

Relative Efficiency of Research Assessment Based on External Peer Review

Evidence from Copenhagen Business School

Nilsson, Ole Stenvinkel; Rovio-Johansson, Airi

Document Version
Final published version

Publication date:
2010

License
CC BY-NC-ND

Citation for published version (APA):
Nilsson, O. S., & Rovio-Johansson, A. (2010). *Relative Efficiency of Research Assessment Based on External Peer Review: Evidence from Copenhagen Business School*. CBS Learning Lab.

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. Jul. 2025

Relative Efficiency of Research Assessment Based on External Peer Review – Evidence from Copenhagen Business School

Airi Rovio-Johansson and Ole Stenvinkel Nilsson

Copenhagen Business School, Learning Lab, Kilevej 14 A, DK-2000 Frederiksberg, Denmark

arj.ll@cbs.dk

osn.ll@cbs.dk

Abstract

In three decades Higher Education Institutions have experienced decreasing trust and increasing demands of accountability from society. The aim is to explore how assessments of departments' research activities can contribute to improvement of research organization, research culture and credibility. The empirical study is based on reports from two assessment rounds of nine research departments, in 1994 to 2007. Qualitative statements and recommendations are transformed to relative quantifiable performance measures on ten different dimensions in order to study development over time and analysis of departments' variation. Results indicate significant improvement among some departments while others have not managed to such extent.

Keywords: Research evaluation, institutional research, external review, time series analysis, mixed methods;

1. Introduction

The contribution of this institutional research project is a longitudinal study of research, its quality, strategy, productivity and culture at a number of university departments over almost fifteen years at Copenhagen Business School (CBS)ⁱ. Assessment of research activities at university departments has been a developing research area, and considerable research assessments have been done for decades focused on subject oriented research. Lindsay (1982, p. 175) asserts: “While quite comprehensive self-study techniques for evaluating institutions of higher education have been developed [...] there are [still] a number of unresolved problems which limit the success of these activities. These include the lack of a precise conceptualisation for ‘institutional performance’ and the difficulties of measuring the elusive and imponderable outputs from the educational process.”

Even if we agree to Lindsay’s comment, first we ask - *Are the departments’ Self-Evaluation Reports examples of ‘a fragmented array of measurements and evaluative judgments?’* - *Are these reports difficult to aggregate to the overall assessment of institution’s research performances?* From the perspective of the CBS, we hesitate to state that maybe it was the case, when the Self-Evaluation report was conducted for the first time by the departments, but not at the second occasion or any longer. Second, we claim that regular Self-Evaluation Reports (SER) are demanded by the departments in the recurrent assessment of quality management processes and in the accreditation processes at CBS (EUA, 2010). Lindsay (1982) states that “the focus is generally on comparing an institution’s outputs with its goals” or with “stakeholders’ needs”, which means it is an assessment of *effectiveness* and not of efficiency. Welsh (1998, p. 158) suggest a definition of efficiency as: “the rising tide of ‘efficiency’ in contemporary education often masks not only a reduction in both the quality of education provided, but also attempts to increase productivity levels in education, particularly in the public sector.” **We use the *relative efficiency* to describe the assessors’ understanding of the department, partly based on the Self-Evaluation Report (SER) and partly based on the results of the visit.**

In 1993 Copenhagen Business School (CBS) launched its first formal strategy document of its vision and mission, and thereby outlining the paths of development for the next 5 years (Handelshøjskolen i København, 1992). Among the strategies for research, PhD programs, undergraduate education, and evaluation, three central basic elements in this strategy were the

notion of the 'Learning University', "Partnership with society" and "Internationalisation". These basic elements demanded continuous focus on quality development in all areas of activity and required concrete initiatives to support continuous development of activities, thus leading to improved quality levels. In terms of research, these strategies meant to introduce the assessments of departments' research by means of using external evaluative Peer Review experts within each research area (Trow, 1994). In both assessments rounds 1994 and 2002, the *Peer Review Evaluation* were performed as an external evaluative process of departments' research activities, by internationally acknowledged researchers of specific research fields, selected by the dean of the faculty based on suggestions from each department. These Peer Review processes were meant to focus on the departments' research quality of various research projects, research profile, research productivity, research environment, specifically for junior researchers and doctoral students, research productivity, networks, ways of publication, and PhD program and the relation between research and education (Foss Hansen, Maskell and Mouritsen, 1995).

Pagani (2002) examined the practice of peer reviews and the related effect of peer pressure in the context of international organisations, particularly the Organisation for Economic Co-operation and Development (OECD). He asserts that Peer Review as a working method is closely associated with OECD. Several other international organisations and programmes made use the Peer Review process at that time such as World Trade Organisation (WTO) under the Trade Policy Review Mechanism and several of the United Nations' (UN's) organisations, European University Association (EUA, 2002; Amaral et al., 2008) as well as several National Higher Education Agencies in Europe, which were established for enhancement of quality assessments of higher education programs. National Higher Education Agencies have used Peer Review as an assessment method for quality management, quality assurance assessments and in accreditation processes (ENQA, 2005). Since Peer Review is carried out in several activity areas, the method is contextually adjusted to the aims and the context of the assessment.

Accordingly, there is no standardised Peer Review mechanism (Pagani, 2002; Amaral, 2008; Haug, 2009), but some structural elements are identified; for instance: (a) decision or a request of assessment; (b) an agreed set of principles/criteria/standards (guidelines) against which, in this case the research, is to be reviewed; (c) designated assessors to carry out the Peer Review; (d) set procedures leading to the final result (report) of the Peer Review process.

In this study, the *Peer Review procedure* includes (1) *the preparatory phase* when the department's work out its Self-Evaluation report (SER); (2) *the evaluation phase*, when the assessors meet to discuss the SER and then go to the department to discuss with representatives of different staff groups, doctoral students and the management of the department; (3) *the assessment phase*, the assessors' ways of using all the information in order to deliver their report and *recommendations* to the department. Those three phases lead to the creation of a fourth phase, (4) *the catalyst phase* for action planning and performance enhancement based on a time schedule. The effectiveness of the Peer Review procedure depends upon the combination of a number of factors such as *value sharing, identifying an adequate level of commitment, mutual trust and credibility* (Pagani, 2002).

The Peer Review process and its different phases as an assessment method has been criticised by representatives of different research paradigms, research perspectives (Bence and Oppenheim, 2004) and of actors taken part in the Peer Review assessments (Haugh, 2009). Critics of the method have meant that several difficulties are embedded in the Peer Review process. One is "falling into the trap of buying the internal perspective", which is presented through SER and through conducted interviews with management, lecturers, researchers and students. From the assessors perspective that is losing the distance to what is actually going on in the department. Another difficulty concerns the data collected and included in the SER, which means the assessors have to operate within the department's frames of reference and they have to accept the information offered. The assessors had the possibility to ask for additional information.

The aim is to explore how assessments of departments' research activities can contribute to improvement of research organization, research culture and credibility.

Here, the core questions are: *What are the conclusions about relative efficiency of research assessments based on external Peer Review? Do such assessments actually support the department in its development? Can general challenges in terms of research quality be deferred from the Peer Review feedback? Or in other words, is the benefits from external Peer Review of the institution's research worth the expenses and effort put into the assessment process?* This paper aims to answer these questions. The paper also makes some attempts to point out the methodological issue on combining qualitative and quantitative

information in order to gain reliable and valid knowledge of complex research questions (Tashakkori and Teddlie, 2003a; 2003b; Greene, 2007).

The main focus of the study is on the assessment of departments' research, in some cases centres or clusters of research projects at the department and never on the individual researcher's performance in round 1 and round 2. However, today all publications by CBS faculty are recorded in a *Research database (PURE)* based on a common system used by all Danish universities. These data are available for conducting quality reviews of a single researcher. During this period of 16 years, CBS has developed and implemented a *Quality Culture* (EUA, 2006), described in a review of the work by Nygaard and Christensen (2009).

Some attempts in the study have been made to draw general conclusions of common institutional challenges between assessments. In the study, we have done some attempts to compare best practice across departments and to make time series analysis of the individual department's development from first to second assessment round.

The paper consists of five sections. After this introduction we present the theoretical framework of the study. In the third section, the setting and the data collection methods are presented. In the fourth section, the analysis of results are discussed. Finally, we draw some preliminary conclusions of the development of the research activities at department level between the two assessments at CBS.

2. Theoretical Framework

In late 1980s, the emergence of the "Evaluative State" (Neave, 1988) in continental Europe put increasing emphasis on quality, control and accountability of public sector and higher education. "Clearly, the Evaluative State emerged from two very different discourses, the one European and political, the other mainly American and economic. And whilst the former tended to predominate in France, Sweden and Belgium and – somewhat later, Spain – the latter held sway in the UK and the Netherlands and tended to take root earlier" (Neave, 1998, p. 271). This movement gave increasing public relevance to quality discussions (Harvey, 2006) and quality enhancement in education system. In 80's, this was the broader political context, "in which professionalism was under challenge and the government was seeking to

introduce into the public sector managerialist ideologies: trust in professional self regulation was felt to be misplaced in the contemporary context (Henkel, 1999, p. 107).

The number of students in Higher Education was growing tremendously at that time and contributed to a rapid development of the former elite systems in European countries in terms of conditions for institutions, researchers, professors, teachers and students (Trow, 1996; Neave, 1988, 1998; Bok, 2003; Teixeira et al., 2004). The implementation of the *Bologna Declaration* (1999) aimed at making higher education *comparable* and *transparent* in order to enhance students' *mobility* and *employability* in Western European countries (EUA, 2001). The reorganised European Higher Education Area (EHEA) promoted the wave of change among European universities. The universities put stronger focus on mission and results as well as on the quality enhancement and the development of a quality culture among institutions in the European Higher Education Area. These changes and activities favoured diversity and creative actions in the universities (ENQA, 2005; EUA, 2001, 2003, 2005).

The transformation of the public institutions and higher education system, favoured the emergence of structures of institutional self-regulation, which were influenced and dependant of a number of factors (Teixeira et al., 2004). New institutionalism has pointed out that the institutional structures were of great importance for the aims and output of activities, and that these were reflected in beliefs, categories, rules, procedures and behaviours taken for granted by the members of the institution (March and Olsen, 1989; Powell and DiMaggio, 1991). Previously, institutional theorists have stressed the importance of culturally developed norms and rules as steering devices of institutional activities and for the society's understanding of activities in public institutions. However, new institutionalism as a theoretical rational does not constitute a coherent body of theories but comprises several different research approaches. In new institutionalism, new institutional economics regards institutions as "instrumentally oriented individuals", whereas organizational theorists (Silverman, 1971) argue that institutions are products of human activity, not necessarily products of conscious design (DiMaggio and Powell, 1991). New institutionalism claims that "rules" are sustained by "trust" and therefore the organization expects appropriate behaviour from members of the organization.

Amaral and Joao Rosa (2008, p. 24) argue that trust between higher education institutions and the state and society "plays an important role in determining some of the major characteristics

of quality assessment systems. The available literature shows a decline of the level of trust in public institutions in general, and in higher education institutions in particular, as well as in professionals". In the last two decades, some countries has favoured the idea of 'new public management', primarily the new managerialism and new governance. New public management represents the idea of "replacing the academic collegiality by fast adventure, carefree, computerised individualism by choice, autonomous enterprises and sudden opportunities" (Ball, 1998, p. 124). At the same time public managers will be transformed to entrepreneurs of new privatised business like institutions, ready to be economic independent of the state and government relying on business and industries. In this perspective and in Higher Education institutions, students are regarded as 'clients' or 'customers' and institutions deliver services that must satisfy the customers' needs. Accordingly, Scott (1998) argues that these ideas have gradually undermined the status of academics and the academy has no longer the prestigious status as earlier to claim its political autonomy. This question can be discussed, since it is affected by many factors and dependent on the national context.

However, the rapid transformation of universities over the last decades has challenged how various stakeholders have positioned themselves on the higher education knowledge market (Rovio-Johansson and Bull, 2006). Quality assurance mechanisms and accountability measures are put in place to ensure quality of offered services (Ball, 1998; Schwarz and Westerheijden, 2004). In this perspective, most European Higher Education institutions work intensively to refine their quality assurance and accreditation systems of various forms of education to the society. These investments include continuous quality assessments of all research, education and continuous assessment of quality management (EUA, 2001, 2004, 2007a, 2007b), administrative quality and quality of services to students and information and knowledge transfer to the society (Morgan, 2004; EUA, 2010).

In *Trends 2010* (EUA, 2010), institutional responses to questionnaires revealed that 60% of Higher Education Institutions considered that establishment of internal quality processes has been one of the most important changes during the past ten years. In relation to the present study it seems important, that this Trends 2010 underlines that: "External quality assurance must seek a balance between autonomy and accountability, take into account internal quality processes and stress the self-evaluation phase as the crucial phase in the process in order to ensure the institution's engagement in the evaluation process and the implementation of recommendations, thus leading to improved quality levels" (EUA, 2010, p.88).

This is a background and a general description on macro level of the large changes of the public sector, which included the Higher Education institutions. Our study takes its point of departure at the micro level, the department level, in one higher education institutions in Denmark.

3. Methodology, Design and Data collection

The Danish University Law of 2003 (Universitetsloven, 2003, Number 403) focused on the transformation and development of Danish Higher Education. Thus it was explicitly settled that the Dean of the Faculty has the overall responsibility of the alignment of the quality between research, education and teaching and that he/she must also secure the overall development of the institution's quality of education, teaching and research. Earlier legislation had no explicit focus on the quality dimension. In spite of this, CBS has had a strong quality focus since the first long-term strategic planning document was launched in 1993 (Handelshøjskolen i København, 1992).

Setting

CBS has undergone a rapid development since formulating its first long-term strategy in 1993. The strategy rested on 3 fundamental pillars: Internationalisation, Partnership with the Business Community, and the Learning University. In terms of internationalization a few figures illustrate the development (we only include figures from 1999 – before that time CBS was a national Danish university with a dominant domestic mindset).

Year	Peer reviewed articles (% English)	Core Faculty	Internat. Faculty incl. visiting	Number of students	Internat. students (full degree and exchange)	Number of English taught programs
1999	162 (66% English)	-	-	-	1190	-
2001	202	423	19	11,714	1,568	8
2003	218	489	36	12,270	2,222	-
2005	243	489	49	13,674	2,074	19
2007	288	500	61	14,292	2,504	26
2009	416 (75% English)	566	125	15,081	2,939	33

In terms of partnership with the business community CBS has developed a number of executive programmes, engaged in partnership with 22 of the largest Danish companies, and

established a foundation solely dedicated to executive education, which earlier in 2010 merged with its nearest competitor in executive education.

As a learning university CBS was the first Danish university to conduct international audits and since 2001 CBS has participated annually in ESMU (European Centre for Strategic Management of Universities) benchmarks on a number of areas. The peer evaluations of the research departments has been a natural element in the culture of quality and continuous improvement that has grown out of the concept of the learning university.

The first Institutional Audit of CBS was conducted 1996 by the Association of European Universities (CRE)ⁱⁱ. In 1994 the research evaluations started, albeit with a remark that no clear strategy of follow up of the evaluations has been developed. Therefore, the question was raised by the auditors whether there was an acceptable balance between local autonomy and central policy interests in the review process (CRE, 1996, p. 11).

In 1994 to 1998 all departments at Copenhagen Business School (CBS) were assessed by the same *three phase model* (see page 4), i.e., Self-Evaluation report (phase one, site visit (phase two) and Evaluation report and recommendations (phase three). In spite of the relatively fixed and well defined format of these Peer Reviews it is characteristic that follow up on the assessments was to a high degree left to the individual department. After the assessment, each department was asked to produce an action plan (phase four) based on the recommendations. Progress based on the action plans was supposed to be reported in the departments' annual report, but no systematic monitoring of development took place. In 1999 the dean presented a short summary report on the results from the assessment reports to the heads of department and to the CBS management (Report from the Dean, 1999). This report resulted in a faculty decision to continue to conduct external peer assessment of research in a second round, which took place from 2002 to 2006.

Departments and data collection

In this study, we have collected following data: The Peer Review Evaluation Reports, from 9 departments included in both first and second round of assessment. Although 13 departments took part in the first round, 4 of these departments were either closed down or merged into new units with characteristics that made it impossible to compare with the departments from the second round. Similarly, some of the units taking part in the second round of evaluations had no natural counterpart in the first round. Therefore, in order to create data material

allowing consistent comparisons over time, the analysis has been confined to 9 departments that existed under (more or less) stable conditions during first and second round.

Methods

In addition to the Self-evaluation reports and the Peer Review evaluation reports the analysis is based on individual interviews with two members of each department, who have been involved in the evaluation process during both first and second round, either as heads of department or major contributors to the evaluation process. So far, in writing moment, we have conducted six interviews.

The purpose of the interviews has been to find factors: (i) important for development of a research quality and research culture; (ii) affecting research productivity; and (iii) to verify possible causal relationships between Peer Review findings and subsequent decisions at department level. At a later stage, (iv) the quantitative analysis of performance data per department will be analysed.

Analysis of collected data

The analysis of data was delimited to the evaluation report from round 1 and round 2, for nine departments. These reports were read by the first ‘external’ author. An independent interpretation of each report was done in the first step. In the second step, assessors’ qualitative statements were identified as 10 variables of various quality related to research, covering areas such as research strategy and management, international research quality (see below). These ten variables (indicators) were based on common quality dimensions included in all review reports. In the third step, a five-level Likert Scaleⁱⁱⁱ was used to evaluate and establish a numeric value on each variable. It is possible to rank different values, but it is not possible to measure, in a meaningful way, the exact distance between numerical values, even if they are treated as having the same distance in statistical analysis.

The 10 indicators:

1. International Research Quality
2. International Research Environment/Culture/Organisation
3. Educational Environment of PhD Education
4. Involvement of Junior Researchers
5. Research Strategy and Management
6. Balance between Research and Other Duties (including Teaching)
7. Recruitment, Development and Retaining of Research Staff
8. Possible Future Enhancement of Research and Resources (Internally and Externally)

9. Development Resources, Internally and Externally
10. Publications

In the statistical analysis of data, these indicators are grouped in accordance with their content and the following areas of analysis are formed: *research culture* (1, 2, 3, 4), *research strategy* (5, 6, 7, 8), and *research productivity* (9, 10). Levels of analysis are identified as *governmental level*, *institutional level* and *departmental level*.

Qualitative and quantitative data are merged in order to secure the relevance in the evaluation of qualitative data and in the transformation process of these data to quantitative data, which further on will be supported by interview data (Tashakkori and Teddlie, 2003a; 2003b; Greene, 2007).

Credibility and trustworthiness of data and results

The results of the first author's interpretations and estimations has been re-evaluated and validated by the second "internal" author, who has detailed knowledge of all the departments. These results are further validated in interviews conducted with acting teachers and researchers, who were department heads and deeply involved in these evaluations of research in round 1 or round 2. A total of two interviews per department will be conducted with teachers and researchers, who actively took part in the first or the second round of evaluations of research at CBS.

Even though, the interview study is not finalised we will draw some tentative conclusions based on the current sample.

Anonymity of departments

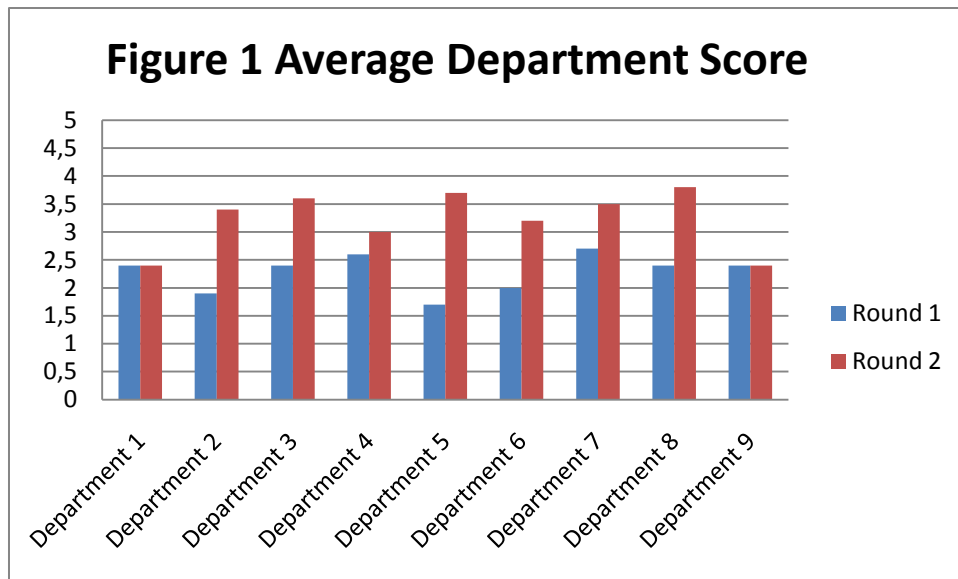
Results from nine evaluated departments are included in this study. All departments are presented with a number, in order to secure the anonymity of each department.

4. Results

Analysis of Departments

Each of the 10 indicators listed above is measured on a 5-level Likert scale, with 5 as the highest score (see Footnote iii).

In order to get a first overview of data we have calculated average indicator score for each department. These averages are presented in Figure 1.



First of all we notice that departments seem to have a different level of departure in the first round, but also that they have benefited to a different degree from the evaluation. Figure 1 demonstrates that while two departments (1 and 9) haven't developed at all other departments have changed significantly in terms of overall score. In particular, department 5, 2 and 8 have all had a tremendous increase in average score.

A simple correlation between average department score in round 1 and 2 clearly demonstrates that the evaluation had different effect from one department to the next. The correlation coefficient is -0.27, indicating a low, and negative, correlation between average score in first and second round. Thus there seems to be no linear relation between the level of departure and the effect of the evaluation.

The immediate observation that departments had different average indicator levels in round 1 is, however, not confirmed by an analysis of variance, see Figure 2. At significance levels less than 0.10 departments can in fact be considered to be at a similar level, and observed differences to be random variations.

A similar analysis for round 2 shows a much higher heterogeneity across departments. This is demonstrated in Figure 3. The p-value is 0.0001, indicating a very strong heterogeneity among departments after the second round, an indication that departments have been able to benefit in varying degree from the peer evaluation.

Figure 2 ANOVA of Department Scores – Round 1

Summary						
<i>Groups</i>	<i>Number</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
Department 1	10	24	2,4	1,6		
Department 2	10	19	1,9	0,766667		
Department 3	10	24	2,4	0,933333		
Department 4	10	26	2,6	0,488889		
Department 5	10	17	1,7	0,455556		
Department 6	10	20	2	0,444444		
Department 7	10	27	2,7	0,677778		
Department 8	10	24	2,4	0,266667		
Department 9	10	24	2,4	0,266667		
ANOVA						
<i>Source of variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit.</i>
Among groups	8,955556	8	1,119444	1,707627	0,109095	2,054882
Within groups	53,1	81	0,655556			
Total	62,05556	89				

Figure 3 ANOVA of Department Scores – Round 2

Summary						
<i>Groups</i>	<i>Number</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
Department 1	10	24	2,4	1,822222		
Department 2	10	34	3,4	0,711111		
Department 3	10	36	3,6	0,266667		
Department 4	10	30	3	0,444444		
Department 5	10	37	3,7	1,122222		
Department 6	10	32	3,2	0,177778		
Department 7	10	35	3,5	0,277778		
Department 8	10	38	3,8	0,4		
Department 9	10	24	2,4	0,266667		
ANOVA						
<i>Source of variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Among groups	22,15556	8	2,769444	4,540992	0,000136	2,054882
Within groups	49,4	81	0,609877			
Total	71,55556	89				

These results indicate that a number of factors apart from the ‘treatment’ (the evaluation) must play a role for the benefit a department gains from participating in the evaluation. Such ‘enabling’ factors could be a sense of urgency, departments’ readiness for change, self-image

(complacency), the department's preparing for the evaluation and the engagement of faculty, the leadership of the department, or general surrounding conditions. Some indication of the influence of these factors is found in the interviews.

The above conclusion is confirmed if one considers the differences in total (or average) score from round 1 to round 2. Since the data represent two sets of consecutive measurements of the same units it makes sense to treat the observations statistically as 'matched pairs'. This analysis is conducted in Figure 4. The analysis demonstrates much clearer than in Figure 2 and 3 how strong the effect of the 'treatment' really is when one eliminates the individual department variation by considering the differences.

Figure 4 ANOVA Analysis of differences across departments

Summary						
<i>Groups</i>	<i>Number</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
Department 1	10	0	0	0,222222		
Department 2	10	15	1,5	1,388889		
Department 3	10	12	1,2	0,622222		
Department 4	10	4	0,4	1,155556		
Department 5	10	20	2	1,555556		
Department 6	10	12	1,2	0,622222		
Department 7	10	8	0,8	0,177778		
Department 8	10	14	1,4	0,711111		
Department 9	10	0	0	0,666667		
ANOVA						
<i>Variationskilde</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Among groups	38,62222	8	4,827778	6,100624	3,9E-06	2,054882
Within groups	64,1	81	0,791358			
Total	102,7222	89				

Analysis of the indicators

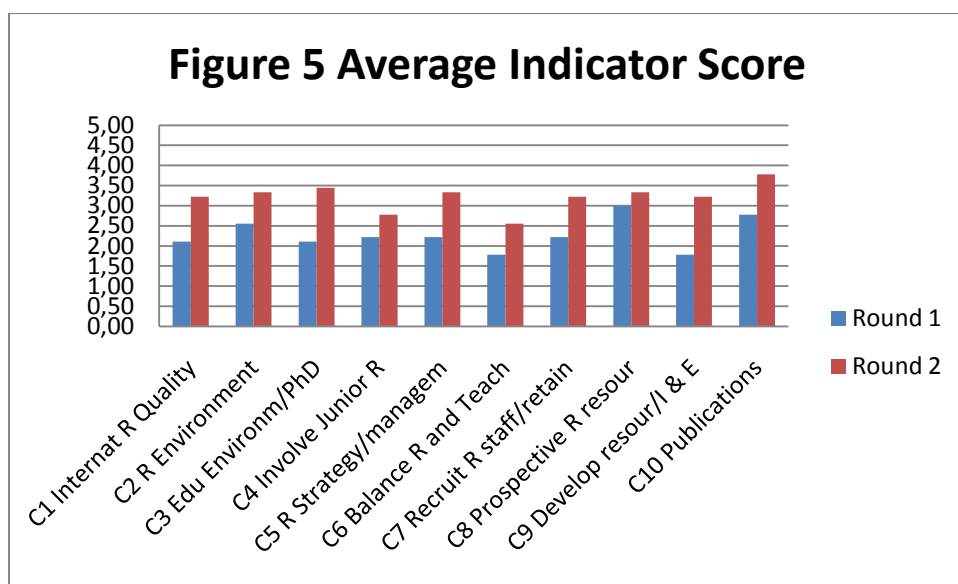
Of course the initial analyses of departments in Figures 1, 2, 3 and 4 only provide a very rough picture of the assessment results since one adds 'apples and pears' (different indicators) together in an illegitimate manner. In order to get further insight in the data material we need more analyses of the different indicators. This is the purpose of this section.

In Figure 5 average indicator scores across departments are presented. We notice an improvement in all indicators from round 1 to round 2, but we also notice a slight tendency to co-variation in indicator score from round 1 to round 2, indicating that general areas of

concern in the first round to some extent were identified also in the second round. This is supported by a correlation coefficient between scores in round 1 and round 2 of 0.58.

At a closer look at the individual indicators and their development the following results appear. The areas scoring lowest in round 1 are C6 (Balance between research and other duties, including teaching), and C9 (Development resources). In some sense this appears natural since these indicators are to a large extent beyond the department's control. In the Self evaluation report it seems natural to blame outside factors for eventual under performance [for eventual declines in research performances or developmental activities], and at the same time use the opportunity to argue for better working conditions. Similarly, the highest score is found in indicator 8, the expected future performance, which is more based on ambitions and plans than actual results.

In round 2 we first of all note that all indicators have increased, and in spite of individual observed differences there are no significant differences among the 10 indicators, meaning that they all support the general picture of positive development, cf. Figure 7.



Secondly, one could ask if the 10 indicators are in fact telling 10 individual stories or if it is possible to reduce the number of indicators to a more limited number while still maintain the relevant information. In order to find out we tentatively classify the 10 indicators in 3 different groups according their semantic content:

Research Quality and Culture
 Research Strategy
 Results/productivity

including indicators 1-4
 including indicators 5-8
 including indicators 9-10

Analysis of Variance of the 10 indicators

Figure 6 – ANOVA of indicators from Round 1 – shows that the 10 indicators are in fact significantly different at the 5% level. However, this was not the case in round 2, cf. Figure 7.

Figure 6 ANOVA Analysis of Indicators – Round 1

Summary						
Groups	Number	Sum	Average	Variance		
C1 Internat R Quality	9	19	2,111111	0,361111		
C2 R Environment	9	23	2,555556	0,277778		
C3 Edu Environm/PhD	9	19	2,111111	0,361111		
C4 Involve Junior R	9	20	2,222222	0,694444		
C5 R Strategy/managem	9	20	2,222222	1,194444		
C6 Balance R and Teach	9	16	1,777778	0,194444		
C7 Recruit R staff/retain	9	20	2,222222	0,444444		
C8 Prospective R resour	9	27	3	0,25		
C9 Develop resour/I & E	9	16	1,777778	1,694444		
C10 Publications	9	25	2,777778	0,694444		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Among groups	12,72222	9	1,41358	2,292292	0,024083	1,999115
Within groups	49,33333	80	0,616667			
Total	62,05556	89				

Figure 7 ANOVA Analysis of Indicators – Round 2

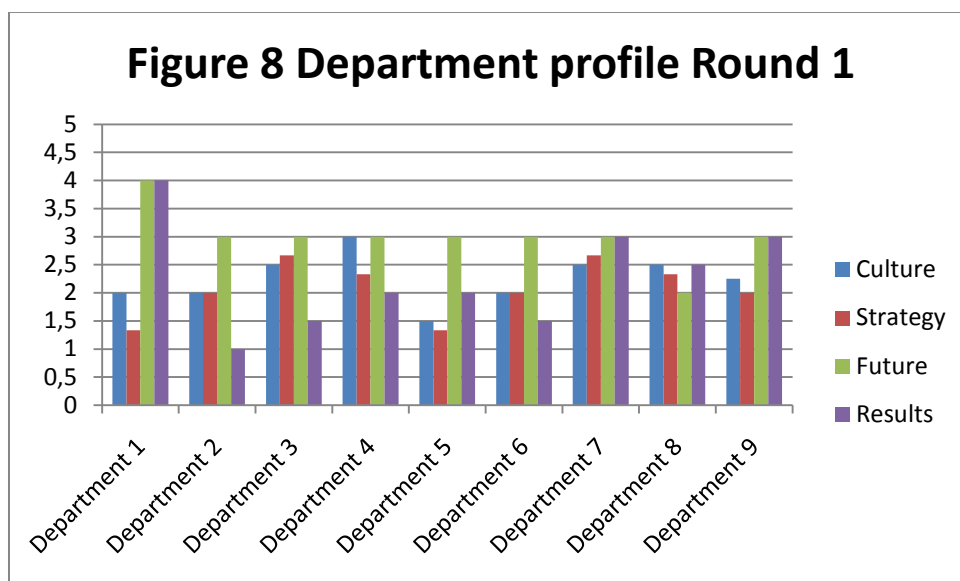
Summary						
Groups	Number	Sum	Average	Variance		
C1 Internat R Quality	9	29	3,222222	1,194444		
C2 R Environment	9	30	3,333333	0,5		
C3 Edu Environm/PhD	9	31	3,444444	0,777778		
C4 Involve Junior R	9	25	2,777778	0,444444		
C5 R Strategy/managem	9	30	3,333333	2,5		
C6 Balance R and Teach	9	23	2,555556	0,277778		
C7 Recruit R staff/retain	9	29	3,222222	0,194444		
C8 Prospective R resour	9	30	3,333333	0,5		
C9 Develop resour/I & E	9	29	3,222222	0,694444		
C10 Publications	9	34	3,777778	0,694444		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Among groups	9,333333	9	1,037037	1,333333	0,23322	1,999115
Within groups	62,22222	80	0,777778			
Total	71,55556	89				

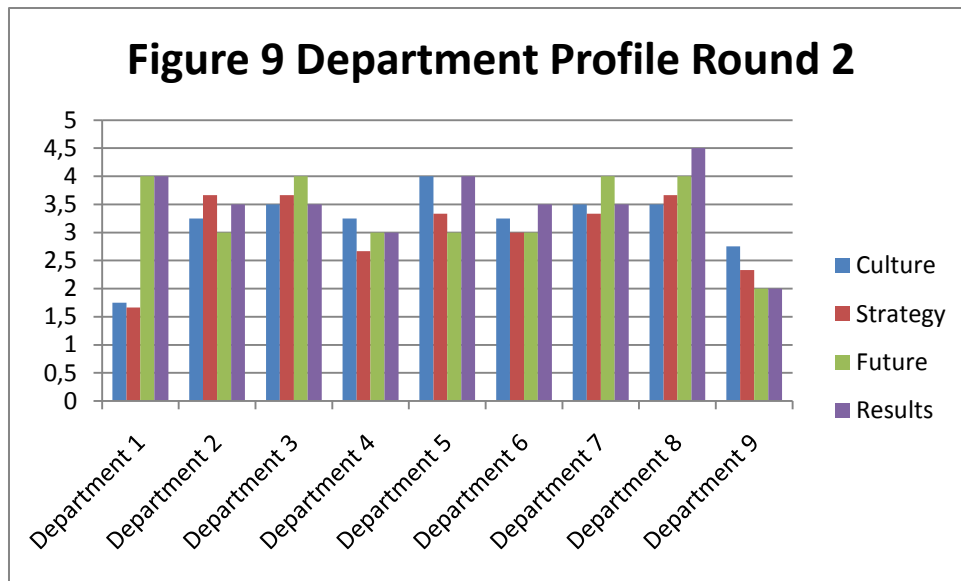
In round 2 all indicators seem to measure a general quality level of the department. One way to interpret this is a hypothesis that during 1st round departments were only at the beginning to develop their research quality and it was somewhat random which area a department had focused on. In the second round, however, all departments were at a higher level of maturity which means that they have met a reasonable standard on all indicators – even though some departments had been more successful than others, which was demonstrated in Figure 4.

Profiles of the 9 departments

Based on these findings it is now possible to establish a department profile for each department, using the 3 aggregate groups listed above; research culture, research strategy, and results/productivity. However, analyses of variance for each group reveal that while the indicators of group 1 and 3 are homogeneous, this is not true for indicators of group 2 – research strategy. But if indicator 8 – possible future enhancement of research resources – is omitted, then the 3 remaining indicators of group 2 can be considered to be homogeneous. To define indicator 8 as a special group seems natural as this is the only indicator that focuses on the future (which is open to more or less optimistic scenarios!).

Based on the four new indicator groups we can present a more detailed picture of the 9 departments in terms of their score on culture, strategy, future expectations and results. This is presented in Figure 8 and Figure 9.





Accordingly, figures 1 – 2 indicate that the nine departments have undergone significantly different levels of improvement. While some departments have not developed at all, others have more than doubled their performance score.

5. Conclusions

This case study mirrors at micro level, the wave of change of the public sector (March and Olsen, 1989; Powell and DiMaggio, 1991), which have happened during the last three decades in Western European countries. As mentioned by Neave (1988, 1998) the *Evaluative State*, at political macro level, became associated with a higher degree of autonomy and with increasing diversity at institutional level. At the same time, we can verify growing expectations and demands from the society on higher education institutions (Teixeira et al.,

2004; Amaral et al., 2008) in terms of economic and social benefits, life-long-learning opportunities, internationalisation and globalisation of higher education.

In the present case, this assessment process of research was initiated by the management of the Copenhagen Business School (CBS). Long discussions in 1994 at CBS ended by a decision, that these systematic reviews have to be initiated as bottom-up processes, in which the departments have to take responsibility and own the change process of the department. Accordingly, this internal process initiated by a Self-Evaluation Report (SER) and supplemented with an external evaluative review (Trow, 1994), finalised by an Evaluation Report belonging to each department. Today, the linkage to resource allocation, control and accountability, is a common feature of these evaluative processes conducted by external experts. In this case and at that time, the management decided that these evaluations have to be *external supportive evaluations*, primarily focusing on the development of *research quality and culture, research strategy and research productivity*. In this paper we have analysed how these systematic reviews of nine departments have been carried out from 1994 to 2007, in two different rounds.

The data collection of the study includes two assessment rounds and data from: Self-Evaluation Reports, Peer Review Reports/Evaluation Reports and interview data from two teachers/researchers per department, who took part in both evaluation rounds. In this paper we have confined data to the Evaluation Reports. The content of these reports are analysed and ten indicators are identified. In each report (18 all together), the ten indicators are evaluated, using five graded scale. The qualitative assessments are transformed to quantitative data in order to make comparisons possible between the reports for each of the nine departments.

The main argument launched in this paper is that there are many possibilities to support development of research and education at a university department. However, the data analysis indicates significantly different levels of improvement among the nine departments in terms of research quality and culture, research strategy and research productivity. The relatedness between context and research activities of these assessments and evaluations has a unique character for each department, which confine comparisons among departments. The meaningful comparison of a department can only be done between the first and the second round for each department.

In this study, we have investigated the *relative efficiency* as the evaluation of partly the departments' self-understanding through its way of offering the reviewers a description of the departments' research activities, and partly their presentation of the research activities to the assessors during their site visit resulting in an evaluation report. The results demonstrate that

- Although departments were to some extent different in terms of their point of departure, overall their performance level at the time of round 1 (1993-1999) was not significantly different, cf. Figure 3.
- Departments vary significantly in terms of the benefit they seem to have obtained from the evaluation, cf. Figure 4 and 5. These differences do not stem from their level of departure (even though there seems to be a – weak – tendency to benefit more if a department comes from a low level of performance, cf. a correlation of -0,27). Rather other enabling factors, such as sense of urgency, departments' readiness for change, self-image (complacency), the department's preparation for the evaluation and the engagement of faculty, and the leadership of the department, seem to play important role for the benefit. These 'enablers' were mentioned in several of the interviews.
- It seems fair to conclude that this type of department evaluation – when conducted across departments – provide information not only on the individual department, but also adds to identify areas of general concern to the institution. This was clear from both rounds.
- Also, it seems valid to state that the institutionalization of peer review evaluation of departments as it was the case at CBS since 1994 has added to the general quality awareness and performance focus that has resulted in the overall quality profile that characterizes CBS today.

In that sense we are able to confirm the findings from Henkel (1999, p. 105) who claims that the research assessment [modernization in UK] triggered substantial changes, for instance “in the management of research functions, in the culture of university departments as well as the relationships between the individual academic, the discipline, the department and the institution. Also, individual professional identity has been affected and concepts of research responsibilities, argues Henkel. We could add “self image” and “self conception” to this list of impacts on individual researchers, according to one of our interviewees.

On page 4 we raised the following research questions: *What are the conclusions about relative efficiency of research assessments based on external Peer Review? Do such*

assessments actually support the department in its development? Can general challenges in terms of research quality be deferred from the Peer Review feedback? Or in other words, is the benefits from external Peer Review of the institution's research worth the expenses and effort put into the assessment process?

The short answers to these questions are 1) Peer reviews do have a (sometimes significant) effect on the relative efficiency of research performance at department level, 2) The effect depends on the existence of a number of 'enabling' factors, 3) General institutional challenges can be deferred from departments' peer reviews, 4) Department peer reviews is one among several factors to create a general quality culture, 5) Peer reviews have prepared the institution and the departments for a more focused research management and regular systems of accountability that seems to be a main tendency in these years. Overall, the peer reviews seem to have been worth their cost in terms of money and time invested in these processes.

References

- Amaral, A., Rovio-Johansson, A., Joao Rosa, M. and Westerheijden, D. (2008). *Essays on Supportive Peer Review*. New York: Nova Science Publishers.
- Ball, S. J. (1998). Big Policies/Small World: An introduction to international perspectives in educational policies. *Comparative Education*, 34, 2, pp. 119-130.
- Bence, V. And Oppenheim, C. (2004). The influence of peer review on research assessment exercise. *Journal of Information Science*, 30, 4, pp. 347-368.
- Bok, D. (2003). *Universities in the Marketplace: The Commercialization of Higher Education*. Princeton, NJ.: Princeton University Press.
- Bologna Declaration of 19 June 1999. Brussels: European Commission.
- CRE [Association of European Universities] (1996). *Institutional Audit of the Copenhagen Business School. CRE Auditors' Report*. Geneva: CRE [Association of European Universities].
- DiMaggio, P. J. and Powell, W., W. (1991) Introduction. In W. Powell and P. DiMaggio (Eds.), *The New Institutionalism in Organizational Analysis*. Chicago: The University of Chicago Press.
- European Association for Quality Assurance in Higher Education [ENQA] (2005). *Standards and Guidelines for Quality Assurance in European Higher education Area*. Helsinki: European Association for Quality Assurance in Higher Education.
- European University Association (2001). *Messages from Salamanca: Shaping the European Higher Education Area*. Brussels, Belgium: European University Association.
- European University Association (2003). *Graz Declaration. Forward from Berlin*. Brussels, Belgium: European University Association.
- European University Association (2004). *EUA's QA policy position in the context of Berlin Communiqué*. Brussels, Belgium: European University Association.
- European University Association (2005). *Glasgow Declaration. Strong Universities for a Strong Europe*. Brussels, Belgium: European University Association.
- European University Association (2006). *Quality Culture in European Universities: A Bottom-Up Approach. Report on the Three Rounds of the Quality Culture Project 2002-2006*. Brussels, Belgium: European University Association.
- European University Association (2007a). *EUA position on Quality*. Brussels, Belgium: European University Association.
- European University Association (2007b). *Lisbon declaration. Europe's Universities beyond 2010: Diversity with a common purpose*. Brussels, Belgium: European University Association.
- European University Association (2010). *Trends 2010: A decade of change in Higher Education*. Brussels, Belgium: European University Association.
- Foss Hansen, H., Maskell, P. And Mouritsen, J. (1995). *Extern forskningsevaluering ved Handelshøjskolen i København. Det Økonomiske Fakultet*. [External evaluation of research at Copenhagen Business School. Faculty of Economics and Business Administration. Internal report]. Copenhagen: Copenhagen Business School.
- Greene, J. C. (2007). *Mixed Methods in Social Inquiry*. San Francisco, CA: Jossey-Bass.
- Handelshøjskolen i København (1992). *Handelshøjskolen i København. Det Økonomiske Fakultet. Langtidsplan 1992-1996*. [Copenhagen Business School. Faculty of Economics

- and Business Administration. Long Term Planning, 1992-1996. Internal report]. Copenhagen: Copenhagen Business School.
- Harvey, L. (2006). Understanding Quality. In Froment, E., Kohler, J., Purser, L. and Wilson, L.: *EUA Bologna Handbook. Making Bologna Work*. B4, 1-1. Berlin: Raabe
- Haug, P. (2009). Evaluating external expert panel evaluation. *Nordisk Pedagogik [Nordic Educational Research]*, Vol. 29, pp. 210-220.
- Henkel, M. (1999). The modernisation of research evaluation: The case of the UK. *Higher Education*, 38, pp. 105-122.
- Lindsay, A. W. (1982). Institutional Performance in Higher Education: The Efficiency Dimension, *Review of Educational Research*, Vol. 52, 2, pp. 175-199.
- March, J. and Olsen, J. (1989). *Rediscovering institutions: the organizational basis of politics*. New York: Free Press.
- Morgan, K. J. (2004). The research assessment exercise in English universities, 2001. *Higher Education*, 48, pp. 461-482.
- Neave, G. (1988). On the Cultivation of Quality, Efficiency and Enterprise: An overview of recent trends in Higher Education in Western Europe, 1986-1988. *European Journal of Education*, Vol. 23, 1/2, pp. 7-23.
- Neave, G. (1998). The Evaluative State Reconsidered. *European Journal of Education*, Vol. 33, 3, pp. 265-284.
- Nygaard, C. and Christensen, B. (2009). Quality work at Copenhagen Business School – how to create a university culture. In Froment, E., Kohler, J., Purser, L. and Wilson, L.: *EUA Bologna Handbook. Making Bologna Work*. C2, 2-3. Berlin: Raabe.
- Pagani, F. (2002). Peer Review as a tool for Co-operation and change. An analysis of an OECD working method. *African Security Review*, Vol. 11, 4.
- Powell, W. and DiMaggio, P. (1991). *The New Institutionalism in Organizational Analysis*. Chicago: University of Chicago Press.
- Report from the Dean (1999). *Forskningsevaluering ved Det Erhvervsøkonomiske Fakultet*. Copenhagen: Copenhagen Business School.
- Rovio-Johansson, Airi and Bull, Tove (2006). Roles and Interactions of Different Players in Quality Assurance. In Froment, E., Kohler, J., Purser, L. and Wilson, L.: *EUA Bologna Handbook. Making Bologna Work*. B4.4-1. Berlin: Raabe
- Schwarz, S. and Westerhijden, D. (2004). Accreditation in the Framework of Evaluation Activities: A comparative Study in the European Higher education Area. In Schwarz, S. and Westerheijden D. F. (Eds.), *Accreditation and evaluation in the European Higher education Area*, (pp. 1-41). Dordrecht: Kluwer Academic Publishers.
- Scott, W. (1991). Unpacking institutional arguments, in W. Powell and P. DiMaggio (1991), *The New Institutionalism in Organizational Analysis*. Chicago: University of Chicago Press.
- Scott, P. (1989). The Power of Ideas. In C. Ball and H. Eggins (Eds.), *Higher Education into the 1990s. New Dimensions*, pp. 7-16. Dordrecht: Kluwer Academic Publishers.
- Silverman, D. (1970). *The Theory of Organisations*. London: Heinemann.
- Tashakkori, A. and Teddlie, C. (2003a). *Handbook of Mixed Methods in Social and Behavioural Research*. Thousand Oaks, CA: Sage.
- Tashakkori, A. and Teddlie, C. (2003b). The past and future of mixed method research: From data triangulation to model designs. In A Tashakkori and C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioural Research*, pp. 671-701. Thousand Oaks, CA: Sage.
- Teixeira, P., Jongbloed, B., Dill, D. and Amaral, A. (2004). *Markets in Higher education. Rhetoric or Reality?* Dordrecht: Kluwer Academic Publishers.

- Trow, M. (1994). *Academic Reviews and the Culture of Excellence*. Report 1994:1. Stockholm: Chancellor's Office: Studies of Higher Education and Research.
- Trow, M. (1996). Trust, Markets and Accountability in Higher Education: A comparative Perspective. *Higher Education Policy*, Vol. 9, 4, pp.309-324.
- Universitetsloven (2003, Number 403). *Betænkning over Forslag til lov om universiteter (universitetsloven). Betænkning afgivet af Udvalget for Videnskab og Teknologi*. København: Ministeriet for Videnskab og Teknologi. [University Law, Number 403, 2003. Copenhagen: Ministry of Science and Technology.]
- Welch, A. R. (1998). The Cult of Efficiency in education: Comparative reflections on the reality and the rethoric. *Comparative Education*, Vol.34, 2, pp. 157-175.

Notes

ⁱ One of the authors has been deeply involved in the assessment process at CBS's management level, first as Programme Director of the M.Sc. programme from 1994 – 2000, subsequently as Dean of the Faculty of Economics and Business Administration from 2000 – 2005 and finally as Director of Accreditation and Quality Assurance from 2006 and until today. The other author has been associated with CBS since 2009, but has not been directly involved in this research evaluation, round 1 and Round 2..

ⁱⁱ CRE (*Conférence Permanente des Recteurs, des Présidents et Vice-Chanceliers des Universités*), known as The Association of European Universities (in English) was founded 1957. CRE merged 2001/2002 with the Confederation of European Rectors' Conferences and established European University Association (EUA).

ⁱⁱⁱ Rensis Likert (1932; *Archives of Psychology*, 140, pp. 1-55) developed the 5-point item measurement instrument, often used in social science research and known as the Likert scale. The Likert scale usually consists of a number of Likert items, most often statements, and the term *Likert scale* is used to indicate the summated scale. A Likert item is a statement, which the respondent is asked to evaluate according to any kind of subjective or objective criteria. A Likert item consists of five-levels of agreement: (1) strongly disagree; (2) Disagree; (3) neither agree nor disagree; (4) Agree; (5) Strongly agree. - The main point is that Likert scaling is a bipolar scaling method, measuring either positive or negative response to a statement. In this study the items in the bipolar five-level scale were the 10 variables (criteria)/indicators, which were evaluated consecutively by the authors.