DOES AN ASSOCIATION BETWEEN STUDENT EVALUATIONS OF RELATED CDIO COURSES EXIST?

Bjarne Kjær Ersbøll, Tamara Sliusarenko, Line Harder Clemmensen

Department of Informatics and Mathematical Modelling, The Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

ABSTRACT

This paper discusses possible uses of student course evaluations on a pair of courses developed to comply with the CDIO concept. It is seen that both similarities and differences in the evaluations can be found. These can in part be used to assess if the CDIO concept has been implemented as it was intended and possible adjustments can be suggested.

The data consist of 33 observations with full information on 8 general course evaluation questions on each of the two courses. The data has been collected over three years (2008, 2009, and 2010). This makes it necessary to consider methods which are able to handle possible differences between years.

We illustrate different ways to analyse the associations between the two courses by utilizing such data. Inferences about the mean differences between the courses are performed using analysis of variance techniques. In this context they may be considered as generalisations of the paired t-test. The generalisation to an analysis of variance makes it possible to handle differences between years. Inferences about the correlation structure of the data are performed using so-called canonical correlation analyses. A possible difference between years of the evaluations makes it necessary to consider adjusting the data for the year effect.

We find that one course generally is evaluated as more satisfactory than the other on five of the questions. Also we find a very strong effect of year, indicating the need to remove the year effect before proceeding with the canonical correlation analysis. The canonical correlation analysis is only significant at a 10% level of significance for these data and resulting associations must therefore be interpreted with caution. The interpretation results in a combination of evaluation questions for one course which correlate well with another combination of evaluation questions for the other course.

KEYWORDS

Associations, student evaluations, related CDIO courses, paired t-test, analysis of variance, canonical correlations

INTRODUCTION

Universities all over the world discuss ways to improve the quality of the teaching and learning processes. As stressed by the CDIO homepage: [1] “The CDIO initiative is an innovative educational framework for producing the next generation of engineers. The framework provides students with an education stressing engineering fundamentals set in the context of Conceiving — Designing — Implementing — Operating real-world systems and products. Throughout the world, CDIO initiative collaborators have adopted CDIO as the framework of their curricular planning and outcome-based assessment."

Teacher evaluations and overall course quality evaluations are widely used in higher education. Students submit their feedback about the teacher and the course anonymously during the course or at the end of the course. Results are usually employed either directly by the teacher(s) or indirectly by management to improve courses for future students and to improve instructor effectiveness. Many researchers have stated that student rating is the most valid and practical source of data on teaching and course effectiveness [2] (McKeachie, 1997). Therefore, research on student evaluations is critical to make improvements in course construction and teaching methods.

Many authors have considered different ways of analysing, interpreting and utilizing evaluation data. Some are on relationships in the questionnaire itself. In [3], Cohen considers the analysis of data from 67 multisection courses and found an association between overall instructor ratings and student achievement. This study was later refined by Feldman [4]. In [5], Althouse et al. consider the relationship between ratings of basic science courses and the "overall evaluation" of the courses. Guest et al. [6] compare the survey responses with the actual examination performance of the student. In [7], Ersbøll considers grouping of the different questions by factor analysis and examining consistency between different years. He also investigates which questions are most related to the grade achieved by the student. Finally, Sliusarenko and Ersbøll [8] consider the relationships between general questions related to the course and general questions related to the instructor.

The CDIO concept was formally introduced in the professional bachelor degree education at DTU in 2008. This paper analyses routine course evaluations performed by students in the computer science related professional bachelor degree educations at DTU. Specifically, a pair of related courses is considered, namely: "Introductory Programming" (course no. 02312) and "Development Methods for IT-Systems" (course no. 02313). Both courses include lectures and lab work. However, the first is slightly more oriented towards traditional lectures while the second is slightly more oriented towards project work in groups. Together the pair of courses cover the CDIO concept and it is the intention that the courses be taken in parallel during the same semester. It is therefore of interest to analyse the evaluations of the courses together utilizing that (some of) the students have evaluated both courses. Here we consider ways of detecting differences between the course evaluations and other possible associations between the evaluations of the two courses.

DATA

The student course evaluation questionnaire used at DTU is standardised across the university. The actual evaluation is performed online through CampusNet (the university intranet) a week before the final week of the course. The questionnaire is split in three parts: form A which considers course related questions, form B which considers teacher related questions, and finally form C which is a free format qualitative feed-back form considering three cases: "What went well?", "What did not go so well?", and "Suggestions for changes". In the present analysis we will only consider form A.

To conduct the analysis we collected the evaluation results from two bachelor level courses from the department of Informatics and Mathematical Modelling at DTU: "Introductory Programming" and "Development methods for IT-Systems" which correspond to 10 ECTS points and 5 ECTS points respectively. Together the pair of courses is considered to fulfil the CDIO concept. Furthermore, it is recommended that students follow the "Introductory Programming" course at the same time as the "Development methods for IT-Systems" course. Therefore some students will have filled in evaluation forms for both courses. The course characteristics (taken from [9] and [10]) are given in Table 1.

Table 1
Course characteristics

Course name	**Introductory Programming (02312)**	**Development methods for IT-Systems (02313)**
Points(ECTS)	10	5
Course type	bachelor	bachelor
Scope and form	Lecture, exercises and a programming project	Lectures and lab work
Duration of Course	13 weeks + 3 weeks	13 weeks
Type of assessment	Oral examination and reports	Oral examination and reports
General course objectives	The goal of the course is to make the student able to use the basic concepts and techniques in an imperative- and object oriented programming language. The course will use a programming language that is used in industries (JAVA). The main purpose of the course is to make the student able to design, implement and test smaller programs	The purpose of the course is to train an engineering approach to developing software systems in small project groups
Learning objectives:	• Understand the different number representations • Use loops and branching. • Understand classes and the anatomy of objects. • Use simple UML notations for classes and associations. • Use arrays. • Use inheritance. • Use simple I/O operations without corresponding exception handling. • Explain simple test methods and use these in simple examples. • Work in groups to design a smaller software system based on a problem description in a predefined task and implement the most important parts of this design. • Use simple time and activity planning of a project progress.	• Plan, control and carry out a small software project in project groups • Describe important roles in a project group • Carry out requirement specifications • Design og programs, processes and modules • Develop smaller programs based on a particular design • Use configuration management • Develop program documentation • Plan, carry out and document user and Unit test • Evaluate own and others work based on review techniques • Prepare a report which documents the product

The actual evaluation questions in form A used for both courses are presented in Table 2. The student has the possibility to rate each question between 1 and 5, where 1 means that the student strongly disagrees with the underlying statement and 5 means that the student strongly agrees with the statement. For question A.1.6 a 1 corresponds to much more and a 5 to much less, while for A.1.7 a 1 corresponds to too high and a 5 to too low. In a sense for these two questions a 3 corresponds to satisfactory and anything else (higher or lower) corresponds to less satisfactory. Therefore we will also consider a transformation of the two variables corresponding to A.1.6 and A.1.7 namely: 5-abs(2x-6). Then a value of 5 means "satisfactory" and anything less means "less satisfactory".

Table 2
Questions in course evaluation Form A.

	Question
A.1.1	I think I am learning a lot in this course (1=disagree strongly, 5=agree strongly)
A.1.2	I think the teaching method encourages my active participation (1=disagree strongly, 5=agree strongly)
A.1.3	I think the teaching material is good (1=disagree strongly, 5=agree strongly)
A.1.4	I think that throughout the course, the teacher has clearly communicated to me where I stand academically (1=disagree strongly, 5=agree strongly)
A.1.5	I think the teacher creates good continuity between the different teaching activities (1=disagree strongly, 5=agree strongly)
A.1.6	5 points is equivalent to 9 hours per week. I think my performance during the course is (1=much more, 5=much less)
A.1.7	I think the course description's prerequisites are (1=too high, 5=too low)
A.1.8	In general, I think this is a good course (1=disagree strongly, 5=agree strongly)

Filling in the questionnaire is not mandatory at DTU. However, students are urged to respond by means of a "nag screen". Unfortunately the response rate is often still low, sometimes as low as 10-15%. For the case considered it ranges between 13% (for 2010) and 42% (for 2008 and 2009). In order to ensure sufficient data for the analyses three years of course evaluations (2008, 2009, and 2010) are combined and analysed together.

Using an anonymous key it is possible to pair the evaluations for every single student between the two courses. This results in a total of 33 observations for 2008 (14), 2009 (14) and 2010 (5) combined. An overview of the numbers of students who could answer, who did answer, answering percentages and number of students, who evaluated both courses, is seen in Table 3.

Table 3
Basic statistics on numbers of students evaluating the courses: "Introductory Programming" (02312) and "Development methods for IT-Systems" (02313)

Year	**2008**		**2009**		**2010**	
Course #	02312	02313	02312	02313	02312	02313
# enrolled	75	94	50	74	62	90
# evaluated	23	35	21	23	8	15
Answer %	30.7%	37.2%	42.0%	31.1%	12.9%	16.7%
# evaluated both	14		14		5	

METHODS

In all statistical tests p-values of 5% or less are considered significant while p-values between 5% and 10% are considered indicative. Parametric analysis more or less implying use of the normal distribution is employed although the data are clearly not normal. However, with a suitably large number of observations this is considered a reasonable approximation.

Descriptive statistics of the sample are given as counts, sample means and standard deviations.

Possible differences in mean level of the answers with respect to “year” and “course” are tested using three-way analyses of variance with “student” as the third factor. The model contains the fixed effects: “year”, “course”, and the interaction “year*course”. The student effect is modelled as random and nested below “year” as “student(year)”. The effects: “course” and “year*course” are tested against residual error, while “year” is tested against “student(year)”. Significance of “year” means the level of the answers differs over the years regardless of course. Likewise significance of “course” means the level of the answers differs for the two courses regardless of year. Finally, a significant interaction between year and course indicates that the mean differs more (or less) than linearly for the combination of year and course. This is sometimes called super- and supra-additivity, respectively.

It is noted that if all observations had been from the course pair from the same year then the above analysis could have been performed using a pairwise t-test.

A so-called canonical correlation analysis [11], [12] is performed in order to assess the degree of association between the questionnaires in the two courses. Canonical correlation analysis is a statistical technique which can be considered as an extension of ordinary linear regression analysis. Ordinary linear regression analysis relates one response variable “y” to a linear combination of a number of “x” variables. Canonical correlation analysis extends this by allowing a number of “y” variables. It works by relating a linear combination of the “y” variables to a linear combination of the “x” variables such that the correlation between them is maximal. This is the first canonical correlation. It is possible to extend this scheme to several canonical correlations. The corresponding pairs of sets of weights can be interpreted as the importance of the different questions in the questionnaire.

RESULTS

As a general overview one may consider simple averages of the scores for the different questions. These indicate that the courses are generally considered satisfactory by the students.

As seen from Table 3 the response rate is low for each course and year. This of course means the response rate for the combination also runs the risk of being low.

From the simple descriptive statistics presented in Table 4 it is evident that there is an overall difference in student rating between the two courses. The “Introductory Programming” course, get lower rates than “Development methods for IT-Systems” course.

Ten three-way analyses of variance, one for each of the questions, with “year”, “course” and “student” as factors were performed. The first null-hypothesis was that there was no year-effect (difference between years). The second null-hypothesis was that there was no course-effect (difference between courses). The third null-hypothesis was that there was no effect of the interaction between year and course. The results of the tests are shown in table 5.

A canonical correlation analysis was performed for the 33 observations at hand. Our interest was to investigate the (correlation) structure of the data between the two courses. Since a difference in mean between years has been detected above, we subtracted off the mean for each year from each answer. (The canonical correlation analysis automatically adjusts for the course means.)

Table 4
Sample descriptive statistics, all years

	Introductory Programming (02312)			**Development methods for IT-Systems (02313)**		
Variable	**N**	**Mean**	**Std Dev**	**N**	**Mean**	**Std Dev**
A.1.1	33	3.48	1.30	33	4.12	0.96
A.1.2	33	3.24	1.39	33	3.88	1.02
A.1.3	33	3.39	1.12	33	3.36	1.11
A.1.4	33	2.45	1.50	33	3.39	1.37
A.1.5	32	2.78	1.36	33	3.85	1.09
A.1.6	33	2.97	0.88	33	3.21	0.82
A.1.6T	33	3.85	1.33	33	3.97	1.33
A.1.7	33	2.94	0.50	33	2.97	0.47
A.1.7T	33	4.51	0.87	33	4.70	0.88
A.1.8	33	3.27	1.18	33	3.88	1.05

Table 5
P-values of effects in three-way analyses of variance (ANOVA) for each question.
Significances at the 5% level shown in boldface.

Question	**Year**	**Course**	**Year*Course**
A.1.1 (Learning a lot)	**0.0003**	**0.0037**	0.2644
A.1.2 (Activation)	**<0.0001**	**0.0015**	**0.0354**
A.1.3 (Material)	**<0.0001**	0.4341	0.0756
A.1.4 (Feedback)	**<0.0001**	**0.0003**	0.1187
A.1.5 (Continuity)	**0.0002**	**0.0003**	0.2122
A.1.6 (Workload)	0.4992	0.5992	**0.0178**
A.1.6T (Workload, transformed)	0.6253	0.4228	0.6233
A.1.7 (Prerequisites)	0.4613	0.8391	0.6353
A.1.7T (Prerequisites, transformed)	0.3451	0.5265	0.2521
A.1.8 (General)	**<0.0001**	**0.0027**	0.4415

A canonical correlation analysis is concerned with the analysis of all variables simultaneously and requires that observations do not contain missing values. In the data set at hand there is a single missing value for question A.1.5 for the “Introductory programming” course. The missing value was substituted by the mean in order to include the observation in the analysis. Alternatively more elaborate methods like imputation might be used.

In order to ease interpretation the transformed values of questions A.1.6 and A.1.7 were used.

The canonical correlation analysis results in one pair of components being indicative (significant at the 10% level but not at the 5% level), with a p-value of 0.0825. This means no firm conclusions should be drawn. However, we may still try to interpret the results. The results are shown in Table 6. The interpretation is performed by considering the largest weights first. These are seen for questions A.1.1 (positive for both courses), A.1.3 (contrast between courses), A.1.4 (negative for both courses), A.1.6T (contrast between courses), and A.1.8 (contrast between courses).

Table 6
Standardized canonical coefficients for the two courses.

	Introductory programming	Development methods
A.1.1 (Learning a lot)	0.58	0.65
A.1.2 (Activation)	-0.23	-0.15
A.1.3 (Material)	-0.44	0.95
A.1.4 (Feedback)	-0.34	-0.63
A.1.5 (Continuity)	-0.23	0.21
A.1.6T (Workload, transformed)	-0.56	0.43
A.1.7T (Prereq., transformed)	-0.12	-0.05
A.1.8 (General)	0.74	-0.38

In order to ease interpretation a reduced set of variables is produced by removing variables with small coefficients one at a time and re-running the analysis each time. With only five variables left all coefficients are greater than 0.5. The result is shown in table 7. We note the same variables as before except variable A.1.8 are represented.

Table 7
Standardized canonical coefficients for reduced variable set for the two courses.

	Introductory programming	Development methods
A.1.1 (Learning a lot)	0.92	
A.1.2 (Activation)		
A.1.3 (Material)		0.94
A.1.4 (Feedback)	-0.71	-0.68
A.1.5 (Continuity)		
A.1.6t (Workload, transformed)	-0.67	
A.1.7t (Prereq., transformed)		
A.1.8 (General)		

DISCUSSION

The overall result of the three-sided analysis of variance (Table 5) was that all questions, except questions A1.6 and A1.7 and their transformed versions, showed that “year” was extremely significant. For “course” the overall result is that all questions except question A.1.3 and questions A1.6 and A1.7 and their transformed versions, were very significant. The relevant mean values for each course may be judged from table 4. It is noted that for questions which are found to be significant, the difference is towards more satisfaction with the course “Development methods for IT-Systems”. For the interaction term only questions A.1.2 and A.1.6 showed significance. The significance is nowhere near that of the main-effects “year” and “course” and for simplicity the interaction effect will therefore not be considered further here.

The standardized coefficients from the full set and the reduced set canonical correlation analyses are shown in Tables 6 and 7 respectively. We can interpret the simplified result as follows: In “Introductory programming” the student thinks she is learning a lot, she does not think she is receiving very much feedback, and she has an unsatisfactory workload. In “Development methods” the same student tends to think the material is good, and she is not receiving very much feedback in this course either. The complete result in Table 6 basically gives the same interpretation, but in more detail.

Thirty-three (33) observations with evaluation information from the same student in both courses in the same year were available for analysis. The low number is unfortunate, but a consequence of the rather low response rates seen in many professional bachelor courses at DTU. Therefore three years of evaluations were analysed together. Furthermore, small courses also run the risk of having very few student evaluations even at high response rates. In this case combining the evaluations from several years may be the only possibility.

Generally it is a benefit that the observations are paired, since this makes it possible to eliminate much of the variation between students. Therefore, inferences about differences between courses can be expected to be more valid, than had different students evaluated the two courses.

CONCLUSION

This paper discusses possible use of student course evaluations on a pair of courses developed to comply with the CDIO concept. It is seen that both similarities and differences in the evaluations of the courses can be found. The similarities and differences can in part be used to assess if the CDIO concept has been implemented as it was intended and possible adjustments can be suggested.

The use of paired data more easily and validly highlights differences between courses with respect to mean value. Here methods like the paired t-test and more generally analysis of variance may be employed. Also it gives the unique possibility of finding associations between course evaluations by means of techniques like canonical correlation analysis.

In the case analysed an obvious and consistent shift in mean between the courses was seen using analysis of variance. Also shifts in mean from year to year were shown to occur. Before further analysis the data was adjusted for this. Finally, insight into the structure between courses was achieved by means of canonical correlation analysis. Both pieces of information are expected to help in further developing the courses and the interaction between them.

REFERENCES

[1] http://www.cdio.org/

[2] McKeachie, W.J., “Student Ratings: Their Validity of Use”, American Psychologist, Vol. 52, 1997, pp 1218-1225.

[3] Cohen, P. A., “Student ratings of instruction and student achievement”, Review of Educational Research, Vol. 51, No. 3, 1981, pp 281-309.

[4] Feldman, K. A., 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. 1989.

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

[6] Guest, A. R., Roubidoux, M. A., Blance, C. E., Fitzgerald, J. T., Bowerman, R. A., 1999. “Limitations of student evaluations of curriculum”, Acta Radiol, Vol. 6, 1999, pp 229-35.

[7] Ersbøll, B. K., ”Analysing course evaluations and exam grades”, Proceedings of: International Conference on Computer Supported Education, CSEDU 2010, Valencia, Spain, 2010, 6pp.

[8] Sliusarenko, T. Ersbøll, B. K., ”Canonical correlation analysis of course and teacher evaluation”, Proceedings of: International Conference on Computer Supported Education, CSEDU 2010, Valencia, Spain, 2010, 4pp.

[9] http://www.kurser.dtu.dk/2009-2010/02312.aspx?menulanguage=en-GB

[10] http://www.kurser.dtu.dk/2009-2010/02313.aspx?menulanguage=en-GB

[11] Hotelling, H., “The most predictable criterion”, Journal of Educational Psychology, Vol. 26, 1935, pp 139-142.

[12] Hotelling, H., “Relation between two sets of variates”, Biometrika, Vol. 28, No. 3-4 1936, pp 321-377.

Biographical Information

Bjarne K. Ersbøll is a full professor at the Department of Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. He is head of the section named “DTU Data Analysis”. His and his sections main focus is on applying statistical analysis and pattern recognition techniques cross-disciplinarily to application areas all over the rest of the university.

Tamara Sliusarenko is a PhD student at the Department of Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark, and she is a member of the section “DTU Data Analysis”. Her PhD project is concerned with analysis and data mining of student course evaluations especially at the Techcnical University of Denmark.

Line H. Clemmensen is an assistant professor at the Department of Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. She is also a member of “DTU Data Analysis”. Her main interests are especially within analysis of large datasets using and developing so-called sparse methods.

Corresponding author

Professor Bjarne K. Ersbøll
Department of Mathematical Modelling
Richard Petersens Plads, B321
Technical University of Denmark
DK-2800 Kgs. Lyngby
Denmark
+45 45253413
be@imm.dtu.dk