



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

*A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)*



Peer Reviewing

EERQI Semantic Meeting,
XEROX, Grenoble,
January 22nd 2010

Ingrid Gogolin & Christian Rudelt



Peer review: an overview

Peer review: self correcting, responsive, efficient, self policing, accountable, fair and dynamic (Wood, 1997)

Peer review: still main tool for quality assessment of scientific output, evaluation of research proposals, academic appointments

Why?

- Only “peers” have enough/same/more prior knowledge to judge on the “true” intrinsic quality of the reviewed subject
- This is hard to replicate by any less direct form of assessment



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

*A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)*



Peer review – indispensable or replaceable?

- Criticism against peer review is not a new phenomenon
- Opinions range from peer review as being “censorship” (Froehlich 2003) to peer review as being “irreplaceable to measure the innovativeness of research in general” (Hornborstel, 2007).
- Peer review as “convenient” quality assessment tool for the editors (as it is not remunerated) (e.g. Rennie, 1993)

Peer review: the downs

But: system inherent flaws unavoidable

➤ Mueller subdivides flaws of peer review into seven categories:

- 1) Time delay
- 2) Costs and efficiency
- 3) Lack of **reliability** (inconsistency/randomness of results)
- 4) No guarantee for error detection (e.g.contentwise)
- 5) Systematic „particular“ bias (i.e. against gender, ideas, authors, institutions)
- 6) Possible fraud and misuse („cronyism, „old-boys-network“, politics)
- 7) Missing **validity** („⁴**expressiveness**“ of results)



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

*A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)*



Peer review and validity

Definition of validity in peer review (PR): „Rightness“, accuracy of reviewers' recommendation, generally regarded as being **high** in PR

But: it remains difficult to measure „real“ validity of PR!

Example: „success“ of reviewed AND published article by means of citation counts later on, e.g.)

And: empirical research on validity of peer review very limited (e.g. Weller 2001: only seven empirical studies analysing the empirical validity of peer review in editorial peer review so far with limited significance)



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

*A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)*



Causes of limited validity in PR

Reduced/limited validity in PR is a serious problem for peer review:

- Reasons: different research approaches/ contexts/traditions of reviewers
- How to apply general criteria (e.g. „rigour“) to a particular review situation in often diverse scientific fields (such as e.g. education?)

Limited validity not restricted to interdisciplinary peer review; also within the same research field criteria may vary:

- Reasons: ongoing internationalization of scientific world and thus pluralization



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)



But: there is hope...

Although empirical studies on validity in editorial peer review remain rather vague...

- Studies on screening Ph'D grants application by NSF **show validity** of PR:

by showing the success/failure of students acquiring a PhD (Chapman/McCauley 1993)

Conclusion: PR **can** work as quality filter and self-monitoring mechanism in scientific community (Caverson, Carter, et al...1997...)



Reliability of peer review

Def.: Reliability in PR as the ***extent*** to which the assessments of the reviewers for **one** piece of work are the same or differ

Reasons for limited reliability: Each individual review process is influenced by several other factors (e.g.: selection of reviewers by the publisher, see Godlee & Dickersin 1999, Rothwell & Martyn 2000)

Lack of consistency remains persistent problem (cp. empirical study by Rothwell & Martyn, 2000)



Reliability of peer review

Main problem: Inconsistency due to inherent subjectivity of individual assessment of each reviewer → difficult to minimize

What makes editorial PR especially vulnerable to lack of reliability:

- Two to three reviews are usually obtained per submitted document
- **Finally:** Decision-making regarding publishing normally takes place in secrecy and non-transparent manner by editor



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

*A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)*



Peer review and reliability: some empirical data

Weller: The average likelihood of overall agreement [of reviewers in editorial PR] = 50 %)

Observations: research up-to-now displays various levels of interreviewer agreement, depending on review context:

- In grant applications up to 82 % of reviewer agreement is reached (e.g. in the DFG organization, in: Hartmann/Neidhardt 1990)
- Reviewer consensus in negatively reviewed (thus denied) research funding applications is higher than in positive ones (76% vs. 54%, in Smith, Weller, 2001)



Research and the „reliability problem“

Generally:

- Various empirical studies show **kappa coefficients** between 0.20 and 0.40 (*fair amount of reviewer agreement* according to Hornborstel 2006, Cicchetti 1991, Weller 2001)

And finally:

- Reviewer disagreement is not merely a negative factor but allows *multiperspectivity* (see Bornmann/Daniel 2003 and Weingart 2005)

Hodgson: „That reviewers don't agree doesn't mean that they can't“ (1997) – disagreement does not question the procedure of PR as such!

Ways ahead: Possible ways to improve validity and reliability of peer review

- Recruiting of *relevant* experts for the review
- Standardization of review criteria
- Establishing of review guidelines (e.g. reviewers working on samples together *before* the review takes place)
- Moderation of the PR process
- Learn from best practices in „peer review management“ (e.g. from books like „How to Survive Peer Review“, Wager et al. 2000)



Peer review – indispensable or replaceable?

- Mueller (2008): No real alternative to peer review can be found as quality assurance system in science
- Peer review as filter function and quality maintainer in science, especially in manifold online publications (e.g. open-access movement, pre-publishing repositories)
- BUT: One dilemma remains: standardization as unavoidable by-product of quality assessment



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

*A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)*



Peer Review and Semantic Analysis

- Agnes/Angela: Semantic Analysis and (automatic) highlighting key sentences containing keywords/key can help the peer reviewer to:
 - a) Focus the reader's attention
 - b) Systematic reviewing
 - c) Saving time by grasping the „core content“ of a text more quickly



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)



Semantic Analysis and Validity/Reliability of PR

- Automatic highlighting of key concepts/sentences is process of standardization/guidance of the reviewer:
- reliability of the PR will most likely *increase*, as reviewer disagreement decreases
- How about validity? Does automatic guidance distract/influence reviewer „too“ much and have unwanted effects? (e.g. is the „real“ quality of the text been measured?)
- More tests of semantic analysis and PR will have to be gathered; empirical data collected to better understand interplay between PR and semantic analysis



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

*A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)*



Thank you for your attention!



EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators

A project funded under the Socio-Economic Sciences
and Humanities Theme (SSH)



Literature

- Benos, D. (2007): The ups and downs of peer review, in: *Advances in Physiology Education*, Vol. 31, pp. 145-152.
- Bornmann, L.; Daniel, H.-D (2004): Reliability, fairness and predictive validity of committee peer review, in: *B.I.F. Futura* Vol. 19, pp. 7-9.
- Bridges, D.: *A note on peer review of academic papers* (not published essay), 2009.
- Cicchetti, D. V. (1991): The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation, in: *Behavioral and Brain Sciences*, Vol. 14, Nr.1, pp. 119-135.
- Curtis, M. J.; Shattock, M. J. (1998): Die Rolle des Gutachters, in: Hall, G. M. (Ed.): *Publish or perish: wie man einen wissenschaftlichen Beitrag schreibt, ohne die Leser zu langweilen oder die Daten zu verfälschen*. Bern, Göttingen, Toronto, Seattle: Huber, pp. 111-119.
- Sandor, Agnes, Vorndran, Angela: Detecting key sentences for automatic assistance in peer reviewing research articles in educational sciences