

## **EVALUATION OF COURSE EVALUATIONS.**

**Prof. Peter Friis-Hansen, Assoc. Prof. Niels Houbak<sup>†</sup> and Prof. Peder Klit**

Department for Mechanical Engineering (MEK),  
Technical University of Denmark (DTU)  
DK-2800 Kgs. Lyngby, Denmark.

### **ABSTRACT**

A general evaluation of any education shall be based on the education's ability to meet defined goals and objectives. If such an evaluation is performed continuously it can be viewed as a relative quality measure. A main task when evaluating the whole education will be an evaluation of the individual courses. This may comprise of several parts: A lecturer evaluation/report, a student evaluation/report of the course and the lecturer(s), the number of students having passed the course, the grade average and the distribution of the grades.

At DTU (The Technical University of Denmark), students have for more than 10 years been evaluating the courses they attend. During the last 5 years, this evaluation has been completed electronically as an integral part of our CampusNet computing and course administration system.

At the MEK department we have a study board in charge of all our educational activities. One obligation is – two times a year – to go through in detail the evaluations for each of our 100+ courses and for each of the 50+ lecturers. A rather tedious job, where the outcome seldom justifies the resources spent.

The electronic version has opened up for further analysis of the evaluation data and extraction of important information; this will be the main focus of this paper. In the evaluation of courses, the students are given seven different questions and for each question they can select between 5 different answers. Each answer is given a certain weight, and by summing up the weights for the selected answers and making an average over all the students, each course obtains a utility value. A similar set of questions and answers exists for all course lecturers.

Sorting the utilities for all courses and also for all lecturers reveals some interesting cumulative curves. One obvious result of the analysis is the possibility to identify both good and bad courses and lecturers. Though, in most cases the analysis only quantifies what is already well known, it is important in particular to carefully search reasons for low values of the utility. There are a large number of reasons for poor performance and we have seen excellent lecturers obtaining poor evaluations due to circumstances out of their control. On the other hand, with the analysis at hand, it is much easier to focus on specific needs for improvement. The effect of 'chasing the bad performance' over the past few years will be shown.

<sup>†</sup> Corresponding author:

## INTRODUCTION

At DTU we have 2 separate engineering educations: a three and a half year Bachelor of Engineering (Diplomingeniør) and a five year Master of Science in Engineering (Civilingeniør). Both of these are divided into different disciplines (Electronics-, Chemistry-, Mechanical- Engineering, etc.). In total there are 7 disciplines in the Bachelor program and 12 entry disciplines and 16 exit disciplines in the Masters program. All in all DTU offers more than 900 courses for the whole curriculum. The reason for this huge number of courses is that we at DTU put emphasis on giving the students a high degree of freedom in their selection of courses for shaping 'their own' education.

This large amount of courses also leaves us with the problem of assuring the quality of the individual courses. This is not an easy task. Over the years different approaches have been taken: Approval of each course by either collegial boards or by the central administration, imposing drastic cuts in resources to certain areas, monitoring the number of students attending the courses, hearing of students in the collegial boards, accepting complaints from students, questioning attending students about the courses etc. etc.

During the last 10 to 15 years all students attending a course have been asked to fill out a questionnaire anonymously at the end of the course. At first, the intention was to obtain feedback on the course format from the students to the lecturer such that improvements could be made the next time the course was run.

Since the very beginning, the questionnaire has had a standard format of 3 sections: section A for evaluating the course, section B for evaluating each lecturer involved in the course, and section C for making verbal comments, criticism, and suggestions for improvements.

The academic year at DTU is divided into two terms (fall and spring term). Both terms consists of three sections: a 13 weeks section for lecturing, a section lasting approximately 10 working days with exams, and finally a 3 weeks section for whole day activity courses. Thus, the course evaluation is all in all done 4 times pr year (at the end of each 13 weeks period and at the end of each 3 weeks period).

In the first many years, the lecturer was responsible for handing out the questionnaire, collecting the answers, and evaluating the evaluations. Condensed results were then typically shown to and discussed with the students during the last lecture in the term. Only the lecturer had access to them afterwards.

At a certain point in time, the questionnaires were compiled in electronic form and today these have become an integral part of our CampusNet. This has resulted in electronic access to the data and that the data no longer are the lecturers 'property' only. The course evaluations are now made available also to the study board, and thus used more actively than previously. Also, the active use of the evaluations has made it possible to use these as a general quality control of the courses and for identifying problems to be solved in due time. Problem solving initiatives can be assisting the lecturer with ideas to reshaping a course, or if the facilities are inadequate to update these. It is important to emphasize that staff decisions never will be executed on the basis of course evaluations alone. In a long-term perspective, these may of course be used in forming a complete a picture.

## THE QUESTIONNAIRE

Section A of the questionnaire contains the 7 questions, see appendix A for details, to each of which the student must select among 4 or 5 possible answers. The questions are:

1. Are the prescribed course prerequisites adequate?
2. How is the course material?
3. Is the form of the course adequate?
4. A standard course has an average workload of 9 hours pr week; how much time did you spend?
5. How many lectures did you attend?
6. What is your general satisfaction with the course?
7. For courses taught in English; did this influence your outcome?

Among these questions the student satisfaction with the course (question 6), satisfaction with the form of the course (lectures, exercises, experiments, homework) (question 3), and course material (question 2) are considered the most important.

Similarly, the questions for section B, evaluation of each of the lecturers involved in the course are:

1. How did the (named) lecturer present the subject?
2. Is the lecturer inspiring?
3. How did you experience the dialog/cooperation with the lecturer?
4. How is the lecturer as supervisor?
5. For group works; did you receive criticism for handed in exercises during the course?

The aim again is to obtain relevant subjective information in an objective manner. In most cases there are 5 possible choices for the answer, graduating the student experience from very bad to very good. The first four questions have been given almost equal weight whereas the last has been given only half weight compared to the others.

In Section C of the evaluation the student can give a written personal evaluation of the course. Each student may express his/her impression by writing an immediate response to the sub-fields 'I appreciate', 'I criticize', and 'I suggest'.

When the students have submitted their answers to the questionnaire, these are in a spreadsheet format prepared by the central administration made available to the study boards for examination.

## WEIGHING QUESTIONS AND ANSWERS

The objective of establishing a formalised procedure for evaluating the course evaluation is to present a single number (we call it utility) that extracts the essence of the student's assessment of the quality of the course and a single number describing the quality of the lecturer. These two numbers allows the study board to rank the courses and lecturers faster and more objectively and thus focus on where to improve.

The benefit of the formalised procedure is, above all, to assure that the metric of priorities remain consistent throughout the examinations of the evaluations. This was not the case earlier when the examinations were performed manually. Secondly, in the establishment of the formalised procedure fruitful discussions of the importance, implications, and relevance of the

individual questions and answers in the questionnaire are initiated. Clearly, it becomes visible that not all questions are equally important and meaningful in conveying knowledge on the quality of the course and the lecturer; neither are the possible answers.

To establish the weighted average (that is the utility) of the course evaluation requires first a definition of the utility of the possible answers the student can select among for the individual questions. The principle here is first to assign a value of 10 to the most preferred answer and a value of 0 to the least preferred. The possible intermediate answers are assigned a value between 0 and 10. This phase gives a more clear understanding of possible ambiguities in the predefined answers.

Table 1 shows the assigned utilities to the individual possible answers. The detailed arguments for assigning the utilities will not be given here. Only is it noted that it is preferred that the student uses more time than prescribed compared to less.

As already mentioned not all questions are of equal importance in evaluating the quality (the utility) of the course or the lecturer. To establish the complete evaluation of the questionnaire the utilities from the answers are weighed according to an assessed importance of the individual questions. The weighing factor on the individual questions is between 0 and 1, and the sum of these weights is 1.

**Table 1: Utility values for the different possible answers in Section A**

1	Prerequisites	Too many	Suitable	Too few	None	
		4	10	0	10	
2	Teaching material	Very bad	Bad	Acceptable	Good	Very good
		0	2,5	5	7,5	10
3	Form	Less suited	Reasonable	Well suited		
		0	5	10		
4	Time consumption	Much less	Less	Normed	More	Much more
		0	4	10	8	3
5	Participation	0-25	26-50	51-75	76-100	
		0	2,5	7,5	10	
6	Satisfaction	Very small	Small	Acceptable	Much	Very much
		0	2	6	8	10
7	English	Very negative	Negative	Not influenced	Positive	Very positive
		0	5	10	10	10

In section A it has been decided that course satisfaction (q. 6), teaching form (q. 3) and course material (q. 2) are the questions that carries most information of the quality of the course. Each of these has therefore been given a weight in the interval from 0.2 to 0.3. The remaining questions have been given a weight between 0.05 and 0.1, except for teaching in English (q. 7) that has been given a weight on zero because presence of foreign students requires this.

**Table 2: Utility values for the answers to section B.**

1	Presentation	Very bad	Bad	Acceptable	Good	Very good
		0	2,5	5	7,5	10
2	Inspiration	No	Only little	To some extent	Yes	Very much
		0	2,5	5	7,5	10
3	Dialog, Cooperation	Very bad	Bad	Satisfying	Good	Very good
		0	2,5	5	7,5	10
4	Supervision	Very bad	Bad	Satisfying	Good	Very good
		0	2,5	5	7,5	10
5	Criticism of work	No	Only little	To some extent	Yes	Very much
		0	2,5	5	7,5	10

Table 2 gives the utilities of the evaluation of the lecturer as a linear scale for all choices of answers.

All questions except the last one have been given an almost equal weight. The last question has been given a weight on 0.1 to reflect that this is not relevant for all courses.

## THE UTILITY EVALUATION

The resulting utility of the course evaluation in questionnaire section A and B can now be established as the weighted average of the utility of the given answers. This is done simply by calculating the expected utility for each question and subsequently multiplying this value by the weight of the question and summing up over all questions in each questionnaire. The result is the Utility value for the particular course or lecturer.

No special account has been made for students refraining answering specific questions. Any reason for doing so is unknown and it is difficult to formulate a rational procedure for this. As already described, students specific critics may be given in writing in questionnaire C.

The Utility value will be in the range from 0 to 10 and in most cases between 5 and 8 to 9. It is important to notice, that the obtained Utility is an indicator and shall, like the impreciseness of the answers, not be considered as some kind of absolute truth! Very similar to the grades we give our students. Our use of the resulting ranking provided by the Utility measure has been to quickly identify excellent and bad courses/lecturers (high or low Utility). Previously, a detailed examination of all the answers to the questionnaires was performed when the study board of the department at semester end evaluated the lecturing. This tedious procedure has now been eliminated since we now primarily concentrate on the courses/ lecturers in either end of the ranking. In many cases acceptable reasons for extreme Utility values could be found: 1. Only a few students (< 5) actually filled out the questionnaire, 2. The lecturer had only a few lectures on the outskirts of a course, or 3. For some reason a lecturer was asked to take over a courser as part of a "rescue operation".

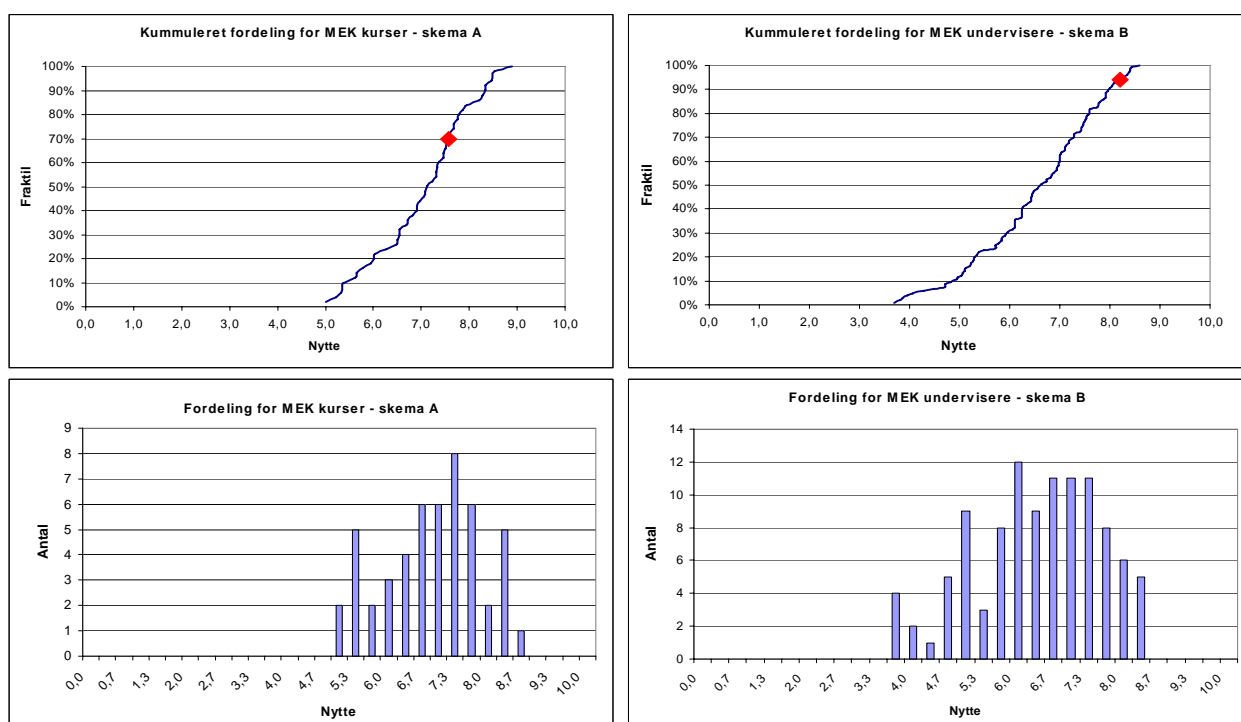
It is our experience, that the same lecturer participating in different courses may obtain very different Utility values for his/her effort. Also, from year to year a course and a lecturer may obtain rather different evaluations. In general, however, we observe a rather limited variation of the ranking of lecturers and courses. Therefore, if we find that the same course or lecturer gets

rather low ratings year after year, benefits to both lecturer and students can be obtained by digging into the problem and understand reasons for this.

## CUMULATIVE CURVES

Further processing of the Utility data provides interesting information. With a range of the Utility value between 0 and 10, this interval can be subdivided to plot histograms of the Utility value. Similarly, the cumulative curve is quite illustrative, where we plot the share of the courses that has a Utility value less than a particular value,. This is done as well for the lecturers.

Figure 1 shows examples of such curves for both MEK courses and MEK lecturers. A specific course or lecturer is plotted into the cumulative graphs thus marking the relative performance of the specific course/lecturer. Each lecturer will get these curves for his/her own course(s).



**Figure 1: Top left: Cumulative distribution for course evaluation. Top right: Cumulative distribution for lecturer evaluation. Bottom figures show the corresponding histograms. The dots represent the ranked location of a course and the corresponding lecturer at MEK.**

With nearly 100 courses offered by the department, these curves are very good indicators, both absolutely and relatively from year to year, on the general performance of the lecturers and the students attitudes towards the courses in general. Naturally, it is desirable to have a very steep curve located as long as possible to the right on the graph; in reality, this need not be obtainable. The shape of the curves may also differ from semester to semester, and the evaluation of the 3 weeks period with whole day activities is very different from the ordinary 13 weeks period with lectures.

## **RESPONSE TO LECTURERS**

The department study board sends out the cumulative curves described above to all involved lecturers. The board also uses this information to study in detail the evaluations of both courses and lecturers in either end of the cumulative curves. In cases, where the board finds it appropriate to recognize an outstanding (good or bad) performance a formal letter is submitted to the person(s) in question.

A good performance is appraised whereas a bad performance is notified with the question: "What can be done to improve the situation?" It is our experience in most cases, that the person in question actually is doing something about the problem. Possibly, this situation can initiate a dialogue on how to improve the teaching or maybe a discussion on how can this course be given in the future for the better of both students and lecturer. Maybe the lecturer needs a pedagogical course himself (or a CDIO-training workshop). Poor performance can only be solved by an increased focus on the problems and since dialogue is the tool for this, the relative course/lecturer evaluation is a good starting point for these discussions.

It is to be noted, that the communications on poor performance is not supposed to contain any threats like: "You will be fired if you don't ...". On the other hand it is clear that a lecturer not being responsive at all to reasonable and well founded criticism from both several students and the study board may have a shorter career. It is though the experience that all of our lecturers want to do a good job!

## **REACTIONS AND CONSEQUENCES**

Initially, the automated evaluations of the course evaluations were primarily made to improve the insight of the situation by the study board. The critical letters the study board send out to colleagues with a (long term) teaching problem were most unwelcome; but they served their purpose. Today we observe that there are fewer serious complaints on specific staff members.

On a more long term basis, it has turned out that sending out the cumulative curves to all lecturers allowing them to see their own ranking compared to all the other colleagues has been a good idea. For most competitive people being below than the average is not nice even though this fact is revealed only to them selves. (Someone will evidently be below the average, by definition!) Similarly, some of the lecturers have realized that it is possible by a more focused effort to climb up the list.

Figures 2 to 6 show the cumulative curves for our courses during the terms of 2002, 03 and 04. It is neither the same course nor the same lecturer, which have been indicated on the different plots. When comparing the cumulative curves for courses they have become slightly steeper and they have moved slightly to the right (a higher Utility value in average). The top of the curves seems to be the pivot of the movement. Though marginal, this effect is most probably caused by a higher awareness among the staff, of that someone cares about the quality of the teaching; the way the courses are run makes a difference to someone (besides the students!). The cumulative curve for the lecturers has not moved much but a tail of very low Utility evaluations has appeared. We believe that the movement of the curves is an un-intended side-effect of making public the cumulative curves both for the courses and for each staff member indicating their personal performance. We have observed that courses with several changing lectures often results in most lectures will be ranked very low. One cause for this is that the student does not see clearly the course responsible person. We have taken follow-up initiatives on these matters, and we look forward to measure the effect.

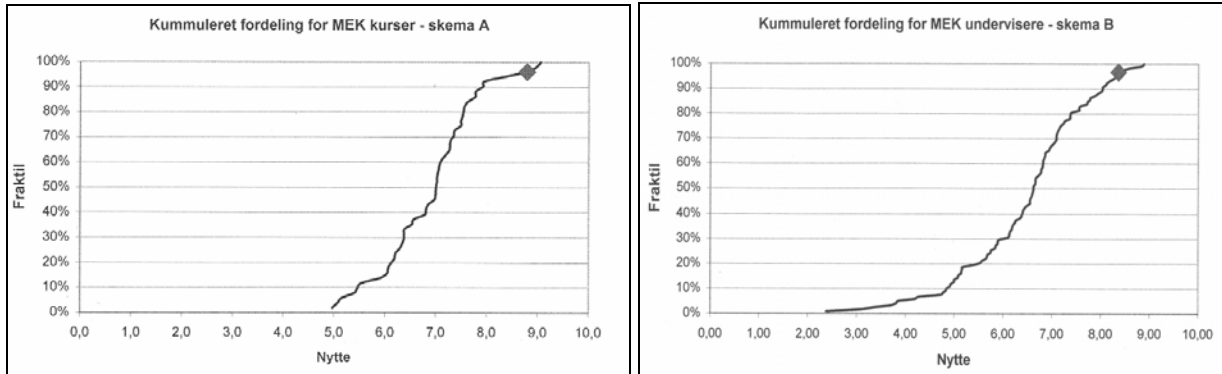


Figure 2: The cumulative curves for MEK courses and lecturers, Fall 2002.

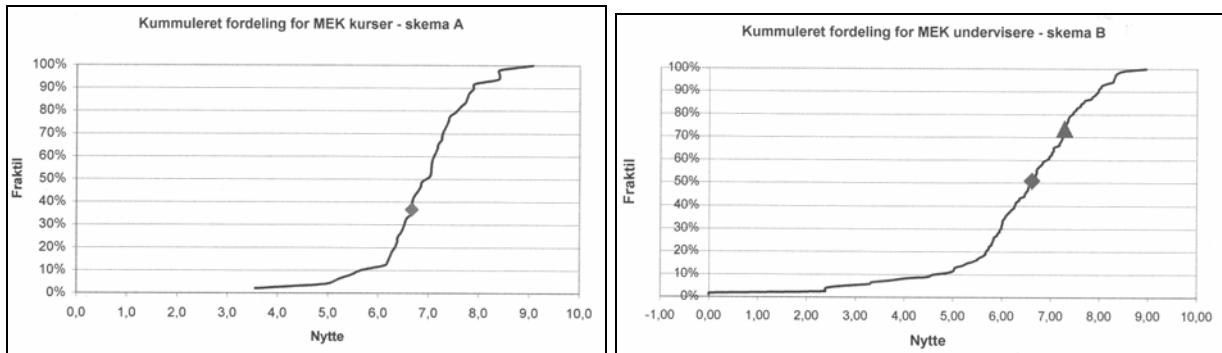


Figure 3: The cumulative curves for MEK courses and lecturers, Spring 2003.

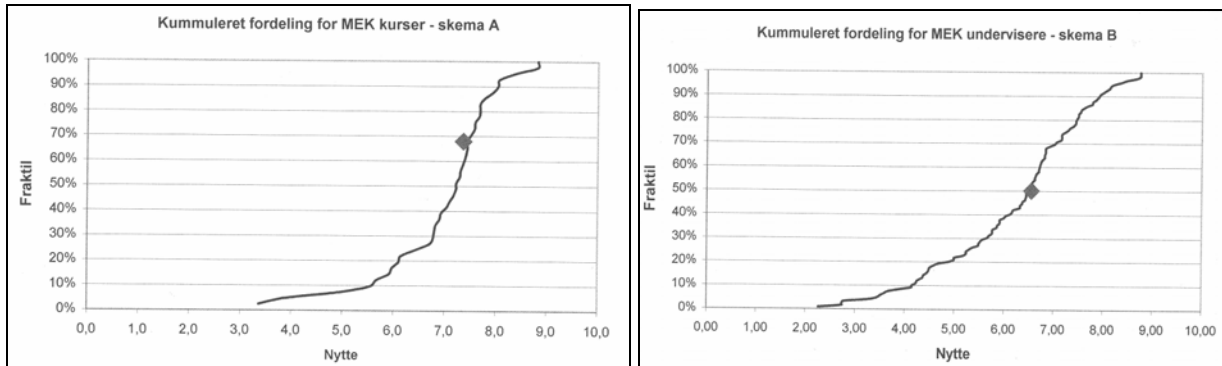


Figure 4: The cumulative curves for MEK courses and lecturers, Fall 2003.



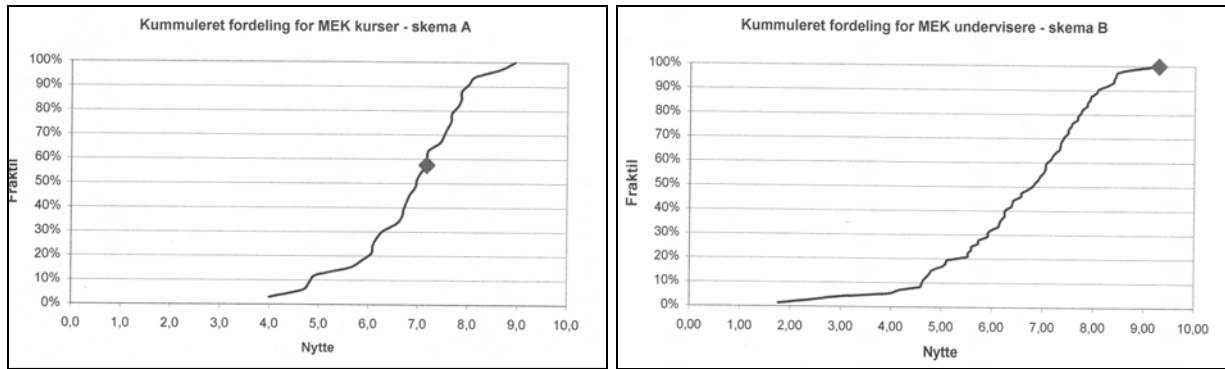


Figure 5: The cumulative curves for MEK courses and lecturers, Spring 2004.

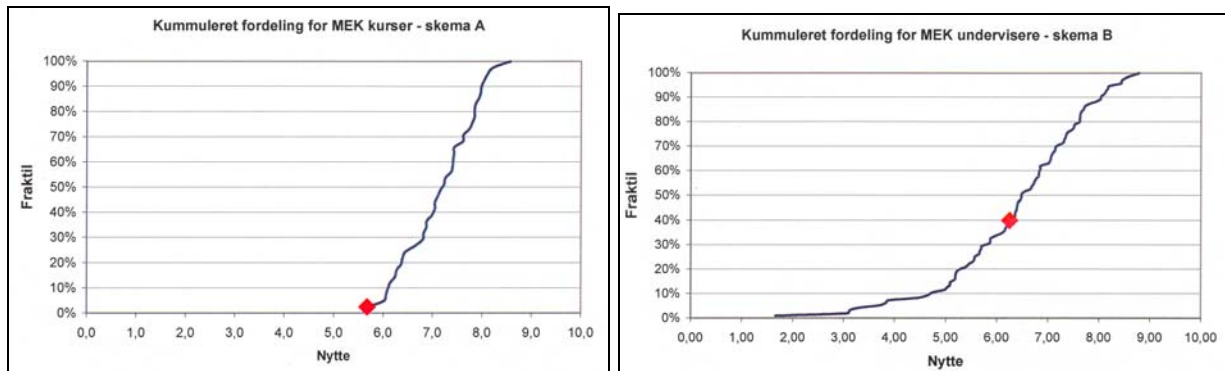


Figure 6: The cumulative curves for MEK courses and lecturers, Fall 2004.

## CONCLUSION

The total performance of a university department is determined by a number of parameters, as for example the number of publications, the number of ph.d.-students, patents, external founding, and the number of students attending the department courses etc. The department management needs to follow all parameters attentively to keep up the department performance. The developed course and teaching staff evaluation system is an important tool for the department head to follow the development in the teaching quality.

It has often been discussed if it should be possible for the lecturer to reformulate the questions in the questionnaire or to add more questions so that it would fit better to the actual course. So far, the study board has been reluctant to open up for this because it is the use of the same questionnaire for all courses and all staff that ensures that individual students has a chance to be objective in the sense that they use the same “quality scale”. However, the study board is very open for suggestions on general modifications of the questions and possible answers in the questionnaires such that these may shed more light on quality of the teaching activities.

Looking at the history of the course evaluations there are some lessons that can be learned:

1. It is always worthwhile to evaluate of what is being done as long as those evaluations do not take too much time.
2. Sometimes evaluations can be used for something differently or one finds out that they actually contain more information than was previously recognized.

3. Through the anonymous comparison with the other lecturers at the department it is possible for the lecturer to judge his own performance relative to the other staff members.
4. All lecturers notice that their evaluation result has been processed. It is not only the poor performers that are noticed.
5. The department study board has, by giving values to the weighing factors, indicated what is important to focus on when trying to improve the teaching quality.
6. The defined "Utility-system" generates a "dynamic normal", i.e. conscientious staff will try to improve the teaching quality to obtain a good position on the cumulative curves.
7. The lecturers can from term to term track changes in the evaluation.
8. Tests have shown that radical changes to the weights are needed to form a noticeable change in the ranking.

The evaluation of the courses has over the years been an important tool for the study board in improving the courses in general. Most feed-back from the students have been positive and constructive and this is therefore taken into account in the process of updating courses. Lately, it has turned out, that we can use these evaluations also for sharpening the competitions between lecturers and we believe that this also will have further impact on the quality of our teaching.