

## Paper title

Evaluating Quality of Higher Education by Assessing its Output – The Swedish Example

## Autobiographical note for author:

Bjarne Bergquist, Professor, PhD<sup>1</sup> and Rickard Garvare, Professor, PhD<sup>2</sup>

Quality Technology & Management

Department of Business Administration, Technology and Social Sciences

Luleå University of Technology, 971 87 Luleå, Sweden

<sup>1</sup> [bjarne.bergquist@ltu.se](mailto:bjarne.bergquist@ltu.se); <sup>2</sup> [rickard.garvare@ltu.se](mailto:rickard.garvare@ltu.se)

Ändrad fältkod

## Abstract

*Purpose:* The purpose of this study is to describe experiences of evaluating quality of higher education based on ~~indicators of~~ output indicators rather than process indicators as main basis for assessment.

*Methodology/Approach:* Deriving from previous experiences this observational study addresses the method for governmental evaluation of quality in Swedish higher education which has recently been fundamentally changed. A previously broad approach including measurement of various process indicators was replaced by a strict focus on output, mainly in terms of to what degree student theses fulfil stated overall aims of the education program or subject area at hand. The new assessment is mainly performed by a second set of teachers from other universities evaluating a random sample of the students' final theses of a given university program. This paper examines the new evaluation approach from a qualitative perspective of its review process.

*Findings:* ~~The new method has been controversial, receiving criticism from several directions. In 2012, the European Organisation for Quality Assurance in Higher Education criticized the new system for, among other things, not being directed toward improving the education but rather to rank programs or to punish programs of low quality. However, much of the critique has been based on theoretical reasoning regarding how the assessment was to be performed, rather than on measurements of actual performance or outcome.~~ Different problems regarding reliability and validity in the review process are discussed, including issues related to consensus meetings, sample sizes, university self-assessment, and of the ranking, reprisal and improvement components of the evaluation results.

*Research limitations/implications:* There is a need to further study the reliability of the university program assessment process, and the individual thesis assessments in particular. The validity of the assessment method, primarily focusing on external reviewers of approved theses, is also questionable and in need of further studies. ~~Focus-~~ The focus of the assessment is now on controlling that the learning goals ~~of set by the Swedish higher~~ Higher education Education Act are implemented. ~~Our~~ Besides reliability and validity concerns, our contention is that more emphasis should also be put on the perspectives of other stakeholders.

*Originality/Value of paper:* Given the importance of an inadequate quality statement and how that would reflect the overall ranking of the program, especially for many small programs, this study ~~argues the importance of developing~~ gives empirical support for many of the ~~the~~ flaws highlighted in the published critiques of the assessment system that primarily were of

conceptual or theoretical nature, based on how the assessment was thought to be performed rather than how the assessment did work once implemented.

*Keywords:* quality assessment, higher education, performance measurement

## Introduction

The ambition to assess and rank the quality of higher education has grown considerably during the last decades. Rising demands of competence from public and private employers drive a climbing percentage of the population to study at universities, leading to an increase in public spending on higher education. Students want good education and there is a civic interest in receiving best possible effects of public spending. However, the growing number of programs and inauguration of new universities makes evaluation of the quality of different programs difficult for potential students and employers. Third party assessment of higher education has therefore gained an increased prominence.

Assessments may have several goals; quality control to secure that rules are followed, that political decisions are implemented, that money spent is put to best use, ranking for externals, e.g. to facilitate for presumptive students or research partners, and to act as tools for improvements for the assessed organization. Cronbach et al. (1980) defined evaluation of educational programs as a “systematic examination of events occurring in and consequent of a contemporary program – and examination conducted to assist in improving this program” (Cronbach et al., 1980, p.14), and they portrayed the evaluator as an educator rather than a referee. The possibility of using the assessment as a tool for improvement lies in the verbal, quantitative descriptions to explain how the organization works to fulfil a certain requirement of the instrument, thereby exposing flaws or neglected areas though working with the instrument.

Weiss saw evaluation as “the *systematic assessment* of the *operation* and/or the *outcomes* of a program or policy compared to a set of *explicit* or *implicit standards*, as a means to contributing to the *improvement* of the program or policy” (Weiss, 1998, p. 4, emphasis in original). Shaw, Greene and Mark conceptualize systematic evaluation as a “social and politicized practice that nonetheless aspires to some position of impartiality or fairness, so that people can contribute meaningfully to the well-being of people in that specific context and beyond (2006, p.5-6). While improvement is a stated purpose of Weiss’ (1998) definition, the referencing to standards speaks for a more positivistic and control based view compared to that of Cronbach’s (1980). Shaw et al. (2006) on the other hand emphasize political motives of evaluation and measurement and impartiality aspects of the procedure.

University rankings using quantitative measures are popular, but tend to differ based on how ranking formulas are designed. Many university rankings are based predominantly on indicators of research output rather than on educational outcomes. Unfortunately, excellent research does not necessarily mean excellent education. Most rankings are, due to simplicity, made on a university level, not of single education programs, and some rankings are based on

non-disclosed criteria. Ranking tables may be quickly grasped, but may also lead to wrong conclusions about the educations concerned if the reader does not weigh in what is measured and how rankings are calculated. It is unlikely that presumptive university students have the time, competence, or even the access to how these indices are created and may thus be tricked into selecting poor programmes at highly prestigious research universities.

### **Assessment of Swedish education**

In 1999, the shared view that the European higher education sector needed to strengthen its competitiveness and attractiveness led to the creation of the European Higher Education Area (EHEA) and the so called Bologna process, and Sweden was one of the 30 countries signing the original treaty. This process is intended to provide a basis for harmonisation of the many different systems of higher education in the European countries, including issues of quality assessment. In 2001, EHEA objectives expanded through the Prague Communiqué from an original aim of making European education systems comparable, to also include development of national quality assurance and national qualification frameworks (ENQA, 2001). The Communiqué highlighted the need for establishing a common framework and to disseminate best practice. One year prior to that the European Organisation for Quality Assurance in Higher Education (ENQA) was founded and ENQA was in 2005 mandated to developing and proposing standards, procedures and guidelines on quality assurance in the EHEA (Kauku, 2012).

Swedish National Agency for Higher Education (SNAHE) has been a member of ENQA since 2000, and following a review, became full member in 2005/06 (ENQA, 2012a). In the year 2001, a new model for assessing Swedish higher education programmes was launched by the SNAHE, with the intent that all higher education programmes should be evaluated once every six years. The stated aims were to secure that the quality of programs were equivalent across the country, that the international reputation of Higher Education Institutions (HEI) of Sweden would remain high, and that the evaluation would develop the teaching forms and the teachers' abilities to keep the quality high despite a massive expansion of the number of student seats during the 1990:s. (Karlsson, Andersson and Lundin, 2002, Wahlén, 2004)

Explicitly, the purposes were *control* over the sector due to public spending for political concerns, that the HEI could use the evaluation for *improvements*, that students and other stakeholders would be able to use the results as *information* related to their choice of program, and for students' and other stakeholders' *comparison* between HEI and their counterparts internationally. A meta-analysis showed that the improvement purpose was hard to combine with the control purpose (Karlsson, Andersson and Lundin, 2002, p.13-14). The HEI:s often used the work with self-assessments for improvement, but were also critical to the administration involved e.g. in delivering statistics for the assessment report (Wahlén, 2004).

In 2013, SNAHE was replaced by the Swedish Higher Education Authority (SHEA). It is estimated that the SHEA has the responsibility to assess 2200 programmes (SHEA, 2013a).

### **A new model for assessment of Swedish higher education**

In the year 2011, the Swedish self-assessment based model was replaced by an assessment system that focuses strictly on results in terms of students' achieved learning outcomes, in particular the quality and relevance of their theses. It replaced a system that had a wider focus on indicators of inputs, processes and resources such as teacher competence or the socio-economic status of new students, see for example SHEA (2013a). In the new system, students' learning outcomes are divided into three main parts; a) knowledge and understanding, b) skills and abilities, and c) assessment skills and attitudes. The incentives for HEI of being given a favourable quality grade are significant. The programs being given the very high quality grade receive additional funding, and the inadequate quality graded programs may get their examination licence revoked. .

The new assessment system was strongly criticized both before and after its introduction (Adamson, 2013). In 2012, it was criticized by ENQA for, among other things, not being directed toward improving the education but rather to rank programs or to punish programs of low quality. Hallén & Nilsson (2012) reported discrepancies between instructions provided and actual assessments performed. The unclear stature and independence of the assessing body, the Swedish Education Authority has also been criticized. ENQA raised doubts regarding SNAHE's operational independence and a system not aligned with the European Standards and Guidelines (ESG) for external quality assurance agencies (ESQA, 2012, p.1).

Following the revision of the Swedish assessment system in 2012, ENQA found that the new system no longer complied with ENQA standards. Problems that were highlighted include a failure to "take into account the effectiveness of the internal quality assurance processes described in Part 1 of the European Standards and Guidelines" (ENQA, 2012, p. 11). The focus on outcome was also considered to be more on control or ranking, rather than assessment and enhancement (ibid. p. 11-12). The assessment of Swedish higher education institutions (HEI) were seen as being made mainly to punish bad programmes by foreclosing their right to examine students, rather than to aid those HEI, which was not the aim of ENQA. As a result of the report, SNAHE was given status as non-compliant of the European Standards and Guidelines, leading to Sweden being removed from full membership of the ENQA list to "ENQA Full member under review" (ENQA 2012b). Such an assessment is comparable to those given by SNAHE to programs not fulfilling current requirements where rights to examine students of a particular program is given a probationary status.

### **Observations**

One of the authors has been part of a group of evaluators (from here on: assessors) assessing the quality of Swedish university programs and exams within a cluster of engineering. The group consisted of university teachers from the senior lecturer or professor level as well as representatives of students and the industry, about twelve persons in total. Also involved was a few representatives from SNAHE who did not partake in the assessment, but acted as facilitators during the meetings. Supporting the group was also about ten so called expert readers who read theses and provided their assessments. The group had representatives of most, but not all, HEI that were to be assessed, and had the responsibility to assess about 40 programs that were considered to reside within the area.

The stated intentions as described to the assessors were that they should aim at assessing the programs' adherence to the program area learning goals. These goals were set for the programs by The Higher Education Act, originating from the so called Dublin descriptors, developed through the Bologna process (Hallén & Nilsson, 2012). These goals state that the students, for instance should be able to show knowledge and understanding of the area of study, can show methodological skills, and having the ability to combine and analyze complex tasks. Students should be able to work independently, have the ability to present their work orally and in writing, and have the ability to participate in research or advanced work in their discipline. Students should also being able make relevant assessment regarding ethical, societal and scientific considerations, show insights into science's possibilities and limits, the role for science in society and awareness of how it is used, and to show ability to identify further knowledge gaps.

The assessment should be based on four different sources of data; 1) Student theses, 2) University self-assessments, 3) Alumni surveys and 4) Group interviews with University representatives and students. However, the theses were in practice the primary basis for judgment of if the program was complying with SNAHE goals. The other sources were used when data could not be validly obtained via examination of theses.

All the goals in terms of educational results were divided into the areas of: *Knowledge and understanding*, *Skills and abilities*, and *Assessment skills and attitudes*. To our knowledge, none of the assessment groups assessed fulfillment of all the learning goals. Based on a discussion within the group of assessors, the group chose to assess the programs according to a subset of goals that differed among the different types of exams (e.g. Candidate, Master of Science and Master of Engineering). The main rationale for this was a perception that the goals to be excluded would have been difficult to measure based on available data, rather than selecting goals based on importance.

The group of assessors operationalized the selected goals into a list of requirements and criteria for discrimination between three quality levels based on goal fulfillment: *Very high quality*, *High quality* and *Inadequate quality*. A training exercise was performed where all assessors involved made assessments of three example theses. The results were then compared and discussed within the group during a consensus meeting, thereby calibrating individual differences in terms of judgment among assessors.

Each program to be assessed had to have had at least five theses examined during the last three years. SNAHE randomly selected and made anonymous about 400 theses from the total pool of theses and then distributed these amongst the assessors. Each assessor was given around 20 theses to grade, and each thesis was only graded by one assessor. The thesis assignments were made to preclude assessors to rate the quality of the theses of their own university. Many of the theses had been written in collaboration between two students, but the information of whether a thesis was written by one or two authors was not presented to the assessors. The information given did, however include statements of the level of the thesis (e.g. Master, Bachelor, or the Swedish degrees *Civilingenjör* and *Magister*).

When the individual assessments were completed, the results were compiled by SNAHE. The assessors then assembled in smaller groups to rate the different programs. The grading was based upon rating of fulfillments of the fulfillments of SNAHE's learning goals by the individual theses. This step was not intended to be primarily mathematical, but quantitative summing of results was nonetheless an important basis for decisions.

### **Discussion and conclusions**

Our results generally agree with those obtained in previous studies. Based on their own experiences of partaking as assessors Hallén & Nilsson (2012) conclude that the current Swedish system for assessment of higher education has several drawbacks, for instance in terms of limited goal coverage and risks of bias, e.g. due to the risk of jeopardizing the job security of involved teachers and other professionals if examination licences are revoked. We also believe that the current system does not really make use of the substantial advantages of focusing on results instead of on processes.

The quality of an assessment system is never better than the quality of its indicators. It is of course fundamental that the measurements used provide a fair portrayal of the situation at hand. To adequately assess the quality of a higher educational program by only measuring its results implies that the indicators used are related and relevant for the intended aims of the program.

Our study indicates that students' theses are the primary basis for judgment of programs. The degree to which a student has received tutoring is not clearly shown, and thus not weighed in, and neither is the fact that a thesis is written by a single or two students. Our experience from tutoring is that those factors may have a strong impact on the quality of the thesis, but not necessarily on the learning outcomes for the student in doing the thesis work. Being single author places requirements that one should be responsible for all the work, while a pair of students may use their complementary skills, and each student will not have to excel in all areas.

The focus of the assessment on the thesis is also problematic, since the thesis work only represents a fraction of the total program. A thesis is intended to show that the student can work individually, on a specific task that is representative of engineers or researchers of the program area. The requirement for subject depth precludes the breadth that should be required for a broader assessment of the skills obtained during the full studies. Results from university self-assessments, alumni surveys and interviews are used only when data cannot be validly obtained via examination of theses. Similar findings were reported by Hallén & Nilsson (2012).

What factors are related to good results in terms of students' theses? Anderson (2013) states that students' socio-economic status is the variable which is strongest correlated to students' study success. She argues that as a result of this, with the new system universities are not inclined to primarily strive for developing good educations but to compete for the best students.

We agree with Hallén and Nilsson (2012) in their criticism of a single reader for each thesis. There is variation both in the time spent by each assessor on each report and on how reports are rated. Given the importance of an inadequate quality statement and how that would reflect the overall ranking of the program, especially for small programs, a system with a second reader would have been favourable with regard to the reliability of the verdict. There is a need to study the reliability of the whole assessment process, and the individual thesis assessments in particular.

The students' theses are complemented with the HEI's self-assessments. One of the authors have been active in producing self-assessments for two programs, and the experience from that work is that the unspoken aim was to get a favourable grading rather than to expose identified areas of improvement. The learning outcome was important, but not nearly as much a focal point of the self-assessment as the focus on warding off possible critique in areas seen as hard to assess from theses work, such as how the students are examined on oral presentation skills. The similarities of the self-assessments suggest that they often become idealized descriptions of the operations and poor instruments for learning outcomes. The supplementary interviews made by SNAHE/SHEA follow the same route. We conclude that the assessments focus on controlling that the learning goals of the higher education act are implemented. More emphasis should be put on the perspective of employees, of students and alumni and of the external society.

While it is certainly achievable to make improvements based on knowledge from rankings, e.g. that the researchers of the university publish remarkably often in prestigious journals, if indeed that is considered an issue for the educational quality. However, the step towards improvement might be longer from an third party ranking than it would have been from an in-house self-assessment.

## References

Adamson, L. (2013). *Kvalitetsutvärdering av högre utbildning – en kritisk granskning av det svenska systemet framtagen på uppdrag av SNS Utbildningskommission.* Stockholm: SNS. *In Swedish.*

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C. et al. (1980). *Toward reforms of program evaluation: Aims methods and institutional arrangements.* San Francisco: Jossey-Bass.

ENQA (2001). *Towards the European Higher Education Area. Communiqué of the meeting of European Ministers in charge of Higher Education in Prague on May 19<sup>th</sup>, 2001.*

Electronic reference:

[http://www.ehea.info/Uploads/Declarations/PRAGUE\\_COMMUNIQUE.pdf](http://www.ehea.info/Uploads/Declarations/PRAGUE_COMMUNIQUE.pdf). Access date: June 10, 2013.

ENQA (2012a). *Swedish National Agency for Higher Education: Review of ENQA Membership.* Report,

Formaterat: Svenska (Sverige)

[http://www.hsv.se/download/18.1c6d4396136bbbed2bd80002238/HSV\\_review-ENQA-Criteria-Report-April2012.pdf](http://www.hsv.se/download/18.1c6d4396136bbbed2bd80002238/HSV_review-ENQA-Criteria-Report-April2012.pdf) . access date April 4, 2013.

ENQA (2012b). *Letter to University Chancellor and Head of the Swedish National Agency for Higher Education (HSV)*. Hopbach, A. Electronic reference: [http://hsv.se/download/18.485f1ec213870b672a680004340/Letter\\_ENQA\\_HSV\\_170912.pdf](http://hsv.se/download/18.485f1ec213870b672a680004340/Letter_ENQA_HSV_170912.pdf). Access date: June 11, 2012.

Hallén, L. & Nilsson, A. (2012). Från tolkning av utbildningsmål till bedömning av måluppfyllelse – Observationer och reflexioner mot bakgrund av Högskoleverkets utvärdering av företagsekonomi 2011-2012. *Företagsekonomiska ämneskonferensen*. Umeå 17-18 October, 2012. *In Swedish*.

Karlsson, O. Andersson, I.M. and Lundin, A. (2002). *Metautvärdering av Högskoleverkets modell för kvalitetsbedömning av högre utbildning – Hur har lärosäten och bedömare uppfattat modellen*. Report 2002\_20 R, Stockholm: Högskoleverket. *In Swedish*.

Formaterat: Engelska (USA)

Kauko, J. (2012). The Power of Normative Coordination in the Bologna Process. *Journal of the European Higher Education Area*. Issue 4. <http://www.ehea-journal.eu/>. Access date June 10, 2013.

Shaw, I.F., Greene, J.C., and Mark, M.M. (2006). *The SAGE handbook of evaluation*. London: SAGE Publications, Ltd.

SHEA. 2013a. Universitetskanslersämbetet. Electronic reference: <http://www.uk-ambetet.se/nyheter/slutsatserefterutvarderingssystemetsforstaar.5.4149f55713bbd9175638000308.html>. (In Swedish). Access date: March 13, 2013.

SHEA. Universitetskanslersämbetet. Electronic reference: <http://www.uk-ambetet.se/nyheter/slutsatserefterutvarderingssystemetsforstaar.5.4149f55713bbd9175638000308.html>. (In Swedish). Access date: March 13, 2013.

Wahlén, S. (2004). Does national quality monitoring make a difference? *Quality in Higher Education*, 10(2), 139-147.

Weiss, C.H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.