

Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history

Sven E. Hug^{1,2,*}, Michael Ochsner¹ and Hans-Dieter Daniel^{1,2}

¹*Professorship for Social Psychology and Research on Higher Education, ETH Zurich, D-GESS, Muehlegasse 21, 8001 Zurich, Switzerland and* ²*Evaluation Office, Muehlegasse 21, University of Zurich, 8001 Zurich, Switzerland*

**Corresponding author. Email: sven.hug@gess.ethz.ch*

How to assess research quality in the humanities is an intricate question. Despite the recent efforts of many initiatives, the measurement and assessment of research quality still faces strong opposition from humanities scholars, indicating that currently used evaluation schemes and tools are not tailored well enough to humanities disciplines. We have collected quality criteria from scholars in Switzerland and at League of European Research Universities (LERU) in the three disciplines, German literature studies, English literature studies and art history with a multi-round Delphi survey. The first Delphi round resulted in a comprehensive list of 19 criteria of good research specified by 70 aspects. Although 10 of these criteria are well known and commonly used in various evaluation schemes, nine criteria are not, or at least not frequently, employed in evaluation schemes. In the second Delphi round, consensual criteria and aspects (i.e. items that were clearly approved by a majority and disapproved by very few scholars) were identified in each discipline. Specifically, 11 criteria reached consensus in all three disciplines (shared criteria), six criteria were consensual in only one or two disciplines (discipline-specific criteria) and two criteria were not consensual in any discipline (i.e. 'productivity' and 'relation to and impact on society'). The results of this study corroborate previous findings that researchers have not yet adopted Mode 2-related assessment criteria. Implications for research assessment are being discussed. The focus lies in particular on the mismatch of criteria between evaluators and scholars as well as on an approach to bridge such a mismatch.

Keywords: research evaluation; quality criteria; Delphi survey; research quality; humanities; English literature studies; German literature studies; art history; LERU.

1. Introduction

How to assess research quality in the humanities is an intricate and still open question. Different initiatives are currently investigating assessment tools for the humanities or are striving to make the quantity and quality of humanities' research visible. Examples include the RESH and DICE databases ([Evaluation of Scientific Publications Research Group 2010: 11–13, 2012](#)), the VABB-SHW

database, ([Engels et al. 2012](#)), the CRISTIN database ([Schneider 2009; Sivertsen 2010](#)), the ERIH project ([European Science Foundation 2011](#)), the MESUR project ([National Science Foundation 2009](#)), the Book Citation Index ([Thomson Reuters 2011](#)), an evaluation of book publishers ([Gimenez-Toledo et al. 2013](#)), Libcitations ([White et al. 2009](#)), a label for peer-reviewed books ([Verleysen and Engels 2013](#)), the Research Rating of the German Council of Science and Humanities

(Wissenschaftsrat 2011a,b), the Quality Indicators for Research in the Humanities (Royal Netherlands Academy of Arts and Sciences 2011), the EERQI project (EERQI Consortium 2011), and the ERA initiative (Australian Research Council 2012).

Despite these efforts, the measurement and assessment of research quality still faces strong resistance and severe criticism from the humanities (e.g. Andersen et al. 2009; Plumpe 2009; Pontille and Torný 2010, 2012; Verband der Historikerinnen und Historiker Deutschlands 2010), indicating that so far, too little attention has been paid to scholars' opposition in the development of tools for assessing research quality in the humanities. Hence, in order to increase the acceptance of research assessments in the humanities, scholars' criticisms must be taken into account. For this reason, we have analysed scholars' objections and identified four crucial points that need close attention when developing assessment tools (for an elaboration, see Hug et al., in press): (1) humanities scholars criticize that the methods employed in research evaluation originated from the natural sciences and are modelled after them; (2) scholars have strong reservations regarding the quantification of research quality; (3) they fear negative steering effects of indicators; and (4) scholars state that there is a lack of consensus on quality criteria within humanities disciplines and, therefore, comparing or assessing research quality is impossible. In order to address these four objections, we devised a framework to explore and develop quality criteria and indicators for assessing research quality in the humanities (for an elaboration, see Hug et al., in press). In brief, the four cornerstones of this framework are:

- (1) Adopting an inside-out approach. The development of criteria and indicators must be rooted in the humanities themselves, ideally in each discipline or subdiscipline, so that the unique quality criteria of each discipline can emerge.
- (2) Relying on a sound measurement approach. A sound measurement approach simply means that research quality is not merely defined by its measures, as is often the case (Brooks 2005; Bazeley 2010; Donovan 2007, 2008), but that indicators must be tied to concepts of quality. In particular, every quality criterion has to be defined and specified by one or more aspects (i.e. conceptualization of research quality) and each aspect has to be linked, if possible, to at least one quantitative indicator (i.e. operationalization of research quality). Such a measurement approach clarifies what is being measured by indicators and it allows for the identification of quantifiable and non-quantifiable quality criteria.
- (3) Making notions of quality explicit. The notions of quality that underlie an indicator, an assessment tool or an evaluation procedure should be made as clear and explicit as possible to disclose the direction

in which research should evolve. Moreover, scholars' tacit and explicit knowledge of quality should be taken into account when developing criteria and indicators so that research will ultimately evolve into the direction of the scholars' understanding of good research.

- (4) Striving for consensus. To account for the lack of consensus regarding quality criteria, an approach should be adopted that either allows for a consensus to be reached in a discipline/subdiscipline or it should at least allow for the detection of criteria that are consensual and those that are not.

We have implemented the above delineated framework empirically using a mixed methods approach. First, we conducted Repertory Grid interviews to explore scholars' criteria of good research and to make them explicit (Ochsner et al. 2012) and, subsequently, we surveyed a large group of scholars with a three-round Delphi. The first Delphi round aimed at completing the quality criteria derived from the Repertory Grid interviews, the second round at rating the criteria's aspects, and the third round at rating quantitative indicators. In the Repertory Grid interviews and the Delphi survey, we focused on English literature studies (ELS), German literature studies (GLS) and art history (AH) because they represent humanities disciplines that elude evaluative bibliometrics. This article presents the results of the first and second rounds of the Delphi survey and examines quality criteria. Due to space limitations, the third Delphi round, which concentrates on ratings of indicators, will be published in a future paper.

There exist a few studies that have investigated quality criteria in the humanities empirically (Hemlin 1993, 1996; Hemlin and Gustafsson 1996; Wooding and Grant 2003; Guetzkow et al. 2004; Oancea and Furlong 2007; Lamont 2009; Bazeley 2010; Gogolin 2011), but none of them focused exclusively on the humanities or treated the humanities as an independent unit of analysis. Instead, the analysis of humanities disciplines was linked with disciplines from the natural or social sciences, thus hampering the emergence of quality criteria particular to the humanities. As postulated in our framework, in order to establish research assessments that are more likely to be accepted by scholars, the humanities' own quality criteria have to be included in assessments. For this reason, we have collected quality criteria in humanities disciplines only. Accordingly, this article seeks to answer the following research questions: (1) What are criteria for good research of scholars in English literature, German literature and art history? (2) How appropriate are these criteria in the eyes of scholars to assess their own research? (3) Is there a set of criteria which is clearly approved by a majority and disapproved by very few scholars in one or more disciplines (i.e. consensual quality criteria)?

2. Methods

We used the Delphi method to investigate scholars' quality criteria. Delphi is 'a method for the systematic solicitation and collection of judgments on a particular topic through a set of carefully designed sequential questionnaires interspersed with summarized information and feedback of opinions derived from earlier responses' (Delbecq et al. 1975: 10). A Delphi survey makes use of experts' opinions in multiple rounds with anonymous feedback after each round in order to solve a problem (Häder and Häder 2000). According to Linstone and Turoff (1975: 4), the Delphi method is able to tackle, *inter alia*, the following methodological issues: 'More individuals are needed than can effectively interact in a face-to-face exchange. The problem [under study] does not lend itself to precise analytical techniques but can benefit from subjective judgments on a collective basis'. Delphi is considered to be a flexible and adaptable method (Hsu and Sandford 2007) and is a valid, widely used and recognized instrument (Landeta 2006). It has also been applied in research on higher education (see Murry and Hammons 1995; Häder and Häder 2000; Palomares-Montero and Garcia-Aracil 2011) and, specifically, it was used to gather criteria and indicators for research evaluation (Nadeau 1995; Lahtinen et al. 2005; Palomares-Montero and Garcia-Aracil 2011).

2.1 Panel

The Delphi panel comprised all research-active faculty members of Swiss universities holding a PhD in GLS, ELS or AH. Swiss universities were chosen because the Rectors' Conference of the Swiss Universities commissioned this study. In addition, to ensure international standards and comparability, the panel included all research-active faculty members holding a PhD in the three respective disciplines from universities that were members of the LERU in spring 2010. LERU describes itself as an 'association of research-intensive universities that share the values of high-quality teaching within an environment of internationally competitive research' (LERU 2012a). The universities of Geneva and Zurich are both LERU and Swiss universities but are solely regarded as Swiss universities in this study. As not all Swiss and LERU universities covered the three disciplines in spring 2010, 8 Swiss and 17 LERU universities were represented in the panel. The panel was set up in spring 2010 for the first Delphi round and consisted of 582 scholars. To account for the turnover of research staff at the 25 universities, the panel was updated in winter 2010–11 for the second Delphi round and comprised 664 scholars at that time (Table 1).

2.2 Questionnaires

The questionnaire for the first Delphi round consisted of 17 criteria for good research and a total of 49 aspects that

Table 1. Composition of survey panel in spring 2010 and winter 2010–11 by discipline, university, and sex

	Spring 2010 (N = 582) n (%)	Winter 2010–11 (N = 664) n (%)
Discipline		
Art History	230 (39.5)	254 (38.3)
English Literature Studies	179 (30.8)	200 (30.1)
German Literature Studies	173 (29.7)	210 (31.6)
University		
LERU	454 (78.0)	505 (76.1)
Swiss	128 (22.0)	159 (23.9)
Sex		
Female	234 (40.2)	264 (39.8)
Male	348 (59.8)	400 (60.2)

specified these criteria. The criteria and aspects were compiled from Repertory Grid interviews with humanities scholars (Ochsner et al. 2012) and from pertinent publications in the Arts and Humanities Research Assessment Bibliography (Peric et al. 2013: ch. 3). As this compilation was subjective to a certain degree and potentially incomplete, the aim of the first Delphi round was to let scholars modify and complete the criteria and aspects. Additionally, in preparation for the third Delphi round, which aimed at rating quantitative indicators, the purpose of the first round was also to collect indicators that can quantify or measure aspects. All of the questionnaire's 17 criteria were structured according to the measurement approach outlined by Hug et al. (in press): each *criterion* was composed of *aspects* that specify and define a criterion and of *indicators* that specify how an aspect can be observed, quantified, or measured. The participants were asked (1) to tick those of the 49 aspects that they personally found fit or matched the criteria, (2) to add aspects to a criterion that, from their personal point of view, were missing, and (3) to name indicators to quantify or measure an aspect. Furthermore, participants were encouraged to name additional criteria as well as corresponding aspects and indicators if they thought that important criteria were missing. As the questionnaire was qualitative in nature, it was administered to a subset of the panel that comprised 180 persons, 30 from each of the three disciplines at Swiss universities and 30 of each from the three disciplines at LERU. An equal distribution of Swiss and LERU universities was chosen in the subset in order to guarantee a sufficient number of answers from Swiss universities. The subset was generated using stratified sampling with each discipline at a university as a stratum. The sampling yielded 77 women and 103 men (42.7 and 57.2% of the subset, respectively). The first round of Delphi was conducted from May to August 2010 employing paper-and-pencil questionnaires as well as questionnaires in the Portable Document Format (ISO 32000-1 2008).

The questionnaire of the second Delphi round consisted of 19 criteria specified by 70 aspects that had resulted from the first Delphi round. The participants were asked to rate all 70 aspects in the form of statements. A statement consisted of a generic part (i.e. ‘My research is assessed appropriately, if the assessment considers whether I...’) and an aspect (e.g. ‘...introduce new research topics.’) of a criterion (e.g. ‘Innovation, Originality’). The scale ranged from 1 to 6, where 1 meant ‘I strongly disagree with the statement’, 2: ‘I disagree’, 3: ‘I slightly disagree’, 4: ‘I slightly agree’, 5: ‘I agree’, and 6: ‘I strongly agree with the statement’. The second round of Delphi was conducted from March to April 2011 by administering an online questionnaire to the whole panel (i.e. 664 scholars).

The questionnaires were administered in English and German. In order to minimize linguistic and cultural bias in the development of the questionnaires, a team approach to translation was taken as is endorsed by Harkness (2003). The team consisted of two native speakers of English with good German skills and two native speakers of German with good English skills. When terms proved difficult to translate, they relied on the cultural decentering approach introduced by Werner and Campbell (1970). In such cases, the team rephrased the terms in the original language until equivalent terms in both languages were found.

3. Results

3.1 First round

Out of 180, 50 questionnaires have been returned which corresponds to an overall response rate of 28%. Among the respondents were 22 scholars of ELS, 18 scholars of GLS, and 10 scholars of AH (corresponding to response rates of 38, 30, and 17%, respectively). Of the respondents, 31 were members of Swiss universities and 19 members of LERU (corresponding to respective response rates of 36 and 20%). Of the respondents, 26 were women and 24 men (corresponding to respective response rates of 34 and 23%). The 49 provided aspects were ticked on average by 67% of the respondents (range: 24–94%); only eight aspects were ticked by <50% of the respondents. This indicates that the respondents found the provided aspects fit or match the criteria. Therefore, none of the provided aspects was dropped for the second Delphi round. However, based on the information given by the respondents, eight aspects were merged reducing the number of aspects to 41. In addition to the provided aspects, the respondents named 29 new aspects. Relying on the comments and suggestions of the respondents, 2 of the 17 provided criteria were merged (*originality* and *innovation*) and three new criteria were added (i.e. *productivity*, *vision of future research*, and *relevance*). This resulted in a comprehensive list of 19 criteria of good research specified by 70 aspects (In Table A.1, the full wording of each aspect

is given). This list formed the basis for the second Delphi round.

3.2 Second round

In total, 196 out of 664 questionnaires have been returned which corresponds to an overall response rate of 30%. Among the respondents were 77 scholars of GLS, 60 scholars of AH and 59 scholars of ELS (corresponding to response rates of 37, 24, and 29%, respectively). Of the respondents, 68 were members of Swiss universities and 128 were members of LERU (corresponding to respective response rates of 43 and 25%). The 71 women and 125 men (corresponding to respective response rates of 27 and 31%) spent on average 22 min answering the online questionnaire.

The second Delphi round was analysed separately for each discipline to preserve the characteristics of each discipline (Table A.2 for detailed results). Scholars of GLS rated the 70 aspects with a mean score of 4.71 on a scale from 1 to 6. The mean scores of the aspects varied between 3.34 and 5.74 (SD = 0.59, median = 4.8). Out of 70 aspects, 10 received average ratings of <4 and two of these aspects had a median of <4 (i.e. $\geq 50\%$ of the GLS’ respondents rated the two aspects negatively). The analysis of AH yielded similar results: 10 aspects received average ratings that were <4, two aspects had a median <4 and the mean score of the aspects was 4.64 (SD = 0.56, Range: 3.15–5.6, median = 4.8). In contrast, scholars of ELS rated 13 aspects on average <4, five aspects had a median <4, while the mean score of the aspects was 4.56 (SD = 0.64, Range: 2.88–5.56, median = 4.7). In all three disciplines, no criterion was rejected altogether (i.e. every criterion had at least one aspect with a mean score >4 as well as a median ≥ 4). However, seven aspects received means <4 in every discipline (name of the respective criterion in parentheses): ‘reputation in society’ and ‘insights are recognized by society’ (*recognition*), ‘continuation of research traditions’, and ‘long-term pursuit of research topics’ (*continuity*, *continuation*), ‘establishing a new school of thought’ (*impact on research community*), ‘responding to societal concerns’ (*relation to and impact on society*), and ‘research has its impact mainly in teaching’ (*connection between research and teaching*, *scholarship of teaching*). The latter aspect was the only one that had a median <4 in every discipline. Summing up, these results indicate that in all three disciplines, every criterion and almost every aspect are seen as appropriate to assess the respondents’ research.

In order to identify aspects that were clearly approved by a majority and disapproved by very few scholars (i.e. consensual aspects), the median as well as the 10th percentile for each aspect were calculated per discipline. Consensus on an aspect was regarded as established in a discipline if at least 50% of the respondents rated an aspect either with a 5 (‘agree’) or a 6 (‘strongly agree’) and if

≤10% rated an aspect with a 1 ('strongly disagree'), 2 ('disagree'), or 3 ('slightly disagree'). In the present sample, every aspect meeting the 10th percentile constraint also met the 50% requirement but not vice versa. Thus, the 10th percentile was the decisive condition for an aspect to be classified as consensual. A criterion was classified as consensual in a discipline if at least one of its aspects met the consensus standard. All the criteria and their aspects are listed in Table 2 with an indication of the discipline in which a given aspect met the consensus standard. Overall, 42 aspects pertaining to 17 criteria reached consensus in at least one discipline. Specifically, 36 aspects pertaining to 16 criteria reached consensus in GLS, 31 aspects connected to 13 criteria did so in AH, and 29 aspects related to 13 criteria in ELS. In the following three sections, criteria shared by all three disciplines will be presented first, followed by discipline-specific criteria and by non-consensual criteria.

3.3 Shared criteria

In total, 11 criteria are shared by all three disciplines (i.e. each of these criteria has consensual aspects in every discipline). Among these criteria are five which are commonly used for assessing or reviewing research (i.e. *scholarly exchange*, *impact on research community*, *innovation/originality*, *rigour*, and *connection to other research*). However, other frequently applied criteria are missing, such as *relevance*, reputation (subsumed under the criterion *recognition* in the present study), *relation to and impact on society*, *productivity*, or promotion of young researchers (subsumed under the criterion *continuity/continuation* in the present study). Besides the five well-known criteria, six criteria reach consensus in all three disciplines that are either new or at least not commonly used to assess research, these are: *fostering cultural memory*, *scholarship/erudition*, *openness to ideas and persons*, *passion/enthusiasm*, *connection between research and teaching/scholarship of teaching*, and *vision of future research*. Among this set of six criteria, one would also expect *reflection/criticism* or *relation to and impact on society* because they are both said to be characteristic of the humanities (e.g. Fisher et al. 2000; Luckman 2004; Nussbaum 2010; Oancea and Furlong 2007), but these two criteria do not reach consensus in all three disciplines simultaneously. However, one has to be aware that the 11 shared criteria vary in each discipline on the level of aspects. For example, *scholarly exchange* has a common denominator in *exchange within a discipline* (aspect consensual in GLS, ELS, and AH) but differs in *interdisciplinary exchange* (aspect consensual in GLS and AH) and *international exchange* (aspect consensual in ELS and AH). Table 3 lists all aspects in their full wording that are common denominators of the three studied disciplines (i.e. aspects consensual in GLS, ELS, and AH simultaneously). These 20 aspects pertaining to nine criteria can be considered the core set of shared criteria/aspects in our

study. In this core set, the shared criteria *connection to other research* and *connection between research and teaching/scholarship of teaching* are not included because they do not exhibit aspects that are simultaneously consensual in all three disciplines. The core set can be summarized in a simplistic manner as 'shared values of good research' as follows:

Participate in the scholarly exchange of your discipline. Be innovative and original in choosing topics, approaches, and in your interpretations. Identify gaps in existing knowledge and generate new paradigms. Work with rigour, that is, use stringent and clear language as well as a clear structure, reflect on your method and present relevant evidence. Renew the understanding of the past and stimulate new research. Have a vision of future research. Be open to ideas and persons different from yourself. Gain rich experience and knowledge by reading and reflecting thoroughly as well as by studying material comprehensively. Finally, share your passion for research and arouse it in others.

3.4 Discipline-specific criteria

Six criteria reached consensus in one or two disciplines (i.e. each of these criteria has consensual aspects in only one or two disciplines). The discipline-specific criteria are (discipline in which a criterion is consensual in parenthesis): *recognition* (ELS), *continuity/continuation* (GLS), *relevance* (GLS), *variety of research* (GLS, AH), *reflection/criticism* (GLS, AH), and *self-management/independence* (GLS, ELS). In contrast to the first three criteria, the last three are, to our knowledge, not employed in evaluation schemes. Considering the discipline-specific criteria, it is obvious that research in ELS, GLS and AH cannot be assessed appropriately solely by relying on the shared criteria presented above. This is also evident on the level of aspects where the three disciplines share a core set of 20 aspects (Table 3) but differ on 22 discipline-specific aspects (i.e. aspects that are consensual in only one or two disciplines). These specific aspects can be summarized for each discipline as 'discipline-specific values of good research' completing the 'shared values of good research' presented above: GLS (16 discipline-specific aspects):

Participate in interdisciplinary exchange. Make your research inter-subjectively verifiable and refutable. Foster cultural memory by documenting aspects of the past. Be self-reflective in your research, look for differences and make complexity visible. Promote young scholars. Contribute towards variety and take risks. Connect to other research by building on the current state of research as well as by re-connecting to neglected research or by engaging in ongoing debates. Realize your own chosen research goals but be open-ended as to the results. Let teaching inspire and influence your research. Work on something that is relevant for the scholarly community. AH (11 discipline-specific aspects): Engage in

Table 2. All criteria and their aspects as well as an indication of the discipline in which a given aspect meets the consensus standard (GLS, ELS and AH)

Criterion	Aspect (abbreviated)	Consensual in discipline		
Scholarly exchange	Disciplinary exchange	GLS	ELS	AH
	Interdisciplinary exchange	GLS		AH
	International exchange		ELS	AH
Innovation, originality	New data or novel combination of data		ELS	AH
	Introduction of new research topics	GLS	ELS	AH
	New approach to topic or data	GLS	ELS	AH
	Generating new paradigms	GLS	ELS	AH
	Contribution of new findings or interpretations	GLS	ELS	AH
	Innovative language or depiction			
	Identification of gaps in existing knowledge	GLS	ELS	AH
Productivity Rigour	Continuous research outputs			
	Systematic and transparent research process			
	Stringent argumentation	GLS	ELS	AH
	Presentation of relevant documents and evidence	GLS	ELS	AH
	Clear language	GLS	ELS	AH
	Clear structure	GLS	ELS	AH
	Reflection of method	GLS	ELS	AH
	Intersubjectivity	GLS		AH
	Adherence to rules of scientific honesty	GLS	ELS	AH
	Discussion of generalizability of insights			
	Reflection of personal relation to research topic			
	Fostering cultural memory	Documentation of aspects of the past	GLS	
Renewal of understanding of aspects of the past		GLS	ELS	AH
Putting the past in relation to the present				AH
Insights into perennial aspects of human nature				
Recognition	Insights are recognized by the research community		ELS	
	Insights are recognized by society			
	Reputation within research community			
	Reputation in society			
	Reputation at own university			
Reflection, criticism	Looking for distinctions	GLS		
	Deconstruction of the illusion of ‘definitive and final truth’			
	Criticizing assertive claims and social norms			
	Criticizing established scholarly approaches			
	Self-critical and self-reflective research	GLS		
Continuity, continuation	Visualizing complexity	GLS		AH
	Promotion of young academics	GLS		
	Continuation of research traditions			
	Long-term pursuit of research topics			
Impact on research community	Stimulating new research	GLS	ELS	AH
	Concluding a debate			
	Establishing a new school of thought			
	Influencing the research community		ELS	
Relation to and impact on society	Topics relevant for society from the scholars’ perspective			
	Responding to societal concerns			
	Conveying findings to a non-academic audience			
	Affecting national or local culture			
Variety of research	Contributing towards variety and diversity	GLS		AH
	Taking risks and working outside of mainstream	GLS		AH
Connection to other research	Building on current state of research	GLS	ELS	
	Re-connecting to neglected research	GLS	ELS	
	Engaging in ongoing research debates	GLS		AH
Openness to Ideas and Persons	Openness to other ideas	GLS	ELS	AH
	Openness to other persons	GLS	ELS	AH

(continued)

Table 2. Continued

Criterion	Aspect (abbreviated)	Consensual in discipline		
		GLS	ELS	AH
Self-management, independence	Realization of own research goals	GLS	ELS	
	Research outcomes are unpredictable	GLS	ELS	
	Research is not directly utilizable			
	Research is not directly targeted at a recipient			
Scholarship, erudition	Rich experience with sources	GLS	ELS	AH
	Knowledge based on own research	GLS	ELS	AH
Passion, enthusiasm	Passionate about research			
	Arouse passion for research	GLS	ELS	AH
	Intrinsic motivation for research activity			
Vision of future research	Pointing out important research for the future	GLS	ELS	AH
Connection between research and teaching, scholarship of teaching	Research-based teaching		ELS	AH
	Teaching-based research	GLS		
	Research has its impact mainly in teaching			
	Building character of oneself and others			
Relevance	Social competency			
	Research is relevant for the research community	GLS		

Table 3. Aspects consensual in all three disciplines (GLS, ELS, and AH)

Criterion	Aspect (full wording)
	'My research is assessed appropriately, if the assessment considers whether ...'
Scholarly exchange Innovation, originality	I participate in discourse regarding my field (exchange within the discipline).
	I introduce new research topics.
	I use a new approach ^a to a research topic or to data.
	I generate new paradigms, hypotheses, theories or methods, or open up new areas of research.
	I produce new findings or contribute new interpretations (knowledge gain within existing paradigms, theories, methods, research areas, established research questions, or hypotheses).
Rigour	I draw attention to questions and areas of research that previous research has ignored, or identify gaps in existing knowledge.
	I present stringent ^b , comprehensible ^c , and convincing arguments.
	I present the relevant documents and evidence (cite sources, document the material, furnish primary texts or original works).
	I use clear and understandable language.
	My texts or presentations have a clear structure.
Fostering cultural memory Impact on research community Openness to ideas and persons	I reflect on my method or choice of method.
	I adhere to the rules of scientific honesty.
	I renew the understanding and the interpretation of aspects of the past.
Scholarship, erudition	I stimulate new research.
	I am open to other, competing ideas, approaches, theories or methods, or acknowledge them.
	I am open to other persons in research and do not exclude them (such as students, junior scholars, emeritus researchers, researchers who think differently, researchers at other institutions/in other countries).
Passion, enthusiasm Vision of future research	I make use of my rich experience with sources, materials, original works (knowledge of the material).
	I make use of rich knowledge and insights that I have obtained through my own research and reflection (my own extensive research, reception, reading or reflection, my own sound, and critical analysis of important works).
	I arouse passion or enthusiasm for the research or the research topic.
	I point out and anticipate research, which may be important for the future.

^aApproach: for example, new/different perspective, new method, new theory, new research question, new interpretation, new hypothesis, different research tradition, different historical context, different cultural phenomena.

^bStringent: for example, coherent, conclusive, of compelling logic and faultless, clear and logical development of the hypothesis, the steps of the investigation are logically ordered, and coordinated.

^cComprehensible: for example, the steps in the argumentation are transparent and verifiable, the findings produced are understandable at all times and based on the explanation of research question and method.

international and interdisciplinary exchange. Discover new material, identify not yet discussed sources and put material together in a new way. Make your research inter-subjectively verifiable and refutable. Foster cultural memory by documenting aspects of the past and by putting the past in relation to the present. Make complexity visible. Contribute towards variety and take risks. Connect to other research by engaging in ongoing debates. Make your research a central part of your teaching. ELS (nine discipline-specific aspects): Engage in international exchange. Discover new material, identify little-known/not yet discussed sources and put material together in a new way. Strive for your insights and research to find acknowledgement and recognition in your community. Try to have an influence on the research community. Connect to other research by building on the current state of research or by re-connecting to neglected research. Realize your own chosen research goals but be open-ended as to the results. Make your research a central part of your teaching.

3.5 Non-consensual criteria

Two criteria are not consensual in any of the three disciplines: *productivity* (i.e. continually or periodically generating research output) and *relation to and impact on society* (i.e. responding to societal concerns, studying topics relevant to society from the scholars' perspective, conveying findings to a non-academic audience, affecting local/regional/national culture). One possible reason why *productivity* does not reach consensus might be that scholars are not striving for continual or periodical output but rather for a few big intellectual achievements. This is in line with Zuccala (2012: 231) who concludes regarding the evaluation of literary work: 'A literary artist or critic may write two to five major works in his or her lifetime, which could influence the thinking and imagination of many generations later. With that said, it is not the *quantity* of performances (publications) that matter as much in literary work but the type of *quality conversations* associated with a published piece and how much *influence* it has over time.' However, it has to be emphasized that productivity is an important criterion in many existing evaluation schemes and social impact is seen as one of the characteristics of the humanities (see, e.g. Hellqvist, 2010; Weingart et al. 1991). Moreover, social impact has lately been included in evaluation schemes which also cover the humanities, for example, the UK's Research Excellence Framework (REF; Higher Education Funding Council for England, 2012), Netherland's Standard Evaluation Protocol (SEP; Royal Netherlands Academy of Arts and Sciences, Association of Universities in the Netherlands and Netherlands Organisation for Scientific Research 2009) or Germany's Research Rating (Wissenschaftsrat 2011b).

4. Discussion

This is the first empirical study that establishes criteria for assessing research quality in the humanities entirely from within the humanities. In order to investigate scholars' quality criteria in GLS, ELS, and AH, we have conducted a multi-round Delphi survey in all Swiss and LERU universities that cover the three disciplines. The first Delphi round resulted in a comprehensive list of 19 criteria of good research specified by 70 aspects. Although 10 of these criteria are well known and commonly used in various evaluation schemes, the following nine criteria are not, or at least not frequently employed, namely: *fostering cultural memory*, *reflection/criticism*, *variety of research*, *openness to ideas and persons*, *self-management/independence*, *scholarship/erudition*, *passion/enthusiasm*, *vision of future research*, *connection between research and teaching/scholarship of teaching*. Two of these criteria are also mentioned in the empirical literature on quality criteria in the humanities, they are *reflection/criticism* corresponding to *reflexivity*, *deliberation and criticism* (Oancea and Furlong 2007) and *passion/enthusiasm* corresponding to *engagement* (Bazeley 2010). The second Delphi round showed that all 19 criteria are appropriate to assess the scholars' own research as only one of the 70 aspects (i.e. 'research has its impact mainly in teaching') was rated negatively by >50% of the scholars in every discipline. These results indicate that the catalogue of criteria and aspects elaborated in this study aptly reflects the scholars' understanding of research quality. Furthermore, the response rates indicate that the first as well as the second Delphi round were received surprisingly well by scholars (response rates of 28 and 30%, respectively) despite the contested issue. Similar studies that have surveyed academics state considerably lower (e.g. Rosa et al. 2012), lower (e.g. Frey et al. 2007; Braun and Ganser 2011;), or similar response rates (e.g. Giménez-Toledo et al. 2013).

In order to determine the criteria for assessing research quality that finds acceptance among scholars, we have identified consensual criteria in each discipline. A criterion was classified as consensual in a discipline if at least one of its aspects was clearly approved by a majority and disapproved by very few scholars. Accordingly, 11 criteria reached consensus in all three disciplines (shared criteria), six criteria were consensual in only one or two disciplines (discipline-specific criteria) and two criteria were not consensual in any discipline (i.e. 'productivity' and 'relation to and impact on society'). Hence, our results show that consensus on a set of quality criteria is possible within a given humanities discipline and that there are consensual criteria shared between different humanities' disciplines.

With respect to research assessment, our results have three implications: (1) a broad spectrum of criteria has to be applied to assess research quality in the humanities appropriately; (2) even though we have identified shared

criteria, humanities disciplines cannot be lumped together for the purpose of research assessment, because each of the three studied disciplines has its own specific set of criteria and aspects; and (3) our study shows that not all criteria commonly used in evaluation schemes are shared among humanities scholars (e.g. social impact, reputation, productivity) and that scholars have knowledge of criteria either not known or not frequently applied in research evaluation (e.g. *fostering cultural memory, reflection/criticism, scholarship/erudition*). Thus, scholars and evaluators (university administrators, funders, evaluation specialists, etc.) seem to have different views on assessing research quality in the humanities.

The last implication is in line with Hellström (2010) who has identified a mismatch of evaluation standards between evaluators and researchers in artistic research. Moreover, there are indications coming from studies on Mode 2-related assessment criteria that such mismatches between evaluators and researchers might also exist in further research fields (as for Mode 2, see Gibbons et al. 1994; Nowotny et al. 2001). On the evaluators' side, a shift towards Mode 2-inspired criteria, such as economic utility or social impact, took place [e.g. in UK's REF, Netherland's SEP, Germany's Research Rating; see also Hemlin and Rasmussen (2006)]. However, social and natural scientists seem not to have adopted Mode 2 criteria (Gulbrandsen and Langfeldt 2004; Albert et al. 2012), which is consistent with the findings of this study. Accordingly, Donovan (2011) speaks about a gulf between the research evaluation community and academics regarding social impact and concludes that there has been too little engagement between evaluation specialists and the broader academic community, which has led to social impact being dubbed the 'height of philistinism' by academics. Furthermore, Vanholsbeeck (2012) reported that Belgian scholars in the field of media and communication sciences have an ambivalent attitude towards the managerial prescriptions ('prescribed quality') relating to the assessment of scholarly publications, which fit only partially into their own definitions of 'quality publication' ('desired quality'). In view of these findings, effects caused by a mismatch of criteria between evaluators and researchers have to be considered when assessing research (e.g. negative steering effects or the potential for conflict inherent to differing views on criteria). Clearly, more research on such mismatches and their effects is needed.

One way to bridge mismatches and to establish a shared view on assessing research between evaluators and researchers would be to make use of 'forums' as put forward by Barré (2010). Barré argues that the definition and interpretation of any indicator is highly debatable and, therefore, need to be validated by the relevant stakeholders in an interactive forum that ensures 'social robustness'. Characteristics of a forum are collective learning capacity, openness, and accountability. It is our opinion that, in order to produce social robustness, not only

research evaluation specialists, research funders and science policy makers should be represented in forums, but also stakeholders from the public, economic and private sector and, naturally, academics themselves, including those 'working at the coal-face of research' as recommended by LERU (2012b). We suggest analysing each stakeholder along the lines of our framework (Hug et al., in press) and to inform forums with the results of these analyses as the framework addresses many points discussed by Barré (2010), especially those related to the collective learning capacity of forums.

Finally, two limitations of this study need to be considered. First, the quality criteria and aspects presented (as well as their ratings) are restricted to GLS, ELS and AH. Second, the ratings of the criteria's aspects are based on the answers of one-third of the panel, which comprised scholars from LERU and Swiss universities. Thus, studies in other humanities disciplines are needed as well as the involvement of more scholars from these and other universities. Furthermore, indicators to measure criteria and their respective aspects need to be determined and assessed in future studies.

Acknowledgements

The authors thank Prof. Dr Christian Simon, Dr John Bendix, and Vanessa McSorley for their contributions to the Delphi study. The authors would also like to thank Esther Germann for proofreading.

Funding

This work was supported by the Rectors' Conference of the Swiss Universities (CRUS) within the framework of the SUK B-05 Innovation and Cooperation Project 'Mesurer les performances de la recherche' (Measuring Research Performance) and by the University of Zurich. It is part of a cooperative project of the Universities of Zurich and Basel entitled 'Developing and Testing Research Quality Criteria in the Humanities, with an emphasis on Literature Studies and Art History'.

References

- Albert, M., Laberge, S. and McGuire, W. (2012) 'Criteria for Assessing Quality in Academic Research: The Views of Biomedical Scientists, Clinical Scientists and Social Scientists', *Higher Education*, 64/5: 661–76.
- Andersen, H. et al. (2009) 'Editorial: Journals under threat: a joint response from History of Science, Technology and Medicine Editors', *Social Studies of Science*, 39/1: 6–9.
- Australian Research Council. (2012) *The Excellence in Research for Australia (ERA) Initiative*, <<http://www.arc.gov.au/era/>> accessed 30 January 2012.
- Barré, R. (2010) 'Towards Socially Robust S&T Indicators: Indicators as Debatable Devices, Enabling Collective Learning', *Research Evaluation*, 19/3: 227–32.

- Bazeley, P. (2010) 'Conceptualising Research Performance', *Studies in Higher Education*, 35/8: 889–903.
- Braun, N. and Ganser, C. (2011) 'Fundamental Knowledge of Sociology? A Written Survey of Male and Female Professors of German Sociology and its Results', *Soziologie*, 40/2: 151–74.
- Brooks, R. L. (2005) 'Measuring University Quality', *The Review of Higher Education*, 29/1: 1–21.
- Delbecq, A. L., Van de Ven, A. and Gustafson, D. H. (1975) *Group Techniques for Program Planning. A Guide to Nominal Group and Delphi Processes*. Glenview: Scott, Foresman.
- Donovan, C. (2007) 'The Qualitative Future of Research Evaluation', *Science and Public Policy*, 34/8: 585–97.
- . (2008) 'Das zweiköpfige Lama zähmen: Die australische Suche nach den besten Evaluierungsmethoden für die Geisteswissenschaften [Taming the Pushmi-pullyu: State of the Art in Evaluating Humanities Research Quality - A View from Down Under]'. In: Lack, E. and Marksches, C. (eds) *What the Hell is Quality? Qualitätsstandards in den Geisteswissenschaften*, pp. 74–98. Frankfurt/New York: Campus.
- . (2011) 'State of the Art in Assessing Research Impact: Introduction to a Special Issue', *Research Evaluation*, 20/3: 175–9.
- EERQI Consortium. (2011) *EERQI - European Educational Research Quality Indicators*, <<http://www.eerqi.eu/>> accessed 30 January 2012.
- Engels, T. C. E., Ossenblok, T. L. B. and Spruyt, E. H. J. (2012) 'Changing Publication Patterns in the Social Sciences and Humanities, 2000–2009', *Scientometrics*, 93/2: 373–90.
- European Science Foundation. (2011) *European Reference Index for the Humanities (ERIH)*, <<http://www.esf.org/research-areas/humanities/erih-european-reference-index-for-the-humanities.html>> accessed 30 January 2012.
- Evaluation of Scientific Publications Research Group. (2010) *Difusión y Calidad Editorial de las Revistas Españolas de Humanidades y Ciencias Sociales y Jurídicas (DICE) [Diffusion and Editorial Quality of Spanish Journals in the Humanities, Social Sciences and Law]*, <<http://dice.cindoc.csic.es/>> accessed 21 June 2012.
- . (2012) *Revistas Españolas de Ciencias Sociales y Humanidades [Spanish Journals in Social Sciences and Humanities]*, (RESH) <<http://epuc.cchs.csic.es/resh/>> accessed 21 June 2012.
- Fisher, D., Rubenson, K., Rockwell, K., Grosjean, G. and Atkinson-Grosjean, J. (2000) *Performance Indicators and the Humanities and Social Sciences*. Centre for Policy Studies in Higher Education and Training, <<http://www.fedcan.virtuo.ca/pdf/364.pdf>> accessed 18 December 2012.
- Frey, B. S., Humbert, S. and Schneider, F. (2007) 'Was denken deutsche Ökonomen? Eine empirische Auswertung einer Internetbefragung unter den Mitgliedern des Vereins für Socialpolitik im Sommer 2006 [What do German Economists Think? Results of an Online Survey Among the Members of the Association of German Speaking Economists]', *Perspektiven der Wirtschaftspolitik*, 8/4: 359–77.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. and Trow, M. (1994) *The new Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. Thousand Oaks: Sage.
- Gimenez-Toledo, E., Tejada-Artigas, C. and Manana-Rodríguez, J. (2013) 'Evaluation of Scientific Books' Publishers in Social Sciences And Humanities: Results of a Survey', *Research Evaluation*, 22/1: 64–77.
- Gogolin, I. (2011) *European Educational Research Quality Indicators. Project Final Report*, <http://eerqi.eu/sites/default/files/Final_Report.pdf> accessed 6 December 2012.
- Guetzkow, J., Lamont, M. and Mallard, G. (2004) 'What is Originality in the Humanities and the Social Sciences?', *American Sociological Review*, 69/2: 190–212.
- Gulbrandsen, M. and Langfeldt, L. (2004) 'In Search of Mode 2: The Nature of Knowledge Production in Norway', *Minerva*, 42/3: 237–50.
- Häder, M. and Häder, S., eds (2000) 'Die Delphi-Methode als Gegenstand methodischer Forschung [The Delphi Method as a Subject in Methodical Research]', *Die Delphi-Technik in den Sozialwissenschaften. Methodische Forschungen und innovative Anwendungen*, pp. 11–31. Wiesbaden: Westdeutscher.
- Harkness, J.A. (2003) 'Questionnaire Translation'. In: Harkness, J. A., van de Vijver, F. and Mohler, P.P. (eds) *Cross-Cultural Survey Methods*, pp. 35–56. Hoboken, NJ: Wiley.
- Hellqvist, B. (2010) 'Referencing in the Humanities and its Implications for Citation Analysis', *Journal of the American Society for Information Science and Technology*, 61/2: 310–8.
- Hellström, T. (2010) 'Evaluation of Artistic Research', *Research Evaluation*, 19/5: 306–16.
- Hemlin, S. (1993) 'Scientific Quality in the Eyes of the Scientist. A Questionnaire Study', *Scientometrics*, 27/1: 3–18.
- . (1996) 'Social Studies of the Humanities. A Case Study of Research Conditions and Performance in Ancient History and Classical Archaeology and English', *Research Evaluation*, 6/1: 53–61.
- Hemlin, S. and Gustafsson, M. (1996) 'Research Production in the Arts and Humanities. A Questionnaire Study of Factors Influencing Research Performance', *Scientometrics*, 37/3: 417–32.
- Hemlin, S. and Rasmussen, S. B. (2006) 'The Shift in Academic Quality Control', *Science Technology & Human Values*, 31/2: 173–98.
- Higher Education Funding Council for England. (2012) *Panel Criteria and Working Methods*, Bristol, <http://www.ref.ac.uk/media/ref/content/pub/panelcriteriaandworkingmethods/01_12.pdf> accessed 9 September 2012.
- Hsu, C. C. and Sandford, B. A. (2007) 'The Delphi Technique. Making Sense of Consensus', *Practical Assessment, Research and Evaluation*, 12/10: 1–8.
- Hug, S. E., Ochsner, M. and Daniel, H.-D. (in press) 'A Framework to Explore and Develop Criteria for Assessing Research Quality in the Humanities', *International Journal of Education Law and Policy*.
- ISO 32000-1. (2008) *Document Management. Portable Document Format. Part 1: PDF 1.7*, <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=51502> accessed 07 July 2012.
- Lahtinen, E., Koskinen-Ollonqvist, P., Rouvinen-Wilenius, P., Tuominen, P. and Mittelmark, M.B. (2005) 'The Development of Quality Criteria for Research: A Finnish Approach', *Health Promotion International*, 20/3: 306–15.
- Lamont, M. (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge: Harvard University Press.
- Landeta, J. (2006) 'Technological Forecasting and Social Change. Current validity of the Delphi method in social sciences', *Technological Forecasting and Social Change*, 73/5: 467–82.
- League of European Research Universities (LERU). (2012a) *League of European Research Universities*, <<http://www.leru.org/index.php/public/home/>> accessed 18 May 2012.
- . (2012b) *Research Universities and Research Assessment Leuven*, <http://www.leru.org/files/publications/LERU_PP_2012_May_Research_Assesment.pdf> accessed 18 December 2012.

- Linstone, H. A. and Turoff, M., eds (1975) 'Introduction', *The Delphi Method. Techniques and Applications*, pp. 3–12. Don Mills: Addison-Wesley.
- Luckman, S. (2004) 'More Than the Sum of its Parts: The Humanities and Communicating the "Hidden Work" of Research'. In: Kenway, J., Bullen, E. and Robb, S. (eds) *Innovation and Tradition. The Arts, Humanities, and the Knowledge Economy*, pp. 82–90. New York: Peter Lang.
- Murry, J. W. and Hammons, J. O. (1995) 'Delphi. A Versatile Methodology for Conducting Qualitative Research', *Review of Higher Education*, 18/4: 423–36.
- Nadeau, G. G. (1995) *Critères et indicateurs de qualité et d'excellence dans les collèges et les universités du Canada: Bilan de trois phases du projet* [Criteria and Indicators of Quality and Excellence in Colleges and Universities in Canada: Summary of the Three Phases of the Project]. Winnipeg: Centre for Higher Education Research and Development, University of Manitoba, Canada.
- National Science Foundation. (2009) MESUR project, <<http://mesur.informatics.indiana.edu/>> accessed 30 January 2012.
- Nowotny, H., Scott, P. and Gibbons, M. (2001) *Re-thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge, UK: Polity.
- Nussbaum, M. C. (2010) *Not for Profit. Why Democracy Needs the Humanities*. Princeton, NJ: Princeton University Press.
- Oancea, A. and Furlong, J. (2007) 'Expressions of excellence and the assessment of applied and practice-based research', *Research Papers in Education*, 22/2: 119–37.
- Ochsner, M., Hug, S. E. and Daniel, H.-D. (2012) 'Four Types of Research in the Humanities: Setting the Stage for Research Quality Criteria in the Humanities', *Research Evaluation*, 22/2: 79–92.
- Palomares-Montero, D. and Garcia-Aracil, A. (2011) 'What are the Key Indicators for Evaluating the Activities of Universities?', *Research Evaluation*, 20/5: 353–63.
- Peric, B., Ochsner, M., Hug, S. E. and Daniel, H.-D. (2013) *Arts and Humanities Research Assessment Bibliography (AHRABi)*, <<http://www.psh.ethz.ch/crus/bibliography/ahrabi.pdf>> accessed 21 July 2013.
- Plumpe, W. (2009) 'Stellungnahme zum Rating des Wissenschaftsrates aus Sicht des Historikerverbandes [Comment on the Research Rating of the German Council of Science and Humanities from the Perspective of the Association of German Historians]'. In: Prinz, C. and Hohls, R. (eds) *Qualitätsmessung, Evaluation, Forschungsrating. Risiken und Chancen für die Geschichtswissenschaften? Historisches Forum, Bd. 12*, pp. 121–6. Berlin: Clio-online und Humboldt-Universität zu Berlin.
- Pontille, D. and Torny, D. (2010) 'The controversial policies of journal ratings: evaluating social sciences and humanities', *Research Evaluation*, 19/5: 347–60.
- . (2012) 'Rendre publique l'évaluation des SHS: les controverses sur les listes de revues de l'AERES', *Quaderni*, 77: 11–24.
- Rosa, M. J., Sarrico, C. S. and Amaral, A. (2012) 'Academics' perceptions on the purposes of quality assessment', *Quality in Higher Education*, 18/3: 349–66.
- Royal Netherlands Academy of Arts and Sciences, Association of Universities in the Netherlands and Netherlands Organisation for Scientific Research. (2009) *Standard Evaluation Protocol 2009–2015. Protocol for Research Assessment in the Netherlands*, <<http://www.knaw.nl/smart>> accessed 9 September 2012.
- Royal Netherlands Academy of Arts and Sciences. (2011) *Quality Indicators for Research in the Humanities*. Amsterdam: Royal Netherlands Academy of Arts and Sciences, <http://www.knaw.nl/Content/Internet_KNAW/publicaties/pdf/20111024.pdf> accessed 18 December 2012.
- Schneider, J. W. (2009) 'An Outline of the Bibliometric Indicator Used for Performance-Based Funding of Research Institutions in Norway', *European Political Science*, 8/3: 364–78.
- Sivertsen, G. (2010) 'A Performance Indicator Based on Complete Data for the Scientific Publication Output at Research Institutions', *ISSI Newsletter*, 6/1: 22–8.
- Thomson Reuters. (2011) 'The Book Citation Index', <http://wokinfo.com/products_tools/multidisciplinary/bookcitationindex/> accessed 30 January 2012.
- Vanholsbeeck, M. (2012) 'Entre qualité prescrite et qualité souhaitable: l'ambivalence des chercheurs en communication face à l'évaluation de leurs publications [Between Prescribed Quality and Desired Quality: The Ambivalence of Scholars in Media and Communication Sciences Regarding the Evaluation of Their Publications]', *Quaderni*, 77: 71–84.
- Verband der Historikerinnen und Historiker Deutschlands. (2010) 'Ausschuss des Historikerverbandes beriet über Forschungsratings in den Geisteswissenschaften und über den Historikertag 2010 [The Committee of the Association of Historians discussed the Research Rating in the Humanities and the Meeting of Historians]', <http://www.historikerverband.de/fileadmin/_vhd/bilder/Pressemitteilung_WR_Rating.pdf> accessed 30 January 2012.
- Verleyen, F. T. and Engels, T. C. E. (2013) 'A Label for Peer Reviewed Books', *Journal of the American Society for Information Science and Technology*, 64/2: 428–30.
- Weingart, P., Prinz, W., Kastner, M., Maasen, S. and Walter, W. (1991) *Die sogenannten Geisteswissenschaften: Aussenansichten: Die Entwicklung der Geisteswissenschaften in der BRD, 1954-1987 [The So-Called Humanities: View from Outside: The Development of the Humanities in the FRG, 1954-1987]*. Frankfurt am Main: Suhrkamp.
- Werner, O. and Campbell, D. T. (1970) 'Translating, Working Through Interpreters, and the Problem of Decentering'. In: Naroll, R. and Cohen, R. (eds) *A Handbook of Cultural Anthropology*, pp. 303–9. Hague: Mouton.
- White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, C. S. and Cole, F. T. H. (2009) 'Libcitations: A Measure for Comparative Assessment of Book Publications in the Humanities and Social Sciences', *Journal of the American Society for Information Science and Technology*, 60/6: 1083–96.
- Wissenschaftsrat. (2011a) *Forschungsrating Anglistik/Amerikanistik*, <<http://www.wissenschaftsrat.de/arbeitsbereiche-arbeit-sprogramm/forschungsrating/anglistikamerikanistik/>> accessed 30 January 2012.
- . (2011b) *Zum Forschungsrating allgemein*, <<http://www.wissenschaftsrat.de/arbeitsbereiche-arbeitsprogramm/forschungsrating/>> accessed 30 January 2012.
- Wooding, S. and Grant, J. (2003) *Assessing Research. The Researchers' View*. Santa Monica, CA: RAND Corporation.
- Zuccala, A. (2012) 'Quality and Influence in Literary Work: Evaluating the Educated Imagination', *Research Evaluation*, 21/3: 229–41.

Appendix

Table A1. All criteria and their aspects in full wording

Criterion	Full aspect
	'My research is assessed appropriately, if the assessment considers whether...'
Scholarly exchange	<p>I participate in discourse regarding my field (exchange within the discipline).</p> <p>I participate in discourse beyond the boundaries of my own field (interdisciplinary exchange).</p> <p>I participate in discourse in an international context (international exchange).</p>
Innovation, originality	<p>I am innovative with regard to the data^a (discover new data, identify little known, neglected, or not yet discussed data, put data together in a new way).</p> <p>I introduce new research topics.</p> <p>I use a new approach^b to a research topic or to data.</p> <p>I generate new paradigms, hypotheses, theories or methods, or open up new areas of research.</p> <p>I produce new findings or contribute new interpretations (knowledge gain within existing paradigms, theories, methods, research areas, established research questions, or hypotheses).</p> <p>I am innovative with regard to language or depiction (depict or explain the research in an original way, explain established knowledge in a new way, push the boundaries of what is sayable).</p> <p>I draw attention to questions and areas of research that previous research has ignored, or identify gaps in existing knowledge.</p>
Productivity	I continually or periodically generate research output.
Rigour	<p>My research process (planning, organization, implementation of the research plan) is systematic and understandable to experts.</p> <p>I present stringent^c, comprehensible^d, and convincing arguments.</p> <p>I present the relevant documents and evidence (cite sources, document the material, furnish primary texts or original works).</p> <p>I use clear and understandable language.</p> <p>My texts or presentations have a clear structure.</p> <p>I reflect on my method or choice of method.</p> <p>My research is intersubjectively verifiable and refutable.</p> <p>I adhere to the rules of scientific honesty.</p> <p>I explain whether and to what extent my insights and materials are representative, and discuss their generalizability.</p> <p>I reflect on my own relation to the research topic.</p>
Fostering cultural memory	<p>I document, preserve, archive or keep alive aspects of the past (including cultural legacy).</p> <p>I renew the understanding and the interpretation of aspects of the past.</p> <p>I put the past in relation to the present.</p> <p>I provide insights into essential and perennial aspects of human nature.</p>
Recognition	<p>My insights or research find acknowledgement, agreement, or recognition in the research community.</p> <p>My insights or research find acknowledgement, agreement, or recognition in society (i.e. outside the universities).</p> <p>I have a good reputation within the research community.</p> <p>I have a good reputation in society (i.e. outside the universities).</p> <p>I have a good reputation in my own institute or at my own university.</p>
Reflection, criticism	<p>I look for differences or distinctions.</p> <p>I deconstruct the illusion of 'definitive and final truth' or do not allow the illusion to arise at all (break down the 'false idyll of unambiguousness', show that there are no unambiguous and fixed interpretations, have or create an awareness of the provisional nature of all knowledge and insights).</p> <p>I criticize assertive claims, things that are a matter of course, social norms or things that are fashionable in the present society.</p> <p>I criticize established scholarly approaches, dominant paradigms, basic assumptions, sources, theories, interpretations, traditions, ideologies, prevailing schools, stereotypes, or research practices.</p> <p>My research is self-critical and self-reflective (e.g. I reflect on my own historical and social viewpoint, I reflect on my own level of abstraction, I reflect on the way of generating findings, I reflect on the language I use, I reflect on value criteria and premises, I recognize limitations).</p> <p>I make complexity visible (e.g. I work against the trend towards reducing complexity, I show the ambivalence, ambiguity, or complexity of a research topic, I take the many facets of a research question seriously).</p>

(continued)

Table A1. Continued

Criterion	Full aspect 'My research is assessed appropriately, if the assessment considers whether ...'
Continuity, continuation	I ensure that there are well-trained junior scholars. I continue research traditions. I deal over the long term with the same research topics, the same research questions, the same theories, or the same methods.
Impact on research community	I stimulate new research. I bring a debate or discussion to a conclusion. I form a new school of thought or establish a research tradition. I have an impact (of any kind) or influence on the research community.
Relation to and impact on society	I deal with topics that, from the researchers' point of view, are relevant for society. I respond to societal concerns and needs as stated by society (e.g. I participate in current, socially important discussions, I respond to central questions of present-day policy, and social discussions). I make my findings understandable to a non-academic audience using understandable language. My research has an effect on national, regional, or local culture/society.
Variety of research	I contribute towards diversity and variety (regarding research topics, approaches, theories, methods, data, and so on). I take risks or work outside the mainstream/the prevailing fashions.
Connection to other research	My research builds on the current state of research. My research connects or re-connects to older, neglected or over-looked research or debates. I engage in ongoing research debates.
Openness to ideas and persons	I am open to other competing ideas, approaches, theories or methods, or acknowledge them. I am open to other persons in research and do not exclude them (such as students, junior scholars, emeritus researchers, researchers who think differently, researchers at other institutions/in other countries).
Self-management, independence	I realize my own chosen research goals. My research is open-ended as to the results (the findings, product or output are not predictable). I conduct research that is not directly utilizable or applicable. I conduct research that is not aimed at any direct target groups or recipients and that can unfold its potential at an unspecified time in the future.
Scholarship, erudition	I make use of my rich experience with sources, materials, original works (knowledge of the material). I make use of rich knowledge and insights that I have obtained through my own research and reflection (my own extensive research, reception, reading or reflection, my own sound, and critical analysis of important works).
Passion, enthusiasm	I have a passion or enthusiasm for the research or the research topic. I arouse passion or enthusiasm for the research or the research topic. I have intrinsic motivation (i.e. for me the only incentive to do the research is the research activity itself).
Vision of future research	I point out and anticipate research, which may be important for the future.
Connection between research and teaching, scholarship of teaching	My research is a central component in my teaching or at least enriches my teaching. My research is affected by my teaching (e.g. my research responds to ideas and questions raised in teaching, in my teaching I check my findings, their importance and how well they can be communicated). My research has its impact mainly in my teaching. With my research, I build the character of those I teach as well as my own character. I possess character and social competency.
Relevance	I am working on something that has relevance for the research community and make it clear why it has relevance.

^aData: for example, sources, material, text corpora, original works, manuscripts.^bApproach: for example, new/different perspective, new method, new theory, new research question, new interpretation, new hypothesis, different research tradition, different historical context, different cultural phenomena.^cStringent: for example, coherent, conclusive, of compelling logic and faultless, clear and logical development of the hypothesis, the steps of the investigation are logically ordered and coordinated.^dComprehensible: for example, the steps in the argumentation are transparent and verifiable, the findings produced are understandable at all times and based on the explanation of research question and method.

Table A2. Ratings of aspects by discipline, grouped by criteria

Criterion	Aspect (abbreviated)	German Literature Studies			English Literature Studies			Art History		
		M	Mdn	P10 ^a	M	Mdn	P10 ^a	M	Mdn	P10 ^a
Scholarly exchange	Disciplinary exchange	5.29	5	4	5.42	5	5	5.12	5	4
	Interdisciplinary exchange	4.99	5	4	4.46	4	3	4.88	5	4
	International exchange	4.74	5	3	5.37	6	4	5.08	5	4
Innovation, originality	New data or novel combination of data	4.99	5	3	5.12	5	4	5.22	6	4
	Introduction of new research topics	5.05	5	4	5.08	5	4	5.27	5.5	4
	New approach to topic or data	5.35	5	4	5.20	5	4	5.22	5	4
	Generating new paradigms	5.27	5	4	5.08	5	4	5.18	6	4
	Contribution of new findings or interpretations	5.52	6	5	5.32	5	5	5.40	6	4
	Innovative language or depiction	4.30	5	2	4.46	5	3	4.25	4	2
	Identification of gaps in existing knowledge	5.32	6	4	5.36	5	5	5.17	5	4
Productivity Rigour	Continuous research outputs	4.68	5	3	4.90	5	3	4.65	5	3
	Systematic and transparent research process	4.47	5	2	4.39	5	2	4.30	4	3
	Stringent argumentation	5.55	6	5	5.25	5	4	5.20	5	4
	Presentation of relevant documents and evidence	5.45	6	5	5.56	6	5	5.43	6	4.5
	Clear language	5.39	6	4	5.25	5	4	5.08	5	4
	Clear structure	5.35	6	4	5.19	5	4	5.12	5	4
	Reflection of method	5.12	5	4	4.88	5	4	5.08	5	4
	Intersubjectivity	5.18	5	4	4.76	5	3	4.98	5	4
	Adherence to rules of scientific honesty	5.74	6	5	5.53	6	4	5.60	6	5
	Discussion of generalizability of insights	4.64	5	3	4.59	5	3	4.90	5	3
Reflection of personal relation to research topic	4.49	4	3	4	4	2	4.33	5	3	
Fostering cultural memory	Documentation of aspects of the past	5	5	4	4.90	5	3	5.03	5	4
	Renewal of understanding of aspects of the past	5.43	6	4	5.20	5	4	5.40	5.5	4.5
	Putting the past in relation to the present	4.90	5	3	4.46	5	2	4.92	5	4
	Insights into perennial aspects of human nature	4.13	4	2	3.95	4	2	3.95	4	2
Recognition	Insights are recognized by the research community	4.31	4	3	4.86	5	4	4.48	5	2
	Insights are recognized by society	3.51	4	2	3.42	3	2	3.83	4	2
	Reputation within research community	4.31	5	2	4.71	5	3	4.22	4	2
	Reputation in society	3.34	4	1	3.24	4	1	3.52	4	1
	Reputation at own university	3.77	4	2	4.24	5	2	4.03	5	2
Reflection, criticism	Looking for distinctions	4.68	5	4	4.59	5	3	4.43	5	3
	Deconstruction of the illusion of 'definitive and final truth'	4.56	5	3	4.08	4	2	4.33	4.5	2
	Criticizing assertive claims and social norms	4.31	4	3	4.42	5	3	3.95	4	2
	Criticizing established scholarly approaches	4.32	4	3	4.27	5	3	4.38	4	3
	Self-critical and self-reflective research	5.09	5	4	4.56	5	3	4.60	5	3
	Visualizing complexity	5.39	6	4	4.85	5	3	5.03	5	4
Continuity, continuation	Promotion of young academics	5.04	5	4	4.53	5	2	4.87	5	3
	Continuation of research traditions	3.95	4	2	3.92	4	2	3.63	4	2
	Long-term pursuit of research topics	3.44	4	1	2.88	3	1	3.15	3	2
Impact on research community	Stimulating new research	4.99	5	4	4.95	5	4	5.17	5	4
	Concluding a debate	3.43	3	2	3.61	4	2	4.00	4	3
	Establishing a new school of thought	3.87	4	2	3.83	4	2	3.98	4	2
	Influencing the research community	4.49	5	3	4.88	5	4	4.68	5	3
Relation to and impact on society	Topics relevant for society from the scholars' perspective	4.10	4	2	3.78	4	2	3.90	4	2
	Responding to societal concerns	3.65	4	2	3.25	3	1	3.67	4	2
	Conveying findings to a non-academic audience	4.39	4	3	4.24	4	2	4.47	5	3
	Affecting national or local culture	3.64	4	2	3.39	4	1	4.00	4	2
Variety of research	Contributing towards variety and diversity	4.83	5	4	4.32	5	2	4.80	5	3.5
	Taking risks and working outside of mainstream	5.13	5	4	4.42	5	2	4.97	5	4

(continued)

Table A2. Continued

Criterion	Aspect (abbreviated)	German Literature Studies			English Literature Studies			Art History		
		M	Mdn	P10 ^a	M	Mdn	P10 ^a	M	Mdn	P10 ^a
Connection to other research	Building on current state of research	4.68	5	4	4.80	5	4	4.67	5	3
	Re-connecting to neglected research	4.91	5	4	4.76	5	4	4.58	5	3
	Engaging in ongoing research debates	4.53	5	4	4.75	5	3	4.60	5	3.5
Openness to ideas and persons	Openness to other ideas	5.19	5	4	5.10	5	4	5.25	5	4
	Openness to other persons	5.21	5	4	5.19	5	4	5.33	6	4
Self-management, independence	Realization of own research goals	4.97	5	4	4.76	5	4	4.82	5	3
	Research outcomes are unpredictable	5.25	5	4	4.78	5	4	4.95	5	3
	Research is not directly utilizable	4.77	5	3	4.32	5	2	4.08	4.5	2
	Research is not directly targeted at a recipient	4.65	5	3	4.68	5	3	4.17	4	2
Scholarship, erudition	Rich experience with sources	5.27	5	4	5.36	5	5	5.35	6	4
	Knowledge based on own research	5.45	6	5	5.39	5	5	5.30	6	4
Passion, enthusiasm	Passionate about research	4.91	5	3	4.49	5	2	4.90	5	3
	Arouse passion for research	4.86	5	4	4.92	5	4	4.85	5	3.5
	Intrinsic motivation for research activity	4.49	5	2	4.34	5	2	4.53	5	2
Vision of future research	Pointing out important research for the future	4.65	5	4	4.73	5	4	4.83	5	4
Connection between research and teaching, scholarship of teaching	Research-based teaching	4.95	5	3	4.80	5	4	5.08	5	4
	Teaching-based research	5.01	5	4	4.58	5	3	4.75	5	3
	Research has its impact mainly in teaching	3.40	3	2	3.02	3	1	3.40	3	2
	Building character of oneself and others	4.55	5	3	3.68	4	1	4.13	4	2
	Social competency	4.60	5	3	3.24	3	1	4.15	4	2
Relevance	Research is relevant for the research community	4.70	5	4	4.69	5	3	4.57	5	3

^aP10 stands for 10th percentile.