#### EHzürich



Validity Issues of Bibliometric and Performance Indicators: What we can Learn from the Case of the Humanities

### Michael Ochsner, Sven E. Hug, & Hans-Dieter Daniel

Nordic Workshop on Bibliometrics and Research Policy, 25.-26. September 2014, Reykjavik

# Outline

- Background of the project
- Concept of validity
- Notions of quality
  - Validity issues
- Quality criteria and indicators
  - Validity issues
- What can we learn from the humanities?



# Background

Cooperation project of the universities of Zurich and Basel "Developing and testing research quality criteria in the humanities"

- Part of CRUS project B-05: "mesurer les performances de la recherche" 2008-2011
- Developing quality criteria
  - that adequately represent research quality in the three disciplines covered, i.e., German literature studies, English literature studies and art history
  - that are accepted by the scholars of the respective disciplines
  - that are applicable in different cultural, linguistic and regional/national contexts

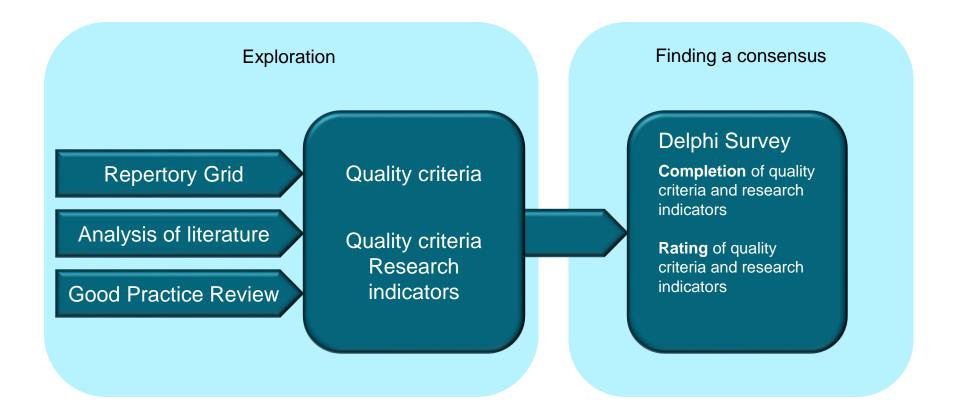
## **Approach and Research Questions**

# Appropriate criteria and indicators from scholars' point of view

- Approach
  - Discipline-specific approach
  - Involving all scholars (bottom-up)
  - Open outcome
  - Linking indicators to criteria
- Research questions
  - What are scholars' crit.&ind. for research quality?
  - How appropriate are these crit.&ind. in the eyes of scholars to assess their own research?
  - Is there a set of shared crit.&ind.? (consensus)

#### DGESS

## **Project overview**

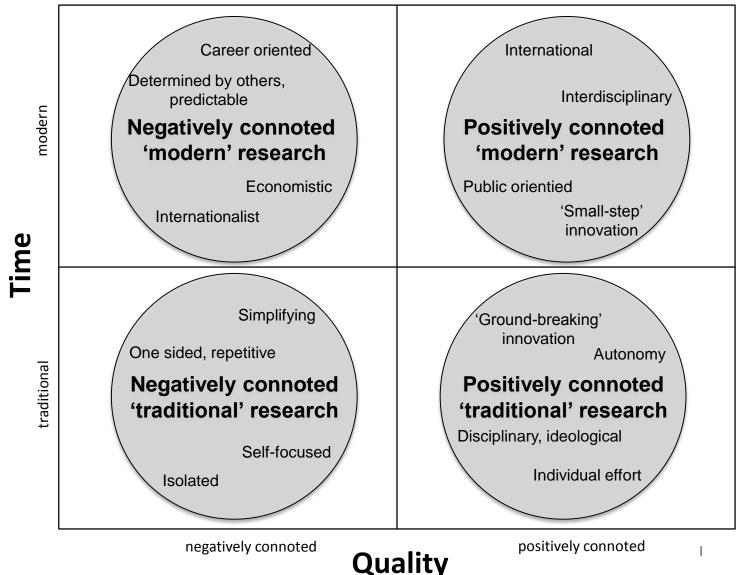


# Validity

- the extent to which a measure (i.e., an indicator) actually measures what it purports to measure (i.e., a concept) (Borsboom et al., 2004, p. 1061)
- Data-driven: "measuring what can be measured" endangers validity, mostly reducing it to correlation.
- Thunder correlates highly with lightning (and there is even a causal relationship). However, lightning cannot measure thunder.

#### ETHzürich

## Validity Check 1: Indicator for Quality or Conception?



| 7

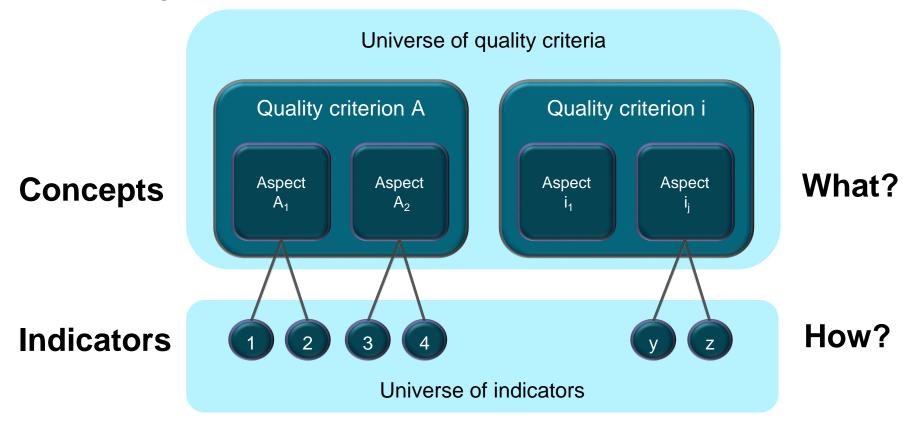
DGESS

## **Connex Criteria and Indicators**

- We captured the humanities scholars' notions of quality
- We know the criteria humanities scholars think are important to differentiate between "good" and "bad" research
- We collected Indicators of about pertinent 100 publications and from humanities scholars
- We can relate them to the scholars' notions of quality and assign them to the criteria

# **Our Measurement Approach**

Measuring in the Social Sciences



# 19 criteria (and 70 aspects, cf. p. 9)

- 1. Scholarly exchange
- 2. Innovation, originality
- 3. Productivity
- 4. Rigour
- 5. Fostering cultural memory
- 6. Recognition
- 7. Reflection, criticism
- 8. Continuity, continuation

- 9. Impact on research community
- 10. Relation to and impact on society
- 11. Variety of research
- 12. Connection to other research
- 13. Openness to ideas and persons
- 14. Self-management, independence

- 15. Scholarship, erudition
- 16. Passion, enthusiasm
- 17. Vision of future research
- 18. Connection between research and teaching, scholarship of teaching
- 19. Relevance

## **Examples for aspects**

#### Scholarly exchange

- Disciplinary exchange
- International exchange
- Interdisciplinary exchange

#### Recognition

- Insights are recognized by the research community
- Insights are recognized by society
- Reputation within research community
- Reputation in society
- Reputation at own university
- Variety of research
  - Contributing towards variety and diversity
  - Taking risks and working outside of mainstream

# Validity Check 2

- How many criteria and aspects can be measured?
  - If no quantitative indicator can be found or thought of,
    → the aspect cannot be measured
- 50% of the aspects relevant to the scholars cannot be measured, e.g.:
  - Rigour: all aspects
  - Innovation: all aspects
  - Connection to other research:
    - Re-connecting to neglected research

## ■ → Indicators do not measure Quality encompassingly

# Validity Check 3 (1/4)

- What do Indicators that are commonly used measure?
- Are these the relevant criteria?

Indicators	Criterion
Citations	Recognition; impact on research community; relevance
Prizes	Recognition; impact on research community; relevance
Third party funding	Recognition; impact on research community; relevance; relation to and impact on society
Collaborations	Scholarly exchange; recognition
Transfers to society and economy	Relation to and impact on society
Publications	Scholarly exchange; productivity
Board memberships	Scholarly exchange; recognition; impact on research community
Recruitment	Continuity, continuation

Table 1: Frequently used indicators and criteria they can potentially measure

#### DGESS

# Validity Check 3 (2/4)

- Measured by commonly used indicators (*bold and italic*)
  - 1. Scholarly exchange
  - 2. Innovation, originality
  - 3. Productivity
  - 4. Rigour

DGESS

- 5. Fostering cultural memory
- 6. Recognition
- 7. Reflection, criticism
- 8. Continuity, continuation

- 9. Impact on research community
- 10. Relation to and impact on society
- 11. Variety of research
- 12. Connection to other research
- 13. Openness ideas and persons
- 14. Self-management, independence

- 15. Scholarship, erudition
- 16. Passion, enthusiasm
- 17. Vision of future research
- Connection between research and teaching, scholarship of teaching
- 19. Relevance

# Validity Check 3 (3/4)

- Consensual Indicators (orange: all three; blue: in two disciplines)
  - 1. Scholarly exchange
  - 2. Innovation, originality
  - 3. Productivity
  - 4. Rigour
  - 5. Fostering cultural memory
  - 6. Recognition
  - 7. Reflection, criticism
  - 8. Continuity, continuation

- 9. Impact on research community
- 10. Relation to and impact on society
- 11. Variety of research
- 12. Connection to other research
- 13. Openness ideas and persons
- 14. Self-management, independence

- 15. Scholarship, erudition
- 16. Passion, enthusiasm
- 17. Vision of future research
- Connection between research and teaching, scholarship of teaching
- 19. Relevance

# Validity Check 3 (4/4)

- Valid measures for research quality?
  orange: three disc.; blue: two disc.; bold and italic: commonly used
  - 1. Scholarly exchange
  - 2. Innovation, originality
  - 3. Productivity
  - 4. Rigour
  - 5. Fostering cultural memory
  - 6. Recognition
  - 7. Reflection, criticism
  - 8. Continuity, continuation

- 9. Impact on research community
- 10. Relation to and impact on society
- 11. Variety of research
- 12. Connection to other research
- 13. Openness ideas and persons
- 14. Self-management, independence

- 15. Scholarship, erudition
- 16. Passion, enthusiasm
- 17. Vision of future research
- 18. Connection between research and teaching, scholarship of teaching
- 19. Relevance

# Validity Check 4 (1/2)

How are the commonly used indicators rated by the scholars?

Put differently:

 How well do these indicators measure the criteria they are potentially capable of measuring?

# Validity Check 4 (2/2)

Indicator group	Criteria	Range	Consens
Citations	Relevance, Impact, Recognition	GLS 3.69; ELS 3.77-3.89; AH 3.76	
Collaborations	Exchange, Recognition	GLS 4.73-4.77; ELS 3.37-3.95; AH 4.67- 4.95	GLS; AH
Panels	Exchange, Recognition, Impact	GLS 4.17-4.21; ELS 3.84-3.95; AH 4.29- 4.36	
Prizes	Relevance, Recognition, Impact	GLS 3.83; ELS 3.93-4.16	
Publications	Exchange	GLS 4.85-4.98; ELS 4.70-5.02; AH 4.95- 5.21	G; E; A
Third party funds	Relevance, Recognition	GLS 4.17; ELS 3.23	
Not so comm	on but still used		
Monographs	Scholarship, Reflection	GLS 3.54-4.23; ELS 4.05; AH: 4.19-4.52	
Peer reviewed	Connection, Relevance, Recog.	GLS 4.17-4.21; ELS 4.63	
Teaching awards	Passion	GLS 3.83; ELS 3.58; AH 3.55	
DGESS		Nordic Workshop on Bibliometrics, Reykjavik	29.09.2014   18

# What can we learn from the humanities

- Think about what you want to measure and why
- Think ahead
  - What happens if a certain indicator is applied in assessments
- Not everything that matters can be put into numbers

## **Solutions:**

- Get a grip on the concept: Quality
- Bottom-up (meaning by discipline & ALL researchers)
- Informed peer review
- Declare measurement: who why which how.



#### **D** GESS

- Project publications:
  - http://www.psh.ethz.ch/crus/publications
- Colloquium:
  - http://www.psh.ethz.ch/crus/kolloquium
- Publication data base:
  - Data base for Literature on Arts & Humanities and Assessment (+/- 1000 Entries) "Arts & Humanities Research Assessment Bibliography" (AHRABi)

http://www.psh.ethz.ch/crus/bibliography

#### **ETH** zürich

# Thank you for your attention!



Innovative language or depiction



Promotion of young academics

Visualizing complexity

the past fubure Disciplinary exchange کو Stimulating new research Pointing out important research for the Research outcomes are unpredictable Documentation of aspects of the past of Engaging in ongoing research debates Igoing research debates Research is relevant for the research community Building on current state of research Openness to other persons Arouse passion for research Interdisciplinary exchange

**Realization of own research goals** 

Knowledge based on own research

Promotion of young academics

Disciplinary exchange

Documentation of aspects of the past

Openness to other persons

Interdisciplinary exchange

## **Second Round: Design**

- Rating of the aspects
  - "My research is assessed appropriately, if the assessment considers whether...[aspect]"
  - Scale: 1-6 (1: I strongly disagree; 2: I disagree; 3: I slightly disagree; 4: I slightly agree; 5: I agree; 6: I strongly agree)
- Approval of aspect:
  - At least 50% rate the aspect as positive (4, 5, or 6)
- Consensus on aspect:
  - At least 50% clearly agree with the aspect (5 or 6)
  - only a few disagree (max. 10% rate the aspect with 1, 2, or 3)

# Third Round: Examples (1/2)

- Indicators can be linked to criteria and aspects
- Criterion: Fostering cultural memory
  - Aspect: Documentation of aspects of the past
    - Number, weighting and duration of documentation or preservation activities
    - Number and weighting of outputs reflecting documentation or preservation activities
    - Number and weighting of activities for the public (e.g., guided tours, public lectures, readings, media appearances, performances)
    - Number and weighting of outputs for the public (e.g., popular books or articles, exhibitions, documentary films)

Criterion	Aspect		ensual i cipline	n
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	Continuous research outputs			
Rigour	Systematic and transparent research process			
0	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		AH
	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		
	Visualising complexity	GLS		AH
Continuity, continuation	Promotion of young academics	GLS		
<b>.</b>	Continuation of research traditions			
	Long-term pursuit of research topics			

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

Criterion	Aspect	Consensual i	n Discipl	ine
Impact on research community	Stimulating new research Concluding a debate Establishing a new school of thought	GLS	ELS	AH
	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 Openness to other persons	GLS GLS	ELS ELS	AH AH
Self-management, independence	Realisation of own research goals Research outcomes are unpredictable Research is not directly utilisable Research is not directly targeted at a recipient	GLS GLS	ELS ELS	
Scholarship, erudition	Rich experience with sources Knowledge based on own research	GLS GLS	ELS ELS	AH AH
Passion, enthusiasm	Passionate about research Arouse passion for research Intrinsic motivation for research activity	GLS	ELS	AH
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 Teaching-based research Research has its impact mainly in teaching Building character of oneself and others Social competency	GLS	ELS	AH
Relevance	Research is relevant for the research community	GLS		

Table 2.All criteria and their aspects with an indication of the discipline in which an aspect is consensual (continued).

Name of group	Criterion	Aspect(s)
Publications	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
References	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
	Scholarship, erudition	Knowledge based on own research
Presentations	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
	Connection to other research	Building on current state of research
Editorship	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
	Recognition	Insights are recognized by the research community
	Impact on research community	Stimulating new research Influencing the research community
	Variety of research <sup>a</sup> Relevance <sup>a</sup>	Contributing towards variety and diversity <sup>a</sup> Research is relevant for the research community <sup>a</sup>
Organized events	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
Collaborations	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
	Recognition <sup>a</sup>	Insights are recognized by the research community <sup>a</sup>
Personal contacts	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
Review activities	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
	Recognition	Insights are recognized by the research community
	Impact on research community	Influencing the research community
Academic associations	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange
	Recognition	Insights are recognized by the research community
	Impact on research community	Influencing the research community
Panels	Scholarly exchange	Disciplinary exchange Interdisciplinary exchange International exchange)
	Recognition	Insights are recognized by the research community
<b>D</b>	Impact on research community	Influencing the research community
Documentation activities	Fostering cultural memories Scholarship, erudition	Documentation of aspects of the past Rich experience with sources
		2 N

Table A.2: Groups of indicators and the criteria and aspects they can potentially measure

Table A.2: Groups of indicators and the criteria and aspects they can potentially measure	
(continued)	

Name of group	Criterion	Aspect(s)
Output of documentation activities	Fostering cultural memories Scholarship, erudition	Documentation of aspects of the past Rich experience with sources
Activities for the public	Fostering cultural memories	Documentation of aspects of the past
Outputs for the public	Fostering cultural memories	Documentation of aspects of the past
Survey: renewal of interpretations	Fostering cultural memories	Renewal of interpretations of aspects o the past
Reviews of the researcher's work	Recognition	Insights are recognized by the research community
	Impact on research community Relevance	Influencing the research community Research is relevant for the research community
Citations	Recognition	Insights are recognized by the research community
	Impact on research community	Stimulating new research Influencing the research community
	Relevance	Research is relevant for the research community
Acknowledgments	Recognition	Insights are recognized by the research community
	Impact on research community	Stimulating new research Influencing the research community
	Passion, enthusiasm Relevance	Arouse passion for research Research is relevant for the research community
Peer-reviewed channels	Recognition	Insights are recognized by the research community
	Connection to other research Relevance	Building on current state of research Research is relevant for the research community
Usage statistics	Recognition	Insights are recognized by the research community
	Impact on research community Relevance	Influencing the research community Research is relevant for the research community
Third party funding	Recognition	Insights are recognized by the research community
	Relevance	Research is relevant for the research community
Prizes	Recognition	Insights are recognized by the research community
	Impact on research community Relevance	Influencing the research community Research is relevant for the research community
	Recognition	Insights are recognized by the research
Appointments to professorship	-	community Influencing the research community

Table A.1: Groups of indicators for research quality in the humanities (continued)

Name of group	Specification <sup>a</sup>	
Attractivity to junior researchers	Attractivity to junior researchers (e.g., number of Ph.D. students I have, postdoctoral researchers and researchers from abroad I have; number of participants in my courses)	
Monographs	Number of my monographs	
Monographs relative to articels	Number of my monographs related to the number of my articles	
Qualification of junior researchers	Qualification of students and junior researchers (e.g., number of bachelor/master/doctoral degrees; success rate (appointments to a professorship) of former students; drop-out rate of students and junior researchers; survey of alumni about the skills/competencies/qualifications they acquired)	
Success of junior researchers	Success of junior researchers (e.g., number and weighting of publications; honours, awards and prizes of my students and junior researchers; number of citations of junior researchers; stay in research of junior researchers; teaching quality among junior researchers I have trained)	
Teaching	What I offer in teaching (e.g., teaching hours, the time that I spend in helping and guiding junior researchers; my participation in a graduate program, graduate school or comparable program; the number and quality of further training courses I offer)	
External education	External education of junior researchers (e.g., research stays of junior researchers at other institutions; number of external further training these junior researchers have attended; the financial resources I make available to them for attending congresses or receiving additional training)	
Collaboration with junior researchers	Collaboration with junior researchers (e.g., co-authorship, co-editorship, joint projects)	
Survey: satisfaction	Survey of junior researchers' satisfaction	
Started initiatives	Number and weighting of what I have initiated or founded (e.g., periodical, book series, research institution or research cluster, post graduate programm, degree programm)	
Research topics	Number of research topics, approaches, theories, methods, materials, disciplinary areas and languages that I use (e.g., evident in the bibliography of my publications and presentations, information on my research website)	
Infrastructure	Number and weighting of infrastructure I have established or I administer (e.g., art collection, specialized library, museum, database)	
Current references	Number of current references (e.g., number of sources not older than for instance 5 years that I quote in my publications)	
Discussions/debates	Number and weighting of participation, organization or moderation of disputes, debates or discussions about research	
Written responses	Number and weighting of written responses (e.g., essay, editorial or newspaper column, open letter)	
Opportunities for junior researchers	Career opportunities for junior researchers (e.g., number of positions for junior researchers, number of publications by junior researchers who have been my students, number of co-authorships with junior researchers)	
Assessed openness	Assessment of my openness by students and junior researchers	
Heterogeneity of junior researchers	Heterogeneity of the junior researchers (e.g., number of exchange students; number of students of other research institutions; number of students from educationally disadvantaged backgrounds, number of minority group students; junior researchers from countries with underdeveloped research structures)	
Assistance	Number of collaborations or publications with researchers from institutions with weak reputations or with researchers from countries with underdeveloped research structures	

Table A.1: Groups of indicators for research quality in the humanities (continued)

Name of group	Specification <sup>a</sup>
Course accessibility	Accessibility to my courses or course of study I offer (e.g., waiving the cost of participation for students, junior researchers and researchers from countries with underdeveloped research structures; public announcement of my courses; access by all to events and courses I offer)
Availability of publications	Availability of my publications (e.g., number of open-access publications in relation to the total number of my publications)
Achievement of own goals	The degree to which I have met the goals I set
Financial independence	The degree of my financial independence (e.g., how much of the research funding is not tied to set targets; the ratio of basic funding to third-party funding)
Absence of requirements	The absence of set targets or output requirements
Sources	Number of sources, materials and original works used in publications or presentations
Research time	Time spent on research (e.g., time spent on research in relation to time spent on teaching and administration; the number of paid hours per year that I can devote to research)
Personal library	Size of my personal library (e.g., size of my personal library, art collection or collection of source material; number of journals subscribed to)
Teaching awards	Number and weighting of awards and prizes for teaching
Survey: enthusiasm – teaching	Survey of students, junior researchers and alumni about whether I arouse passion or enthusiasm for the research or the research topic
Survey: enthusiasm – public	Survey of the public about whether I arouse passion or enthusiasm for the research or the research topic
Strategies	The existence of goals and strategies how to reach these goals
Utilizing sources	Examining and utilizing sources (e.g., number and weighting of historical critical editions; number of appraisals of museum collections; number and weighting of collections or texts in digital database and documenting these collections or texts)
Congruence research - teaching	Degree of congruence between my research and my teaching (e.g., drawing comparisons between my course titles and my publication titles, my course descriptions and my research summaries, the table of contents of my lecture notes and my publications, bibliography of my teaching material and my publications)
Research orientation of teaching	Student satisfaction with the research orientation of the courses
Students' publications	Number and weighting of scholarly publications or presentations by my students
Acknowledging junior researchers	Number of times I refer to my students or junior researchers I have trained in my acknowledgments
Collaboration with students	Collaboration with students (e.g., co-authorship, co-editorship, joint projects)
Publication of course material	Number of publications resulting from revising material I have used in teaching
Invited lectures	Number and weighting of invited lectures

Note: Sometimes the group of indicators can measure more than one aspect (see table A.2). In such a case, the indicators must be slightly adapted to the aspect. For example, the indicator group 'publications' can measure *disciplinary exchange*, *interdisciplinary exchange*, or *international exchange*. In the first case, only the disciplinary publications will be examined, in the second case only interdisciplinary publications will be included, and in the third case only international publications will be counted.

<sup>a</sup> The specification of the groups of indicators is displayed in the original wording of the questionnaire for the rating of the indicators. We used the first person singular to establish a close association of the aspects and indicators with the respondent's own research. The scholars had to rate the indicators as to their potential to give good indications of the occurrence of the aspect in their own research. This ensures the link between the ratings and the notions of quality the scholars use in their everyday work.