

Mining Assessment and Teaching Evaluation Data of Regular and Advanced Stream Students

IRENA KOPRINSKA
University of Sydney, Australia

This paper investigates the effect of the stream (regular or advanced) on the student evaluation of teaching and the course marks. It presents a case study in a third year Computer Science course at an Australian university. The results show that there were no significant differences between the two groups in their perception of teaching and learning. However, the two groups significantly differed in their assessment results in all assessment components and also in the most important predictors of the final mark.

1. INTRODUCTION

Some Australian universities offer the same Units of Study (UoS) in two versions: regular and advanced. In the advanced stream the material is more demanding and the students have higher previous academic results. The aim of this study is to investigate how the differences between the two streams (previous academic performance and UoS demands) affect the student evaluation of teaching and the UoS assessment results.

Our study was conducted in a third year Computer Science course. The two streams had shared lectures (for the common material), separate labs/tutorials (for the more challenging material for the advanced stream) and common and different assessment components (more challenging and open ended for the advanced stream). The number of all enrolled students was 48, 18 advanced and 30 regular.

2. EFFECT OF THE STREAM ON STUDENT EVALUATION OF TEACHING

At the end of the semester students completed a survey similar to [USE] which measures their perception of teaching on a five-point Likert scale. Student evaluation of teaching has been an active area of research [Richardson 2005]. There is convincing evidence that it is reliable, valid and stable over time, and also relatively independent of the class size and expected grades. We extend previous research by studying if there are differences in the perception of the advanced and regular students doing the same UoS. Some possible differences are in the workload, clarity of explanation and adequacy of previous UoSs. E.g., it may be difficult to find the right balance between the common and different content and assessment, resulting in much higher workload for the advanced stream. Also, the teaching methods and resources may not be equally efficient for both groups.

Using the Mann-Whitney test we found that overall there were no statistically significant differences between the two streams in their perception of teaching, with both groups being positive. It was particularly encouraging to see that the workload imposed by the more challenging assignments for the advanced stream was not perceived as too high. The findings can be used by teaching management to justify the existence of two streams within a single UoS as an alternative to two independent streams.

3. EFFECT OF THE STREAM ON ASSESSMENT RESULTS

We investigate two questions: 1) if there are significant differences between the marks of the two groups in the individual assessment components and the final mark and 2) which assessment components are the best predictors of the final mark for each groups. Previous work on mining assessment data includes McNamara [2004] whose goal was competency mapping and Pechenizkiy et al. [2008] who investigated if test questions and feedback on them matched the individual student needs.

We used the student marks on all three assessment components: homeworks, assignments and exam. The weighting of these components was 9%, 26% and 65%. The

weekly homeworks involved small problem solving tasks. The two assignments were project-based and involved writing a computer program to solve a real-world problem and a report to discuss the results. The end-of-semester exam involved problem solving tasks organised in 9 questions. The difference in assessment between the two streams was 25% in total: 0% in the homeworks, 10% in the assignments and 15% in the exam.

As the weighting of the assessment components was different, all marks were converted into percentages of the individual components. Table I shows the mean and standard deviations for the assessment components. The advanced group performed better than the regular group in all three assessment components and the differences were statistically significant. While the differences in the homework and exam marks were consistent (19%), there was a huge difference in the assignments mark (33%). A closer examination showed that the assignments engaged better the advanced students who implemented additional features in their programs and analysed the results much better. Another interesting observation is that there was significant difference between the two groups in the common part of the exam but not in the different part. This can be interpreted as a well chosen different part, not disadvantaging any of the two groups.

Table I. Assessment data (** - Mann-Whitney stat. significant differences, $p < 0.05$)

Mark	Homeworks **	Assignments **	Exam common **	Exam different	Exam total **	Final **
Reg.	58.9±27.8	58.2±24.9	65.6±16.2	68.9±28.3	66.4±17.4	63.6±18.3
Adv.	77.3±26.6	91.2±11.4	81.5±13.0	71.9±22.4	79.3±13.7	82.2±10.4

To examine the predictive power of the individual assessment components on the final mark we used correlation analysis. We found that for the regular stream all assessment components correlated highly with the final mark ($r=0.806-0.951$). For the advanced stream the exam marks correlated highly ($r=0.702-0.931$), the homework mark correlated moderately ($r=0.604$) and the assignment mark correlated weakly ($r=0.229$). All correlations except the last one were statistically significant at $p < 0.05$.

To investigate the predictive power of the individual exam questions on the final mark we used multiple regression. The exam is the most comprehensive assessment component; it tests all topics while the other assessment components are focused on selected topics. Initially all 9 questions were entered as independent variables and at each step the most non-significant independent variable was removed ($p > 0.05$) until only significant predictors were left. The results showed that for the regular stream all exam questions except two relatively easy questions were good predictors of the final mark. For the advanced stream the most important predictors were two non-trivial questions.

In summary, we showed how correlation and regression analysis can be used to gain a better understanding of the assessment results. The advanced and regular streams differed in the assessment results and the predictors of the final marks. As expected the advanced students achieved higher marks in all assessment components and this difference was highest in the project-based assignments. The results can be used to improve future offerings of the course and provide timely feedback to students during the semester, e.g. by predicting the final mark based on a progress mark during the semester.

REFERENCES

- USE. <http://www.itl.usyd.edu.au/use/questionnaires.htm>
- MCNAMARA, R.A. (2004). Evaluating assessment with competency mapping. In *Proc. ACE'2004*.
- RICHARDSON, J.T.E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment and Evaluation in Higher Education* 30(4):387-415.
- PECHENIZKIY, M., CALDES, T., VASILYEVA, E., AND DE BRA, P. (2008). Mining the student assessment data: lessons drawn from a small scale case study. In *Proc. EDM'2008*, 187-191.