Gogolin, Ingrid (2012). Identificación de la calidad en las Publicaciones de Investigación Educativa: Proyecto Europeo sobre los Indicadores de Calidad en la Investigación Educativa (EERQI). *Revista de Investigación Educativa, 30* (1), 13-27.

IDENTIFICACIÓN DE LA CALIDAD EN LAS PUBLICACIONES DE INVESTIGACIÓN EDUCATIVA: PROYECTO EUROPEO SOBRE LOS INDICADORES DE CALIDAD EN LA INVESTIGACIÓN EDUCATIVA (EERQI)

Ingrid Gogolin
University of Hamburg

RESUMEN

El proyecto EERQI surge porque la noción internacional de calidad científica (como principal determinante sobre el que se financia y apoya cualquier investigación) puede causar efectos secundarios no deseados si no se responde adecuadamente la pregunta de ¿Cómo se interpreta y mide la calidad? *Los instrumentos actuales de "medición" de la calidad a través de la contabilidad de citas y otros métodos similares* causan graves efectos secundarios, *pues están muy sesgados y resultan inadecuados para la investigación en las Ciencias Sociales y Humanidades.*

El proyecto EERQI ha desarrollado un método para detectar la calidad de los textos de investigación —con la investigación educativa como modelo para las Ciencias Sociales y Humanidades (CCSSHH)— mediante la aplicación de una combinación inteligente de diferentes enfoques que se complementan entre sí. Es esto lo que llamamos el Marco Prototipo EERQI, y se compone de productos y métodos que pueden servir como alternativas en los procesos de evaluación de la calidad en la investigación en las CCSSHH. La posibilidad de asistencia multilingüe de evaluación elaborada por el motor de búsqueda multilingüe y análisis semántico automático de EERQI están hechos a medida para que se produzca un fortalecimiento del espacio europeo de investigación. Los productos y métodos EERQI son los siguientes:

- *Una base de contenido con textos de investigación educativa en cuatro idiomas europeos que fueron la muestra incluida en el proyecto EERQI: inglés, alemán, francés y sueco.*

Correspondencia:
Ingrid Gogolin, University of Hamburg. Mail: sekgogolin@erzwiss.uni-hamburg.de

- *Un motor de búsqueda multilingüe que incluye la ampliación de consultas: una herramienta eficaz dedicada a la investigación educativa en general, capaz de encontrar textos de investigación educativa en la Web en los cuatro idiomas de EERQI.*
- *Un análisis semántico automático para la detección de frases clave en un texto. El método es aplicable a las publicaciones de investigación en educación en al menos los cuatro idiomas de EERQI.*
- *Una combinación de métodos bibliométricos / webométricos para la detección de indicadores "extrínsecos" de calidad (herramienta aMeasure).*
- *Las primeras pruebas de un método de análisis de citas que tiene el potencial de desarrollarse para la aplicación a textos de investigación educativa (y otras CCSSHH).*
- *Un conjunto de indicadores de texto inmanente para la detección de la calidad en las publicaciones de investigación educativa que se ha presentado a la comunidad científica y evaluado positivamente.*
- *Un cuestionario de revisión por expertos del que se ha testado su fiabilidad y funcionalidad.*
- *Un conjunto de escenarios de casos de uso que aportan consejos sobre cómo utilizar cualquier combinación de las herramientas antes mencionadas.*
- *Primeros intentos para detectar las interrelaciones entre indicadores de calidad "extrínsecos" y "intrínsecos".*

Palabras clave: *calidad científica, asistencia multilingüe, indicadores intrínsecos de calidad, indicadores extrínsecos de calidad, revisión por pares / expertos.*

IDENTIFICATION OF QUALITY IN EDUCATIONAL RESEARCH PUBLICATIONS: THE EUROPEAN EDUCATIONAL RESEARCH QUALITY INDICATORS (EERQI) PROJECT[1]

ABSTRACT

The EERQI project was motivated by the fact that the international notion of scientific quality as being the main determinant on which research is funded and supported may cause undesired side effects, if the questions of how quality is interpreted and how it is measured *are not adequately answered. Current instruments for 'measuring' quality via citation counting and similar methods do cause such side effects, as they are strongly biased and largely inadequate for research in the Social Sciences and Humanities.*

The EERQI project developed an approach to detect the quality of research texts – with educational research serving as model case for Social Sciences and Humanities (SSH) – by applying an intelligent combination of different approaches that complement each other. This is what we call the EERQI Prototype Framework. It consists of products and methods that can serve as alternatives in processes of assessment of quality in SSH research. The possibility of multilingual assistance of assessment processed by EERQI's multilingual search engine and automatic semantic analysis are tailor-made for strengthening the European research space. The EERQI products and methods consist of:

1 This article is based on the EERQI Project's Final Reporting to the European Commission. The project was funded under the EU's 7th Framework, Social Sciences and Humanities Scheme, from 2008 to 2011. See for further information: www.eerqi.eu

- *A content base with educational research texts in four European languages that were exemplary included in the EERQI project: English, German, French and Swedish.*
- *A multilingual search engine that includes query expansion: an effective tool dedicated to educational research in general, capable of finding educational research texts in the Web in the four EERQI languages.*
- *An automatic semantic analysis for the detection of key sentences in a text; the method is applicable to educational research publications (in at least) the four EERQI languages.*
- *A combination of bibliometric/ webometric approaches for the detection of 'extrinsic' quality indicators (tool aMeasure).*
- *First tests of a citation analysis method that has the potential to be further developed for the application to educational research (and other SSH) texts.*
- *A set of text-immanent indicators for the detection of quality in educational research publications that has been presented to the research community and was positively evaluated.*
- *An accompanying peer review questionnaire that was tested for reliability and practicality.*
- *A set of use-case scenarios that give advice on how to use which resp. combination of the above-mentioned tools.*
- *First attempts to detect interrelations between 'extrinsic' and 'intrinsic' quality indicators.*

Key words: *scientific quality, multilingual assistance, extrinsic indicators of quality, intrinsic indicators of quality, peer review.*

1. CONTEXT AND OBJECTIVES OF THE PROJECT

All across the world, the structures and control mechanisms of publicly funded research are changing dramatically in the last decade. There are many widely discussed causes of these developments. The set of causes on which the EERQI-project concentrated is based on the evocation of the 'ability to compete internationally' – a request that is expressed vis-à-vis national research landscapes in Europe as well as the whole European Research Area.

A metaphor that is either explicitly used or implicitly resonates in the existing discourses, in the decisions on new governance mechanisms, and in new modes of research funding is *quality*. The discovery, improvement and promotion of research quality and the quality of research outcomes – such as publications – are the driving motives for the tendency to re-evaluate and redevelop structures for the research area, for redesigning the funding of research institutions and projects, and for instituting control and legitimization systems that are (or intend or pretend to be) helpful for decision-makers.

In the framework of these developments, the questions of how quality is interpreted and how it is measured are of fundamental importance. Analyses dealing with this question supplied the starting point for the development of the research project 'European Educational Research Quality Indicators (EERQI)'.

The project was developed by a truly interdisciplinary European research consortium, a unique composition of experts from educational science, biblio- and webometrics,

information and communication technologies, computational linguistics and publishing houses. [2]

The focus of the analysis prior to the project was on special questions such as: What constitutes and marks the current quality control systems that are applied in contexts of governance and funding, irrespective of the genre and the type of research that is at stake? And what are possible (desired and undesired) effects of these systems on research that is conducted in the European research area, especially in the domains of the social sciences and the humanities?

Educational research is particularly suitable for considerations and research on such questions, because it can be considered to be prototypical for vast areas of the whole field of the social sciences and humanities. Educational science and research combine a wide spectrum of theoretical and methodological approaches – from primarily philosophical-historical methodologies as used in the humanities to psychologically or sociologically based empirical observations of individual development, education, training or Bildung (formation); from hermeneutical interpretation, single case studies to the generation and statistical analysis of great amounts of survey data. This manifests relevant characteristics of knowledge production, which are also found in other disciplines in the social sciences and humanities.

Another reason why it was appropriate to chose educational research as a model is that the visibility of education and learning as a policy space and its emergence as a significant area of policy are not matched by useful analyses of its operation. Policy in education and educational research is no longer the sole domain of the nation-state, but has become a key feature of a 'Europeanizing' process. 'New Learning' through social innovation is central to the knowledge economy, allowing education to be compared, promoted, researched and improved in its European role as a key part of the knowledge economy and as a distinctive element in the particular mission of Europeanization within globalization. However, the contribution of European education research is hampered by the way it is organized in Europe. Distinctive and fruitful traditions of work are locked into national intellectual resources and it is a slow process of enabling them to move across borders. Thus, there is a need to intensify networks and agree on common standards paving the way to a shared discourse space for European educational researchers. The EERQI project aimed to contribute to this development.

Before the project started, we carried out a firm review on state of the art research of quality assessment. We focused on the appropriateness of instruments and strategies for quality assessment that are actually applied to educational science (and SSH in general). Our review resulted in a generic judgment that can briefly be articulated as follows: The existing instruments do not lead to valid results, because they do not measure what they claim to measure. An example to illustrate this statement is quality assessment based on citation indices and journal rankings. As yet, this has been the most common approach in vast areas of quality assessment in Higher Education.

2 It was my privilege to act as the EERQI project coordinator. Among other partners, the European and some national educational research associations participated in the project. I would like to stress here that what I report on is the harvest of a joint and cooperative endeavor of the EERQI team, not my individual work.

The central and joint quality criterion that is used in these instruments is 'international visibility' of research findings. This is expressed by the placement of the publication, namely in journals with a good reputation, and by the number of citations of a publication. This approach is – for example – characteristic of the Social Science Citation Index, a commercial instrument, owned by the US American publishing group Thomson Reuter. Its results often play an important role in reporting systems on research achievement. A closer look at the documentation of the journals represented by this index reveals (for the field of educational science according to the "Journal Citation Report") the following:

In total, 201 educational research journals were incorporated in the rankings of 2009. More than half of these journals are published by US American publishers. An additional 25% derive from British publishing houses. The next 'largest' nations in this ranking are the Netherlands (with 4% of cited journals) and Germany (with 3% of publications). Altogether, 15 nations across the world are represented in the ranking.

Another perspective on the Journal Citation Report reveals that 89% of the publications that were ranked in 2009 are in English. The next 'largest' languages with 2.5% and 2% respectively are German, Spanish and Turkish. In total, eleven languages are represented by the index. A language such as French is not included.

Although the owners of these instruments are constantly striving to improve their methodologies, and although these methodologies differ more or less extensively between the instruments, the general problem of their validity is illustrated by these (and the further) findings of the preparatory EERQI survey: These kinds of approaches do not produce valid information in the sense that they pretend to do. The illustration shows that the intended international relevancy of the included publications cannot be proven. The rankings are heavily biased: they essentially refer to US American or UK publications and publications written in English. International visibility as a quality criterion must be translated here to: visibility of products from a narrow selection of national research spaces to the rest of the world. The provided information is perfectly suitable to substantiate the dominance of a 'minority' of regional and linguistic research areas.

This means in fact, that these methodologies do not reflect an adequate coverage of (not only) European scientific publications, in particular in the social sciences and humanities. Not only ignoring of most other languages than English and of publication production outside the US and United Kingdom is relevant here. Moreover, the exclusion of other types of publication, namely books, means a bias to actual publication practice in educational research as well as the whole field of SSH. Another criticism is that mere citation counting may, if at all, indicate research quality in natural sciences, where a 'cumulative model' of citing is broadly practiced: a citation indicates that the citing author positively refers to and builds on the work of a cited author. In the social sciences and humanities however, this is not the main tradition and function of citations; instead, citations are often used for contradiction or rejection of another position, or for negotiation. Thus, if citations are simply counted rather than analyzed in their intention, the significance of the result is weak.[3]

3 That is why one of the EERQI approaches dealt with citation analysis rather than citation counting; this cannot be presented here, for further information see Frederik Åström (2009): Models of Scholarly Com-

Hence, if European science or institutions are exposed to these evaluation methods, not only are individual researchers and institutions, but also complete subject domains and language areas widely ignored.

Based on such analyses, the motivation for the development of the EERQI project was, in a nutshell, the observation that the strategies of assessment that were developed in 'hard science'-contexts have to be heavily criticized for their methodological weakness and lack of validity – not only from a social sciences and humanities point of view. At the same time, there was a genuine desire to develop approaches that can serve better for the aim of detecting research quality in our domains. This desire unites the research community as well as relevant stakeholders from other spheres, such as publishing houses, research funding and political decision making.

The EERQI team's general intention was to develop useful tools that support the *process* of quality detection. An intelligent combination of such tools – that was our assumption – would be useful to assist and support the researcher who has to form an opinion on research publications, aiming at the determination of their quality. The application of these tools should meet two aims:

a) it should raise the transparency and value of the process of quality detection itself;
b) it should make the task better manageable and less time consuming.

In order to meet these aims it was *not* EERQI's objective to develop one single method, such as another index that can compete with the existing ones. The aim was the development and testing of a set of tools with different functions that can support and accompany the process of detecting research quality in texts - from the moment of identification of a text to the moment of the conclusive determination of quality.

The new tools to be developed were compiled in a broader prototype framework, each tool addressing a specific part of the assessment process:

Before being able to assess educational research documents for their quality they have to be identified and gathered. Therefore, when searching for a specific term it was the task to identify relevant educational research documents and make them available to the user. The EERQI-project developed a specific search engine that is able to identify educational research texts (see http://makalu.xrce.xerox.com/eerqi/). The stock that the search engine comprises of needed to contain a wide range of documents, those being freely available in the World Wide Web and those being in the possession of publishing houses or research institutions – which are normally not freely accessible. Since the relevance of the harvested documents plays a crucial role, it was the task of the developers to refine the search engine in order to gather only educational research documents relevant to a respective search term. Taking into consideration the European context in which educational research documents are published in different languages the search engine needed additional multilingual functions so that it was able to deliver results to the search term in several languages.

---

munication and Citation Analysis. ISSI 2009: The 12th International Conference of the International Society for Scientometrics and Informetrics. BIREME/PAHO/WHO & Federal University of Rio de Janeiro. www.gbv.de/dms/tib-ub-hannover/60745721x.pdf

As the EERQI project could only develop a prototype method, we included four languages as examples: English, German, French and Swedish.

After retrieving the documents, their quality was assessed with different approaches. This involved the improvement of 'classic' indicators that are used in the above mentioned approaches (e.g. amount of citations; classification of a journal). But furthermore, the project developed other approaches. Very soon in the research process, the EERQI team – supported by a number of relevant experts – decided that the development should differentiate between two different types of indicators:

- one type that is external to the text, such as bibliometric and webometric features. We called these 'extrinsic indicators'; and
- another type that is internal in the text – namely the signals that are given within the words, graphs, metaphors of which the text is composed. We called these 'intrinsic indicators'.

In order to assist an evaluator (a reader) with the detection of quality, the extrinsic and intrinsic indicators have to be applied to the text.[4] Tools had to be developed to assist with their easy detection. For the intrinsic and the extrinsic indicators, different approaches have to be used: Measuring extrinsic characteristics of research publications involves the harvesting of relevant pieces of information from different search engines such as Google Scholar, Google Web Search etc. The detection of the intrinsic indicators is a much more complex process requiring assistance if large amounts of texts have to be assessed. This assistance can be provided by automatic semantic analysis, a tool developed by the EERQI team members from Xerox Laboratories in Grenoble, France. Another relevant tool is the EERQI peer review questionnaire; this instrument comprises of the operationalized items that indicate internal features of the quality of a text. The reliability and acceptance of this questionnaire was tested with a positive result; responsible were the team members from University of Hamburg and the European Educational Research Association.

The combination of methods and approaches to assist a reader in the process of quality detection is what we called the EERQI Prototype Framework. The following chart illustrates the details of this framework:

4 It has to be mentioned here that the whole EERQI procedures can only be applied to texts that are electronically available.

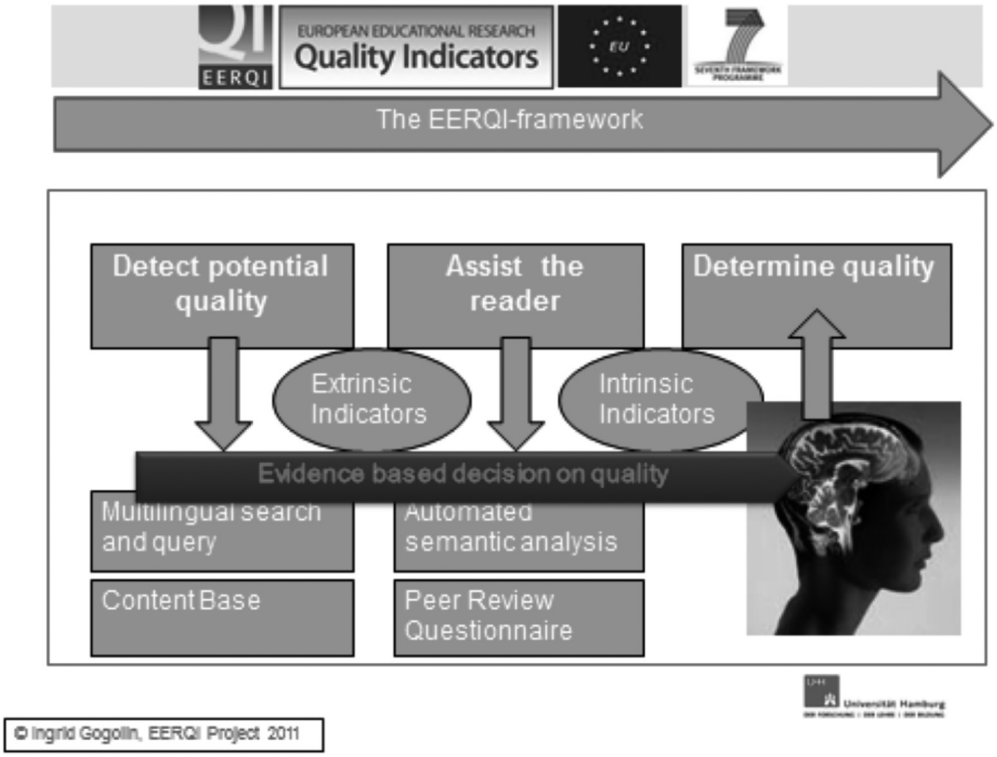

FIGURE 1
THE EERQI-FRAMEWORK

The EERQI Prototype Framework consists of:

— a content base and the search and query engine that support the detection of potential quality via identification of relevant educational research texts in different (electronically available) sources (developed by EERQI partners RRZN, Hannover, and ISN, Oldenburg)
— a tool called 'aMeasure' that identifies extrinsic characteristics of research publications by using Google Scholar, Google Web Search, MetaGer, LibraryThing, Connotea, Mendeley, and citeulike (developed by EERQI partner Humboldt University Berlin).
— a linguistic technology that allows for the automatic identification of key sentences to indicate parts of documents to which reviewers should pay particular attention (Automated semantic analysis, developed by EERQI partners Xerox and DIPF).
— a Peer Review Questionnaire that contains a tested operationalization of the intrinsic indicators that were developed by the EERQI project (developed by University of Hamburg and EERA).

In the following chapter, I will explain same examples of the EERQI products.

2. EXAMPLE NO. 1: 'INSTRINSIC INDICATORS' AND PEER REVIEW QUESTIONNAIRE

As was mentioned already, the EERQI consortium decided very early in the project's process to distinguish between intrinsic and extrinsic indicators. Our hypothesis was that both types of indicators may be relevant for evaluation and assessment processes, but the type of information on the performance of a text differs and may or may not be cor- or interrelated. The identification of cor- or interrelations between the two types of indicators belongs to the project's aims. Although the funding period is finished, the data analysis for this purpose has not yet been terminated. First attempts show that on a holistic basis there are only weak ties between both types of indicators. Multilevel analysis however, is still ongoing.[5]

The development of the intrinsic indicators was an iterative process that started with an expert conference in Leuven, 2008. Here, a preliminary list of indicators was established that was further developed and refined in a continuous process of consultation and evaluation, included were several hundred experts that were nominated by EERA and the participating national Educational Research Associations. On this basis, a comprehensive set of qualitative indicators emerged. The preliminary 'Leuven list' of intrinsic indicators was step by step structured, modified and condensed to the following five generic indicators:

- Rigour
- Originality
- Significance (for other researchers, policy and practice)
- Integrity (including considerations of authenticity, honesty and ethical requirements in the conduct of research)
- Style (including clarity, communicability, eloquence and elegance).

This list was unanimously positively evaluated by the consulted expert as generally relevant for the assessment of educational research quality.

The indicators were operationalized, again in a consultation process, and transferred into the 'EERQI Peer Review-Questionnaire' that was tested in three waves. In each wave, statistical analysis was carried out in order to test the reliability of the items. Moreover, qualitative questions referring to the suitability and validity of the instrument were included and analyzed in order to optimize it. According to the results of the statistical analysis and the qualitative information, the following modifications of version 1 of the questionnaire were introduced:

- The number of indicators was reduced to the three main indicators rigour, originality and significance. This was due to the fact that the selectivity of items referring to 'integrity' and 'style' was low because of interrelations with the scales for rigour and originality. As they did not function as individual scales, items concerning 'integrity' and 'style' were included in the scales for rigout and originality.

5 The final results will be published in: Ingrid Gogolin/ Frederik Åström (eds., 2012): Assessing Quality in European Educational Research: Indicators and Approaches. To be published: Wiesbaden: VS-Publishers/ Springer.

- The items' scaling was enlarged to in order to receive a better variance of the ratings.
- Additional items on demographic data of the reviewers were included in order to receive more detailed information on the match of expertise of a reviewer with the reviewed text on one hand, with the overall results of a review on the other.

The results of the third wave (carried out from December 2010 to January 2011) lead to a further refinement of the questionnaire in terms of its validity and practicability. In the third wave, the elicitation of qualitative feedback from members of the educational research community with respect to the acceptance of the questionnaire was included. The final version of the questionnaire contains three scales with respect to the indicators rigour, originality and significance with all in all 16 items.

All test and item characteristics show good to very good values, which could be approved for subsamples of different areas of educational research with different cultural and linguistic backgrounds. The following table illustrates the values for the scales:

TABLE 1
OVERVIEW FINAL SCALE VALUES[6]

Scale	Subscale	Number of items	Reliability	Mean value for item validity
Rigour		9	.92	.76
	Methods & Approaches	3	.83	.72
	Results	2	.94	.64
	Discussion	4	.90	.82
Originality		3	.91	.78
Significance		4	.91	.78

Especially helpful for the reviewing process were the reviewers' comments on the relevance of the indicators and the practical use of the questionnaire. The acceptance and indication of usefulness of the questionnaire was generally very high; it was especially high in reference to educational research texts that derive from empirical studies. The statistical results, however, show that the questionnaire can well be applied across other areas of educational research such as historical and philosophical research in education, international comparative and intercultural research. The analysis of the qualitative

6 The reliability is measured by using Cronbachs Alpha. The item validity is measured by using a procedure presented and tested by Yousif, Koopmann & Amelang (Yousfi, S., Koopmann, B., & Amelang, M. (2005). Correlates of item validity. On the eminent importance of global self-ratings. Unpublished manuscript. – For a detailed information about the methodology that was applied for testing the questionnaire see the EERQI final report on www.eerqi.eu or the above mentioned book publication, Gogolin & Åström forthcoming (2012).

responses substantiates that the questionnaire includes the most important indicators in the field of quality assessment in educational research publications.

The process of the development and the evaluation results of the questionnaire as well as the final version of the instrument were presented to the educational research public at different occasions such as various international research conferences and the EERQI Final Conference in Brussels (March 2011). At all occasions, the responses and feedback were entirely positive. Different suggestions for an extended use of the questionnaire, for instance a further development for the purpose of training (especially new) researchers in assessment tasks, were made. These possibilities should be further explored within the educational research communities in the future.

To sum up: EERQI's intrinsic indicators and their operationalization as shown in the questionnaire were successfully tested. The acceptance of the instrument in the educational research community appears to be high according to our investigations. The instrument is now available on the EERQI website for implementation and further development. This will hopefully support the intentions to facilitate and to raise the transparency of assessment processes in educational research, and thus: enhance the *quality* of quality assessment procedures as such.

3. EXAMPLE NO. 2: AUTOMATED SEMANTIC ANALYSIS

EERQI concentrated on providing evidence for the invalidity of one of the underlying assumptions, namely the idea that research communication is merely cumulative in the sense that one piece of research is cited by another piece of research in order to build on it for accumulating knowledge. This assumption gives each citation the same weight as a quality indicator in citation indexes. EERQI argues that citation typing provides evidence for the fact that research communication – especially in social sciences and humanities – can also be negotiating. In this case new knowledge does not directly build on the cited knowledge and possibly even contradicts it. Consequently, in these two cases, research quality can not be indicated in the same way.

To identify the type of citations matching the author's motivations, the XEROX Incremental Parser was used with support of DIPF and HU-Berlin.

2. Automatic semantic analysis: reading assistance and search engine enhancing (with tests in the four EERQI languages) – main contributor: XEROX, supported by DIPF and HU-Berlin who provided access to the EERQI content base.

The goal of this is twofold:

(a) Providing reading assistance to peer-reviewers.

EERQI considers the intrinsic indicators of quality as defined in WP4 (significance, rigour, originality, integrity and style) as revealable by human reading only. Human reading, however, is time-consuming and hard intellectual work. One technique of assisting the reader is highlighting key messages automatically, i.e. by focusing the reader's attention to the main message. Following the definition of the intrinsic quality criteria, the following two types of sentences are highlighted as key sentences: sentences that describe the research problems the article treats (1) and sentences that describe the

author's goals, claims and conclusions (2). The effectiveness of highlighting was tested in a peer-review exercise in which peer-reviewers were asked to write brief references to the articles they read. These sentences were then compared to the highlighted sentences with the expectation that the content of the highlighted sentences corresponds with the content of the 'human' reference. This process was accompanied by further research on highlighting salient messages in research texts for supporting reading comprehension and peer review processes and the potential of integrating semantic text analysis in search engine rankings.

(b) Enhancing the search engine.

The process was further accompanied by the attempt to integrate semantic text analysis in search engine rankings which incorporates the consideration of key sentences in relevance ranking (1) as well as the usage of key sentences in the result snippets (2).

3. Genre analysis: automatic genre recognition – contributors: XEROX, supported by DIPF

The approach was carried out with the intention to automatically recognize the genre of a publication. This has been proposed in order to be able to refine indicators relevant to special sub-domains of educational research.

RESULTS

1. Citation analysis:

LUB-LU applied the retrieval strategies in the context of general analyses of citation impact of education research journals that are compiled in the EERQI content base. It was investigated in which other fields of research those journals had been cited. Furthermore, some general traits in terms citation impact could be identified, showing few citations spread over a relatively large number of cited articles. Citation structures related to EERQI material was further investigated by analyses of Google Scholar data on EERQI journals and articles. The 'life span' of EERQI material was tested by analysing the yearly distribution of citations to EERQI articles published 2000-2003. In addition to this, the distribution of citations between EERQI articles was analysed and the nature of documents citing EERQI articles was also investigated in order to determine language structures and the exchange of citations between (a) documents in different languages (b) and the kind of documents EERQI articles are cited in. Finally, citation structures were investigated by analysing documents cited by EERQI articles through co-citation mapping. The results of this testing were presented at different international conferences as well as EERQI workshops and Final Conference and will also be published.

2. EERQI Automatic Text Analysis Methodology[7]

7 This part of my contribution is based on a text written by Àgnes Sándor from EERQI partner Xerox (France) for the EERQI final report in March 2011.

The aim of the EERQI approach 'automatic semantic analysis' is to assist human evaluators to recognize intrinsic indicators of quality. This endeavor was carried out by investigating in two directions: key sentence detection and citation analysis (see Ágnes Sàndor & Angela Vorndran [2009]: Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries. Association for Computational Linguistics, Singapore, pp 36-44. aye.comp.nus.edu.sg/nlpir4dl/presentations/Sandor-Vorndran.pdf).

(a) Key sentence detection

For Key sentence detection, a tool was developed that works on the basis of the Xerox Incremental Parser. This tool detects key sentences in educational research articles in English, French, German and Swedish.

The application of key-sentence detection is twofold in EERQI: It aims at providing reading assistance for peer-reviewers; and it shall contribute to improving the search and query engine by highlighting key sentences and thus helping relevance evaluation for the users as well as integration into the mechanism that ranks the results of search engine.

For pursuing the aim of key sentence detection, the approach consists in highlighting salient sentences that provide textual evidence for the peer reviewers in order to back up their evaluation. This means that the tool does not 'evaluate' automatically, but suggests to a reader which are the parts of the texts that are relevant for founding the judgment on.

The approach is based on a consensus in the EERQI team according to which an evaluation by peers is supposed to judge the relevance of the topic, the clarity of the problem statement, the coherence of the argumentation and the well-foundedness of the conclusions.

These are criteria of judgment that are inherent in the evaluation that can be carried out with the EERQI Peer Review Questionnaire. Following these evaluation criteria the tool that was developed highlights key sentences that describe research problems, purposes and conclusions related to the topic of the article as indicated by keywords.

The underlying supposition that drives this approach is that by highlighting this information, a relevant and coherent dimension of the representation of the flow of the article is added to it, which complements and completes the representation provided by the structure of the article, i.e. the title, the summary, its outline that is sometimes, but not always made explicit by the section headings. Whereas the summary, when present, gives concise information of the overall issues in the article, it is not sufficient for evaluation. Section headings and article structure play an important role for synthesizing the development of the arguments, however, in the domain of educational sciences - as in social sciences and humanities in general - they do not follow general patterns and thus are in many cases not indicative of the underlying argumentation.

Several tests have been carried out in order to evaluate the appropriateness and usefulness of the results of the automated semantic analysis procedures, namely highlighted key sentences. One of these tests concerned the time that was needed for evaluation procedures, using exemplary texts with and without highlighted sentences. Our conclusions are the following: Highlighting allows the support of peer reviewers'

evaluation process according to the scales of significance, originality and the items for style. Evaluation with respect to the items referring to 'integrity' and the scale for rigour however, can as yet not be supported by the highlighted sentences; here, further development of the method would be necessary. Another relevant result is that the highlighting makes it possible to rapidly filter out bad quality. In this respect, processing of the highlighted texts took four times shorter than the text without highlights.

(b) Peer-review test

Another testing concerned the question if the highlighted sentences do in fact cover the most relevant contents of the articles. Peer-reviewers who participated in the testing of the questionnaire were asked to briefly summarize the reviewed article in their own words. They were asked to state the main subject of the reviewed article, the research problem, the main conclusions or results and open questions. The summaries were written in the same language as the reviewed article. The sentences of these summaries were compared to the highlighted sentences. The following results have been obtained:

- 84% of the human summary sentences correspond to a sentence from the article.
- 56% of the corresponding sentences fulfill the criteria of salient sentences.
- 68% of the salient sentences were detected by XIP, the parser developed by Xerox.
- On average, the automatically highlighted sentences showed four times more of the nouns from the text than occurred in the human summary sentences.

These results show that the automatically detected sentences cover a considerable proportion of human summary sentences. This suggests that automatic highlighting does indeed have the potential of providing key sentences for peer reviewers, and thus allow for less time consuming processes of quality detection.

4. Concluding remarks

The EERQI results as well as all products are presented in more detail on the website www.eerqi.eu. They were presented at several occasions to the scientific community for verification and acceptance, not only in Europe. These presentations addressed educational research societies, experts in the field, representatives of research funding agencies, and promotion and evaluation bodies at the national and European level. Presentations took place at EERQI workshops, expert consultations and international conferences. Since it was the aim to apply the new methods to other SSH disciplines, the indicators and prototype framework were tested for transferability to political sciences.

All in all we can conclude that the aim of the project – the development of a prototype framework that consists of tools which give support to 'human' assessment procedures, but do not replace the human judgment – was met. A special value of the project's processes and results is that we actuated a disciplinary discourse on shared notions of quality as they occur in educational research publications. Moreover, the project could show ways to integrate European multilingualism in attempts to identify texts that might be relevant to a researcher who does not master the respective languages. This points to possibilities of strengthening the ties among European educational researchers and allow for more international cooperation and collaboration, even if the different partners' languages are not known to the others. Models of cooperation with

one working language, but inclusion of research that is carried out in several different languages, should be better feasible in future.

Anyhow, the EERQI project also shows that there is still a lot of work to be done. In the life span of the project, we were able to work out prototypes of the tools and methodologies that we considered to be promising. The research community is now invited to take up the results – the EERQI team welcomes all interested researchers to make use of the data and products that we developed – and advance the approaches.

Fecha de recepción: 25 de octubre de 2011.
Fecha de revisión: 25 de octubre de 2011.
Fecha de aceptación: 06 de noviembre de 2011.