the partner can be possible. Vice versa may also hold. The evaluated may misinterpret the criteria as an indication of expected future behavior. An impressive publication list is required, but half of the job may involve teaching.

### **Cohorts and country specific backgrounds need to be taken into account: an argument for a life-cycle perspective**

Individuals develop their careers in different periods. They orient on criteria that rule their work and career progress at one point, but by the time they live up to these criteria, these may have been replaced by different ones. They may end up in a Catch-22 situation. Similarly, and within a Europe of open borders increasingly more important, individuals who move between countries, may face problems because their move exposes them to different evaluation and career regimes than the ones they grew accustomed to. (Section 3.4)

One cannot change one's life's work with every change in criteria. Evaluations should take this into account. One way to do this is simply not to change important criteria too drastically in re-occurring types of evaluations. Another way is to formulate alternative criteria for different cohorts or individuals who have moved countries as part of their training or career or life trajectory.

---

<sup>52</sup> See ACUMEN (2010) for a description.

## Take invisible work into account

Chapter 4 presents a list of activities that interviewees identified as important for their work but not evaluated. This list of invisible work, summarized in Table 22 (p. 56) is used to check the elements of the design of the ACUMEN Portfolio.

ACUMEN points to the role of on-line activities that have grown in the past two decades since the arrival of the World Wide Web. It should be noted that our interviewees did not mention on-line activities such as blogging, tweeting, on-line discussions and YouTube as important for their work or career but not evaluated. The list of unevaluated work does include activities that may involve on-line work, such as teaching and activities related to societal relevance. Although researchers did not mention on-line activities, these activities in the future may become more important.

Another indication for future demand, is that a Dutch HR manager who responded to a different question, explicitly mentioned media and when asked which media, started her answer with social media and continued with newspapers and television. She saw these mostly in terms of visibility and more or less equated that to societal relevance. Her institution keeps track of media-appearances, so, at present university managers may be more keen on alt-metrics than researchers are.

## 6.3 Opportunities for ACUMEN portfolio in informal evaluation

### Informal evaluations

A considerable number of evaluations that played a role in important developments in interviewees' career and a large number of influential evaluations concern informal job applications, and invitations to for example participate in a project, undertake a PhD project after the MA, undertake a postdoc after the PhD, apply for a job, or write a chapter in a book. (Chapter 5)

Such evaluations can be characterized as 'up close', 'in full' and 'in view'. The individual or materials under evaluation (the application or other work that the evaluated has written or done) are closely examined, the evaluator may know the evaluated from earlier collaborations or interactions, and may take more materials into account than the actual application if there is one.

A striking aspect of many informal evaluations is that the evaluated, and in some cases also the evaluators, are not anonymous. Put differently, these evaluations have little to do with the model of double-blind peer-review, which opens the evaluation up for use of a portfolio because portfolios are anything but anonymous.

Obviously, in some situations, a portfolio presentation is not likely to have a place in the evaluation. For example when a supervisor invites his/her PhD student to do a postdoc, the supervisor will already have enough experience with the student that a portfolio presentation may not add much.

Still, in other situations a portfolio presentation could play a role. Possible examples include the following: the aforementioned supervisor may need the Portfolio on record for formal reasons; when a project team is looking for someone for a particular task and team members need to present potential candidates to each other; a book editor meets someone at a conference and wants to consider him/her as an author for a chapter; two researchers meet at a workshop and want to know a little more about one another. At present, one would invite a LinkedIn connection or check a staff page.

Another striking aspect of informal evaluations is that they are, indeed, informal. Although many interviewees talked positively about the informal evaluations they reported, we should keep in mind that the outcomes of these evaluations were positive<sup>53</sup>. However, as argued in the conclusions of Chapter 5, there is a dark side to informal evaluations because bias, departmental politics, animosity and window dressing remain unchecked through formalization. In the same chapter, we also argue that it will be

---

<sup>53</sup> Indeed, a few interviewees did mention evaluations with negative outcomes. See Section 4.3



impossible to completely remove such phenomena from evaluations by pressing for pre-set, measurable, scientific criteria for success and for better rules of conduct.

We do see an additional role for the ACUMEN Portfolio when it comes to informal parts of formal evaluations, which we will discuss in the next section.

### **ACUMEN Portfolio in informal parts of formal evaluations**

If<sup>54</sup> we consider the formal evaluations that interviewees have identified as important to their work and career (i.e. job applications, grant applications, and applications for promotion) , we can describe these with a simple model consisting out of three phases. 1) a stage where the evaluation criteria and the evaluation process are being discussed and designed; 2) a stage after publication of the criteria where applications are collected and compared to the criteria. A number of applications can easily be rejected based on these criteria; 3) a last stage in case the remaining applications outnumber the possibilities to grant. Actors need to further discuss additional criteria and how to proceed. The evaluations concerned here occur on a regular basis, so to there is a cyclical character to the stages.

All three stages provide room for informal evaluations to occur, however, once measurable criteria are established and published, all actors can refer to them, be they evaluators doing their work or applicants crafting their applications and perhaps objecting to outcomes<sup>55</sup>.

The ACUMEN Portfolio can contribute to a reduction of the dark sides of informal evaluations. In stage one they can help and inspire the articulation of important aspects of the evaluation and the identification of measurable criteria. In stage three<sup>56</sup>, the applicants' portfolio's may fuel the discussion when evaluators need to decide about additional criteria. It means they need to further specify what they require from the applicants or applications. In terms of the famous saying: during the second stage the oranges and the bad apples were rejected, leaving a pallet of apples of different kinds, sizes and colors to choose from. Which apples should be rejected and which accepted?

In such situations, the Portfolios can be used to immediately check whether new additional criteria that are being proposed in the discussion make sense considering the remaining candidates. Secondly, because of their broad set-up<sup>57</sup> the portfolios may inspire the evaluation committees. Through their portfolios, applicants propose their criteria for success to evaluators. The decision will still be difficult, but at least the applicants can present themselves in full.

---

<sup>54</sup> A longer version of this argument can be found in Section 5.8.

<sup>55</sup> Or praising the outcomes.

<sup>56</sup> In stage two they can be an aid in the comparison, but this is left out here because it is not an informal part. Otherwise, at present, we see no particular role for the Portfolio to reduce informal evaluating of applications in this stage.

<sup>57</sup> This can only happen if this width is allowed to enter the process. Some institutions provide very strict and restrictive formats for applications.

## 7 References

- ACUMEN. (2010). *Academic Careers Understood through Measurement and Norms. Annex I - 'Description of work'*.
- Aizenman, J., & Kletzer, K. (2008). *The life cycle of scholars and papers in economics - the 'citation death tax'.* Working paper. Cambridge MA: Naitonal Bureau of Economic Research.
- Antonijevic, S., Dormans, S., & Wyatt, S. (2013). Working in virtual knowledge: affective labor in scholarly collaboration. In P. Wouters, A. Beaulieu, A. Scharnhorst & S. Wyatt (Eds.), *Virtual knowledge. Experimenting in the humanities and the social sciences* (pp. 57 - 88). Cambridge (MA), London: MIT Press.
- Baldwin, R. G., & Blackburn, R. T. (1981). The academic career as a developmental process: Implications for higher education. *52*(6), 598 - 614.
- Bazely, P. (2003). Defining 'early career' in research. *Higher Education*, *45*, 257 - 279.
- Chubin, D. E., & Connolly, T. (1982). Research trails and science policies: local and extra-local negotiation of scientific work. In N. Elias, H. Martins & R. Whitley (Eds.), *Scientific establishments and hierarchies* (pp. 293 - 311). Dordrecht, Boston, London: D. Reidel Publishing company.
- Dahler-Larsen, P. (2011). *The evaluation society*: Stanford University Press.
- Deuten, J. J. (2003). *Cosmopolitanising technologies. A study of four emerging technological regimes*. Twente University Press, Enschede.
- Enders, J. (2001). Between state control and academic capitalism: A comparative perspective on academic staff in Europe. In J. Enders (Ed.), *Academic staff in Europe, changing contexts and conditions* (pp. 1-23). Westport, London: Greenwood Press.
- Felt, U. (Ed.). (2009). *Knowing and living in academic research : Convergences and heterogeneity in research cultures in the European context*. Prague: Institute of Sociology of the Academy of Sciences of the Czech Republic.
- Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists. *48*(6), 781-795.
- Glänzel, W., & Wouters, P. (15 July 2013). *The dos and don'ts in individual-level bibliometrics*. Paper presented at the 14th ISSI Conference. Retrieved from <http://www.slideshare.net/paulwouters1/issi2013-wg-pw>
- Gläser, J., & Laudel, G. (2009). Identifying individual research trails.
- Goodwin, T. H., & Sauer, R. D. (1995). Life cycle productivity in academic research: Evidence from cumulative publication. *Southern Economic Journal*, *61*, 728 - 743.
- Havighurst, R. J., McDonald, W. J., Maeulen, L., & Mazel, J. (1979). Male social scientists: Lives after sixty. *19*(1), 55-60.
- Hellsten, I., Lambiotte, R., Scharnhorst, A., & Ausloos, M. (2007). Self-citations, co- authorships and keywords: A new approach to scientists' field mobility? *Scientometrics*, *72*(3), 469-486.
- Hermanowicz, J. C. (2007). Argument and outline for the sociology of scientific (and other) careers. *Social Studies of Science*, *37*(4), 625 - 646.
- Horlings, E., & Gurney, T. (2012). Search strategies along the academic lifecycle. *Scientometrics*.
- Kaulisch, M., & Enders, J. (2005). Careers in overlapping institutional contexts. The case of academe. *Career development international*, *10*(2), 130-144.
- Kostoff, R. N. (1994). Assessing research impact. US Government retrospective and quantitative aporaches. *Science and Public Policy*, *21*(1), 13 - 22.

## Appendix 7 - References

- Latour, B., & Woolgar, S. (1979). *Laboratory Life : The social construction of scientific facts*. Beverly Hills, CA: Sage.
- Laudel, G., & Gläser, J. (2008). From apprentice to colleague: The metamorphosis of early career researchers. *Higher Education*, 55, 387-406.
- Levin, S. G., & Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, 81(1), 114 - 132.
- Luukkonen, T., & Ståhle, B. (1993). *Evaluation of Research Fields*. Stockholm, Copenhagen: Nordic Council of Ministers.
- Must, Ü., Otsus, O., Denton, M., Mustajoki, A., & Tiirik, T. (2012). *Peer review practices: ACUMEN*.
- Nussbaum, M. C. (1986). *The fragility of Goodness: luck and ethics in Greek tragedy and philosophy*. Cambridge: Cambridge University Press.
- Orvis, L. K. (1996). *The influence of critical university events and marker life events on Career Development, Productivity, and Satisfaction: a Study of Faculty in a Liberal Arts Department*.
- Philips, D. C., & Turney, J. (1988). Bibliometrics and UK science policy. *Scientometrics*, 14(3 - 4), 185 - 200.
- Rauber, M., & Ursprung, H. W. (2008). Life cycle and cohort productivity in economic research: The case of Germany. *German Economic Review*, 9(4), 431 - 456.
- Riley, M. W. (1998). A life course approach. In J. Z. Giele & G. H. Elder Jr. (Eds.), *Methods of life course research* (pp. 28 - 51). Thousand Oaks, London, New Delhi: SAGE Publications.
- Shapin, S., & Shaffer, S. (1985). *Leviathan and the air-pump : Hobbes, Boyle, and the experimental life*. Princeton: Princeton University Press.
- Spaapen, J., Dijstelbloem, H., & Wamelink, F. (2007). *Evaluating research in context : A method for comprehensive assessment*. The Hague: Consultative Committee of Sector Councils for Research and Development.
- Ståhle, B. (Ed.). (1987). *Evaluation of research : Nordic experiences. Proceedings of a Nordic workshop in Saltsjöbaden, Sweden, 3-5 September, 1986*. Copenhagen: Nordic Science Policy Council.
- Star, S. L., & Straus, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work*, 8, 9 - 30.
- Strauss, A. (1985). Work and the division of labor. *Sociological Quarterly*, 26(1), 1-19.
- Ter Bogt, H. J., & Scapens, R. W. (2009). *Performance measurement in universities: A comparative study of two A&F groups in the Netherlands and the UK*.
- Thakur, M. (2007). The Impact of Ranking Systems on Higher Education and its Stakeholders. *Journal of Institutional Research*, 13(1), 83 - 96.
- Usher, A., & Savino, M. (2006, January). *A world of difference: A global survey of university league tables*. Virginia Beach, Toronto, Melbourne: The Educational Policy Institute.
- Van der Meulen, B. (1992). *Beoordelingsprocessen in wetenschap : de constructie van kwaliteit door wetenschap, industrie en overheid. (Evaluation Processes in Science: the construction of quality by science, government and industry)*. University of Twente, Enschede.
- Van der Most, F. (2010). Use and non-use of research evaluation. A literature review. (unpublished manuscript).
- Verbree, M., Weijden, I. v. d., & Besselaar, P. v. d. (2011). *Generation and life cycle effects on academic leadership*. Den Haag: Rathenau Instituut.
- Whitley, R. (2008). Changing governance of the public sciences: The consequences of establishing research evaluation systems for knowledge production in different countries and scientific fields. In R. Whitley & J. Gläser (Eds.), *The changing governance of the sciences : The advent of research evaluation systems* (Vol. 26, pp. 3 - 27). Dordrecht: Springer.

## 8 Abbreviations

Abbreviation	Explanation
AA	Astronomy and Astrophysics
ACUMEN	Academic Careers Understood through MEasurements and Norms
AI	Academic Individual. The group of people who were interviewed about their personal experiences with evaluations in their careers. (See also DH)
DH	Deans, Human resources manager or department head. The group of people who were interviewed about the organizational perspective on evaluations and careers (See also AI)
EC	Early Career researcher: someone who has worked five to ten years in academia since his/her PhD.
EE	Environmental Engineering
HoD	Head of Department
HR	Human Resources
MC	Mid Career researcher: someone who has worked 11 to 25 years in academia since his/her PhD.
MID	Most Important Development
MIEval	Most Important Evaluation
P+	Philosophy Plus the history and philosophy of science
PH	Public Health
PhD	Doctor of Philosophy. In this report used to indicate all doctor-level degrees.
SC	Senior Career researchers: someone who has worked 26 years or more in academia since his/her PhD.
WP	Work Package



## **Appendix 1: Interview protocol for researcher interviews**



# Interview protocol for researcher interviews ACUMEN Task 1.2 b

## Final

Frank van der Most<sup>1</sup>

e-Humanities Group / DANS, KNAW

6 September 2012

## 0 Introduction

The body of the researcher interview contains 4 parts.

The first part is the introduction and the check of the study and working history. This history will be presented on one or two pages A4 and contains multiple rows: one for a period description, one for study/job, one for affiliation, one for living location, one for task description, and a few others. For more details see the subsection on the researcher interviews in the project document. Possibly, a graphical version of it can be made as well or in stead of a text column version. In any case, space should be available to add columns or important information during the interview, so that the overview starts to 'live' for the interviewee and to convey that things are somewhat open for her/his view her/his career.

The history is prepared before the interview but will not be sent in advance because it may not 'land' well.

Discussing this history has two functions. First is to complete and check the collected data. Second, but equally important, is to hopefully stimulate the interviewee's memories about his/her career so far.

The second part of the interview will focus on the question to the interviewee: **'Which are the three most important developments in your career?'** Subsequently, each of these three changes will be discussed in detail, including any evaluation that may have played a role.

The third part of the interview will focus on the question to the interviewee: **'Which ten evaluations had the biggest effect on change processes?'** Here, each of the top three of these evaluations will then be discussed in detail.

The basic idea is that these two main questions put the role of evaluations into perspective. Evaluations may play modest roles in important career developments and large roles in relatively small changes/continuations. And vice versa. If the occasions discussed in the second part are the same as the

<sup>1</sup> [frank.vandermost@inter.nl.net](mailto:frank.vandermost@inter.nl.net), Telephone: +31 6 1428 8768, [www.researchaffairs.net](http://www.researchaffairs.net)  
The e-Humanities Group, P.O.Box 94264, 1090 GG Amsterdam, The Netherlands



occasions in the third part, then evaluations played major roles in major developments and their 'impact' can be considered high.

The fourth part rounds off with general questions about evaluations and their coverage of important activities in practical research life.

### Interview duration

The design is aiming for interviews of about one hour and a half. In case the answers to the questions of part 2 and 3 overlap, the interview may be much shorter. For such a scenario, additional questions may be available, or one can spend more time on the detailed questions.

Please notice that very long interviews are not necessarily preferable to very short ones. They also take much more time to process.

A rough time schedule for the interview:

10 minutes introduction

20 minutes part 1

30 minutes part 2 (3 x 10 minutes per career development mentioned)

30 minutes part 3 (3 x 10 minutes per evaluation mentioned)

10 minutes part 4 - rounding off with questions

This adds up to 1 hour and 40 minutes.

Perhaps instead of 3 we could ask for two examples in parts 2 and 3. To gain some time, but also because answers could be lengthy. If they are, then I suggest to go with the flow: I would rather have one clear and detailed case than 3 superficially described.

### Language issues

- 'Evaluation'

We mean to talk about evaluation in a wide sense, i.e. not only about peer review of papers and applications, rankings and disciplinary evaluations.

Other words and phrases : assessment - usually refers to activities that can be quantified; examination; annual appraisals or reviews; course evaluations; student evaluations; job applications; self-evaluations; judgements; ratings; rankings; reviews; informal evaluations

- 'Impact'

In the phrasing of the questions, it is important to steer away from 'impact' lingo. This is why the question of the third part speaks of 'roles' in stead of 'impact' or 'influence'.

Other words and phrases that may refer to mutual impact or sound less 'impact': role; 'mutual' impact or influence; how do evaluations play out; evaluation as a/one factor - but not 'impact factor').

- 'Career change'

Secondly, it is important not to stress career changes because it may suggest that we are only interested in changes of jobs, which is why the question of the second part speaks of 'developments'. Arguably, this suggests changes, but I could not come up with a better word or a better question. Evaluations may also play a roles in continuations. The interviewee may think about changes, but that should be his/her interpretation of 'important'.

Other words and phrases: developments; changes; continuations; promotion; demotion; functional promotion; function change; different tasks; reputation boost; reputation blow

### Differences with previous version

Compared to the version of 11 May 012, this version has a slightly different numbering of questions and in the fourth part, the order of questions was adapted to actual interview practice.

## 0 Introduction for the interviewee and the interview 00:00

- Switch on the recording device and do this clearly visible. Then ask if it is o.k. to record. If it is not o.k. clearly switch it off and put it away.
- Explain the background of the interview: part of the ACUMEN project, we are interested in the role of evaluations in academic careers.
- Explain how we will process the results: transcribe and code
- Explain that we will anonymize the data when we publish and report and that quotes will be anonymized too.
- We may circulate data within the ACUMEN consortium but it will be difficult to anonymize (because our partners know already our search strategy and our selection of universities). If the interviewee does not want that then he/she can let us know that via the consent form or otherwise. Also if she/he has second thoughts at a later stage after the interview.
- Refer to the invitation e-mail where all this is explained as well - this is also our backup in case you forget to mention something
- Explain the outline of the interview as we hope it will go.
- Explain that when we ask about evaluations, we mean evaluations in a wide sense of the word. For example also including job application reviews, annual appraisals, research assessment exercises, institute evaluations, course evaluations and informal evaluations.
- Does the interviewee have any questions? She/he will probably ask anyway.
- Hand over the consent form and ask the interviewee to sign it.

Sometimes the interviewee or the situation does not allow you to do all this neatly at the start. Then pick the first possible moment. It would be a waste of momentum to interrupt the flow.

## **1 First part: Check of the study and work life history<sup>2</sup>** **00:10**

### **Q 1.1 [intro first] Please check the career history we prepared. Could you indicate improvements please and fill in important blank spots?**

Explain how we prepared the career history. In the communications before the interview we will ask for a detailed CV, even if a CV is online available. That we looked at it should be no surprise, but we need to remind the interviewee that we did and that we codified it.

The history should indicate through highlighting what important blanks are.

## **2 Second part: Important career developments** **00:20**

### **Q 2 Which are the three most important developments / changes / breaks / alternations / continuations in your career?**

What if: If the interviewee feels, she/he cannot identify three most important developments, because he/she feels too little has happened yet in his/her short career, then point out that there still may be three 'most' important developments in relation to this short career even if these may not be that earth shattering to more experienced researchers.

**For each of the three developments ask the following sub-questions (next page).**

Most likely, some will be answered at the same time as others are answered, but check that all questions are answered.

---

<sup>2</sup> The chapter title mentions 'study and work' not merely career, because depending on one's notion of 'career', study may not be included. I want to make clear that it should be included.

**Q 2.1 What was important about it for you, and why was it among the most important developments?**

Possible answers: it was an important promotion within the organization, it allowed the researcher to pursue a research goal, it made her/him widely recognized, he/she opened an entire research field, it made the researcher rich (patent), the change was or implied a huge social change: met his/her spouse or moved abroad, essential expansion of research skills.

**Q 2.2 What triggered it? What happened, where, how and when? Who were involved? (Let's say, the facts)**

Possible answers: a chance meeting with a person, a brilliant research question, identification of a niche in the research landscape, remarkable finding, development of a new method, increased student numbers, decreased student numbers, sudden media attention, a crisis averted or rightly predicted, university cutbacks.

**Q 2.3 Which other factors played a role? And how would you weigh all factors in mutual comparison?**

The question should be asked as an open question but if the interviewee does not really know how to answer the question then we could suggest categories like family, politics, economic situation and the like.

The weights do not have to be roughly or precisely quantified. It would be nice if the interviewee could, but a sort order, or qualitative indications are equally welcome. The question is trying to get from the individual story to an analysis or protoanalysis by the interviewee.

What if: Here, interviewees may mention evaluations. We should not immediately actively fish for that when posing this question. But once the interviewee has finished answering this, we can ask while summarizing to confirm that evaluations played no role of significance. If the interviewee confirms this, then we have established a clear 'no'. If the interviewee on second thoughts agrees/admits that they did play a role, then we have a 'soft' confirmation, but possibly an answer to please us.

What if: If the interviewee mentions evaluations (of any kind), assessments, appraisals, judgment, gauging, ratings, rankings, comparisons, tests, examinations, or review then we want to know a few things about it. See Q 3.4 . It may take some time.

**What if: What to do if the interviewee does not go along with the triggers and factors questions?**

**Q 2.4 How would you qualify or typify this development?**

This could be things like: personal, institutional, inside science, outside science, political, economical, life circumstances. It could be one of these, but they are not so different from what we ask above. I find it difficult to imagine possible answers.

**Q 2.5 Could you further detail the story of what happened?**

**What if: If needed, fill in the career history for the period/moment of the development.**

What if: If a development refers to a change that is still not filled in in the career history overview, then make sure all the columns are filled for the concerning period. Be aware that the answer of the interviewee may go outside these boxes, which is o.k.

### **3 Third part: Evaluations with high impact on (possibly unimportant) change processes in the career**

**00:50**

**Q 3 Of all evaluations you encountered, which ten had the biggest effect on your career?**

**Q 3.1 Please, rank them from most to least influential**

We ask for the 10 most influential to induce the interviewee to think somewhat better and prevent them from stopping after the first 3 have come to mind and to make sure that we 'catch' those important evaluations that relate to the three most important career developments. It is okay if the interviewee can not mention ten.

A short introduction to this question could be

"So far, we discussed the three most important changes in your career. I would now like to discuss the three most influential evaluations in your career."

What if: If the interviewee claims there have not been any evaluations, then remind him/her that there can be other forms than 'peer review' or things that are officially called evaluations.

As far as I am concerned, groups of evaluations can be considered 'one' as long as the answers to the subquestions are the same for each evaluation in the group. For example, peer reviews of conference papers.

What if: In case the respondent understands the question but finds it hard to decide because she/he finds it difficult to determine the relative weight of an evaluation in comparison to other evaluations or in comparison to other elements in the constellation of events, then try to make the respondent articulate what is difficult about it. The background of this question is that we are partly delegating the problem of measuring impact of evaluations to the researchers/interviewees.

**Q 3.2 Why is it difficult to determine the weight of an evaluation?**

Possible answers: I don't know all the factors - some were out of my sight; I do know all the factors, but they are of such different nature that it is difficult to weigh them (Ask the interviewee to list them)

Question 3.2 may be of some help to the interviewee. A negative evaluation can stop a chain of events quite abruptly, especially when unexpected.

**Q 3.3 Which evaluations prevented you from making a certain change? Does any of these go in the set of three?**

This question is meant as an add on to Q 3 . There may be certain categories of evaluation-event follow ups that people tend to keep hidden. For example failed job applications or requests for promotion. This question may make the interviewee reconsider the set of three and if he/she gives a positive answer to this question than the ranking should be checked again.

What if: If an interviewee gives an answer to this question and you are pressed for time, then make sure to go into this answer in detail at the cost of others because I expect people to down

Interview protocol for researcher interviews. ACUMEN Task 1.2 b

Version: 6 September 2012

**Make a selection of three of these:**

- Does the top include evaluations that have been discussed in Part 2 then check

**For each of the first three evaluations in the list ask the following questions (next page):**

**Q 3.4 What kind of evaluation was it?**

The answer to this question will be categorized. The list of evaluation categories is not yet made, but once it is developed, it may lead to a number of more precise questions or issues for the interviewee. Think of 'Was it a blind/double blind?', 'Who did the evaluation?', 'What was the official purpose?', etceteras. Answering this question may take a lot of time when the interviewee starts describing it.

**Q 3.5 What triggered the evaluation? What happened with/around the evaluation, where, how and when? Who were involved?**

Let's say, the facts of the link between the evaluation and the (possibly small) change in the interviewee's career.

Possible answers:

- "well, I did my studies and wrote my papers and then I had my viva and passed with major revisions. The opponent was a professor from Göttingen whose work I read, but whom I'd never met. Neither had my supervisor, so he took a bit of chance there."
  - "I heard later, it was a close call between me and one other candidate. My supervisor told me that the other had a completely different background, so it was not easy to choose. I did not ask how they decided to ask me, but I guess that sometimes you're lucky and sometimes the other is."
- "I don't know. I just entered the poster in the competition, just like everybody else did. There was a jury of some hot shots from the field and they choose my poster. Of course, I put the prize on my cv, and I guess it did not hurt."
- "What do you mean? Obviously, it was one of the best research proposals. Otherwise, they wouldn't have selected mine for the grant. I knew two of the committee members, but most applicants probably knew a few, after all most of them were the best in their field and well connected."

**Q 3.6 What else played a role? How would you weigh all these 'factors' in mutual comparison?**

See Q 2.3

**Q 3.7 How come this evaluation had such a large effect on the change process of which it was a part? (in other words, why did you list this evaluation?)**

This may have become clear from the previous question, but maybe not. In any case, the interviewee should articulate this:

**Q 3.8 What did the evaluation not take into account, but should have?**

**Q 3.9 How did the evaluation relate to your career or to your daily work / how did it have an impact?**

What if: If the evaluation played a role in the answers of part two (about important changes in the career) then this question can be skipped or only checked. Perhaps the interviewee has something more to add.

Possible answers: it made me consider a completely different career for a while but in the end I decided to stay in academia; it provided some suggestions for my research agenda but I mostly stuck to what I already had planned; it forced me to take on more teaching (could be a bad research evaluation or a good course evaluation); it improved my standing in my field a little.

## 4 Rounding off

01:20

**Q 4** What is important for your work or what do you have to do, that is not addressed in any evaluations but should be?

I.e. what never gets any acknowledgement? For example, writing letters of recommendation for students.

**Q 4.1** Why should this/these be addressed?

**Q 4.2** How should this/these be addressed?

**Q 5** Which activities are evaluated in the wrong kind of evaluation?

**Q 5.1** Where should they go and why?

**Q 6** Which activities are evaluated too much or inappropriately ?

**Q 6.1** Why? How much should be enough or would be appropriate?

**Q 7** Have you ever deposited your research data, or have you ever reviewed a data set? Could you describe the occasions in a few words?

**Q 7.1** In which evaluations is data deposition, creation, maintenance, or review addressed?

Present this as a question that we have a special interest in, with a reference to the work of the DANS institute.

**Q 8** Which metaphor comes to your mind when overviewing your career? Why this one?

This will give an idea of how the interviewee sees his/her career: random, up the career ladder, shooting star, 'it's a job to earn money', serial temporary positions, distracted, etceteras.





## **Appendix 2: Interview protocol for deans and managers**



# **Interview protocol for Deans/manager interviews ACUMEN Task 1.2 b**

## **Final**

Frank van der Most<sup>1</sup>

e-Humanities Group / DANS, KNAW

6 September 2012

## **0 Introduction**

The aims for these interviews are two-fold. One is to collect information about career and evaluation regimes. The other aim is to collect data about the impact of evaluations on careers which we are not likely to receive from individual researchers: they have no active role in a change decision or they may be less willing to talk about certain events. Below, the aims are further elaborated.

For these aims, potential interview candidates exist at different levels in the organization and in two groups. One group consists of academics in leading positions, ranging from group leaders, to department<sup>2</sup> heads, deans, and central university managers. The other group consists of non-academics / administrative staff in human resources departments or sections and in managing positions. They are likely to be found at faculty/department level and higher. The two groups may overlap and whether they do may differ from country to country. In the Netherlands for example, the professional dean has replaced the primus-inter-pares dean, i.e. a professor who takes on the function of dean. In a number of cases - probably most cases - the primus-inter-pares dean has become the professional dean, but in some professional managers were attracted from outside the universities. They may have an academic background but that was not the primary criterion for their selection. The only constants are probably that group leaders are academics and that administrative HR staff is not academic, although they may be academically trained.

Of all these groups interviewing group leaders would be the best option to reach both aims, since they are intimately familiar with local evaluations of sorts and sizes and equally familiar with the careers of group members and former group members. They lack overview at the faculty or university level, but we are not primarily interested in that.

<sup>1</sup> [frank.vandermost@inter.nl.net](mailto:frank.vandermost@inter.nl.net), Telephone: +31 6 1428 8768, [www.researchaffairs.net](http://www.researchaffairs.net)

The e-Humanities Group, P.O.Box 94264, 1090 GG Amsterdam, The Netherlands

<sup>2</sup> The word 'department' sometimes refers to 'faculty', i.e. a unit operating at the primary division level of universities. Sometimes, the word refers to a research group, and sometimes it refers to an intermediate level. Here departments are understood to be at that intermediate level, i.e. aggregates of research groups but smaller than faculties. Depending on the university or faculty the level may or may not exist.

Unfortunately, since we are dedicated to the four disciplines and since three of these are rather small, interviewing group leaders of groups whose researchers we are also interviewing seems not opportune and likely to invalidate all interviews from the start: both group leaders and researchers will not feel at ease. Interviewing leaders of other groups has the drawback that we have to revert to other disciplines, although this may be less of a problem in case of public health.

This drawback exists for all other potential candidates as well, which means that the information regarding the second aim can only be collected at department, faculty or university level. On the one hand it is a real drawback, because we don't arrive at data that closely complements the individuals' stories. It could be partly repaired by continuously checking whether answers hold across all the disciplines or sub-disciplines represented in the organizational unit.

On the other hand, the data gathered gives a complementary perspective, aggregating multiple disciplines in one university, faculty or department, which we could use when working on WP 6. The basic question would be 'What do the evaluation induced/related career changes mean to the organizational unit?'

If group leaders are ruled out as candidates, then the second best group consists of department heads if there are departments, and otherwise deans. Preferably, we want to talk to academics rather than the non-academics/administrative staff because they are better aware of the scientific evaluations and scientific aspects of career changes. Non-academics/administrative staff may be better aware of the details of career regimes. If the department head or dean is a non-academic, then who do we want to interview? The academic person 'nearest' in the hierarchical tree while moving upwards? In the end, it probably boils down to who we will be able to find and contact and who not.

Whichever individual is approached at the higher organizational levels, it should be checked that the individual is not by chance closely related to the researcher groups of the researcher interviews.

This protocol has two aims.

1) To collect information about the local career and evaluation regimes, now and in the past.

What 'local' means, depends on the location of the interviewee in the the organizational structure. If he/she was identified correctly, then she/he should be at the lowest 'local' level : i.e. if the person is a dean of a faculty, the answer should not be 'it depends on the group or department in this faculty'. Then we are talking to a potentially sub-optimally informed person. Hopefully, he/she can still provide the details.

Hopefully, the person has been with the organization long enough to tell about the past.

2) To collect anonymized or aggregated experiences about evaluations and their relation to or impact on careers

There is an unacceptable loss when asking researchers exclusively, viz. those cases where the things happening to an individual's career are related to evaluations but outside his/her decision power or influence. For example, a dean fires or demotes an individual when there is negative evaluation<sup>3</sup>. Or, an individual's 'head is hunted' so to say and offered a job elsewhere. Such stories will probably not be told about particular individuals because of privacy reasons, but HR managers or people in similar positions may be able to say things in general or without referring to individuals.

### Present and past

We are not only interested in the present situation but also in former practices. Probably, this will be the hard part to get from the interviewees. It requires an old enough interviewee, with a good overview and a good memory. But at least try to get some information and remind the interview time and again also to think about the past. Possible questions to ask 'Since when is this so?', 'How was it before that?'

<sup>3</sup> Keep in mind that this virtually never happens because of an official evaluation. Usually a decision to fire is due or already made based on informal evaluations, but an official form is needed to back up and legitimize the decision.

### **Interview duration**

The design is aiming for interviews of about one hour and a half.

For the first aim, written information has been requested beforehand. If that was available and provided, then the respective part in the interview may not need to take very long. On the other hand, the written information may give rise to specific questions.

Please notice that very long interviews are not necessarily preferable to very short ones. They also take much more time to process.

### **Language issues**

- 'Evaluation'

We mean to talk about evaluation in a wide sense, i.e. not only about peer review of papers and applications, rankings and disciplinary evaluations.

Other words and phrases : assessment - usually refers to activities that can be quantified; examination; annual appraisals or reviews; course evaluations; student evaluations; job applications; self-evaluations; judgements; ratings; rankings; reviews; informal evaluations

- 'Impact'

In the phrasing of the questions, it is important to steer away from 'impact' lingo. This is why the question of the third part speaks of 'roles' in stead of 'impact' or 'influence'.

Other words and phrases that may refer to mutual impact or sound less 'impact': role; 'mutual' impact or influence; how do evaluations play out; evaluation as a/one factor - but not 'impact factor').

- 'Career change'

Secondly, it is important not to stress career changes because it may suggest that we are only interested in changes of jobs, which is why the question of the second part speaks of 'developments'. Arguably, this suggests changes, but I could not come up with a better word or a better question. Evaluations may also play a roles in continuations. The interviewee may think about changes, but that should be his/her interpretation of 'important'.

Other words and phrases: developments; changes; continuations; promotion; demotion; functional promotion; function change; different tasks; reputation boost; reputation blow

## 0 Introduction for the interviewee and the interview 00:00

- Switch on the recording device and do this clearly visible. Then ask if it is o.k. to record. If it is not o.k. clearly switch it off and put it away.
- Explain the background of the interview: part of the ACUMEN project, we are interested in the role of evaluations in academic careers.
- Explain how we will process the results: transcribe and code
- Explain that we will anonymize the data when we publish and report and that quotes will be anonymized too.
- We may circulate data within the ACUMEN consortium but it will be difficult to anonymize (because our partners know already our search strategy and our selection of universities). If the interviewee does not want that then he/she can let us know that via the consent form or otherwise. Also if she/he has second thoughts at a later stage after the interview.
- Refer to the invitation e-mail where all this is explained as well - this is also our backup in case you forget to mention something
- Explain the outline of the interview as we hope it will go.
- Explain that when we ask about evaluations, we mean evaluations in a wide sense of the word. For example also including job application reviews, annual appraisals, research assessment exercises, institute evaluations, course evaluations and informal evaluations.
- Does the interviewee have any questions? She/he will probably ask anyway.
- Hand over the consent form and ask the interviewee to sign it.

Sometimes the interviewee or the situation does not allow you to do all this neatly at the start. Then pick the first possible moment. It would be a waste of momentum to interrupt the flow.

# 1 First part: Career and evaluation regimes

00:10

## **Q 1.1 Could you describe the academic jobs and positions, people can have in your organization?**

Here we can refer to the 'career ladder', i.e. the career regime in our terms, for as far as we have the data already available. We could bring the table, show it and let the interviewee verify it with the local situation and correct it. A preliminary example of such a table can be found in the reading materials folder for the meeting of 6 December 2011. The file is named '20111118 - 25 draft table national career regime.doc'. The table contains data for a few countries, and is not yet ready for use in the interviews.

If it does not and has never deviated from the national one, then quickly go to the next question. Otherwise, this question boils down to the question to describe the local system. Later we will try to codify this in the career regime table.

### **Q 1.1.1 Have there been important changes in the past, say the past 20 to 30 years?**

## **Q 1.2 Which evaluations - of all possible kinds - are being done at your university/department?**

In the researchers' case, the evaluations that they are talking about may be one time experiences or one particular one in a series. Here we also want to know about the frequency, or in which years evaluations were done.

### **Q 1.2.1 Have there been important changes in the past, say the past 20 to 30 years?**

## **Q 1.3 Please characterize the evaluations (see answer form)**



## 2 Second part: Important career developments

### 00:20

**Q 2    Which three types of developments in careers, posts or functions are most important for the organizational unit?**

We are interested in not only changes of jobs, but also promotions, demotions, change of function and change of tasks or hours/task.

Here, I am not sure whether we should indeed ask about important developments 'for researchers' or the 'for the organizational unit'. In the first case, if the interviewee is an academic, she/he might use her/his personal experiences. If he/she is a non-academic, he/she may not know exactly, but still have an idea. On the other hand, if we ask for important developments for the organizational unit we get less of the researchers' perspective but in a way, that is what we are after here: those developments that researchers are less likely to talk about. Hopefully, these overlap with what is important for the organization. Are there other ways to ask a question to better target the likely missing data from the researchers?

An additional benefit is that we arrive at some contrast with the researchers' views, as mentioned in the introduction.

**For each of the three types of developments, ask the following sub-questions**

**Q 2.1    Why is the change mentioned most important for the organizational unit?**

**Q 2.2    What usually happens in such a case and who (in terms of functions or roles) are involved?**

**Q 2.2.1    In particular what influence can the researchers have?**

**Q 2.3    What could trigger such a development?**

**Q 2.4    Which other factors play a role in this change? And how would you weigh these in mutual comparison?**

The weights do not have to be specifically quantified. It would be nice if the interviewee could, but a sort order, or qualitative indications are equally welcome. The question is trying to get from the individual story to an analysis or protoanalysis by the interviewee.

Here, interviewees may mention evaluations. We should not immediately actively fish for that when posing this question. But once the interviewee has finished answering this, we can ask while summarizing to confirm that evaluations played no role of significance. If the interviewee confirms this, then we have established a clear 'no'. If the interviewee on second thoughts agrees/admits that they did play a role, then we have a 'soft' confirmation, but possibly an answer to please us.

**Q 2.5    Characterize new evaluation types**

If the interviewee mentions evaluations then we want to know a few things about it. See p. 5 of the form for interview notes.

Interview protocol for manager interviews. ACUMEN Task 1.2 b

Version: 6 September 2012

### **3 Third part: Evaluations with high impact on (possibly unimportant) change processes in the career**

**00:50**

**Q 3** Of all the evaluations being held now and in the past within your organizational unit, which three types have the biggest effect on researchers' careers?

Probably, interviewees will talk about the present and not remember much quickly about the past.

**Q 3.1** Please, also think of those evaluations that may prevent career changes

**For each change ask the following questions (next page):**

**Q 3.2 What kind of evaluation was it?**

The answer to this question will be categorized. The list of evaluation categories is not yet made, but once it is developed, it may lead to a number of more precise questions or issues for the interviewee. For the time being, see p. 7 of the form for interview notes for a preliminary list. Answering this question may take a lot of time when the interviewee starts describing it in too much detail.

**Q 3.3 What usually happens with/around such an evaluation, where, how and when? What triggered the evaluation? Who are involved?**

Let's say, the facts of the link between the evaluation and the importance for the researchers' careers.

**Q 3.3.1 In particular, what influence can the researchers whom the evaluation concerns have on the course of events?**

In case the official answer is 'none', we may push for the lobbying and informal actions of researchers: 'Is there no way researchers can take action, say 'off the record'?'

**Q 3.4 What else could play a role? How would you weigh all these 'factors' in mutual comparison?**

The weights do not have to be specifically quantified. It would be nice if the interviewee could, but a sort order, or qualitative indications are equally welcome. The question is trying to get to an analysis or protoanalysis by the interviewee.

One likely answer is that one policy or another played a role, perhaps even to continue working with a researcher in spite of a negative evaluation.

**~~Q 3.5 How come these evaluations - this type of evaluation - have/has such a large effect on the change processes of which it is a part? (in other words, why did you list this type of evaluation?)~~**

~~This may have become clear from the previous question, but maybe not. In any case, the interviewee should articulate this.~~

**Q 3.6 What do these evaluations not take into account but should?**

**Q 3.7 How do these evaluations relate to researchers' careers / how do they have an impact?**

If the evaluation played a role in the answers of part two (about important changes in the career) then this question can be skipped or only checked. Perhaps the interviewee has something more to add.

## 4 Rounding off

01:20

Q 4 What is important for researchers or what do they have to do, that is not addressed in any evaluations but should be?

I.e. what never gets any acknowledgement? For example, writing letters of recommendation for students.

Q 4.1 Why should this/these be addressed?

Q 4.2 How should this/these be addressed

Q 5 Which activities are evaluated in the wrong kind of evaluation?

Q 5.1 Where should they go and why?

Q 6 Which activities are evaluated too much or inappropriately ?

Q 6.1 Why? How much should be enough or would be appropriate?

Q 7 In which evaluations is data deposition, creation, maintenance, or review addressed?

Present this as a question that we have a special interest in, with a reference to the work of the DANS institute.

Q 7.1 Is your institution promoting data deposition and data review? Since when? How?



## **Appendix 3: Information sheet and consent form**

The following pages show the consent form. The information sheet had the same text and layout style but was distributed over two pages and lacked the boxed part that the interviewee was asked to sign.





## **Background information to the requests for interviews for the ACUMEN project**

**and**

## **Certificate of consent for interviewees**

### **1 Information sheet**

#### **About the ACUMEN project**

ACUMEN stands for 'Academic Careers Understood through MEasurements and Norms'

"Evaluation of individual scholars happens at different careers stages and comes in different forms, among others: job interviews, annual performance assessments, journal peer review of researchers' manuscripts, and reviews of grant applications. These evaluations have a tremendous influence on all aspects of knowledge production. Moreover, the very criteria of what counts as excellent and relevant research for the next generation of researchers will be strongly influenced by their current experiences in the regular evaluation exercises to which they are subjected. It is therefore urgent that the criteria used in evaluations at the individual level have a clear and well-understood relationship with the requirements that scientists and scholars will need to meet in the near future.

ACUMEN is a European research collaboration aimed at understanding the ways in which researchers are evaluated by their peers and institutions, and at assessing how the science system can be improved and enhanced." (ACUMEN homepage)

One of the final deliverables is a portfolio instrument for academics to use to be able to present themselves for various types of evaluations. The portfolio takes new work practices and requirements into account.

ACUMEN is an FP7'-funded project with partners in Denmark, Estonia, Germany, Israel, the Netherlands, Spain, the United Kingdom. Professor Paul Wouters (CWTS, Leiden) is the principal investigator.

More information about the project and its partners is available on the website at <http://research-acumen.eu/>



### About the subproject in which the interview is being used

One of the tasks within the project is to investigate the current state of affairs regarding evaluation practices and the role of evaluations in researchers' careers. We do this through a literature study, a survey about peer review of research funding applications, and a series of interviews.

The interviews are being conducted and analysed by researchers from the e-Humanities Group and the Data Archiving and Networked Services Institute (DANS). Both are part of the Royal Netherlands Academy of Arts and Sciences (KNAW). Dr. Frank van der Most will perform most of the work, while Professor Sally Wyatt and Dr. Andrea Scharnhorst supervise. Their staff pages are accessible through the ACUMEN project website and through <http://ehumanities.nl/people/>

### About the interviews

ACUMEN focuses its empirical work on four disciplines: public health, astronomy plus astrophysics, environmental engineering, philosophy plus philosophy and history of science. For the interviews, the United Kingdom and Germany have been selected in view of their evaluation and career systems. In the course of the project one or two other countries will be investigated, to be selected at a later stage. For each discipline within each country we identify universities with groups in these disciplines and aim to interview people in three categories of seniority, viz. early career, mid-career and senior academics. We try to establish seniority through resources available on-line, such as staff pages and personal websites.

In addition to interviews with individuals about their own experiences and opinions, we aim to interview deans, department heads and HR managers about the general organizational aspects of evaluations and career development.

### Processing and anonymity

In order to prepare for the interviews, we will ask interviewees for a recent version of their curriculum vitae and we will use resources available on-line.

We intend to record the interviews if the interviewee agrees to this. The recordings will be transcribed or summarized and then coded and analyzed. We will use database and software tools for this. Transcripts may be circulated among ACUMEN project partners via a secure website/server, provided the interviewees give permission. We will use consent forms to arrange these matters.

In reports and publications all data will be anonymised, that is, data will be presented in aggregated form or through presentation of anonymised interview quotes.

### Voluntary basis and withdrawal

Interviewees are asked to participate on a voluntary basis. They receive no compensation for their participation.

Upon request, interviewees can receive the transcript of their interview and they can edit any of their utterances. The interviewees can do so at any time, but their edits can not be taken into account for publications or reports that may already have been published at the time they inform the KNAW project partners mentioned above.

### Data Archiving and Networked Services (DANS)

DANS promotes sustained access to digital research data. Please visit [www.dans.knaw.nl](http://www.dans.knaw.nl) for more information and contact details.

Anna van Saksestraat 10, 2593 HT The Hague, the Netherlands  
P.O. Box 93067, 2509 AB The Hague, the Netherlands  
T +31 (0)70 3446 484, F +31 (0)70 3446 482, E [info@dans.knaw.nl](mailto:info@dans.knaw.nl)  
KVK 54667089  
DANS is an institute of KNAW and NOW

*Driven by data*

Interviewees can completely withdraw their participation at any time except for publications or reports that may have been published at the time of their withdrawal.

May 2012

## 2 Certificate of consent

I have been invited to participate in research about the role of evaluations in my career / the relation between evaluations and careers on the one hand and their relation to the running of my faculty, school or institute on the other\*.

I have read the above information. I have had the opportunity to ask questions about it and any questions I have asked have been answered to my satisfaction. I consent voluntarily to be a participant in this study.

I agree / do not agree\* to the recording and digital processing of the interview.

I agree / do not agree\* to the circulation of transcripts among the ACUMEN project partners at the KNAW (Wyatt, Scharnhorst, Van der Most)

I agree / do not agree\* to the circulation of the transcripts among all ACUMEN project partners

I understand that I can withdraw my participation at any time except for publications and reports already published at the time of my withdrawal.

Name of participant: \_\_\_\_\_

Signature of participant: \_\_\_\_\_

Date: \_\_\_\_\_  
Day / month / year

\* Please delete as appropriate

### Data Archiving and Networked Services (DANS)

DANS promotes sustained access to digital research data. Please visit [www.dans.knaw.nl](http://www.dans.knaw.nl) for more information and contact details.

Anna van Saksestraat 10, 2593 HT The Hague, the Netherlands  
P.O. Box 93067, 2509 AB The Hague, the Netherlands  
T +31 (0)70 3446 484, F +31 (0)70 3446 482, E [info@dans.knaw.nl](mailto:info@dans.knaw.nl)  
KVK 54667089  
DANS is an institute of KNAW and NOW

*Driven by data*



## **Appendix 4: Time-line format**

Interviewee : 482748

Gender:

Year of birth :

Nationality:

Year	'87	'88	'89	'90	'91	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08	'09	'10	'11	'12	'13
Age																											

Residence																											
Family situation																											
Diplomas																											

Job / position																											
• Group + Affil.																											
• Perm. / temp.																											
• full/parttime/%																											
• % research																											
• % teaching																											
• % education																											
• % management																											
• % other																											

Job / position																											
• Group + Affil.																											
• Perm. / temp.																											
• full/parttime/%																											
• % research																											
• % teaching																											
• % education																											
• % management																											
• % other																											

Job / position																											
• Group + Affil.																											
• Perm. / temp.																											
• full/parttime/%																											
• % research																											
• % teaching																											
• % education																											
• % management																											
• % other																											

## Appendix 5: Event codes

Event code group	Event code	Notes
<b>Changes in context – non-work related or family related</b>	Change (relatively sudden) in circumstances – family circumstances	
	Change (relatively sudden) in circumstances – interviewee's health	
	Change (relatively sudden) in circumstances – event occurring in nature	
<b>Changes in context – organization or work related</b>	Local reorganization	
	Change in personnel	
	Getting in close collaboration with a particular individual	
	Unexpected change in a project	
	Change (relatively sudden) in funding	
	Change (relatively sudden) in circumstances – publication culture	
	Changes in higher education (student numbers, treatment of students, college fees)	
	More collaborative work is needed	
	Internationalization	
<b>Changes in main tasks</b>	Shift towards research	
	Shift towards research coordination	
	Shift towards teaching	
	Harder work in the course of time	
<b>Diplomas</b>	MA equivalent (Diploma)	
	Dr. equivalent (Diploma)	
	Habilitation equivalent (Diploma)	
	Other (Diploma)	
	Unknown (Diploma)	
<b>Doing a PhD (Except diploma)</b>	Doing and finishing the PhD	
	Doing the PhD studies/work	

## Appendix 5 - Event codes

Event code group	Event code	Notes
	Starting on PhD	
<b>Evaluations leading to an invitation to something</b>	Evaluation – informal – after which one is invited to lead a project	[informal]
	Evaluation – formal/informal unknown – after which one is invited to lead a project	[informal]
	Evaluation – informal – after which one is invited to participate in a project	[informal]
	Evaluation – formal/informal unknown – after which one is invited to become head of department	
	Evaluation – informal – after which one is invited to apply for a job	[informal]
	Evaluation – informal – after which one is invited to apply for promotion	[informal]
	A supervisor's informal evaluation of MA thesis/work eventually leading to an offer for a PhD research	[informal]
	Evaluation – formal/informal unknown – after which one is invited to participate in a review committee	
	Evaluation – formal/informal unknown – after which one is invited to membership of a board	
	Evaluation – formal/informal unknown – after which one is invited to become editor	Could be editor for a book, a journal, a special issue, a review series
	Evaluation – formal/informal unknown – after which one is invited to accept a permanent position	
	Evaluations – informal – after which one receives job offers	
	Evaluations – formal/informal unknown – after which one is invited to become expert for the government/a ministry	
	Evaluation – informal – after which one is invited to membership of a board	[informal]
	Evaluation – informal – leading to invitation to write a paper in a prestigious publication	[informal]
<b>Evaluations of teaching</b>	Evaluation of teaching (Peer evaluations of lectures and courses)	
	Evaluation of teaching (Student evaluations	

Event code group	Event code	Notes
	of lectures and courses)	
	Evaluation of teaching (Student plus peer evaluations of lectures and courses)	
	Evaluation of teaching (unspecified)	
<b>Feedback</b>	Feedback – formal/informal unknown – from 'client' organizations	
	Feedback – informal – from colleagues – in general	[informal]
	Feedback – informal – from colleagues – other	[informal]
	Feedback – informal – from a friend – other	[informal]
	Feedback – formal/informal unknown – from patient groups	
	Feedback – formal – from job coach	
<b>Grant applications</b>	Grant application – irrespective of success	
	Grant application – success unknown	
	Feedback – formal – on grant applications	
	Grant application – successful	
	Grant application – successful – informal feedback from colleagues INside department	[informal]
	Grant application – successful – informal feedback from colleagues OUTside department	[informal]
	Grant application – unsuccessful	
	Grant applications – irrespective of success	
	Grant applications – successful	
	Grant applications – unsuccessful	
	Participation in application reviews	The interviewee is the evaluator, rather than the evaluated
<b>Grants</b>	PhD scholarship grant (national)	
	PhD scholarship grant (international)	
	Career grant (national)	
	Career grant (international)	
	Project grant (national)	
	Project grant (international)	



## Appendix 5 - Event codes

Event code group	Event code	Notes
	Visiting scholarship grant (national)	
	Visiting scholarship grant (international)	
	Unknown (national)	
	Unknown (international)	
	Unknown (unknown nature)	
<b>Job</b>	Non-academic	
	Student assistant	Academic job before the MA level diploma
	Scientific assistant	Academic jobs after the MA level diploma but before Dr. level. In some countries, this position is used to do research leading to a doctoral diploma.
	PhD Worker	When 'doing a PhD' means, one has a paid position. This happens in, among other countries, Belgium, the Netherlands and Sweden.
	Post doctoral worker	Post doctoral researcher, teacher, assistant. Also for post doctoral level fellowships
	Assistant prof. level	Lecturer in UK. Universitair Docent in the Netherlands
	Associate prof. level	Reader or Senior lecturer in UK. UHD in the Netherlands
	Professor level (job)	
	Probation end (i.e. moving on in same job)	

Event code group	Event code	Notes
	after probation period)	
	Promotion	
	Other (job)	
	Acquiring/ starting of a job	
	Getting tenure / a permanent position	
	Getting a tenure track position	
	Quitting a particular job	
	Earning an income	
	Unknown (job)	
<b>Job applications</b>	Job application – irrespective of success – for a professor position	
	Job application – irrespective of success	
	Job application – successful	
	Job application – successful – informal	[informal]
	Job application – unsuccessful	
	Job applications – irrespective of success	
	Job applications – successful	
	Job applications – unsuccessful	
	Job application – the interviewee's assessment of the potential employer	[informal]
<b>Job-related evaluations (Not job applications)</b>	Reference letter from boss/supervisor	
	Reference letters from bosses/supervisors	
	Annual (or otherwise regular) appraisal talks	
	Annual evaluation (other than appraisal talks)	For example annual lists of production of all department members
	Application for sabbatical	
	Annual (or otherwise regular) rewards or bonuses	
	Probation evaluation (i.e. evaluation at end of probation period)	
	Promotion application (independent of level)	
	Promotion applications (successful)	
	Promotion (without application, independent of level)	
	Evaluation of individual work as part of a	

## Appendix 5 - Event codes

Event code group	Event code	Notes
	reorganization / redundancy	
	Application for tenure track as a promotion	In the Netherlands, tenure track applications can be a kind of promotion application when someone already has a temporary or permanent position.
	Tenure track intermediate evaluations	
<b>Other</b>	Choice for academia	
	Leaving science	
	Founding of own company	
	Owner of company	
	Valorization of research	
	Developing oneself in a broad way, inside and outside academia	
	Visiting scholar	Independent of job level
	Prize, winning a prize	
	Project	
	Reputation	
	Student evaluation (the students are being evaluated)	
	Change in research orientation	
	Publish (as an activity)	
	Exams during studies	
	Course application	
	Visibility	
	Other	
	Other (multiple)	
<b>Other informal evaluations</b>	Other (informal evaluation)	[informal]
	Informal evaluation leading to copying of a method or procedure	[informal]
<b>Other official or formal evaluations</b>	Evaluation for presidency of professional /	

Event code group	Event code	Notes
	disciplinary association	
	National evaluation scheme	
	National or international rankings	
	Institute evaluation	An evaluation of an entire institute or group of institutes
	Ex post evaluation of a particular project	
<b>Paper and book reviews</b>	Paper review – informal feedback from colleague	[informal]
	Paper/book review – one particular paper – successful	
	Paper/book review – one particular paper – unsuccessful	
	Paper/book review – success unknown	
	Paper/book reviews – multiple – successful	
	Paper/book reviews – multiple – independent of success	
	Paper/book reviewed by interviewee – rejected	
<b>Titles (not referring to a diploma)</b>	Professor level (title)	
	Lecturer level (title)	