

Rudelt, Christian

Peer review: Necessary evil or last resort? An overview of the current research status on the quality of editorial peer review

Erziehungswissenschaft 21 (2010) 40, S. 17-25



Empfohlene Zitierung/ Suggested Citation:

Rudelt, Christian: Peer review: Necessary evil or last resort? An overview of the current research status on the quality of editorial peer review - In: *Erziehungswissenschaft 21 (2010) 40, S. 17-25* - URN: urn:nbn:de:0111-opus-27535

in Kooperation mit / in cooperation with:



<http://www.budrich.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Erziehungswissenschaft

**Mitteilungen der Deutschen Gesellschaft
für Erziehungswissenschaft (DGfE)**

Heft 40

21. Jahrgang 2010

ISSN 0938-5363

Verlag Barbara Budrich

INHALTSVERZEICHNIS

Editorial	7
-----------------	---

Beiträge

Lutz R. Reuter

Erziehungswissenschaftliche und lehrerbildende Studiengänge im Akkreditierungssystem. Ein Praxisbericht	9
--	---

Christian Rudelt

Peer review: Necessary evil or last resort? An overview of the current research status on the quality of editorial peer review	17
---	----

Christoph Wulf

Die Deutsche UNESCO-Kommission	27
--------------------------------------	----

Beiträge der Tagung *Lehrerbildung in den neuen Studienstrukturen*

Lothar Wigger

Das Kerncurriculum Erziehungswissenschaft in der Lehrerbildung – Erfahrungen und Probleme	33
--	----

Lutz R. Reuter

Zu den Schnittmengen und zur Polyvalenz erziehungswissenschaftlicher Studiengänge	41
--	----

Josef Keuffer

Reform der Lehrerbildung und kein Ende? – Eine Standortbestimmung	51
--	----

Eva Arnold

Kooperationen zwischen der ersten und zweiten Phase der Lehrerausbildung	69
---	----

Wilfried Schubarth

Lohnt sich Kooperation? Erste und zweite Phase der Lehrerbildung
zwischen Abgrenzung und Annäherung 79

Alexander Gröschner, Cordula Schmitt

Wirkt, was wir bewegen? – Ansätze zur Untersuchung der Qualität
universitärer Praxisphasen im Kontext der Reform der Lehrerbildung 89

Andreas Keller

Gleichermaßen hochwertig und einheitlich lang: Gewerkschaftliche
Anforderungen an die Reform der Lehrerbildung im Bologna-Prozess 99

Norbert Ricken

Von Bremen über Bologna nach Bremen – Zur Neustrukturierung
der LehrerInnenbildung an der Universität Bremen 109

Olga Zlatkin-Troitschanskaia, Klaus Breuer

Wirtschaftspädagogisches Studium an der Uni Mainz. Ein polyvalenter
Bachelor und Master of Science in Wirtschaftspädagogik 125

Alexa Tegeler

Leistungsbewertungen, Prüfungen, Verschulung.
Ein Beitrag aus studentischer Sicht 135

Mitteilungen des Vorstands

Stellungnahme der Deutschen Gesellschaft für Erziehungswissenschaft
(DGfE) zu den Empfehlungen des Forschungsgutachtens zur Ausbildung
von Psychologischen PsychotherapeutInnen und Kinder- und
JugendlichenpsychotherapeutInnen 145

Berichte aus den Sektionen

Sektion 1 Historische Bildungsforschung 149

Sektion 2 Allgemeine Erziehungswissenschaft 154

Sektion 4 Empirische Bildungsforschung 170

Sektion 5	Schulpädagogik	172
Sektion 6	Sonderpädagogik	176
Sektion 7	Berufs- und Wirtschaftspädagogik	178
Sektion 10	Pädagogische Freizeitforschung und Sportpädagogik	180
Sektion 11	Frauen- und Geschlechterforschung	184
Sektion 12	Medienpädagogik	189
Sektion 13	Differenzielle Erziehungs- und Bildungsforschung	192

Notizen

<i>Aus der Forschung</i>	197
--------------------------------	-----

Aus der Wissenschafts- und Bildungspolitik

Beschluss des Instituts für Pädagogik der Universität Duisburg-Essen zum Ausstieg aus dem CHE-Ranking	203
--	-----

Beschluss des Fachbereichs Sprach-, Literatur- und Medienwissen- schaften der Universität Siegen zum Ausstieg aus dem CHE-Ranking	203
--	-----

Zugangsvoraussetzungen für die Ausbildung zum Psychologischen Psychotherapeuten und Kinder- und Jugendlichenpsychotherapeuten. Stellungnahme zu einigen Empfehlungen des Forschungsgutachtens zur Ausbildung zum Psychologischen Psychotherapeuten und zum Kinder- und Jugendlichenpsychotherapeuten	205
--	-----

Fachbereichstag Soziale Arbeit (FBTS): Die sozialberufliche Kompetenz in der Kinder- und Jugendlichenpsychotherapie stärken. Resolution	212
--	-----

World Education Research Association (WERA) – Weltweite Organisation für erziehungswissenschaftliche Forschung gegründet	213
---	-----

Inhaltsverzeichnis

Ausschreibungen, Preise

Julius-Klinkhardt-Preis zur Förderung des Nachwuchses
in der Historischen Bildungsforschung 215

Reisestipendien für Nachwuchswissenschaftler/innen
der Sektion Historische Bildungsforschung 216

Tagungskalender 217

Personalia 229

Nachruf auf Hans Tietgens 231

Impressum

Peer review: Necessary evil or last resort? An overview of the current research status on the quality of editorial peer review

Christian Rudelt

Zusammenfassung

Der vorliegende Artikel gibt einen kurzen Überblick über einen Ausschnitt des aktuellen Forschungsstandes zur methodologischen Qualität des Peer Review-Verfahrens, der Begutachtung durch unabhängige FachkollegInnen, in diesem Bericht bezogen auf das Peer Review-Verfahren, das im Bereich des wissenschaftlichen Publizierens Anwendung findet. Die Kriterien Reliabilität und Validität werden als zentrale Merkmale von Qualität in diesem Peer Review-Verfahren genauer untersucht. Es zeigt sich, dass (fehlende) Validität nicht das Problem ist, da die Übereinstimmung der Begutachter im Verfahren oftmals gegeben ist und durch genaue Anleitungen für die Durchführung des Begutachtungsprozesses erhöht wird. Geringere Reliabilität ist schon eher ein Defizit, dem jedoch durch Standardisierung in den Verfahren entgegenzuwirken versucht wird. Ein Dilemma bei der Verwendung von Peer Review wird sichtbar, das auch in anderen Verfahren zur Qualitätsmessung von wissenschaftlichem Output generell auftritt: Standardisierung ist ein ständiges Begleitprodukt. Trotz vorhandener methodischer Schwächen wird Peer Review sicherlich auch weiterhin als Werkzeug zur wissenschaftlichen Qualitätsmessung und Selektion von zu veröffentlichenden Artikeln in Zeitschriften verwendet werden. Mit dem Aufkommen von anderen, weniger reglementierten Publikationsplattformen (wie z. B. denen der Open-Access-Bewegung) erhöht sich allerdings der Legitimationsdruck auf das Peer Review-Verfahren, weshalb sicherlich eine weiterführende und vertiefende Betrachtung der Reliabilität und Validität dieses Verfahrens unerlässlich ist.

1 Introduction

“The good ones in the crock, the bad ones in the crow.”

Science has organized the selection process regarding the quality of scientific output with respect to this verdict – taken from „Cinderella“ – for over 300 years now. The name for this procedure is called peer review. It basically means that all scientific activities and results are evaluated by a third party, in most cases other scientists from the same discipline, the so called “peer” (Hornborstel 1997). The basic assumption is that peers with knowledge of the respective discipline/field are more likely able to evaluate the quality of its scientific output than anyone else and with more precision. Peer review can be used to evaluate all different kinds of scientific output: the evaluation of submitted drafts which are supposed to be published in an academic journal, the competence of academic staff applying for a professorship, the assessment of an application for a research project – just to mention a few. The main aim of peer review is to assess the quality of the entity to be reviewed. Quality is difficult to assess and can mean different things in different disciplines. This is why the procedure and the criteria for evaluation can vary widely between fields in which peer review is used and cannot be explicitly defined. The same holds true for the selection of the reviewers who are requested to carry out the review. So – almost self-explanatory to say – since peer review is the most commonly used procedure in the assessment of academic performance, it is the most criticized and disputed method as well (Hornborstel 1997, Mueller 2008).

What makes things even more complicated is the fact that competition in the scientific community increased tremendously on the national and also the international level. International rankings and the attraction of third party funds (for example EU research grants) play an important role not only for the scientific reputation of a faculty, but merely decide who gets which (financial) resource and who does not. Peer review is very much used as a tool to decide on ranking positions and funding decisions, although it has inherent methodological flaws. As a matter of fact, these flaws are subject of scrutiny for many recent research reports as the visibility of the importance of peer review has increased but the method as such is still disputed. With the advent of alterations or possible alternatives to peer review, for example the new requirements regarding transparency and traceability of the open-access movement, the method of peer review has been assessed critically again, which is reflected in the current research status on the subject itself. (Hornborstel/Simon 2006).

2 Current criticism against peer review

Mueller subdivides the criticism against peer review into the following seven categories:

- 1) Time delay;
- 2) Costs and efficiency;
- 3) Inconsistency and randomness of results (reliability);
- 4) Lack of competence of the peer review system to detect errors in the reviewed manuscripts;
- 5) Systematic bias (i.e. against authors, gender, ideas, etc.);
- 6) Possible fraud and misuse;
- 7) Missing validity (cf. Mueller 2008).

In the following overview on the current research status regarding peer review, I will mainly focus on what scholars have written regarding the criteria of validity and reliability of peer review procedures as these elements very much comprise the quality of peer review processes. To insure validity is mainly a managerial problem, as a suitable research design must be found which a very complex task in itself is. For our case this means that peer review is ideally performed along the lines of standardized criteria which are hard to find even within the same scientific discipline. Reliability means consistency of the peer review procedure in itself. In our case, peer review has a high reliability if the different reviewers conclude at merely the same reviews, no matter when and where the review is being undertaken, as long as the review criteria are not being modified (Bridges 2009; Meier, Brudney, Bohte 2006).

3 Peer review and validity

Missing or reduced validity is one of the most serious problems for peer review, as the main task of peer review – the assessment of the quality of the reviewed article – is seriously limited by it. Different research approaches and traditions make interdisciplinary peer reviews difficult, at least reducing its validity significantly. This fact is even enlarged by the ongoing internationalization and pluralisation of the scientific world. But even if one sticks to the same research field, the discrepancies due to the use of different criteria might be huge. David Bridges argues that this is not an intrinsic problem in peer review as such, but more a managerial problem relating to the design of the review process (Bridges 2009). However, it remains truly difficult to design the “right” peer review for each discipline.

Despite these arguments, there is no reason to give up on peer review. According to Hornborstel, empirical evidence of various studies does show kappa coefficients¹ between 0.20 and 0.40 which respond to a relatively fair level of reviewer agreement (cf. Cicchetti 1991, 1997; Weller 2001, in Hornborstel/Siekermann 2006) for peer reviews in general. Furthermore, Hornborstel argues that reviewer disagreement (and thus reduced validity) is not merely a negative factor, but furthermore a positive one in evaluating a manuscript from a number of different perspectives (Daniel 2005, in Hornborstel/Siekermann 2006). This indication is underpinned by various other studies on the validity and fairness of peer review (for example in Bornmann/Daniel 2004 and Weingart 2005).

There is also a visible tendency in current research to improve validity of peer review by giving guidelines and explicit help for the persons responsible for the design of the review (in most cases the editors). In order to standardize review criteria, books giving best practices on “peer review management” such as Irene Hames’ “Peer review manuscript management in scientific journals” (Hames 2007) and titles like “How to survive peer review” (Wager et al. 2002) were written, trying to help reviewers and editors alike by suggesting standards of peer review. The relevance of these books in improving validity of peer review has not been picked up by research so far.

4 Peer review and reliability

Reliability describes the probability that peer review for two submitted manuscripts with the same content would yield the same results regarding a decision relating to publication. Actually, each individual case is influenced by several factors such as for example the selection of reviewers by the publisher. This is subject of scrutiny in many research articles on peer review, such as in Watkins 1979, Godlee & Dickersin 1999 or Rothwell & Martyn 2000 (in Mueller 2008). An empirical study in brain science analysed the consistency of the reviews for the same submitted manuscript and found out that the degree of congruence was only slightly higher than it would have been by random assignment (in this case by throwing a coin) (Rothwell & Martyn 2000, in Mueller 2008).

A revealing statement regarding the influence of other factors was made by Richard Smith who has been the publisher of the famous British Medical Journey (BMJ) for quite some time. He stated that placing an article in a

1 Kappa or Cohen’s kappa is a statistical measurement for inter-rater agreement for qualitative items and categories. It ranges from 0-1, whereas the range from 0.2-0.4 could be interpreted as a fair amount of agreement.

journal like Nature or Cell would be like winning the jackpot in the lottery because getting the submitted document through the peer review process is very much influenced by chance (Mueller 2008).

This inconsistency with respect to the quality of the review is very much due to the inherent subjectivity of the individual assessments of each reviewer. It is difficult to minimize it, as “the ideal reviewer should be totally objective, in other words, supernatural” (Ingelfinder 1974, in Mueller 2008).

The above mentioned idea to give reviewers guidance and teach them best practices “how to do the review” makes sense to some degree, also by spelling out the criteria in detail and explain what the criteria mean (Bridges 2009). According to Day & Peters and Curtis & Shattock these criteria which could be used to assess a manuscript could be “originality and innovativeness, relevance to previous work, building on and relevance to body of knowledge, evidence and objectivity; clarity of writing, quality of argument; theoretical and practical implications and meets editorial objectives” (Day & Peters 1994). All in all, these criteria very much go in line with the commonly used list of criteria for the assessment of scientific quality (Mueller 2008).

However, the problem of subjectivity will prevail regardless of how well instructed the reviewers may be. This holds especially true due to the fact that normally only two to three reviews are obtained per submitted manuscript. Finally, the decision on which article is published and which is not is in the hands of the reviewer, and the governance and decision-making by the editor/publisher mostly takes place in secrecy and not in an open and transparent manner. The above mentioned criteria and best practices list do not seem to be able to change much of this (Mueller 2008).

5 Quo vadis, peer review?

It is not a new phenomenon that peer review as means of scientific quality assurance and its fundamental utility is frequently being criticized. Research has discussed this for decades now. The most serious allegation is that peer review in form of a quality assessment for pre-published manuscripts is another form of censorship (Fröhlich 2003, Mueller 2008, Casadevall & Fang 2009). Beside this viewpoint alluding to an ethical objection against the use of peer review, research displays various nuances of how effective peer review is or is not.

The attitudes range from fundamental opposition to an understanding of peer review as being necessary and irreplaceable. The first attitude is for example articulated by the opinion of Richard Smith of the British Medical Journal who comments on peer review as being “slow, expensive, profligate

of academic time, highly subjective, prone to bias, easily abused, poor at detecting gross defects, and almost useless in detecting fraud” (Smith 1999, in Mueller 2008). Hornborstel thus states that “altogether and despite the numerous problems associated with peer review, this system is nevertheless irreplaceable: no one but peers will be able to judge the originality or the degree of innovation of a research proposal” (Hornborstel/Siekermann 2007). Drummond Rennie, deputy editor of the Journal of American Medical Association JAMA assumes that the fact that popularity of peer review in scientific quality assessment is even growing, despite the numerous problems mentioned above, is also due to the fact that investigators and research institutions see peer review as a convenient quality control mechanism, as they usually do not have to pay for it. Finally, he uses a historical quotation from Churchill, stating that “peer review is like democracy, which is [...] the worst form of government except all those other forms that have been tried from time to time” (Rennie 1993). Mueller goes in line with this argumentation that peer review is far from being perfect and states that another quality assurance system in science would generally be possible, however it would require very much time for the moment. Furthermore, he argues that peer review nowadays has a very important filter function in order to be able to ensure scientific quality of the manifold publications which are published online. According to him, this also holds true for open-access journals and he cites Stevan Harnad, one of the pioneers of the open access movement: “The refereed journal literature needs to be freed from both paper and its costs, but not from peer review whose ‘invisible hand’ is what maintains its quality” (Harnad 2000, in Mueller 2008.).

Benos et al. close on a more critical note by stating that subsequent research on peer review should not so much examine the methodological flaws of the review process as such, but they mainly take the purpose of peer review and its basic assumptions into account. One question which arises is if the reviewing peer must necessarily be an expert in the field of the submitted article. Furthermore, they provocatively ask if “unfettered publishing [in the web, for example, C.R.] of findings leads to more efficient or faster progress in science” (Benos et al. 2007).

6 Peer review and the EERQI-project

The EERQI-project uses peer review as a testing method to validate its created set of quality assessment indicators. EERQI stands for European Educational Research Quality Indicators and is funded by the EU Commission within the FP7 programme. The main aim of the project is to develop new, adequate indicators for the quality assessment of European research

publications and thus to enhance the visibility and competitiveness of European scientists and scientific research. This competitiveness is so far being limited by the still prevailing dominance of the bibliometric Social Science Citation Index (SSCI) as quality assessment tool, which concentrates mainly on Anglophone research.

Currently, there are quite a growing number of European projects and initiatives which are dealing with making especially European Research – particularly in the Educational Sciences, the Humanities and the Social Sciences – more visible on a global level. The project “Towards a Bibliometric Database for the Social Sciences and Humanities – a European Scoping Project” is an example. The main aim of the project is to explore the possibility of developing a bibliometric database for capturing the full range of research outputs from the Social Sciences and Humanities and to help to assess their impact. The underlying ratio is similar to the EERQI-project: in times of growing pressure from policy-makers and research funders for accountability and transparency in the scientific “output” process, performance measurement is one of the key topics, especially in areas such as the Humanities, in which established quality indicators to measure research output barely exist. With respect to the operationalization of such a database, the project initiators are discussing various strategies, for example a top-down approach which would involve creating a database at the European (or other international) level with established standardized rules and set of criteria to ensure full comparability of nationally provided data. An essential but difficult step would be the establishment of a basic threshold criterion (or criteria) to evaluate which research output is “worth” including into the database. Here peer review as a method comes back into play. It is envisaged that “a basic or minimum threshold criterion could focus initially on scholarly articles in peer-reviewed national and international journals, and on scholarly books that have been subject to a peer-review process.”² This project has been mentioned here to explain the context and the recent debates that surround the EERQI-project as well as the central position peer review still occupies as a quality assessment tool also in projects which are looking for new quality assessment strategies.³

As stated before, EERQI aims at combing bibliometric analysis and reference linking, semantic and linguistic analysis of full texts and citations, text strings and metadata correlations. These methodologies combined will determine new indicators for the evaluation of quality in research publications.

2 Martin, Ben et al.: “Towards a Bibliometric Database for the Social Sciences and Humanities”, Sussex, England, June 2009.

3 For more detailed information on the SSH – European Scoping Project compare: *ibid.*

Additionally, a content base consisting of research publications in four European languages (English, French, German and Swedish) is created and included into the project. The soon to start verification process of the project will be carried out by the EERQI team in close collaboration with experts in the field – with peer review being one of the central testing methods. Experts from Educational Science are invited to use the developed indicators in a peer review process to test their reliability and validity. Thus, the procedure of peer review as such is an important part of the heuristic process of evaluation and calibration of the EERQI set of quality indicators.

With regard to the EERQI-project, the project team seems to be well aware of both constraints and the value of the method of peer review as such. The positive as well as the negative aspects of using peer review – as described above – were taken into account when the testing procedure has been designed. The main concerns with peer review mainly revolve around the question of limited reliability of the method as such. However, there is no other viable alternative than the peer review procedure to test the quality (understood as having both reliability and validity) and the utility of the EERQI set of quality indicators.⁴

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. (2009): A note on peer review of academic papers (not published essay).
- Casadevall, A./Fang, F. (2009): Is Peer Review Censorship? In: *Infection and Immunity*, American Society for Microbiology (Eds.), edition Apr. 2009, New York: Blackwell, pp 1273–1274.
- 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.
- Cicchetti, D. V. (1997): Referees, editors, and publication practices: improving the reliability and usefulness of the peer review system. In: *Science and Engineering Ethics*, Vol. 3, pp. 51–62.
- 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.

4 For further information on EERQI and project updates cf. <http://www.eerqi.eu>.

- Daniel, H.-D. (2005): Publications as a measure of scientific advancement and of scientists' productivity. In: *Learned Publishing*, Vol. 18, pp. 143–148.
- Day, A./Peters, J. (1994): Quality Indicators in Academic Publishing. In: *Library Review*, Vol. 43 (7), pp. 4–72.
- Fröhlich, G. (2003): Anonyme Kritik: Peer Review auf dem Prüfstand der Wissenschaftsforschung. In: *medizin-bibliothek-information*, Vol. 3 (2), pp. 33–39.
- Godlee, F./Dickersin, K. (1999): Bias, subjectivity, chance, and conflict of interest in editorial decisions. In: F. Godlee/T. Jefferson (Eds.): *Peer Review in Health Sciences*. London: BMJ Publishing Group, pp. 57–78.
- Hames, I. (2007): *Peer Review and Manuscript Management in Scientific Journals. Guidelines for Good Practice*. London: Blackwell Publishing.
- Harnad, S. (2000): The invisible hand of peer review. In: *Exploit interactive*, Vol. 5.
- Hornborstel, S./Simon, D. (2006): Wie viel (In-)Transparenz ist notwendig? Peer Review revisited. In: *iFQ-Working paper No. 1*, Bonn.
- Hornborstel, S. (1997): *Wissenschaftsindikatoren. Bewertungen in der Wissenschaft*. Opladen: Westdeutscher Verlag.
- Meier, K./Brudney, J./Bohte, J. (2006): *Applied Statistics for Public and Non-profit Administration*. Richmond (USA): Wadsworth Publishing, 6 edition.
- Mueller, U. T. (2008): *Peer-Review-Verfahren zur Qualitätssicherung von Open-Access-Zeitschriften – Systematische Klassifikation und empirische Untersuchung*, (Dissertation). Berlin: Humboldt-Universität.
- Rennie, D. (1993): Editorial peer review: its development and rationale. In: <http://resources.bmj.com/bmj/pdfs/rennie.pdf>.
- Rothwell, P.M./Martyn, C.N. (2000): Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone?. In: *Brain*, 123 (9), pp. 1964–1969.
- Smith, R. (1999): Opening up BMJ peer review: A beginning that should lead to complete transparency. In: *British Medical Journal*, 318, pp. 4–5.
- Wager, E./Godlee, F./Jefferson, T. (2007): *Erfolg im Peer Review. Wissenschaftliche Begutachtungen durchführen und überstehen*. Bern: Huber.
- Watkins, M.W. (1979): Chance and Interrater Agreement on Manuscripts. In: *American Psychologist*, Vol. 34 (19), pp. 796–797.
- Weingart, P. (2005): Das Ritual der Evaluierung und die Verführbarkeit der Zahlen. In: *Die Wissenschaft der Öffentlichkeit* (P. Weingart Ed.). Weilerswist: Velbrück, pp. 102–122.
- Weller, A.C. (2001): *Editorial Peer Review. Its Strengths and Weaknesses*. New Jersey: Information Today.

*The author is Research Assistant of EERA
(European Educational Research Association).*