

Research Quality Indicators: Two large-scale surveys

Introduction to the Surveys¹

This paper focuses on two surveys which generated long lists of indicators; these are potentially relevant for evaluating research quality in the humanities and social sciences. One, generated for a survey of German psychologists, contains 117 “evaluation criteria,”² while the other reports the result of surveying national research councils in the humanities.³ It contains 82 quantitative “measures of research quality.”⁴ They are referred to here as the Krampen survey and the HERA survey.

These are not criteria or measures per se, however, implying yardsticks to measure research quality directly. Instead, the Krampen survey can be characterized as an effort to synthesize the responses in a particular discipline to a list of “criteria” regarded as desirable, disregarding their use (or absence) in practice. The HERA survey, in turn, was interested in gathering cross-national information, in pre-determined categories, about humanities funding agency “practices” with an intent to compare what funding agencies said they did. Neither survey clearly distinguished between what is quantitatively measurable from what is qualitatively evaluated.

In the Krampen survey, the 265 of 555 German psychologists who responded were asked both for their personal reactions were, as well as what they assumed the judgment of most of their German-speaking colleagues was, to a pre-determined list of 117 “criteria.” A

¹ Originally written Dec. 2010 as part of humanities practices surveys I conducted on behalf of the Swiss Rector’s Conference study of establishing quality criteria for the humanities. Some editorial changes to the text were made in early 2026 prior to uploading the text to my website.

² In addition to identifying criteria, this survey asked respondents to provide information about the use of scientific information sources and types of literature, about their work and publication practices and their working conditions. Information, for correlational purposes, was also gathered on age, gender, year the doctorate and *Habilitation* were accepted, year first appointed, and number of accepted and rejected job offers. The source is: Krampen, Günter und Leo Montada. 2002. *Wissenschaftsforschung in der Psychologie* (Göttingen: Hogrefe), pp. 16-48. Göttingen: Hogrefe.

³ The terms agency/agencies and councils are used synonymously here (with some preference for ‘council’), to indicate a functional equivalence: these are evaluators and funders of humanities research. However, as the HERA survey indicates, the name of the institution doing so in a given nation may be council, academy, agency, center, foundation, fund, organization or society. A further complication, not entered into here, is that the agency/council involved may have a narrower brief (humanities and social sciences, say, as in Ireland) or may cover a much wider area (scientific research, as in France).

⁴ HERA is a partnership between the European Science Foundation and 15 Humanities Research Councils across Europe. The source is: HERA. 2006. Report on Impact and Quality Assessment. This was previously available at www.heranet.info. A discussion of this HERA survey can be found in Carl Dolan’s 2007/2008 feasibility study, conducted for the AHRC, entitled “The Evaluation and Benchmarking of Humanities Research in Europe,” Section 3. Available at: https://www.aqu.cat/doc/doc_34869796_1.pdf (accessed 16 Jan. 2026).

9-point scale from 0 (not positive) to 8 (highly positive) was used for both personal reaction and assumed judgment. One might call both desirability rankings. Mean values were reported on both scales, because the aim was to compare subjective judgment with (what was projected as) peer judgment. The thematically-ordered categories included the “criteria” of publication, research activity, appointment/employment and honors, teaching, academic self-governance, reviewing activity and professional association involvement.

Krampen’s survey resulted in identifying four “conformity types” among German psychologists. One aim was create quantitative summaries of individual judgments so as to establish where peer consensus might exist. However, respondents engaged in a theoretical exercise, inasmuch as they were asked to assess *possible* evaluation “criteria” outside of any particular context. They were also provided with a pre-determined list of “criteria” rather than being asked to generate them individually. The analysis then provided was limited to a general, summary description. Here, I’ve tried to conduct a finer disaggregation of the data provided, but one should bear in mind that the source of that data comes entirely from personal reactions rather than objectively gathered information. My analysis suggests the ultimate result of this exercise is to reconfirm a *collective auto-stereotype*.

In the HERA survey, the stated focus was on the “impact and quality assessment practices of humanities funding agencies in Europe and further afield.” In particular, the survey wanted to examine the *ex-post* evaluation of research outcomes, in part to disentangle it from *ex-ante* evaluations found or solicited in funding applications. Because of differences in conceptualization as well as national reporting practices, this survey also wanted to distinguish monitoring “practices” – meaning the successful carrying out of projects relative to stated aims and objectives – from “systematically assess[ing] the quality of ... outcomes.” In other words, HERA wanted to hear from the funding agencies themselves about what they did rather than ask about the significance of one indicator relative to another. The basis of the information provided was thus quite different than in the Krampen survey. The specificity of the questions made it possible to distinguish between quantitative and qualitative indicators, so HERA could establish which indicators are being used more (say, by 10 of 16 funding agencies) or less often (say, by 1 of 16).

The Comparison

The Krampen “criteria” and HERA practices and measures were collected for quite different purposes, in a different manner, and from different kinds of respondents. A prospective list of criteria generated for the purpose of studying a single social science, with

“criteria” generated top-down by a member of the profession conducting the survey is quite different than an international bottom-up survey of *ex-post* evaluation practices for funded research reported by national funding agencies in the humanities. The first is top-down, the second bottom-up; the first is based on individual perceptions and projections, the second on reported practices; the first is in the social sciences, the second in the humanities, and so forth.

Nevertheless, in a comparison of difference, one looks for similarities. If “research quality” is a characteristic independent of time, place, circumstance and person, then one could expect to find “criteria” or “measures” appearing in both studies. Or similar indicators could be found at the same higher or lower ranks in both. Or a specific, abstracted, lesson could result, say, that the production of a video is regarded – by academics and funding agencies alike – as inferior to publishing an article in a peer-reviewed academic journal.

Gathering ideas is a quite different reason to examine these lists of “criteria” or “practdices.” In such long lists, it is likely that item will appear that simply haven’t been considered before, especially if the purpose is to find ways to evaluate the quality of research in the humanities. Indeed, being able to consider what is judged most important (by the practitioners of one social science discipline) or what is most frequently mentioned (by the funding agencies) may be a way to identify what is most central and best rewarded. The more peripheral mentions are valuable for the opposite reason: original research is often regarded as innovative, and less common yardsticks may then be needed to adequately acknowledge this.

The Krampen “criteria”

The Krampen survey separates its many criteria into 19 groups (see column A, Krampen Criteria Groups, below), with varying numbers of criteria in each (column B). A value of 4 given to a particular criterion falls exactly in the middle of the 0-8 scale. Knowing how many of the criteria in a particular criteria group fell above or below a value of 4 provides a quick sense of whether a particular criteria group was valued highly or not (compare, for example, columns C and D).

As one can see from the list of means (column E), the tendency is clearly to value up rather than down; only two groups (f. and p.) have means below 4. In fact, across all 117 criteria, only 16 of them had a mean value less than 4 (column E) – a possible artefact of a study in which criteria are evaluated independently of one another. Hence, one needs to draw subtler distinctions, including over what the range of values in a group of criteria (column F) shows. A narrow range can be read as greater consensus, a broader one the lack of it.

Krampen Criteria Groups

A	B	C	D	E	F
Group of criteria	Number of criteria	Value > 4	Value < 4	Mean	Range
a. Wirkung der Publikationen	8	6	2	5.03	3.6-5.9
b. Status des Autors	3	3	0	4.93	4.8-5.2
c. Publikationsort	8	5	3	4.67	2.2-6.9
d. Rang des Publikationsortes	3	3	0	5.27	4.0-6.8
e. Kategorie der Publikation	7	7	0	5.80	5.0-6.4
f. Multimediale Produkte	8	3	5	3.37	2.5-4.0
g. Herausgeberschaften	4	3	1	4.72	3.2-6.1
h. Qualität der wissenschaftl. Arbeit	11	11	0	5.59	4.4-7.3
i. Höhe der eingeworbene Drittmittel	8	8	0	5.38	4.6-6.4
j. Teilnahme an natl. wiss. Kongresse	6	5	1	5.08	3.9-6.2
k. Teiln. an internatl. wiss. Kongresse	8	8	0	5.67	4.4-7.0
l. Anzahl gemeins. Forschungsprojekte	5	5	0	6.00	5.1-6.6
m. Berufungen und Ehrungen	11	11	0	5.16	4.0-5.9
n. Gastprofessuren	6	6	0	5.35	4.4-6.1
o. Lehre	6	6	0	6.26	5.5-6.6
p. Ämter in der akad. Selbstverwalt.	4	2	2	3.80	3.4-4.3
q. Häufigkeit der Gutachtertätigkeit	6	5	1	4.96	3.9-6.1
r. Wichtige Funktionen in Fachgesell	3	2	1	4.20	3.1-5.2
s. Anzahl Publ. & Publ./Berufsdauer	2	2	0	5.80	5.5-6.1

In this form, these many criteria are not easy to interpret. As discussed below, it is of some significance that the highest means (column E) are found for the group of criteria listed under *Lehre* (o. = 6.26) and *Anzahl gemeinsame Forschungsprojekte* (l. = 6.00), and the lowest means are found for *Ämter in der akademischen Selbstverwaltung* (p. = 3.80) and *Multimediale Produkte* (f. = 3.37).

A closer analysis of the ranges (column F) is instructive, as in some groups, the range from lowest to highest may be narrow (b. = 0.4) or considerable (g. = 2.9). Range differences could certainly be an artefact of the limited number of criteria in a group, but it more often reflects whether the group of criteria is a cluster (f.), a scale (g.), or a mix of different domains in a group (h.).

To illustrate, one can compare groups with a narrow range (difference of 0.4 to 1.1 between lowest and highest values) to those with a broad range (difference of 2.8 or 2.9 between lowest and highest). In the case of a narrow range, there does not appear to be much difference in judgment in the following groups of categories:

Narrow range criteria groups (b., p., and f.)

b. *Status des Autors* (range: 0.4)

- 4.8 Alleinautorenschaft
- 4.8 Coautor renommierten Kollegen
- 5.2 Erstautorenschaft

p. *Ämter in der akademischen Selbstverwaltung* (range: 0.9)

- 3.4 Mitglied im Fakultätsrat
- 3.5 Mitarbeit in anderen Gremien
- 4.0 Fachsprecher/Institutsdirektor
- 4.3 Dekan und vergleichbare Positionen

f. *Lehre* (range: 1.1)

- 5.5 Praxis-/Anwendungsorientiert
- 6.2 Breite der Lehrerfahrung
- 6.4 Individuelle Betreuung der Studierenden
- 6.4 Aktualität der Lehrveranstaltungen
- 6.5 Engagement/Commitment in der Lehre
- 6.6 Didaktische Aufbereitung und Qualität

That is, teaching is very highly regarded (of all 19 groups of criteria, in fact, *Lehre* (o. = 6.26) has the highest mean value), and *all* administrative positions, whether they carry much responsibility (*Dekan*) or relatively little (*Fakultätsratsmitglied*) seem to be equally poorly regarded (this group of criteria, in fact, has the second lowest (p. = 3.80) mean value or all 19 groups). Interestingly, among German psychologists it seems to make little difference (b.) whether you are the sole author, first author or co-author with a well-known colleague.

It is a different story in groups of criteria where the range is broad. In some cases (see d. and g. below), the range reflects a genuine scale and clear difference in judgment between best and worst. In other cases, the range (as in h.) reflects a jumbling together of criteria that simply don't belong in the same group of criteria:

Broad range criteria groups (d., g., and h.)

d. *Range des Publikationsortes* (range: 2.8)

- 6.8 Renomee der Zeitschrift
- 5.0 Renomee d. Hrsg. Eines Sammelbandes
- 4.0 Renomee des Verlages

g. *Herausgeberschaften* (range: 2.9)

- 6.1 von Zeitschriften
- 4.9 von Sammelbänden
- 4.7 von Buchreihen
- 3.2 von Newsletter

h. Qualität der wissenschaftliche Arbeiten (overall range: 2.9)

- 7.3 Originalität der Fragestellung
- 6.3 Sorgfalt der Stichprobengewinnung
- 6.3 Aufwand der empirischen Untersuchung
- 6.3 Kreativität der Interpretation
- 6.4 Benützung modernern Analysemethoden
- (h1.) sub-mean: 6.4

- 4.9 Alter bei Abschluss der Habilitation
- 4.8 Alter bei Abschluss der Promotion
- 4.7 Alter bei erster Berufung
- (h2.) sub-mean: 4.8

- 5.5 Anzahl der betreuten Habilitationen
- 5.2 Anzahl der betreuten Promotionen
- 4.4 Anzahl der betreuten Diplomarbeiten
- (h3.) sub-mean: 5.0

In the first two cases, the range clearly reflects substantial differences in judgment. Publication with a prestigious press is clearly substantially less important than publication in a prestigious journal (group d.), reflecting the importance today in German psychology of journal article rather than book publication. Similarly, being the editor of a newsletter (a lowly 3.2 ranking in group g., well below 4) is unimportant – as compared with being the editor of a specialist journal (at 6.1, well above the mean).

Unfortunately, in the criteria group (h.) that is of greatest interest, the “quality of the scholarly work/research,” the range is not helpful. This group mixes criteria of external judgment (e.g., creative interpretation) with demographic information (e.g., age at dissertation) and training of students and future scholars (e.g., number of doctorates mentored). For that reason, I’ve separated this category into three thematically coherent sub-groups (h1, h2, h3). If one takes only h1 (sub-mean = 6.40; range = 1.5), one finds much greater consensus (e.g., smaller range) than if one takes all of h. together (mean = 5.59; range = 2.9). The broad range of h. is an artefact of mixing disparate criteria or indicators together.

The widest range (4.7 from lowest to highest) among all 19 groups of criteria is the place of publication (c.). This group too is composed of two quite distinct clusters (c1; c2):

c. Publikationsort (overall range: 4.7)

- 6.9 Zeitschrift
- 6.2 Monographie
- 5.9 Lehrbuch
- 5.6 Enzyklopedie
- 5.0 Sammelwerk
- (c1. sub-mean = 5.9, range: 1.9)

- 3.2 Kongress- oder Tagungsband
- 2.4 „Graue Literatur“
- 2.2 Newsletter
- (c2. sub-mean = 2.6, range: 1.0)

But it is more truthful to divide this list differently – and the values for the assumed judgment of peers is included for comparison purposes:

c.1 <i>Good Places to Publish</i> (>4)			c.2 <i>Bad Places to Publish</i> (<4)		
personal judgment		assumed peer judgment	personal judgment		assumed peer judgment
6.9	Zeitschrift	[6.5]	3.2	Kongress- oder Tagungsband	[4.0]
6.2	Monographie	[6.3]	2.4	„Graue Literatur“	[2.6]
5.9	Lehrbuch	[6.0]	2.2	Newsletter	[2.7]
5.6	Enzyklopedie	[6.1]			
5.0	Sammelwerk	[5.2]			

For those outside of this discipline, it may be surprising to see that German psychologists rate textbooks, encyclopaedia articles and edited volumes as highly as they do.

Analysis of the Krampen list

The large number of criteria in this survey can be reduced by eliminating all the groups of criteria that do not focus specifically on research. That means setting aside criteria groups related to service (p., q., r.) and teaching (o., h3), criteria groups specific to psychology (group e., for example, contains criteria about whether an investigation is empirical, experimental, theoretical, etc.), and criteria included for statistical analysis (h2).

The special case of group f. (multimedia products) is discussed separately below.

This winnowing leaves one with two separate clusters:

I. publication criteria

- a. (citation or impact)
- b. & g. (authorship and editorship)
- c. & d. (place and reputation of publisher),
- h1. (quality of publication), and
- s. (total publications and publications divided by time in profession)

II. esteem indicators

- j. & k. (participation in national and international congresses)
- l. (leadership of or cooperation in research)
- m. (honors, awards, appointments)
- n. (guest professorships), and
- i. (third-party funding).

It is revealing to then take the single most highly ranked criterion in each (e.g. in I. and in II.) and list them in descending order:

I. Publication Criteria

h1	7.3	Originalität der Fragestellung
c1.	6.9	Publikationsort: Zeitschrift
d.	6.8	Renomee der Zeitschrift
g.	6.1	Herausgeberschaft von Zeitschriften
s.	6.1	Publikationen/Berufsdauer
a.	5.9	Anzahl der Erwähnung in systematische Reviews
b.	5.2	Erstautorenschaft

II. Esteem Indicators

k.	7.0	Eingeladener Hauptvortrag an internationalen wiss. Kongressen
l.	6.6	Gründung eines Forschungsschwerpunktes/Betreiben eines Forschungszentrums
i.	6.4	Höhe der eingeworbenen Drittmittel von der DFG
j.	6.2	Eingeladener Hauptvortrag an nationalen wissenschaftlichen Kongressen
n.	6.1	Gastprofessuren im Ausland
m.	5.9	Wissenschaftliche Preise

If one then formulates this in prose, one arrives at an auto-stereotypical view of (or the guidelines for) what is believed to be the path to success as a German academic psychologist:

One poses original questions, publishes in highly regarded journals (perhaps is an editor of one), and has been in the academic world for a while. One gives invited keynote addresses at conferences abroad and at home (the former are clearly more prestigious), runs a research center or has founded a research program, has brought in a large amount of money from the German Research Council, is invited to guest professorships abroad, and wins academic prizes.

Mutatis mutandis, this summary could apply to many academic disciplines... Indeed, it reads very much like a job description for a senior academic position.

Note what this formulation does not say, though:

- 'quality,' even if it remains unclear what exactly that means, is embodied by those who are regarded as at the top of the discipline
- 'publication' is important, though the number of key criteria and indicators (where published, reputation of journal, number of citations, invitations to speak abroad and at home, role in advancing the discipline, etc.) that have to do with impact suggest this is at least as important, if not more so, than publication itself

- ‘esteem’ is a combination of valuation by peers in the discipline as well as recognition from those outside the discipline
- the status of those who do not meet these criteria is by implication lower (e.g., position and hierarchy are determinative)

What Matters to German Psychologists Themselves

It is clear German psychologists have little good to say about “service to the university and to the discipline” (other than if you act as a *DFG-Gutachter*, which receives a value of 6.1). But limiting the analysis as above – in other words, by focusing only on research – does not capture the full scope of what respondents themselves believe to be important. In fact, if one instead lists individual criteria with the highest absolute scores (see Highest Mentions, below), then it is clear that in the minds of German academic psychologists (and perhaps surprising to outside evaluators?), *the quality of teaching – preparation, engagement, individual mentoring, timeliness of what is being taught – ranks nearly as highly as publication and esteem.*

Highest Mentions (values above 6.0)

7.3	Originalität der Fragestellung
7.0	Eingeladener Hauptvortrag an internatl. wiss. Kongressen
6.8	Renomee der Zeitschrift
6.6	Gründung eines Forschungsschwerpunktes
6.6	Betreibung eines Forschungszentrums
6.6	<u>Didaktische Aufbereitung und Qualität der Lehre</u>
6.5	<u>Engagement in der Lehre</u>
6.4	<u>Individuelle Betreuung der Studierenden</u>
6.4	<u>Aktualität der Lehrveranstaltungen</u>
6.4	Höhe der von der DFG erworbene Drittmittel
6.3	<i>Sorgfalt der Stichprobengewinnung - criterion specific to the discipline</i>
6.3	<i>Aufwand der empirischen Untersuchung – criterion specific to the discipline</i>
6.3	Kreativität der Interpretation
6.2	<u>Breite der Lehrerfahrung</u>
6.2	Eingeladener Hauptvortrag an nationalen wiss. Kongressen
6.1	Publikationen/Berufsdauer
6.1	Gastprofessuren im Ausland

If one accepts that the Krampen survey is at least in part about what impact means to German psychologists, one can also try to disentangle “impact” as it is reflected in these highest mentions. High values, in terms of impact, seem to reflect:

- new insights (criteria of judgment rather than quantitative indicators per se, with originality ranking higher than creativity)
- direct transmission (keynote addresses at international and national conferences, engaged and up-to-date teaching, guest professorships abroad)
- indirect transmission (publication in respected journals, and over a longer time period), and
- institutional frameworks (founding/directing a new research direction/center, receiving outside funding).

These distinctions could serve, for example, as a guideline for expectations for junior scholars or more generally for careers. By and large, impact is directed externally, though a few discipline-internal aspects related to professional standards were also highly ranked (as in criteria related to the care taken in sampling and the effort empirical investigations involve).

One can also look at this from the other end and examine the lowest mentions:

Lowest mentions (below 4.0)

- 2.2 Publikationsort: Newsletter
- 2.4 Publikationsort: Graue Literatur
- 2.5 Multimediale Produkte: Softwarebesprechung
- 2.7 Multimediale Produkte: Berichte über Kongresse
- 2.7 Multimediale Produkte: Buchbesprechung
- 3.1 Wichtige Funktion in der BDP
- 3.2 Herausgeberschaft von Newslettern
- 3.2 Publikationsort: Kongress- oder Tagungsband
- 3.2 Multimediale Produkte: Editorials
- 3.4 Mitglied im Fakultätsrat
- 3.5 Mitarbeit in anderen Gremien
- 3.7 Multimediale Produkte: Schulungs- und Trainingsverfahren
- 3.9 Gutachtertätigkeit für Kongressbeiträge
- 3.9 Poster an einer nationalen wiss. Kongress

Of the four components of ‘impact’ noted above, neither insight nor direct transmission appear among these lowest ranked criteria and instead we find *undesirable* forms of indirect transmission (e.g., multimedia products and certain channels such as newsletters) and undesirable institutional frameworks (e.g., that involve service provided to the institution and the profession). Stated more broadly, some *contexts* are clearly judged to be better (keynote addresses) or worse (publishing in a newsletter) for having an impact, and the same is true of *channels* (publishing in a respected journal vs. in the ‘grey literature’). The prestige of certain kinds of administrative work is clearly higher in some cases (such as if one directs a research center) than in others (such as serving on the faculty council).

What the neglected middle reveals

The discussion above emphasizes the extremes (values over 6 or under 4), obscuring ‘standard’ criteria that are neither exceptionally high nor exceptionally low. The following presents a list of criteria that range from 5.1 to 5.9. Some criteria groups are absent here, either because that group has no values in this range (e.g., h2., f., g., or p.) or because the criteria group is discipline-internal (e.) or Germany-specific (i.).

5.9	Anzahl der Erwähnung in syst. Reviews Lehrbuch Wissenschaftliche Preise	(a. Wirkung der Publikationen) (c. Publikationsort) (m. Berufungen und Ehrungen)
5.8	Anzahl der Erwähnungen in Lehrbüchern Benützung modernern Analysemethoden Anzahl der Forschungsprojekte/Lebensalter	(a. Wirkung der Publikationen) (h1. Qualität der wiss. Arbeiten) (k. Teilnahme an internationalen wiss. Kongressen)
5.7	Anzahl der (Fremd-) Zitationen Rang d. zitierenden Publikationsquellen Forschungsreferat Zitierung in anderen Disziplinen Anzahl fremdfinanzierter Auslandsaufenthalte	(a. Wirkung der Publikationen) (a. Wirkung der Publikationen) (k. Teilnahme an internationalen wiss. Kongressen) (l. Anzahl gemeinsamer Forschungsprojekte) (n. Gastprofessuren)
5.6	Enzyklopedie Berufung in Akademien Rang der berufenden Institution Reputationseinstufung durch Fachkollegen	(c. Publikationsort) (m. Berufungen und Ehrungen) (m. Berufungen und Ehrungen) (n. Gastprofessuren)
5.5	Praxis-/Anwendungsorientiert Anzahl der betreuten Habilitationen Gutachtertätigkeit für Zeitschriften- manuskripte Anzahl der Publikationen	(o. Lehre) (h3.) (q. Häufigkeit der Gutachtertätigkeiten) (s.)
5.4	Überblicks- oder Positionsreferat Organisation einer Arbeitsgruppe Anzahl der Berufungen auf Professuren Ehrendoktorwürde	(j. Teilnahme an nationale wiss. Kongresse) (k. Teilnahme an internationalen wiss. Kongressen) (m. Berufungen und Ehrungen) (m. Berufungen und Ehrungen)
5.3	Berufung in Gremien von Wissenschafts- organisationen Gutachtertätigkeit für Projektanträge Internationale wissenschaftl. Gesellschaften	(m. Berufungen und Ehrungen) (q. Häufigkeit der Gutachtertätigkeiten) (r. Wichtige Funktionen in Fachgesellschaften)
5.2	Anzahl der Cited Core Documents Erstautorenschaft Anzahl der betreuten Promotionen Angemeldeter Hauptvortrag Stipendien Anzahl d. Forschungsaufenthalte im Ausland	(a. Wirkung der Publikationen) (b. Status des Autors) (h3.) (j. Teilnahme an nationale wiss. Kongresse) (m. Berufungen und Ehrungen) (n. Gastprofessuren)
5.1	Forschungsreferat Anzahl der forschungsaktiven Mitarbeiter	(j. Teilnahme an nationale wiss. Kongresse) (k. Teilnahme an internationalen wiss. Kongressen)

Anzahl nationaler Kooperationen	(l. Anzahl gemeinsamer Forschungsprojekte)
Berufung in wissenschaftliche Beiräte	(m. Berufungen und Ehrungen)
Gastprofessuren im Inland	(n. Gastprofessuren)

More than half (21 of 37) of these are found in criteria groups a., k., m. and n, meaning most of these in the mid-range are *esteem* indicators (including appointments, honors and guest professorships) or indicators of *impact* (including various forms of participation at international congresses and citations in various locations).

As above, though the distinctions here are finer, one can look at them in terms of ranges, if not even of dichotomies. They suggest, for example, that receiving a guest professorship in Germany (5.1.) is a good deal less prestigious than receiving a guest professorship abroad (6.1), or that to lecture about one's research at an international scientific congress (5.7) carries more weight than doing so at a comparable national congress (5.1). These are the kinds of differentiations "everyone knows" in the profession. By the same token, because this is tacit knowledge or privately communicated (say, from professor to doctoral student), it is rarely as clearly visible – or quantified – as in this survey.

There is even a discernable quantitative split between external (5.6 – 5.9) and internal (5.1 – 5.5) impact criteria. That is, being named to an academy (5.6.) or cited in other disciplines (5.7), the number of research projects one has relative to one's age (5.8) or the prizes one has won (5.9) are criteria which reflect aspects of presenting the discipline to outsiders. Criteria such as the number of *Habilitationen* one has mentored (5.5) or organizing a working group at an international conference (5.4), serving as an evaluator for project proposals (5.3), whether one was primary author (5.2), or presenting research results at a national conference (5.1) are all activities oriented towards colleagues in the discipline.

These examples (e.g., in the 5.1 to 5.5 range), are activities that can be called less prestigious (e.g., than those in the 5.6 to 5.9 range), and it is a differentiation that even extends to criteria groups which have activities that are both internal and (potentially) external to a discipline. Different kinds of evaluation activity (*Gutachtertätigkeit*), for example, are quite differently valued – the worst for papers at congresses (3.9; internal), better for book manuscripts (4.7; mostly internal?), even better for article manuscripts (5.5; potentially external?) and best if it is for the German Science Foundation (6.1; definitely external). A similar scale is evident among guest professorships, which are apparently least esteemed if they are simply replacement posts (4.4; internal), better if they are proper guest professorships in Germany (5.1; internal, more status), even better if they are for a stay abroad paid out of third party sources (5.7; external), and best if one has a guest professorship abroad (6.1;

external, highest status). Interestingly, only two groups (1. *Anzahl gemeinsamer Forschungsprojekte* and o. *Lehre*) ranked every criterion at 5.1 or above. As a group, and as a matter of collective self-image, *German psychologists seem to place considerable importance on reaching out and presenting their discipline to outsiders* – at least in what they answered to this opinion survey.

The HERA indicators

These were generated from a 2006 survey of impact and quality assessment practices used by research councils in HERA member countries. Some comparable centers in countries not part of the HERA consortium were included, additional research proved necessary in certain nations, and information was incomplete for some of the questions asked. While 17 responses were received, comprehensive information was not available.

List of research councils surveyed

<i>Acronym</i>	<i>Name (country, if not otherwise evident)</i>
AHRC	Arts and Humanities Research Council (UK)
AKA	Academy of Finland
ARRS	Slovenian Research Agency
ASCR	Academy of Science Czech Republic
CNR	National Centre for Research (Italy)
CNRS	National Centre for Scientific Research (France)
ETF	Estonian Science Foundation
FWF	Austrian Science Fund
FWO	Research Foundation Flanders
GACR	Czech Science Foundation
IRCHSS	Irish Research Council for the Humanities and Social Sciences
NWO	Netherlands Organisation for Scientific Research - Humanities
NSF	National Science Foundation (U.S.)
RCN	Research Council of Norway
RSNZ	Royal Society of New Zealand
SRC	Swedish Research Council

The information gathered allowed for an initial categorization of the main practices used in evaluating humanities research quality:

Data collection

- | | |
|---------------------------------------|-----------------------------------|
| 1. Final reports/End of Award reports | All |
| 2. Post-award surveys/interviews | 5 (FWF, IRCHSS, ASCR, RSNZ, AHRC) |
| 3. Citations | 5 (FWF, ASCR, GACR, RSNZ, CNR) |

Project-level assessment

- | | |
|---|--|
| Peer review* of final reports/end of award reports/interviews | 7 (FWF, FWO, ASCR, GACR, ETF, SRC, AHRC) |
| Case studies | 3 (ARRS, AHRC, NSF) |

Programme/instrument-level assessment

- | | |
|-----------------------------------|---|
| 1. Peer review ^o panel | 7 (FWF, FWO, ASCR, AKA, NWO, RSNZ, RCN) |
| (a) with international peers | 5 (FWO, ASCR, AKA, NWO, RCN) |
| (b) using bibliometric indicators | 4 (FWF, ASCR, AKA, RSNZ) |
| (c) involving site-visits | 2 (AKA, NWO) |

Assessment of disciplines

- | | |
|-----------------------------------|---------------------|
| 1. Peer review ^o panel | 3 (RCN, AHRC, CNRS) |
| (a) with international peers | 2 (RCN, ARRS) |
| (b) using bibliometric indicators | 1 (ARRS) |
| (c) involving site-visits | 2 (RCN, ARRS) |

International benchmarking

- | | |
|--|--------------------------|
| 1. Using informed peer review | 3 (AKA, RCN, CNRS) |
| 2. Using bibliometric/other quantitative indicators only | 4 (FWF, CNR, RSNZ, ARRS) |

* peer review here means where peers are asked to make some assessment of the quality of the outcomes.

^o peer review here means only when there is an assessment of the quality of the outcomes of the programme

There is little agreement at the level of practice across these councils. All gather end of award reports, but only 5 of 16 conduct surveys or check citations. At the level of project assessment, only 7 conduct peer-reviewed quality assessments of the outcomes, and only 3 do case studies. The story is similar for program assessment. Only 7 – not identical with the 7 that conduct project assessment – have peer review panels, and of them, only 5 use international peers, 4 used bibliometric indicators, and 2 site visits. The story is similar for discipline assessment and international benchmarking.

The use of bibliometric (or other quantitative indicators) is clearly a small minority phenomenon among these 16 councils, whether for international benchmarking (4), discipline assessment (1), or program assessment (4). While the use of peer review is clearly more popular, it is also not used by even half the councils, whether for international benchmarking (3), discipline assessment (3), and either program or project-level assessment (both 7). Only a few use both bibliometric indicators and peer review (the FWF, ASCR, AKA and RSNZ in program, the ARRS in discipline, assessment).

Of particular interest are the more specific measures of performance used. The survey collected information on (I.) qualitative and (II.) quantitative measures as well as on (III.) wider social and economic impact and on (IV.) indicators.

I. Qualitative Indicators

The description of qualitative indicators provides a mixed picture, due in part to a lack of clarity in the answers given by the councils themselves, as they mixed statements about on-going evaluation with statements specific to *ex-post* evaluation, and partly to situations in which councils use multiple modes that include evaluation by national experts, site visits and interviews with researchers. There is a mode used by the majority – *self-evaluation plus external expert evaluation* – but this varies considerably by nation.

Some modes of evaluation are rarely used as qualitative measures. They include evaluation committee reports (1), surveys of researchers (3), project outcome peer review (4) and case studies (4). The AHRC is the only council which conducts “subject reviews of academic quality.” Somewhat more popular are interviews with researchers (6) and site visits (7). By far the most popular methods of qualitative assessment are researcher self-assessments (10) and evaluations by national (11) and/or international experts (9). Even so, that leaves out 5 of the councils surveyed.

Qualitative Indicators

Expert evaluations	National (11): FWF, ASCR, AKA, DFG NWO, RCN, ARRS, SRC AHRC (of funding, not of research quality), NEH (ex ante), NSF (ex ante)	International (9): FWF, FWO, ASCR, ESF (ex ante & ex post), AKA, CNRS, NWO, ARRS (sometimes), SRC
Project outcome peer review	National (3): FWF, GACR, ETF	International (1): FWF
Case studies (4):	ASCR, ARRS, NSF & AHRC (uses national and international peer review, subject reviews of academic quality in law, archaeology, music, and French have been carried out. Planned: information science, modern history, medieval history).	
Site visits (7):	FWF, AKA, NWO (some programs), RSNZ, ARRS, NEH, NSF	
Interviews with researchers (6):	FWF, ASCR, AKA, RSNZ, RCN, ARRS	
Researcher self-assessment (10):	ASCR, AKA, NWO, RSNZ, RCN, SRC, NEH & DRA, ICHRSS, AHRC (these three on end of award form, peer reviewed)	
Surveys of researchers (3):	ICHRSS (annual & biannual), NWO, ARRS	
Report of evaluation committee:	CNR (Italy)	

Some councils provided more detail about the internal process. In some agencies, evaluation results in giving an overall grade (FWO, ETF). In others, the goal is to provide information useful for future programmatic decisions (including about impact, as in the FWF) or for future investments in research (ASCR, DRA). Qualitative assessments may also provide information that is complementary to quantitative indicators (AKA).

Details from national research councils

FWF (Austria)

In the case of large projects which are carried out within the National Research Networks, evaluation is based on the following criteria: i. the scientific success of the project; ii. the development of human resources in the course of the project; iii. effects beyond the scientific discipline (in the sense of applications in or impacts on social, cultural, ecological, medical, economic and/or technological areas); iv. the running of the project with regard to use of available resources; and the future perspectives of the research work. Each of these criteria are given a numerical grade (1-100) and an overall grade is assigned to for the project's performance as a whole.

FWO (Belgium/Flanders)

All evaluation is carried out using international boards of referees. The research is then graded using the following categories: i. Good; ii. Conditional (with remarks or questions. Can change after a reply); iii. Unsatisfactory (can change after a rebuttal).

ASCR (Czech Republic)

Qualitative data is collected using end-of-award reports, and the interviews and surveys of researchers carried out by the AAC every 5 years. The research is ranked using 5 grades (A-E) under the following headings: i. Potential impact; ii. Scientific potential of the applicant/institution; iii. Practicability; iv. Research team; v. Infrastructure; vi. Time schedule of research plan; vii. Financial support. The project is then given an overall grade ranging from A (outstanding) to E (unacceptable). The result of the evaluation – the cumulative rank output decided by the Academy Appraisal Committee – has significant influence on distribution of institutional finances in ASCR.

DRA (Denmark)

The EoAs contain information on work carried out, results and publications, and new research issues which arose in the course of the investigation. The results are assessed in relation to the funding given and the original application. The forms are not routinely analysed, but are used when decisions are made about investments in new research.

ETF (Estonia)

The expert commission provides ratings in its project evaluation. Experts rate whether the project achieved its aims (Y/N), the projects' scientific outcome's originality (1-5), the quality of the project's publications (1-5), the outcomes' practical value (if an applied research project), the quality of graduate training (1-5), the quality of the project report (1-5). Based on these ratings, the expert commission gives its overall evaluation of the grants.

AKA (Finland)

Qualitative assessments complement the picture drawn by standard quantitative indicators (including bibliometrics, measures of research esteem, electronic outputs, performance and visual media outputs, indicators of the mobility of researchers, and other statistical data derived from KOTA database and Statistics Finland).

The aim is to reveal kinds of scientific impact that have been hard to establish on the strength of quantitative indicators alone. The entire range of quantitative information is available.

II. Quantitative Indicators

It is interesting how many councils (DRA, ESF, IRCHSS, RCN, NSF) do *not* use quantitative indicators or bibliometric data to measure research quality or outputs. Instead, they “take a holistic view of progress achieved during the period of support,” as, for example, the Irish IRCHSS does.

Two council gave explicit reasons why they do not. The ASRC states that “citation analyses (impact factors, etc.) are not used in evaluations of the humanities due to the perceived limitations of citation databases for these research fields.” The AHRC also makes a still finer distinction:

It is AHRC policy not to use citation indices and bibliometric analyses for the research it funds. AHRC policy is underpinned by differentiating *metrics of activity; impact assessment; and assessment of quality*.

Metrics of activity - such as citations, bibliometric analyses and patents – that describe the nature and scale of research activity and output are very different from impact assessment and assessments of quality. Metrics of activity have tended to be used in some other subject areas and disciplines as proxies for impact and quality. The validity of such proxies has been challenged where such proxies distort or misrepresent the real extent of impact and quality. The AHRC is currently negotiating...the precise nature of the indicators it will present as part of its performance management framework. This has led AHRC to explore possible quantitative measures that might legitimately provide indicative proxies for quality. These include:

- processes and sources of journal and book publication (as opposed to citation rates);
- range and diversity of personal interactions at the international level;
- doctoral student flows;
- visiting academic flows;
- access to and use of infrastructure.

Summarizing larger categories of quantitative indicators councils claim to use is difficult because the information provided is incomplete and national practices unique. Some humanities research councils (in Austria, the Czech Republic, Finland, New Zealand, and Estonia) report using more indicator categories than do others, though it is difficult to know whether this reflects genuine differences in practice or merely differences in reporting: more detailed study would be necessary to establish which. For some countries (notably Germany, France and Italy), information had to be established from sources other than the HERA questionnaire, so this information is not directly or fully comparable. In addition, it is not always clear whether indicators are being gathered for the humanities as a whole or whether they are (or can be) broken down by individual discipline.

Quantitative Measures Utilized

	Research Inputs	Research Outputs	Bibliometric Data	Research Esteem ⁵	Researcher Mobility	Web & Media*
FWF	Yes	Yes	Yes	Yes	Yes	Yes
ASCR	Yes	Yes	Yes	Yes	Yes	Yes
GACR	Yes	Yes	Yes	Yes	Yes	No
AKA	n.a.	Yes	Yes	Yes	Yes	Yes
RSNZ	Yes	Yes	Yes	Yes	Yes	No
ETF	n.a.	Yes	Yes	Yes	Yes	n.a.
NWO	Yes	Yes	Some	(No)	No	Yes
SRC	n.a.	Yes	Yes	n.a.	n.a.	Yes
CNRS	n.a.	Yes	No	(Yes)	Yes	n.a.
FWO	n.a.	Yes, but by universities	No	Some	Some	No
NEH	n.a.	Yes	n.a.	(Yes)	n.a.	Yes
CNR ^o	Yes	Yes	Yes	No	No	n.a.
ARRS [^]	n.a.	Yes	Yes	No	No	n.a.
Total Yes	6	13	9	6 (+2)	7	6

* refers to whether digital information (websites, digital archives, databases, datasets) and performance/visual media outputs are collected

^o for the special case of CNR, see Appendix 1

[^] for the special case of ARRS, see Appendix 2

Note: FWF outsources bibliometry (cf. Evidence Ltd); SRC also collects information on conference papers and popular scientific publications; NEH uses an ‘outcomes database’ which includes “electronic outputs, exhibitions, performances and visual outputs, as well as books, journal articles, critical and scholarly editions.”

In terms of categories, all councils use research outputs and most (9 of 13) also use bibliometric data. All other categories are only used by about half (6 or 7 of 13) of the councils. One can also distinguish between those which say they collect information across nearly all of these six categories (FWF, ASCR, GACR, AKA, RSNZ) and those that collect far less. The two categories in which the least information was available or mentioned were research inputs and web and media.

III. Wider Social and Economic Impact

The evaluating of the wider social and economic impact of the humanities is still in its infancy among the councils surveyed here. Indicators have not been developed in 11 (FWF, FWO, DRA, IRCHSS, NWO, RSNZ, NEH, ASCR, ETF, ESF, and ARRS), and only 3 councils (DFG, SRC, RCN) have plans to do so. The Norwegians (RCN) think this will include “elements of stakeholder opinion and the societal value of research – ‘valorisation,’ while the Finns (AKA) claim to have done so but provided no specifications, nor did they state such indicators were used in evaluation or pilot schemes.

⁵ A qualitative proxy or judgment of what can be quantitatively assessed (e.g., how many articles in leading journals, how many books published in renowned presses, etc.). Esteem is not otherwise readily quantifiable.

The only specific efforts to create robust indicators of wider social and economic impact were mentioned by the AHRC. In their ‘Sector Interaction Studies and Case Studies’ they were interested in understanding the impact of Arts and Humanities Research outside of academe, beginning with end-users and with respect to knowledge pathways, and have launched four ‘Impact Fellowships’ to provide “medium-term reflective thinking focusing on private-public sector support for the arts; the social impacts of cultural arts on individuals and communities; economic methods and valuation of arts and culture; and the impact of higher education on the creative industries.”

The GACR, RCN and NSF each specifically stated they have not collected data on wider social and economic impact, but – in increasing degrees of complexity – the ETF, ESF, NWO ASCR, RSNZ and ARRS have. However, much of this appears to be unsystematic, voluntary, or assessed on the basis of surveys of the media or the press:

ETF: only in end of award reports

ESF: data on media, non-academic appearances, outreach and so forth (provided voluntarily and only the EUROCORES programs)

NWO: collects information on broadcast and print media coverage using constant press surveys, press clippings and other ad hoc methods, including researcher-provided information

ASCR: in end-of-award forms, annual surveys of researchers, 5-year AAC evaluations, daily surveys of press and media

RSNZ: gathers a wide range of information using end of award reports and structured interviews as well as surveys of press and media conducted by a contracted media monitoring system.

The most extensive system, though not specific to the humanities, is undertaken by the Slovenian ARRS (see Appendix 2). It measures indicators of the wider social and economic impacts for all scientific disciplines and gathers wider social and economic indicators through annual reports. Researchers financed by the Slovenian Research Agency must also define the wider social and economic impact of their research. In categorizing the “impact of applied research results,” the ARRS lists the following, at least some of which have clear links to the kind of research typically undertaken in the humanities:

- Development of higher educational training
- Economical development (new product, lowering the costs, spreading, competitiveness, lowering the consumption of energy and materials, higher share of export, profit gain, rise of educational level of employed, new jobs, new investment momentum, etc.)
- Social development (rising quality of life, improvement of management and administration, improvement of public administration, development of social activities, development of civil society)
- Preservation and development of national cultural and natural heritage and identity
- Environment protection
- Development of social infrastructure
- Protection of health

IV. Indicators

The list of indicators the HERA survey provided is divided into six sub-categories. As in the Krampen survey, each sub-category contained from 5 to 20 specified indicators.

Councils varied a great deal in how many indicators they reported using in specific sub-categories. The category ‘inputs’ in the ASR used 6 of 8 indicators, the AKA used only 2, and the DRA none at all. In trying to establish quantitative means, it thus makes sense to distinguish between means based on all 15 of the councils surveyed, and means based only on those reporting using indicators in that category. Thus, while it was theoretically possible for 300 individual cells to be filled in the ‘outputs’ category (e.g. 20 indicators x 15 councils), the actual number mentioned, at 96, was only about one-third as large. If one then divides the number of councils actually mentioning using indicators in this sub-category (12 agencies), one arrives at a figure of 8.0. In other words, those councils which used output indicators employed a relatively large number, especially when compared to research inputs, though less than half of the entire array of indicators mentioned in this sub-category (e.g. 8 of 20).

The following highlights that the most popular sub-categories of indicators are

Sub-category	No. councils w. indicators	No. of indicators	Total poss. mentions (x 15 councils)	Actual mentions	Mean/council listing indicators
Outputs	12	20	300	96 (32%)	8.0 (96/12)
Bibliometrics	11	17	255	98 (38%)	8.9 (98/11)
Research Esteem	10	18	270	76 (28%)	7.6
Researcher Mobility	9	5	75	22 (29%)	2.4
Research Inputs	7	8	120	29 (24%)	4.2
Wider Impact	6	15	225	33 (15%)	5.5

‘Outputs/Bibliometrics’ and ‘Research Esteem.’ ‘Wider Impact’ is the least popular, used by only 6 of 15 councils.

There is heavy skew here, since in the ‘Wider Impact’ category, the ASCR is responsible for 14 (42%) of the 33 mentions. There is also a council skew overall in the information. Nearly half (47%, or 168) of the 354 indicators actually mentioned come from only *three* councils: the ASCR (73), the GACR (41) and the AKA (54). And the ASCR and GACR are located in the same country...

One way to counter-balance at least some of this imbalance is to look at the frequency with which individual indicators are mentioned across all councils. Four councils, though they provided other information, listed practically no information (0 to 2 indicators), so the following is based on the 11 that listed more than 2 indicators:

Number of Councils	Indicator
10	Total publications Total published books/monographs Total articles in refereed journals
9	Conference papers Total articles in books
8	Patents/citations of patents Website Awards, prizes and other honors Participation of researchers (outbound) Postgraduate degree completion rate

Taking this as a popularity context, *cross-nationally in humanities councils, it is the quantity of output that is paramount for research agencies*. Issues of impact (website, awards, outbound research, postgraduate degree completion) only begin to appear further down the list, though they are relevant for the majority of these 11 humanities councils.

In the middle ranks, the picture is diffuse, if with distinctions similar in form to those in the Krampen survey. A conference paper is more significant than a conference abstract (9 vs. 7), a refereed article more than a non-refereed one (10 vs. 7), publishing a book is more often used as indicator than editing a book (10 vs. 6). Interestingly, mounting an exhibition, while less significant than giving a conference paper (6 vs. 9), is no less important than giving a plenary address, editing a book series, or co-authoring a journal article. In that sense, again only in terms of conducting *ex-post* reviews, one has a rough guide to the relative importance of different kinds of ‘research-related’ activities:

7	Conference abstracts Conferences dedicated to specific research Total articles in non-refereed journals Research Funding Participation of researchers (inbound)
6	Personnel costs Exhibition or other presentation Performance, recording, film or broadcast Participation at international conferences Invitations to be conference chair/panelist Invitations to give plenary or keynote presentations Inflow of postgraduate students Co-author of article in journal (total) Editor, book (total) Editor, book series (total)
5	Intellectual property/IP items to be commercialized Contract research International invitations to lecture Joint editor, book (total) or book series (total) Critical edition

At the bottom are the rarely mentioned indicators. The bias here is similar, with a few councils mentioning many indicators, many councils mentioning few ‘outlier’ items. In a given country, on the other hand, this may be quite relevant for funding, for example if the given humanities discipline includes catalogues, artifacts, designs, compositions, or translations in its own list of output to be evaluated. Some impact-associated indicators feature among the little-mentioned: partnerships and advisorships, public lectures, radio and television appearances. A relatively large number of esteem-related indicators are found here, including international academic networks, editorial board membership, learned society membership nationally or internationally, invitations to review research or write textbooks, receiving invitations to be a visiting academic either nationally or internationally.

At least for the purposes of *ex-post* evaluations by national research councils of the humanities, *what seems to matter most internationally is output quantity and certain forms of visible/high impact (websites, prizes). What matters (relatively) less or least are certain discipline-specific products (e.g., compositions, catalogues), esteem (e.g., learned society membership) and reaching out to the public (e.g. public lectures, TV appearances).*

Conclusion: Comparing the Lists

This paper began with the assumption that one might be able to find similar criteria or measures in these two long lists, or that their rank order might show some similarities. The analysis yields two answers.

One is that similarities certainly exist in the larger thematic categories. The general categories of publication and research activity employed by Krampen have their counterparts in the categories of outputs (and bibliometric measurement) in the HERA survey. Measures of research esteem, such as invitations to be a keynote speaker or teach abroad, are also found in both cases.

These similarities also exist in terms of analytic categories. If German psychologists ever think of impact in terms of new insights, direct and indirect transmission, and institutional frameworks, then there are equivalents for this – not surprisingly – to be found in the HERA survey. Seen the other way, the HERA-derived list can be divided in ways that correspond to some of the more subtle distinctions in the Krampen survey, for example to distinguish between external and internal impact criteria.

The other answer is that differences are striking if one compares the two “most popular” lists, the one which lists the criteria with the highest values among German

psychologists, the other the indicators used most frequently by research councils. In the first, **originality, international reputation, highly ranked journal publications, impact in the discipline and teaching** matters most. In the second, it is the **total publications and larger impact** (through total publications, conference papers, websites, prizes, etc.) that matter.

One could not have a clearer indication that what matters to researchers within a discipline and what matters to funding agencies outside the discipline differ. *In the case of the former, while imperfect, these are efforts to identify quality; in the case of the latter, these are efforts to identify effect(iveness).*

This is a place for a meeting of the minds, for external funders and evaluators to understand better what criteria within disciplines look like – and for the researchers within disciplines to understand better what evaluators and funders value. The underside to this is that academics may then work to ‘game the system’ by providing what they know is wanted from without (as it will also result in funding...), even though that may only partly align with what is esteemed within a field. The better solution would be for both sides to bend a little towards one another; funders and evaluators might be persuaded that they can reach their goal differently by refashioning what they understand as academic work.

This returns to the distinction between contexts and channels that was made in passing. In the context of an academic employed at a university, issues of teaching commitment (valued positively) and administrative service (valued negatively) play a significant role, at the very least because of the time commitments involved and the need to balance out these demands against the demands of research and publication. This context, however, is of relatively little interest to funding agencies, though at least indirectly acknowledged in the “research inputs” category. Some funding agencies in fact explicitly include considerations of the conditions for conducting research in their evaluations.

Something similar can be said of the channels, though here the story is more complex. Researchers themselves make fine and subtle distinctions between various channels, in part as a proxy for quality assessments, in part to signal to one another what a ‘good’ rather than ‘bad’ channel is. But not only do differing valuations exist by discipline, there is some flux over time as to the valuation itself – the uncertain status of web-based and multimedia productions are cases in point. The channel itself, in other words, is important.

To research agencies, while channels are important, output *itself* (or even the impact of that output) loom larger. The hierarchy of channels depends on the mandate of the research agency itself or the reputation of a national scholarly community (e.g., are our scholars being invited abroad?) rather than on questions such as which publisher has printed a particular

scholarly monograph. Curiously enough, the one place where there is convergence is at the bottom. Research agencies seem to care little about wider social and economic impact – and at least among German psychologists, activities that go outside the discipline (e.g., *Anzahl der Erwähnung in Massenmedien* or *Editorials*) have similarly low status (3.9 and 3.2).

Last, but certainly not least, one should note that **quality**, in its meaning as “excellence” or “being superior,” is *at best* being assessed by proxies. German psychologists rate “originality” or “creativity” highly, but whether that also means quality (or more to the point, whether that will get your paper published in the ‘best’ journals) is quite a different matter. Being invited to give the keynote speech at an international congress, like being the head of a research center, is more a reflection of importance to, or impact in, the discipline than a measure of quality – or one instead reflects *status* in the discipline. One hopes status is related to quality, but the connection is unsure and indirect; even highly esteemed scholars periodically produce shoddy work...

On the other side, research councils often only inadequately address quality, largely because they don’t seem to much understand what motivates academics or how rewards are distributed. In other words, a research council which values the number of publications rewards productivity, and if it values larger impact, then some kind of assessment about their ‘relevance.’ Fine, but that is not at all the same thing as valuing the originality of a colleague, the importance of teaching (or training future colleagues), or indeed the reputation boost that a department receives when one of its members is invited to be a guest professor.

That is also reflected in two different uses of peer review: funders want peers to focus on impact, but disciplinary colleagues want peers to uphold certain standards internal to a field. It is worse if those external funders focus on what they can count (such as total publications), that not only encourages academics themselves to ‘game the system’ to meet that demand, but it also means not saying anything meaningful about the content of those publications (which is, in turn, an internal assessment among colleague-peers make). In the end, it is an argument for why research councils should more adequately weight how and by which criteria academics themselves assess what it is they do – and why their own, disciplinary-internal standards matter.

Appendix 1: CNR

In assessing of the quality of humanities research, the CNR uses the following indicators:

1. Total publications
2. JCR publications
3. Mean impact factor (IF)
4. Other international publications
5. Personnel costs
6. Research funds
7. Self-financing capacity
8. International publications per researcher
9. National publications per researcher
10. Teaching posts per researcher
11. Trainees per researcher
12. Number of contracts with the EU
13. Average EU contribution
14. Mean Observed Citation Rate (MOCR)
15. Mean Expected Citation Rate (MECR)
16. Relative Citation Rate (RCR)

Appendix 2: ARRS

The ARRS Uses weighted “Indicators of Researcher Efficiency” as well as “Indicators of Developmental and Management Efficiency.”

ARRS INDICATORS OF RESEARCH EFFICIENCY (point system)

1. Scientific articles indexed in SCI Expanded:
 - First quarter of journals: 80 points
 - Second quarter of journals: 60 points
 - Third quarter of journals: 40 points
 - Fourth quarter of journals: 20 points
2. Scientific articles indexed in SSCI:
 - Above median of corresponding sci. journals: 80 points
 - Below median of corresponding sci. journals: 40 points
3. Scientific articles indexed in A&HCI: 20 points
4. Scientific articles not indexed in ISI, but in international bibliographic data base: 10 points
5. Scientific articles published in Slovenian research journals: 5 points
6. Short scientific contributions are evaluated at 80% of what scientific articles receive in corresponding scientific journals.
7. Book published with international scientific publisher: 100 points
8. Book published with domestic scientific publisher: 50 points
9. Book published with other publishers: 30 points

II. INDICATORS OF DEVELOPMENTAL EFFICIENCY

10. Transfer of knowledge into economy and social sphere
11. Integration of research in university study programmes
12. Publishing of faculty handbooks
13. Patents or selling of patent rights
14. Research and developmental work in support for development of databases, indicators, dictionaries, glossaries, lexicons etc.
15. Development of systematic, normative, programmatic, methodological and organizational solutions, including evaluations, reviews, expert report
16. Published expert works

III. INDICATORS OF MANAGEMENT EFFICIENCY

17. Efficiency and successfulness in previous periods
18. Comparison of research goals with available infrastructural capacities
19. Integration of researchers with domestic business sector and local social networks
20. International integration of researchers