ANALYSING COURSE EVALUATIONS AND EXAM GRADES AND THE RELATIONSHIPS BETWEEN THEM

Bjarne Kjær Ersbøll
Department of Informatics and Mathematical Modelling, Richard Petersens Plads, B321
Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

Keywords: Course Evaluation Questionnaire, Exam Grades, Multivariate Analysis, Factor Analysis, Stepwise Regression.

Abstract: Course evaluation data from courses at higher education is often given by students. Commonly the evaluations are given as questionnaires with discrete answers on a Likert scale. At the Technical University of Denmark this is done on a constant basis. However, the data is not used optimally. The standard way of displaying these data is as a histogram or frequency table of each question separately. The paper discusses various ways of enhancing the amount of information, which can be extracted. We consider factor analyses for grouping of the questions and regression analyses in order to relate questionnaire data to student outcome in the form of exam grades.

1 INTRODUCTION

Courses in higher education are commonly evaluated by the participating students sometime during the course or at the end of a course. Typically such evaluations are performed by means of a questionnaire with questions related to the course curriculum, the learning outcomes, the teacher(s), and the organisational aspects of the course.

Many studies have been performed on such evaluations. Some have been on analyses and interpretation of relationships in the questionnaire itself. (Cohen, 1981) considers the analysis of data from 67 multisection courses (same course given by several instructors) in 40 studies. Defining a large number of factors derived from the data, Cohen found an association between overall ratings of instructor ratings and student achievement. He also found large correlations between "skill" (of instructor) and student achievement and "Structure" (instructors ability to structure course) and student achievement. (Feldman, 1989) refined and extended the synthesis of Cohen's data. A main very important conclusion is that students ratings of teachers is correlated with student achievement. (Abrami et al., 1997) performed confirmatory factor analysis using including oblique rotation. They also emphasise the analysis of multisection courses. Based on a meta analysis of 17 studies they extract what they call "common dimensions of teaching". Here 4 factors are identified. These have been interpreted as: factor 1: "instructor viewed in an instructional role", factor 2: "instructor viewed as a person", factor 3: "instructor viewed as a regulator". For factor 4 no interpretation is offered. In a recent study (Sadoski and Sanders, 2007) analysed student course evaluations in medical school for 5 different courses for students after 1 and 3 years of study. These were analysed for "common themes" using principal component analysis on each course. They found the following consistent items which loaded most heavily together with an "overall quality" item, namely: "course organisation", "clearly communicated goals and objectives", and "instructional staff responsiveness". Another such study is (Althouse et al., 1998) who consider the relationship between ratings of basic science courses and the "overall evaluation" of these courses. Items most often found to be significant were described as: "engaged in active learning", "quality of lectures", and "administrative aspects of course". (Guest et al., 1999) conducted a study where survey responses are compared with the actual examination performance of the student. The study found that student perceptions of "value of curriculum" were poorly associated with external measures of performance like the grade. On the other hand, "perceived lecture organization", "stimulation to read", and "interest in

Ersbøll B. (2010).
ANALYSING COURSE EVALUATIONS AND EXAM GRADES AND THE RELATIONSHIPS BETWEEN THEM.
In *Proceedings of the 2nd International Conference on Computer Supported Education*, pages 313-318
DOI: 10.5220/0002801703130318

subject" was found to affect "perceived overall learning" and "perceived value of lectures". Finally, an interesting validation study giving a word of caution in interpretation of student evaluations is (Billings-Gagliardi et al., 2004). They describe how students think/interpret the course evaluation questions. This was assessed by performing think-aloud interviews with 24 students. Not all terms used in a questionnaire turn out to be uniquely understood or interpreted in the same way by the students. For instance the term "independent learning" was understood differently by different students. Also, ratings for certain questions were "adjusted" (raised or lowered) by the students when thinking of other aspects like "effort of teacher".

The overall conclusion from these and many other studies show a good association between student course evaluations and student outcome.

The present study considers student course evaluations at the Technical University of Denmark (DTU). Here an online course evaluation is usually performed in the week preceding the final week of the course. Effectively this means most courses are rated after 12 out of 13 possible lectures and/or exercises. The courses will typically be 5 or 10 ECTS points, corresponding to a nominal workload of either 120 or 240 hours. The questionnaires are used for courses at all levels from introductory to advanced. Normally, the results from the questionnaires are simply summarised as simple histograms and percentages for each question. No attempt is made to assess the multivariate structure of the data. Hereby valuable information is lost, because possible correlations between answers is completely disregarded.

In this paper we will report findings related to a course in Multivariate Statistics. Two different types of analyses and interpretations of these are given. The first considers grouping of the different questions by factor analysis and investigates the consistency between two different years. The second relates the achieved grades to the questionnaire and analyses which questions might be most informative of student outcome.

2 MATERIALS AND METHODS

2.1 Data

The current evaluation form at DTU which is implemented and maintained by a university spin-off: Arcanic A/S, www.arcanic.dk, has been in use since the fall of 2007. It is reasonably standardised in that most of the questions are generic, but a number of questions can be removed and/or further questions can be included in the evaluation by the course responsible before the students are asked to perform the rating.

2.1.1 Questionnaire

The questionnaire has three parts: Form A contains questions related to the course (one form per course); Form B contains questions related to the teacher (one form per teacher/instructor) Forms A and B give answer possibilities on a 5 point Likert scale.

Finally, form C gives the possibility of qualitative feed-back to the three cases: "What went well?", "What did not go so well?", "Suggestions for changes". An example of a questionnaire can be seen in the appendix.

2.1.2 Exam Grades

By means of an anonymous code it is possible to relate the grade obtained by the student to the answers in the questionnaire. The present grading system which complies with the European ECTS grading scale has also been in use since the fall of 2007. The scale is numerical and designed to make it possible to make grade averages. It takes the values: "12", "10", "7", "4", "2", "0", "-3" corresponding to "A", "B", "C", "D", "E". The last two grades: "0" and "-3" both correspond to "fail", "Fx" and "F" respectively. A more detailed explanation of the different grades is given in Tabel 4 in the appendix.

Questionnaire data from a course in Multivariate Statistics at DTU for the autumn semesters in 2007 and 2008 are available. The course is generally taken by students in the last half of their studies. For the autumn semester of 2007 the grades obtained by the students at the exam are also available for the analyses.

2.2 Types of Analyses

2.2.1 Factor Analysis

Factor analyses were performed using principal factor analysis on the correlation matrix of the questionnaire data. The number of factors retained was determined by the commonly used rule of having a variance greater than one. In order to assure an easier interpretation this was followed by a so-called varimax rotation. The varimax rotation tends to simplify the structure and ease interpretation of the resulting factors. A good general reference is (Hair et al., 2006).

3 RESULTS

3.1 Factor Analyses

The factor analyses are performed for the autumn semesters of 2007 and 2008. We choose only to analyse form A, which corresponds to the part of the questionnaire concerned with the course itself.

3.1.1 Autumn Semester 2007

For the factor analysis for the autumn semester of 2007 29 form A questionnaires were available for factor analysis. The analysis resulted in 3 factors having the required minimum variance of one. The resulting three varimax-rotated factors are shown below with the variables associated with each factor in order of importance judged by factor weight (given in parenthesis).

Factor 1 (of 3).
- A1.8 (0.87): In general, I think this is a good course
- A1.5 (0.86): I think the teacher/s create/s good continuity between the different teaching activities
- A1.1 (0.86): I think I am learning a lot on this course
- A1.2 (0.83): I think the teaching method encourages my active participation
- A1.3 (0.79): I think the teaching material is good

This is interpreted as: “overall quality of the course”

Factor 2 (of 3).
- A1.7 (0.93): I think the course description’s prerequisites are
- A1.4 (0.60): I think that throughout the course, the teacher/s have clearly communicated to me where I stand academically

This is interpreted as “academic standing”.

Factor 3 (of 3).
- A1.6 (0.85): 5 points is equivalent to 9 hrs./week. I think my performance during the course is
- A2.1 (0.67): I think my English skills are sufficient to benefit fully from this course

This is interpreted as “student involvement”.

3.1.2 Autumn Semester 2008

For the factor analysis for the autumn semester of 2008 31 form A questionnaires were available for factor analysis. The analysis resulted in 2 factors having the required minimum variance of one. The two factors are shown below. Again the variables in each factor are listed in order of importance judged by factor weight (given in parenthesis).

Factor 1 (of 2):
- A1.8 (0.91): In general, I think this is a good course
- A1.5 (0.85): I think the teacher/s create/s good continuity between the different teaching activities
- A1.2 (0.84): I think the teaching method encourages my active participation
- A1.1 (0.79): I think I am learning a lot on this course
- A1.4 (0.78): I think that throughout the course, the teacher/s have clearly communicated to me where I stand academically
- A1.3 (0.54): I think the teaching material is good

This is interpreted as: “overall quality of the course”

Factor 2 (of 2):
- A1.6 (0.77): 5 points is equivalent to 9 hrs./week. I think my performance during the course is
- A2.1 (0.60): I think my English skills are sufficient to benefit fully from this course
- A1.7 (-0.57): I think the course description’s prerequisites are

This is interpreted as “student involvement and prerequisites”.

3.2 Grades

The grades are available for the 2007 autumn semester only. By means of the anonymous code it is possible to link the grades to the course evaluation questionnaires. An initial illustrative overview of the grades is displayed in Figure 1. Here the distribution of the 48 grades is given depending on whether the student answered the course evaluation or not. The immediately obvious difference is the large proportion of students who neither evaluated (=“Silent”) nor took the exam (=“EM”). The students who passed (grade 2 or above) and who answered the course evaluation seem to have higher grades on average, but this is not significant with the present data.

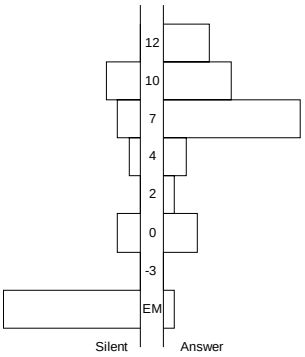

Figure 1: Distribution of the 48 grades for the autumn semester 2007: did not answer course evaluation questionnaire (="Silent", left) vs. answered course evaluation questionnaire (="Answered", right).

3.3 Stepwise Regression of Course Evaluations on Grades

3.3.1 Form A

For the course related questions a stepwise regression of exam grades vs. student ratings of the course evaluations for form A gave the following results:

- A1.2 I think the teaching method encourages my active participation. (positive weight, significant)
- A1.3 I think the teaching material is good (negative weight, however not significant)

It is encouraging to note that the significant item in the questionnaire is related to "active participation". This corresponds well with current understanding of good teaching and learning. The non-significant item on "teaching material" relates to the fact that the students find the lecture notes a bit difficult and too concise. This is revealed by looking at the open questions in form C.

3.3.2 Form A and B

If student ratings of the course evaluations for both form A and B are included in the stepwise regression, it turns out there is one significant question as an outcome:

- B2.2 I think the teacher is good at helping me understand the academic content. (positive weight, significant)

This result is also very encouraging, since it is well known that a good teacher really makes all the difference for student outcome.

4 DISCUSSION

4.1 Factor Analyses

The factor analyses from the two different years show expected similarities. Regardless of the fact that there are 3 selected factors selected in the 2007 data and only two factors selected in the 2008 data, we note an interesting grouping of the questions.

For 2007 factor 1 might be interpreted as "quality of course". Similarly for factor 2 "understanding own standing", and finally for factor 3 "students own effort".

For 2008 factor 1 similarly can be interpreted as "quality of course". It is noted that factor one for both years contain the same questions in nearly the same order except for A1.4 on academic standing which was not included in 2007. This is probably due to the different number of factors retained.

Compared to 2007 factor 2 becomes less comparable because of the different number of factors. We can however, reasonably interpret factor 2 as "student involvement".

In all cases we note a high degree of consistency with the literature.

4.2 Grades

In Figure 1 an interesting difference between students who answer or do not answer the questionnaire is seen. From the data analysed one may conjecture that students who do not respond to the questionnaire also tend to avoid the exam. This important finding was previously unknown, simply because of the obstacle in merging the grade database with the questionnaire database.

4.3 Grades and Questionnaire Data

The result of the stepwise regression of grades and both form A and B confines with what may be expected.

For the course evaluation against grade, question A1.2: "I think the teaching method encourages my active participation" was significant. It is well known that active learning generally is preferable. A runner-up is A1.3 "I think the teaching material is good". This comes in with a negative weight, but is not significant. However, it can be related to the fact that the students tend to find the lecture notes a bit too concise.

Finally, relating both forms A and B to the achieved grades resulted in a significant item from form B related to the teacher.

5 CONCLUSIONS

The work considered concerns the analysis of questionnaire data from student-course evaluations from two time-periods, and also the connection between course evaluations and student outcome in the form of exam grades. We have demonstrated consistency of such evaluation data over time. Furthermore, we have shown relationships between student outcome in the form of exam grades the questionnaire data.

ACKNOWLEDGEMENTS

Christian Westrup Jensen from the administration of DTU and the team at Arcanics A/S are gratefully acknowledged for helping providing the data. Comments and suggestions from the reviewers are also kindly acknowledged.

REFERENCES

Abrami, P.C., d'Apollonia, S, Rosenfield, S., 1997. The dimensionality of student ratings of instruction: what we know and what we do not. In: *Perry, R. P., Smart, J. C., editors: Effective teaching in higher education: research and practice.* New York: Agathon Press.

Althouse, L. A., Stritter, F. T., Strong, D. E., Mattern, W. D., 1998. Course evaluations by students: the relationship of instructional characteristics to overall course quality. *Paper presented at: The Annual Meeting of the American Educational Research Association; 1998 April 13-17; San Diego, CA.*

Billings-Gagliardi, S., Barrett, S. V., Mazor, K. M., 2004. Interpreting course evaluation results: insights from thinkaloud interviews with medical students. *Med. Educ.*; 38: 1061-70.

Cohen, P. A., 1981. Student ratings of instruction and student achievment. *Review of Educational Research*; 51(3): 281-309.

Feldman, K. A., 1989. The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, Vol. 30, No. 6.

Guest, A. R., Roubidoux, M. A., Blance, C. E., Fitzgerald, J. T., Bowerman, R. A., 1999. Limitations of student evaluations of curriculum. *Acad Radiol*.; 6:229-35.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., 2006. *Multivariate Data Analysis.* 6th ed. Prentice Hall.

Sadoski, M., Sanders, C. W., 2007. Student course evaluations: Common themes across courses and years. *Med Educ Online* [serial online] 2007;12:2. Available from http://www.med-ed-online.org

APPENDIX

Table 1: Example of questions in evaluation form A.

	Question	Answer possibilities
1.1	I think I am learning a lot in this course	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
1.2	I think the teaching method encourages my active participation	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
1.3	I think the teaching material is good	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
1.4	I think that throughout the course, the teacher/s have clearly communicated to me where I stand academically	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
1.5	I think the teacher/s create/s good continuity between the different teaching activities	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
1.6	5 points is equivalent to 9 hrs./week. I think my performance during the course is	Much less=5, 4, 3, 2, 1=Much more
1.7	I think the course description's prerequisites are	Too low=5, 4, 3, 2, 1=Too high
1.8	In general, I think this is a good course	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
2.1	I think my English skills are sufficient to benefit fully from this course	Strongly agree=5, 4, 3, 2, 1=Strongly disagree

Table 2: Example of questions in evaluation form B.

	Question	Answer possibilities
1.1	I think that the teaching gives me a good grasp of the academic content of the course	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
1.2	I think the teacher is good at communicating the subject	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
1.3	I think the teacher motivates us to actively follow the class	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
2.1	I think that I generally understand what I am to do in our practical assignments	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
2.2	I think the teacher is good at helping me understand the academic content	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
2.3	I think the teacher gives me useful feedback on my work	Strongly agree=5, 4, 3, 2, 1=Strongly disagree
3.1	I think the teacher's communication skills in English are good	Strongly agree=5, 4, 3, 2, 1=Strongly disagree

Table 3: Example of questions in evaluation form C.

	Question
1.1	What went well – and why?
1.2	What did not go so well – and why?
1.3	Which changes would you suggest for the next time the course is offered?

Table 4: Definition of grades in the Danish 7-step grading scale.

Grade 7-step scale	Description	ECTS scale
12	For an excellent performance displaying a high level of command of all aspects of the relevant material, with no or only a few minor weaknesses.	A
10	For a very good performance displaying a high level of command of most aspects of the relevant material, with only minor weaknesses.	B
7	For a good performance displaying good command of the relevant material but also some weaknesses.	C
4	For a fair performance displaying some command of the relevant material but also some major weaknesses.	D
2	For a performance meeting only the minimum requirements for acceptance.	E
0	For a performance which does not meet the minimum requirements for acceptance.	Fx
-3	For a performance which is unacceptable in all respects.	F