



EUROPEAN EDUCATIONAL RESEARCH  
**Quality Indicators**



## **An Analysis and evaluation of existing methods and indicators for quality assessment of scientific publications**

European Educational Research Quality Indicators (EERQI)

A project funded by the European Commission, DG Research, Socio-economic sciences and humanities in the 7th Framework Programme

Prepared for the 1st EERQI Workshop June 20-21, 2008, Leuven

Report Work Package 3, 12.06.2008

**Angela Vorndran**

**Alexander Botte**

German Institute for International Educational Research, Frankfurt / M.

Vorndran@dipf.de

**AN ANALYSIS AND EVALUATION OF EXISTING METHODS AND INDICATORS FOR QUALITY ASSESSMENT OF SCIENTIFIC PUBLICATIONS** **1**

**EUROPEAN EDUCATIONAL RESEARCH QUALITY INDICATORS (EERQI)** **1**

**A PROJECT FUNDED BY THE EUROPEAN COMMISSION, DG RESEARCH, SOCIO-ECONOMIC SCIENCES AND HUMANITIES IN THE 7TH FRAMEWORK PROGRAMME** **1**

<b>1. Introduction.....</b>	<b>3</b>
<b>2. Analysis of text references, citations and links.....</b>	<b>4</b>
2.1 New Content.....	4
2.2 Link analysis: adapting citation analysis to electronic documents.....	4
2.3 New statistical methods based on citation analyses.....	5
2.4 Semantic text analysis emanating from citation analysis.....	6
<b>3 Analysing elements of the production/editorial process .....</b>	<b>6</b>
3.1 Expanding criteria for production quality.....	6
3.2 Editorial analysis of non-journal document types.....	7
<b>4 Semantic text analysis.....</b>	<b>8</b>
4.1 Full text analysis.....	8
4.2 Metadata analysis.....	9
4.3 The aspect of multilinguality.....	9
<b>5 New retrieval and clustering techniques.....</b>	<b>9</b>
5.1 Concept Similarity Analysis.....	9
5.2 Co-citation Analysis.....	10
5.3 Bibliographic Coupling.....	10
5.4 Co-word Analysis.....	10
5.5 Cluster Analysis.....	10
5.6 Social Network Analysis.....	11
5.7 The aspect of multilinguality.....	11
<b>6 Online Usage Metrics.....</b>	<b>11</b>
6.1 Log file analysis.....	12
6.2 Journal usage statistics.....	12
6.3 Download statistics.....	12
6.4 Link Analysis.....	12
<b>7 Educational research-specific procedures.....</b>	<b>12</b>
<b>8 Summary.....</b>	<b>13</b>
8.1 Approaches based on characteristics of the document.....	14
8.2 Approaches based on the production process.....	14
8.3 Approaches based on use and impact of a document.....	15
8.4 General Suggestions for the Workshop.....	15
<b>Bibliography.....</b>	<b>16</b>
<b>Annex: Key numbers from the Journal Citation Report in Educational Sciences.....</b>	<b>19</b>

## 1. Introduction

Quality assessment in academic research is to a huge extent focused on the methods of citation analysis employed by the Institute for Scientific Information, today Thompson Scientific, in their *Science Citation Index (SCI)* with its by-products, the *Social Science Citation Index (SSCI)*, the *Arts & Humanities Citation Index (AHCI)* and the *Journal Citation Report (JCR)*. These bibliometric instruments cover about 5,800 journals and provide tools for statistical analyses of citation frequencies related to articles, authors and their affiliations. Especially in the subject areas of sciences, technology and medicine, citation analysis is the standard system of quality assessment. As the publication cultures in these subjects adopted the habit of mainly publishing in English language journals, the *SCI* is able to deliver what can be seen as feasible results. For other subject areas, foremost social sciences, humanities, and education, where the focus of the EERQI project is, the publications in the *SSCI* and the respective impact factors prove of little use. The diversity of publication channels and strategies, as well as the national, cultural and linguistic orientation of educational sciences create barriers which make it inappropriate to apply the citation analysis method on European educational scholarly publications. Only a small number of journals which are important in educational research are covered by ISI (see Annex). Most educational scientists publish in their national languages which are not covered by the *SSCI*. Furthermore monographs and edited volumes play an important role in educational research, but these are not considered in ISI's citation analysis process.

In this situation it becomes obvious that there is a strong need for new indicators to provide judgment on the quality of European publications with a special focus on publishing cultures in the social sciences and educational research. But, apart from the above mentioned criticism of citation analysis, the use of citation analysis has some appealing advantages which cannot be easily equalled: the indicators derived from citation analyses are based on external (peer) judgment and they are highly scalable. Furthermore, citation analyses make use of criteria learned from both the publishing process (selection of journals) and from the use/reception of the publication (number of citations, impact factor).

The EERQI project aims at creating a basis for the assessment of research quality in the social sciences in the European context. It is doubtful that this aim can be reached by a single new method of assessment. A multi-faceted approach to developing new indicators and measurements which work in combination as a significant improvement to the current state is considered necessary by the EERQI partners.

The following approaches will be evaluated for the specific purposes of this project:

1. Improving the current state of citation analysis by expanding the corpus of referenced publications and by extending the methods of analysis
2. Analysing elements of the production/editorial process (Metadata, Peer-Review)
3. Semantic analysis of full text and rich metadata (professional keywords, abstracts, affiliation)
4. New retrieval and clustering techniques
5. Statistical analyses of the web (webometrics)
6. Educational research-specific procedures.

## **2. Analysis of text references, citations and links**

Bearing in mind that citation analysis is a well-established approach which still attracts the majority of bibliometric research, the project should make an effort to emulate this method by removing acknowledged barriers which arise when citation analysis is applied on educational sciences literature.

### **2.1 New Content**

One of the main drawbacks of *SCI* is the exclusion of too many publications relevant for educational research. EERQI will build up a new content base of relevant publications and citation analysis will be applied on the new content of the EERQI database. Not only journal articles, but also monographs, edited volumes, online documents and websites will be included in full text or as metadata.

In the *SSCI*, all citations to ISI journals, non-ISI journals and books are indexed but only those to *SSCI* journals are linked to or are included in any citation counts. Citations to other material are neglected in ISI's analysis but could be interpreted by new forms of 'expanded' citation analysis which have been started e.g. by Moed and Visser (cf. Moed 2005: 153f.). The ISI competitor *Scopus* by Elsevier already provides better opportunities to follow citations going beyond the core spectrum of *Scopus* journals.

A web-based corpus of digitized documents with traditional references offers opportunities to improve citation analysis, as all bibliographic data available on the web can be used to locate and resolve references. Based on the documents gathered by Google Scholar a new bibliometric service called 'Publish or Perish' (<http://www.harzing.com/resources.htm#/pop.htm>) offers different indicators of citation analysis for authors of Google Scholar documents. This service developed by the Australian professor Anne-Wil Harzing gives a preview of the opportunities for analysis which digitized scientific documents allow.

So, citation analyses in the traditional form, but based on a larger body of publications can be relevant for the purposes of EERQI namely by providing a lot of data about cited publications which can be analyzed in order to detect the true culture of referencing in educational research in Europe. The analyses could answer the question if and how researchers in Europe refer to each other across country borders, as well as across language borders within a single country.

### **2.2 Link analysis: adapting citation analysis to electronic documents**

In webometrics link analysis is often interpreted as a further development of citation analysis adapted to web documents (cf. Ingwersen 2006, Thelwall/Ruschenburg 2006, Thelwall 2007:8). The analysis of link structures defines inlinks that are directed at a website from outside and outlinks which are linking from a website to another one. Ingwersen (2006) points out that links differ from citations in that they are functional in nature, no conventions for linking exist and links are seldom valued negatively. The evaluation of links directed to a website can be standardized e.g. by dividing their number by the number of web pages or employees at the institution. Analogous to citation analysis Ingwersen (1998) also employed a "web impact factor". It adapts citation analysis to the web context when measuring a website's impact by counting the number of websites directing links to it. Internal links on websites are compared to self-citations. In contrast to the Impact Factor only the number of websites with links to one website are counted not the intensity of the links. Another issue relevant in link analysis is the continuity of web documents and their included links.

Van Impe and Rousseau (2006) diversify their analysis of links when focussing on "web-to-print" citations. These special citation types describe references from the web to printed

sources like journal articles, conference proceedings or books. Their research based on data from five journals of research fields in the humanities did however not yield results as the number of citations proved to small for reasonable conclusions.

For the purposes of EERQI it will be most important to appraise links directed at scientific documents on the web rather than at normal websites. Internal links are also not a valid indicator for the impact of an electronic document.

As several researchers value links as comparable to citations, link analysis offers an opportunity to expand citation analysis to all scientific texts on the Web instead of *SSCI*'s limitation to selected journals. The number of links directed at texts thus indicates their impact and can serve as an evaluation criterion similar to citation analysis.

### **2.3 New statistical methods based on citation analyses**

Based on the normalized *SSCI* data about numbers of citations in a limited set of selected journals a growing number of secondary indicators were derived. It may be worthwhile to look at them.

Such indicators deriving information from processing citation counts include:

- The h-index: relates the number of publications to the times these are cited. An index  $h$  prescribes that  $h$  papers of an author have received at least  $h$  citations each, while the other papers have received less than  $h$  citations (Hirsch 2005).
- The g-index also relates citations to numbers of publications and is a variation of the h-index. It defines in a ranking of articles with decreasing citation counts a set of top articles which on average have received at least  $g$  citations (Egghe 2006).
- The Journal citation identity counts the ratio of the total number of citations of an author and the number of different journals they are published in. Thus the outreach of an author over different journals is measured (Bonnieve-Nebelong 2006). This approach is based on the author citation identity proposed by White (2001)
- The Journal Citation Image (Bonnieve-Nebelong 2006) applies a co-citation analysis between different journals to measure quality. The quality of the co-cited journals can be determined by recurring to the Journal Impact Factor or other criteria. The Journal Citation Image is also derived from a citation analysis based on single authors, namely the citation image of authors depicting who they are co-cited with (cf. Cronin/Shaw 2002).
- The Journal Diffusion Factor (Rowlands 2002) counts the ratio of number of different citing journals and number of citations in a given time frame. Thus it measures the breadth of the impact of a citation in a subject field. For a certain journal and a selected time window a journal diffusion factor can be calculated which indicates in how many different journals citations from this journal are published.
- The New Journal Diffusion Factor (Frandsen 2004) employs the impact of a journal i.e. the Journal Impact Factor as one part in the calculation of the Journal Diffusion Factor. It builds the ratio of the number of different citing journals and the number of citable units published in a given time frame.
- The Eigenfactor derives a ranking of journals from using data from the *Journal Citation Report*. In this method citations from high-quality journals

are more valuable than citations from other journals. A citation is evaluated by the ratio of influence of the citing journal and the total number of citations in that journal. Another indicator, the Article Influence of a journal, is calculated by dividing the Eigenfactor by the number of articles (Bergstrom 2007). The data of several journals can be obtained from <http://www.eigenfactor.org>.

Given the fact that all these indicators are based on citation analyses, the argument must be respected that the reliability of these indicators depend very much on the document base from which the citation analyses have been extracted from.

## **2.4 Semantic text analysis emanating from citation analysis**

A method called “Reference-Directed Indexing (RDI)” is employed to extract feasible index terms from sentences enclosing citations. When an author includes citations in his text he also describes their content and key ideas. RDI makes use of these author statements about cited literature to extract keywords. From analysing a high number of citations a set of most used keywords can be defined and considered as candidates for index terms. Ritchie defines a previously established window of a fixed number of words surrounding the citation to submit to RDI (Ritchie 2006, Bradshaw 2003).

A similar method is employed by Schneider (2006) who extracts noun phrases from citation contexts to detect “concept symbols” which stand for ideas exemplified by the cited passages. According to the frequency of a noun phrase in different citing texts the most prominent concepts are identified and can form concept groups in the field and serve for naming clusters.

The applicability of this method is obviously bound to the availability of a digitized full text corpus in which it is technically possible to process semantic text analyses (e.g XML documents).

## **3 Analysing elements of the production/editorial process**

The following methods to gather indicators of quality from characteristics of the editorial process are not new, in fact many of them are preliminary examinations done by ISI before they accept a journal to be added to their Citation Indexes. But as already suggested for the case of citation analyses (Chapter 1) these methods can be expanded on more types of documents and the sources of information and the criteria of quality can be extended.

### **3.1 Expanding criteria for production quality**

In respect to journals ISI uses very strict criteria with a special emphasis on the aspect of international relevance of a journal. Even though the European Reference Index for the Humanities (ERIH 2007) does not impose the same formal restrictions as ISI (e.g. availability of English title translation and abstract) the ERIH criteria used for the assessment are also very much focused on international relevance and make great demands on the publications included:

- a genuine, varied and regular international cohort of contributors and readership
- a consistently high-quality scholarly content
- a broad consensus within the field concerning international status and visibility, insofar as possible.

As further attributes of journals of high quality they name:

- Active international advisory board

- Openness to unsolicited contributions
- Highly discriminating and selective in the choice of articles published
- Publication on time and to an agreed schedule (ERIH 2007).

For the purposes of EERQI it would be necessary to account national conditions and national relevance of publications considerably more than do these criteria. Reputation and representation within a national scientific community and publication culture should be estimated more. Guidelines and assessment criteria of national evaluation boards should be examined for aspects which could be integrated in the new EERQI Prototype Framework.

For journals, alternative approaches to the ISI impact factor have been developed to assess quality. One up-to-date example of this is the “Publication Power Approach” proposed by Holsapple (2008). He uses the authors’ decisions as to in which to publish as indicator for journal quality. His ‘intensity score’ measures how often on average a journal is used as outlet for each researcher in the subject area. The ‘breadth score’ informs about the share of all researchers of the subject area who have published at least once in the journal in question. High results on both scores indicate high quality of the journal.

### 3.2 Editorial analysis of non-journal document types

Non-journal document types can also be ranked by editorial processes that were combined with their publication, even though the detection of these processes may be more complicated for books and the different types of online documents. If selection criteria or quality assessment are applied in the publishing process, the conditions of this editorial process can be used for a general evaluation of research quality. Moed also includes a reference to “the size and composition of the universe of citation sources” in an evaluation of document quality (2005:314).

Information about the editorial process and about the authors can be gathered from the analysis of formal metadata. Formal characteristics regarding the publication type (monograph, edited volume, journal article, web etc.), author(s) (affiliation, awards, membership in committees etc.) or the publishing institution (renowned publisher, peer-reviewed journal, research series etc.) can be collected here. The evaluation of these data is not always easy, because they are not absolute measures and an assessment of the respective quality is not necessarily objective and clear.

Another approach to quality assessment in the social sciences and humanities with a focus on the heterogeneous publication types existing in these research areas has been developed in Norway. Publications are evaluated concerning their publication type and the publication channel. Regarding the publication type three categories exist: article in an ISSN-title, article in an ISBN-title and ISBN title. These are allocated different values according to the estimation of their quality. These scores are in turn combined with the ratings for the publication channel. These are established in a peer rating process. In two categories journal and book publishers are rated by researchers in the respective research areas. Only 20 % of the publishers are allowed to receive the top score 2, while all other publications are rated 1. Both attributes are thus rated according to the following table:

Publication type	Level 1	Level 2
Research article in a journal or series (ISSN)	1	3
Research article in a book (ISBN)	0,7	1
Scientific book (ISBN)	5	8

(cf. Sivertsen 2007)

The data necessary for this quality assessment process is collected in the database FRIDA (<http://wo.uio.no/as/WebObjects/frida.woa/wa/omtjenesten?la=en&om=kompkat>).

The EERQI project may also test methods to detect editorial characteristics of a document by a semantic analysis of title pages and introductions.

Ratings of different publication channels (publishers, serial publications apart from journals, associations, scientific institutions, repositories) can be based on ‘reputation’ surveys among researchers in the subject area. Publishing houses or digital repositories may be judged by peer rating in the same way as journals have been for ERIH.

## 4 *Semantic text analysis*

### 4.1 Full text analysis

Semantic text analysis is a powerful tool in that it enables an automatic treatment of texts instead of or in addition to intellectual indexing. Methods can be employed which can discover semantic patterns in a text and use these results for retrieval purposes. In most contexts, linguistic keyword search is the prevalent retrieval instrument. Here different approaches to normalize keywords by lexical indexes or stemming processes are applied. But when entering into semantic text analysis far more interesting results can be derived on the basis of semantic attributes of keywords in contrast to the perception of words as alphanumeric strings. These types of sophisticated semantic analyses can be more powerful, especially when they are directed towards the detection of qualitative aspects of a text.

The first step of semantic text analysis is the recognition of some categories of descriptors which are relatively easy to detect by an information extraction tool, namely names, dates, terminology, etc. These terms are described as “core information” (Sandor 2006). An advanced level of semantic text analysis can be achieved by establishing a concept-matching framework. In this process important concepts in the text are identified according to the respective subject area. These concepts can be novelty, contradictions etc. Each of these concepts is supplied with a range of keywords. Those can be detected in the “peripheral information” in the text, which consists of rhetorical expressions and frequent terms embodying the previously defined concepts. A parser then detects co-occurrence of existing concepts in a text and allows for recognition of relevant texts.

The following sentence gives an example how concepts are detected in a sentence:

**“Models** of neurodegenerative disorders **are challenging the classical** defining **role** of tangles in neurotoxicity.” Keywords for defined concepts are printed boldly. “Models” and “role” represent a concept of ‘idea’, “challenging” a concept of ‘contrast’, and “classical” detects a concept of ‘past’. The co-occurrence of these different concepts in the sentence qualify the text as relevant. With complex concepts it is also possible to detect co-occurrence of concepts across sentence boundaries (Sandor 2006).

Another approach to semantic text analysis emanates from the detection of cue phrases in a text. These “semi-fixed cue phrases” (Abdalla/Teufel 2006) e.g. “This paper proposes a novel approach” have to meet certain criteria to be able to identify relevant parts of documents. They are subject to varying syntactic relationships and can show lexical variants. In order to achieve a list of cue word which can be part of cue phrases Abdalla/Teufel employ a lexical bootstrapping algorithm. Like the verb “propose” and the noun “approach” in the example, the words in cue phrases must be syntactically and semantically related. Furthermore semantic constraints are introduced to ensure the relevance of the phrases in question, e.g. reference to

the current and not older publications. In order to detect cue phrases only those words which show a syntactic relationship, key words, and fit the semantic constraints are considered.

## **4.2 Metadata analysis**

If rich metadata (affiliation, abstract, formal descriptions of the text, subject keywords) are available, semantic analyses can be improved, because the structure of metadata inherently provides a semantic allocation of terms and data. No relevant bibliometric studies in this field seem to have been done thus far – which may be an area for the EERQI Project to study.

## **4.3 The aspect of multilinguality**

The need of the EERQI project to solve the problems of different languages in Europe will probably not result in specific multilingual quality indicators, but in methods how to apply approaches of semantic analyses on publications in different languages. For subject metadata (keywords) some multilingual educational thesauri are available. The vocabulary of the thesauri and their scope-notes (explanations of different meanings in different languages) may also be of considerable help for translating key concepts which are derived from semantic analyses.

# **5 New retrieval and clustering techniques**

The following approaches, only scarcely detailed, are in fact part of linguistic approaches, but they were developed primarily for purposes of retrieval. The purposes of retrieval are not identical with the purposes of quality detection, but a successful retrieval reveals findings of high relevance. And the relevance of a document is an essential element of its quality. On that condition good instruments of retrieval are a precondition of detecting quality documents. Metadata which characterise the content of a document very precisely can be used to pre-select the corpus of scientifically relevant documents. If subject metadata are not available, semantic processes to develop the framework of scientific concepts are necessary to support these essential processes of pre-selection of the relevant document corpus.

For the automatic detection of subjects treated, specific forms of text recognition can be applied. A common basis of texts represented by using the same concepts, terms or citing the same authors can be interpreted as an allocation to thematic groups. The texts belonging to one subject area can thus be assigned to spots on a topic map of the subject. Research can thus be evaluated on the one hand on a cognitive basis when similarities are found concerning concepts or co-word use or on the other hand on a level of communication structures when common referencing is considered (cf. Van Raan 2004).

## **5.1 Concept Similarity Analysis**

For the analysis of similar concepts in texts to develop a mapping of a research field the relevant concepts have to be identified and supplied with keywords. A parser then extracts nouns and noun phrases from a document collection and detects the most frequent terms as indicators of concept affiliation. Subject maps thus are established by measuring the similarity of texts in terms of how many concepts those have in common (cf. Van Raan 2004). The analysis results in a similarity matrix which is converted into a bibliometric map clarifying the different subfields of a subject area and the documents belonging to them. The map can also give an overview over research activities of researchers or research groups.

## 5.2 Co-citation Analysis

Co-citation analysis refers to documents which are related because they are cited together by several other documents. The relation is not established actively by the documents in question but is created by third documents referring to them. By interpreting those connections between research papers a mapping of the research area can be achieved.

Co-citation analysis requests a provision of source citation data and thus is not as easily applicable as an analysis based on word occurrence. But it can depict the relationships between different research areas and researchers more obviously. It is mainly focussed on the past structures of a research field. Van den Besselaar/Heimeriks (2006) report the well known criticism referring to the often insufficient coverage of papers and to the overestimation of one dimension of behaviour, namely citing.

## 5.3 Bibliographic Coupling

Bibliographic Coupling, sometimes also called Bibliometric Coupling, is related to co-citation analysis in that it compares references in different documents to detect co-occurrences, too. The focus of this method lies on the active combination of two documents by citing both in source texts. Two papers are related because they cite the same third text. Contrarily in co-citation analysis two papers are related because they are both referenced in a third paper. Bibliometric coupling thus is more feasible to detect research fronts (cf. Van den Besselaar/Heimeriks 2006, Garfield 2001, Glänzel/Czerwon 1998).

## 5.4 Co-word Analysis

Co-word analysis can be applied to titles, abstracts or full texts. It can result in different levels of granularity of maps covering a whole discipline or certain subfields. It is related to concept similarity analysis in that it derives information from unconscious actions leading to relationships between texts, in contrast to conscious citing of other documents. The results achieved offer information about topics worked with and concepts applied. A further approach to co-word analysis by Van den Besselaar/Heimeriks (2006) combines co-word analysis with citation analysis and interprets the words used in a text with reference to the documents used as references. Thus the meaning of words used can be clarified by its reference framework and changes in the meaning of words can be discovered. Zhao/Strotmann (2007) also takes into consideration that co-word analysis could offer the possibility to expose future research topics in contrast to citation analysis which naturally refers to topics discussed in the past.

## 5.5 Cluster Analysis

Cluster analysis strives to evaluate research output in a certain subject field according to basic topics and characteristics. First the raw data of the publications under investigation like citations, concepts or words are represented in a matrix. Then similarities of these raw data are calculated for detecting co-occurrences using arithmetic averages. Those are often visualized in a tree diagram. The similarity matrix resulting from these calculations can be used to define clusters by employing a cluster algorithm. These clusters correspond to subfields of a discipline. They can be presented in Multidimensional Scaling, i.e., graphs indicating the similarity by distances between dots representing the respective subfields (cf. Van Raan 2004, Fernández-Cano/Bueno 2002). Clusters can be designed in such a way that they can be used to detect quality. They visualize data concerning their similarities by representing them in a graph. The proximity of several data on the graph forms a map where clusters of similar qualities can be detected. In respect to the data provided, judgements regarding quality, subject affiliation or other attributes of documents can be illustrated.

Fernández-Cano/Bueno (2002) attained a qualitative clustering of Spanish educational research journals by undertaking a survey where journals should be rated in quality with scores from 1 to 12. Appraising the ratings of 112 participants and 22 journals they conferred the results to a similarity matrix and thus created a cluster structure for Spanish educational research journals. Areas of specialization and a number of basic journals could be identified as well as three clusters identified in similarity of quality assessment by the participants.

Cluster analysis in most cases is applied to identify subfields of a discipline, publications belonging to them and changes in research interest over time. It is not a quality indicator in itself but it can be applied to different contexts like surveys which deal with quality assessment or co-word or co-concept analyses which are more focused on the thematic positioning of a publication in a research field.

Bhattacharya/Basu (1993) interjected earlier that statistically determined clusters or word pairs must not always show similarities regarding content.

## **5.6 Social Network Analysis**

Social Network Analysis as a means to describe the relationship between different persons, institutions and states via their websites is a subfield of link analysis. It uses data from the Web to investigate social relationships, e.g., communication methods, use of media or productivity among researchers. Relationships between different entities are summarized in a matrix. Social Network Analysis enables the researcher to identify a central website which has the shortest connection to most of the other websites of the network. Network nodes are connected by inlinks and outlinks. Further attributes within the network are ‘betweenness centrality’ which measures the number of times a node links two other nodes and ‘closeness centrality’ which identifies the website with the shortest connection to all others in the network. ‘Negopy centrality’ quantifies the average number of hyperlinks to reach all other websites from one website (cf. Park/Thelwall 2003). Other attributes concerning the structure of the network can be measured so that a detailed analysis of the communication structures between websites and the organisations represented there can be executed.

## **5.7 The aspect of multilinguality**

To tackle the problem of multilinguality in the research contexts of the different European countries, a method will have to be employed to allow multilingual retrieval in the EERQI database. As the content of the database incorporates texts in different national languages, a query will have to be translated to retrieve all relevant texts. Cross Language Information Retrieval (CLIR) allows the user to apply queries in one language to retrieve texts in another language. This system is based, on the one hand, on dictionaries and thesauri which are applied to translate the query. On the other hand a ‘Comparable Corpus Translation system’ (COCOT) can be used. In this case collections of topic-related texts in different languages are gathered. By employing statistical methods similarities between the text corpora are detected which create knowledge about the meaning and use of a word. Texts are thus first submitted to a translation with a machine-readable dictionary. Words with multiple alternatives for translation are processed with CLIR, where an analysis of co-occurrence of the word in the corpus of the target language is conducted to determine the right expression in the context (cf. Oard et al. 2008, Talvensaari et al. 2007).

## **6 Online Usage Metrics**

The field of Webometrics offers several opportunities to derive quality statements for web publications from the data gathered. Using statistical methods on the Web can not only give

information about the interdependency of web offerings, like link analysis, but also provide data on the usage of electronic documents. The usage data can offer interesting information about numbers of downloads but are not always easy to interpret.

## **6.1 Log file analysis**

Log file analysis provides the bibliometric researcher with data derived from real actions executed by users. It has the advantage of offering data which is not filtered or influenced either by the actions of the users or by the preparation of the data. It is independent of the users' perceptions about their actions or intentions and focuses on the actual action. It can also be expanded to cover a long time frame. The disadvantages of a log file analysis lie mostly in the uncertain providence of the user data. It is difficult to identify a user through IP addresses which might not even indicate the country where the user works. It also is important to distinguish visits by search engines from those of actual users. Furthermore, the log file does not provide any evidence regarding the extent of use. The length of time logged in at a specific site is not always available and reloading of pages from a local cache remains outside of such statistics. (cf. Jamali 2005).

## **6.2 Journal usage statistics**

Journal usage statistics have become a lot easier to collect and analyse since most journal publications have become available online. Unlike re-shelving studies in libraries, the record of access to the electronic article gives more precise information about which part of a journal volume was accessed and how many times an article was looked at. Shepherd (2007) introduced a 'Journal usage factor' which is based on the usage data provided by the project COUNTER. It relates the total usage for a specified period to the total number of articles published online in this period. The project COUNTER stands for "Counting Online Usage of Networked Electronic Resources" and is an international initiative to provide standards for online usage reports produced by vendors. To standardise user data in format and information structure and to facilitate access, comparability and exchange of this data, COUNTER has issued a 'Code of Practice' which has been adopted by the member organisations.

## **6.3 Download statistics**

Download statistics represent a part of Journal usage statistics. The numbers of downloads are an objective indicator which can be assigned to a single document and analysed in reference to time, size and provenance (through log file analysis). Download numbers cannot give information about the actual use of an article, but they at least signal interest in the publication. Employing the download data as a ranking criterion Darmoni et al. (2000) propose a 'Reading Factor (RF)' which can be applied in collections of electronic texts like digital libraries. It relates the download frequency of a journal to the total number of downloads of journals offered in the collection. Bollen (2005) goes a step further and establishes "reader generated networks", emerging from texts downloaded in the same session by one user.

## **6.4 Link Analysis**

See 2.2

# **7 Educational research-specific procedures**

Education as a subject field belonging to the social sciences does not traditionally rely to a high degree on the quality assessment instruments provided by the ISI Citation Indexes and

other citation analysis methods as explained above. As a subject-specific evaluation method therefore, peer rating is the method most often applied to publications in educational science. Quality assessment systems like the British Research Assessment Exercise (RAE 2006) and the European Reference Index for the Humanities (ERIH 2007) are examples of peer review-based quality judgments applied in the social sciences and humanities.

Judgements and ratings by peers can be organized in different modes and can be discussed under discipline specific aspects. The task of the reviewer can also cover different kinds of decisions like the evaluation of quality or a judgment with respect to the article fitting into the scope of the journal. While peer review has the advantage of being highly specialized and involving competent researchers judging the quality of articles in their fields, it also reveals some disadvantages. The quality judgment is naturally subjective, and it can be influenced by external factors like social or political circumstances. Furthermore, studies have proved that peer review judgments show a low degree of agreement (Moed 2005, Bornmann 2003). Judgments can be influenced by the reputation of the author, bibliometric information about high citation rates or a reviewer's bias towards his own subject area. Another drawback is that many peer judgement processes are not transparent and reasons for a judgment are not made public.

If EERQI considers the application of a module of peer judgement several discipline specific aspects may be discussed:

- The estimation of different scientific 'schools' which may accommodate with statistical evaluation modes more or less
- The role of publications which transfer research results into fields of application (see trend in England reported about in the paper of David Bridges).

## **8 Summary**

This basic listing of approaches shows that the quality of research output can be evaluated in different ways. Already at the beginning of this project, it can be foreseen that nearly all alternative methods and indicators that can be suggested by EERQI will depend on the availability of digitized content (full text or standardized metadata) in an as complete form as possible. During the project, we can focus on exemplary publications which converge to these prerequisites, but we have to be aware that the EERQI content base will not be representative for all educational research publications in Europe. It will serve as a preliminary content base for mainly electronic fulltexts in educational research and offer a first basis for testing and developing new quality indicators. Ideally a content base and search engine which includes all relevant educational research would be necessary to make objective judgements regarding research quality. Such an instrument is difficult to obtain due to time restrictions in the content gathering phase, copyright restrictions and the heterogeneity of content in the educational sciences.

In the follow-up of the Hamburg meeting Ton Mooij has suggested a conceptual framework for the configuration of new indicators which refers to three main concepts ( or 'anchors'), to which a bibliometric method or indicator can be attached:

1. characteristics of the document itself
2. characteristics of the production (editorial) process
3. characteristics of the use of the document.

We will follow Ton's framework in order to make some preliminary suggestions as to which concepts should be tested and which pre-requisites and possible limitations are connected to

these approaches. Ton's framework may also be a scheme for a plenary discussion of the working group results!

## **8.1 Approaches based on characteristics of the document**

The concept of EERQI is focused on methods referring to the characteristics of the publications themselves. These approaches have mainly been described in chapters 4 and 5 (Semantic and Cluster analyses). For EERQI the semantic analyses will be central according to the work programme. Semantic text analysis can be very helpful on the basis of a clearer perception of the contents of a text and keywords to facilitate information retrieval. To improve quality assessment, a clear notion about the contents of a text is necessary and crucial to the successful process of information retrieval. Core information and keywords can be extracted from the text and connected to lists of synonyms. A semantic text analysis (see chapter 4) can in itself further quality assessment if certain concepts in a text can be detected which are predefined as signs for high quality.

The methods connected with clustering techniques and strategies to create maps of scientific fields could be helpful with respect to information retrieval and the provision of alternative keywords and synonyms for selected fields. However, they provide more of a sociological overview on a research field and results achieved from these methods are not primarily relevant for the quality assessment strived for in the EERQI project. Texts could be evaluated in the context of their providence from the core of a subject area or rather the periphery, but that does not necessarily influence the quality judgment.

All approaches in these concept fields are based on the electronic processing of data. Semantic analyses require the electronic availability of the full text or in some cases at least rich metadata.

## **8.2 Approaches based on the production process**

Quality assessment via the conditions of the publication process (see chapter 3) should be a necessary element of our framework for relevance assessment. EERQI should consider the impact and prestige traditionally attributed to publications in the different publication types which still are very distinct in the social sciences. To facilitate the evaluation regarding the conditions of the publication process it would be useful to name publishers and institutional editors whose names are signals for quality. The practices used by renowned publishers could serve as suggestions for innovative methods of analysis which could be developed and tested here. In this context peer review processes and other editorial instruments such as an active advisory board and a high selectivity of contributions guaranteeing varied contributors and openness to unsolicited contributions among others should be considered.

Furthermore, information about the author and his context, his affiliations, qualifications and other activities in the research context should be considered as a hint for his importance in the research field. This data could partly be gathered from metadata or special databases.

Any efforts to derive quality judgements from publication processes will be dependent on the European publication culture and its readiness to adjust to the needs of quality management by the scientific community. Methods based on peer review and editorial standards refer to a certain editorial process and a regular publishing schedule as it is best guaranteed by journals. An editorial reliability adhered to by the professional or scientific publishing board accepting certain standards is a prerequisite to many effective methods of quality assessment.

### 8.3 Approaches based on use and impact of a document

This concept offers two main lines of analyses for EERQI: citation and link analyses on the one hand (Chapter 2) and usage metrics on the other (Chapter 6).

Expanded citation analysis is a new method of evaluation which can be applied here successfully, since the EERQI database will contain a new range of literature specifically focused on educational research and including books, book chapters, and journal articles, commercially published articles, open access items and other research items freely available on the web. This new content base allows citation analysis over a collection of documents relevant for educational sciences in contrast to the very few relevant journals of this subject area integrated in the *SSCI*. Citation analysis can be manipulated by self-citations or citation circles, but in general it gives good evidence concerning the impact of a text, although this does not directly imply a judgment of quality. Citation analyses in the context of EERQI can also serve as corrective and comparative measures for the reliability of other new indicators.

Most forms of expanded citation or link analyses and the selective forms of new statistical analyses (chapter 2.3) should be applied on the digitalized EERQI content base. The workshop should discuss whether ‘expanded’ citation analyses should not only be done by using EERQI’s own content base, but also by the use of existing content bases like *Scopus*, *Scirus* or *Publish or Perish* (Google Scholar) etc. in order to extend citation analyses on a wide range of documents (see chapter 2.1). As far as we can see this would be a new task not yet part of the work packages in the EERQI project.

Statistics of usage based on online data (downloads, library orderings, reading analyses) are in accordance with the EERQI project and should be considered as very relevant. Data from as many external sources as possible and reasonable should be gathered. A cooperation with COUNTER is already envisaged.

The role of discipline-specific surveys or criteria (Chapter 7) which give evidence of the reception and reputation of publications within EERQI must be discussed. The possibilities range from carrying out an exemplary survey to conceptual work in this field.

### 8.4 General Suggestions for the Workshop

This report on the state-of-the-art concerning existing indicators of scientific quality assessment was mainly compiled against the background of various scientometric studies. A first analysis of a state-of-the-art on evaluation practices in European countries based on a survey conducted by the EERQI project is supposed to back up this information with empirical data and might indicate scientific and policy demands for new indicators. For the forthcoming workshop it will be necessary to focus the basis for decision-making on the requirements of the scientific community.

When the Workshop will address the different possible new approaches for quality assessment the participants will possibly have to make their mind up to the following questions:

What is the relationship between ‘Quality’ and ‘Relevance’ of a publication? Can we accept relevance as a highly significant attribute of quality?

Can we define formal characterizations of a document which signify quality/relevance?

Can we define content areas (‘concepts’) of scientific activity which indicate quality/relevance? Which topic areas in educational research are genuinely European and should be visible internationally?

Besides instruments to detect and classify quality of scientific publications another requirement the workshop might reflect is the demand for better international visibility of educational research in Europe. Visibility is an essential prerequisite of European cooperation in educational research irrespective of the question if a publication is assessed correctly in a national or international evaluation. The practical consequences of a 'visibility approach' may result in a similar role of English as THE language of scientific communication as it is already accepted in other disciplines. And successful European cooperation in the field of educational research may need a European scientific publication culture constituted by scholarly associations and specialist publishers.

So a last question for the workshop might be: Is there a realistic chance that publication cultures in European educational research can be changed in such a way that the cross-national visibility and exchangeability of approaches and results of research can be enhanced? Which infrastructures would we need to support such a change?

### **Bibliography**

Abdalla, Rashid., S. Teufel: A bootstrapping approach to unsupervised detection of cue phrase variants. In: Proceedings of ACL/COLING 2006.

Bergstrom, Carl: Eigenfactor: Measuring the value and prestige of scholarly journals. C&RL News 68 (5) 2007,  
<http://www.ala.org/ala/acrl/acrlpubs/crlnews/backissues2007/may07/eigenfactor.cfm>, 29.5.08.

Bhattacharya, Sujit, P. K. Basu: Mapping a Research area at the micro level using Co-word analysis. *Scientometrics*, 43 (3) 1998, pp. 359-372.

Bollen, Johan, Herbert Van de Sompel, Joan A. Smith, Rick Luce: Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing and Management*, 41 2005, pp. 1419–1440.

Bonnevie-Nebelong, Ellen: Methods for journal evaluation: Journal Citation Identity, Journal Citation Image and Internationalisation. *Scientometrics*, 66 (2) 2006, pp. 411–424.

Bornmann, Lutz, Hans-Dieter Daniel: Begutachtung durch Fachkollegen in der Wissenschaft. Stand der Forschung zur Reliabilität, Fairness und Validität des Peer-Review-Verfahrens. In: Schwarz, Stefanie, U. Teichler (Eds.): *Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung*. Campus Verlag Frankfurt/New York 2003, pp. 207-225.

Bradshaw, Shannon: Reference Directed Indexing: Redeeming Relevance for Subject Search in Citation Indexes. In: *Research and Advanced Technology for Digital Libraries*. Springer Berlin / Heidelberg 2003, pp. 499-510.

Cronin, Blaise, Debora Shaw: Identity-creators and image-makers: Using citation analysis and thick description to put authors in their place. *Scientometrics*, 54 (1) 2002, pp. 31-49.

Darmoni, S. J., F. Roussel, J. Benichou: Reading factor as a credible alternative to impact factor: A preliminary study. *Technology and Health Care*, 8 (3-4) 2000, pp. 174-175.

European Reference Index for the Humanities (ERIH), Standing Committee for the Humanities,  
[http://www.esf.org/fileadmin/be\\_user/research\\_areas/HUM/Documents/ERIH/ERIH-11-2007.pdf](http://www.esf.org/fileadmin/be_user/research_areas/HUM/Documents/ERIH/ERIH-11-2007.pdf), 22.4.08.

Egghe, Leo: Theory and practice of the g-index. *Scientometrics*, 69 (1) 2006, pp. 131-152.

Fernández-Cano, Antonio, Ángel Bueno: Multivariate evaluation of Spanish educational research journals. *Scientometrics*, 55 (1) 2002, pp. 87–102.

Frandsen, Tove F.: Journal diffusion factors - a measure of diffusion? *Aslib Proceedings*, 56 (1) 2004, pp. 5-11.

Garfield, Eugene: From Bibliographic Coupling to Co-Citation Analysis via Algorithmic Historio-Bibliography. 2001.  
<http://www.garfield.library.upenn.edu/papers/drexelbelvergriffith92001.pdf>, 29.5.08.

Glänzel, W., H. J. Czerwon: A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics*, 37 (2) 1996, pp. 195-221.

Hirsch, Jorge E.: An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102 (46) 2005, pp. 16569-16572.

Holsapple, Clyde W.: A publication power approach for identifying premier information systems journals. *Journal of the American Society for Information Science and Technology*, 59(2) 2008, pp. 166-185.

Ingwersen, Peter: The calculation of Web impact Factors. *Journal of Documentation*, 54 (2) 1998, pp. 236-243.

Ingwersen, Peter: *Webometrics : ten years of expansion*. 2006.  
<http://eprints.rclis.org/archive/00006264/01/ingwersen.pdf>, 24.4.08.

Jamali, Hamid R., David Nicholas, Paul Huntington: The use and users of scholarly e-journals: a review of log analysis studies. *Aslib Proceedings: New Information Perspectives*, 57 (6) 2005, pp. 554-571.

Moed, Henk F.: *Citation Analysis in Research Evaluation*. Springer Heidelberg 2005.

Oard, Douglas W., Daqing He, & Jianqiang Wang: User-assisted query translation for interactive cross-language information retrieval. *Information Processing and Management*, 44 2008, pp. 181-211.

Park, Han W., M. Thelwall, Mike: Hyperlink analyses of the World Wide Web: A review. *Journal of Computer-Mediated Communication*, 8(4) 2003.

Project COUNTER: <http://www.projectcounter.org/index.html>.

RAE (2006): UOA 45, Education. <http://www.rae.ac.uk/pubs/2006/01/docs/k45.pdf>, 29.4.08

Ritchie, Anna, S.Teufel, S. Robertson: How to find better index terms through citations. In: *Proceedings of the Workshop "Can Computational Linguistics Improve Information Retrieval?"*, at ACL/COLING, Sydney, Australia 2006.

Rowlands, Ian: Journal diffusion factors: a new approach to measuring research influence. *Aslib Proceedings*, 54 (2) 2002, pp. 77-84.

Sandor, Agnes: Using the author s comments for knowledge discovery. *Semaine de la connaissance, Atelier texte et connaissance*, Nantes, June 29, 2006.

Schneider, Jesper W.: Concept symbols revisited: Naming clusters by parsing and filtering of noun phrases from citation contexts of concept symbols. *Scientometrics*, 68 (3) 2006, pp. 573-593.

Shepherd, Peter T.: The feasibility of developing and implementing journal usage factors: a research project sponsored by UKSG. *Serials*, 20 (2) 2007, pp. 117-123.

Sivertsen, Gunnar: Bibliometri for svensk utdanningsvitenskap. Et pilotprosjekt. *Vetenskapsradet (Hrsg.)*. 2007. [http://www.cm.se/webbshop\\_vr/pdf/O\\_131.pdf](http://www.cm.se/webbshop_vr/pdf/O_131.pdf), 15.4.08.

Talvensaari, Tuomas, M. Juhola, J. Laurikkala, K. Järvelin: Corpus-based cross-language information retrieval in retrieval of highly relevant documents: Research Articles. *Journal of the American Society of Information Science and Technology*, 58 (3) 2007, pp. 322-334.

Thelwall, Mike, T. Ruschenburg: Grundlagen und Forschungsfelder der Webometrie. *Information Wissenschaft und Praxis* 57 2006, pp. 401-406.

Thelwall, Mike: Bibliometrics to webometrics. *Journal of Information Science* 34 (4) 2007, pp. 1-18.

Van den Besselaar, Peter, G. Heimeriks: Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics*, 68 (3) 2006, pp. 377-393.

Van Impe, S., R. Rousseau: Web-to-Print Citations and the Humanities. *Information Wissenschaft und Praxis*, 57 (8) 2006, pp. 422-426.

Van Raan. Anthony F. J.: Measuring Science Capita Selecta of Current main issues. In: Moed, Henk F., W. Gänzel, U. Schmoch (Hg.): *Handbook of Quantitative Science and Technology Research*. Kluwer Academic Publishers Dordrecht 2004, pp. 19-50.

White, Howard D.: Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52(2) 2001, pp. 87-108.

Zhao, Dangzhi, A. Strotmann: Can Citation Analysis of Web Publications Better Detect Research Fronts? *Journal of the American Society for Information Science and Technology*, 58 (9) 2007, pp. 1285-1302.

## **Annex: Key numbers from the Journal Citation Report in Educational Sciences**

### **National providence of the journals included in the JCR (2006, ISI)**

<b>Country</b>	<b>Number of Journals</b>	<b>Percentages (%)</b>
United States of America	102	64.97
England	37	23.57
Netherlands <sup>1</sup>	7	4.46
Germany <sup>2</sup>	4	2.5
Australia	1	0.6
China <sup>3</sup>	1	0.6
Japan <sup>4</sup>	1	0.6
New Zealand	1	0.6
Portugal <sup>5</sup>	1	0.6
Russia <sup>6</sup>	1	0.6
South Africa	1	0.6
Total	157	

### **Language of the journals included**

<b>Language</b>	<b>Number of Journals</b>	<b>Percentages (%)</b>
English	149	94.9
Multilingual	3	1.91
German	3	1.91
Japanese	1	0.6
Russian	1	0.6
Total	157	

<sup>1</sup> Higher Education, Instructional Science, Minerva, Research in Higher Education, Research in Science Education (Springer), Learning and Individual Differences (Elsevier), School Effectiveness and School Improvement (Routledge Journals, Taylor & Francis Ltd.) (all publishing in English)

<sup>2</sup> Psychologie in Erziehung und Unterricht, Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie, Zeitschrift für pädagogische Psychologie, Zeitschrift für Pädagogik (3 publishing in German, 1 multilingual)

<sup>3</sup> Chinese Education and Society (English)

<sup>4</sup> Japanese Journal of Educational Psychology (Japanese)

<sup>5</sup> European Journal of Psychology of Education (English)

<sup>6</sup> Voprosy Psikhologii (Russian)