® EMPOWERING STUDENTS TO BE THE JUDGES OF THEIR OWN PERFORMANCE THROUGH PEER ASSESSMENT

BILGIN Ayse, FRASER Sharon
Macquarie University
Australia

Alternative assessment methods are becoming increasingly common in higher education with the aim of increasing the potential learning of students. This paper presents an application of an alternative assessment method: peer assessment of oral presentations for postgraduate students within a statistics department. Even though the assessment of peers is a valuable workplace skill, such an activity is rarely an integrated part of university education. With a new emphasis in universities on the development of generic skills, it is appropriate to explore means of assessment that are valued in the marketplace. The aim of the peer assessment intervention reported here was to increase the critical thinking skills of students and enable them to develop their ability as independent decision makers. The advantages and disadvantages of the intervention and peer assessment in general are discussed and suggestions are made for possible improvements.

INTRODUCTION

At university, students are given the opportunity to learn new concepts and skills to enable them to become capable members of society. A crucial aspect of learning and teaching therefore, is assessing (measuring) what and how much students have learned. In general, assessment is the determination of the amount, level, value or worth of "something", which in education, translates to a judgement about the extent to which students have learned the "knowledge and skills" that are taught during a course or degree. Assessment is therefore an important part of curriculum development, in fact *"from our students' point of view, assessment always defines the actual curriculum."* (Ramsden, 2003, p.182).

The assessment process should be authentic and aligned with learning outcomes. Peer assessment of oral presentation enables learners to use what they have learned to judge the quality of a presentation by critically thinking about both the subject content and the ways in which it has been presented. In this paper, we present an investigation of the outcomes of the peer assessment of oral presentations within a postgraduate class of thirty five students studying decision support systems at the Department of Statistics, Macquarie University.

WAYS OF ASSESSING STUDENT LEARNING

In education, assessment has both a formative and summative purpose, and both goals can be achieved through a range of traditional techniques including multiple choice tests, essays and short or long answer exams, as well as by more alternative means including portfolio assessment and self, peer and group-assessment. To ensure validity and reliability within the assessment process, a three staged process should be followed: setting the criteria for assessing the work; selecting the evidence that enable judgement against these criteria; and making judgement about the extent to which these criteria have been met (Biggs, 2003). Traditionally, the lecturers have responsibility for each of these three stages, with students taking no responsibility for the design of the experience or for judging the quality of the learning. Why shouldn't students be more involved in the assessment process?

PEER ASSESSMENT

Peer assessment is a strategy that Paul Ramsden (2003) lists in his fourteen rules for better assessment in higher education. It involves students assessing "the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status." (Topping, 1998, p.250) against a set of clear criteria and standards. Students learn from working together and assessing their own and each other's work (Bloxham & West, 2004). This process makes assessment more transparent for students as they are exposed to varying standards and approaches to the task and enables lecturers to help students make decisions about their learning (Gibbs, 1981). Peer assessment can range from students marking multiple choice questions test

against a template, or providing review and editorial comments throughout an extensive written task, through to judging the quality of an oral presentation or peer engagement in group work. By involving students in the process of assessing their own learning and/or giving them the responsibility to judge the outcomes of their learning, they can practice much-needed skills for the work place. To be effective, however, it is essential that guidance is available with regard to the criteria and standards that will inform the students' decisions (Bostock, 2001).

ADVANTAGES AND DISADVANTAGES OF PEER AND SELF ASSESSMENTS
        In recent years, peer assessment has been the focus of much investigation, both in terms of its process (Van Den Berg, Admiraal & Pilot, 2006) and its outcomes (Cassidy, 2006). The literature documents both advantages and disadvantages of the process. The benefits (Cassidy, 2006) include increased student responsibility and autonomy; student insight into assessment procedures, harder working students, opportunities for increased levels of feedback and a context that encourages deep learning (Brown, Rust & Gibbs, 1994). Students develop the ability to work cooperatively, reflect upon and be critical of others' work and receive critical appraisals of their own work (Murray-Harvey, Silins & Orrell, 2003). Their active involvement in the assessment experience "promotes the acquisition of life-long skills" (Ballantyne, Hughes & Mylonas, 2002, p.428), enabling them to learn on their own after graduation (Boud & Falchikov, 2006; Stefani, 1994). It is important therefore to introduce students to the process early on their studies and build upon this experience to better prepare them for the workplace (Joliffe, 1997).
        Concerns about the validity and reliability dominate the literature relating to the disadvantages of the peer assessment process. Cassidy (2006) suggests that reliability is an issue as students don't have the skills or understandings to make judgements about the work of others or to provide useful feedback. Overall the literature indicates there is a perception in the academic community that there is a lack of objectivity in the peer assessment process (Brindley & Scoffield, 1998). Brown and Knight (1994, as cited in Langdon et al., 2005) documented the biases in peer assessment, including gender differences and over-marking of friends. Sluijsmans, Brand-Gruwel and Van Merrienboer (2002) advocate however, that peer assessment is a skill that can and should be 'learned' through focussed training, and without this as an embedded component of the curriculum, the process may fail.

A CASE STUDY
        This case study of peer assessment of oral presentations is set in a large metropolitan university (Macquarie University) in Sydney, Australia, in the Division (Faculty) of Economic and Financial Studies (EFS). The Division has around 7000 full-time equivalent students, around half of which come from other countries, predominantly from China, Hong Kong, Indonesia and Thailand.
        The unit which forms the context of this study (MIST812 – STAT820 Decision Support Systems) is an elective postgraduate unit available to students studying a range of coursework masters degrees including Master of Applied Statistics. Since it is an elective subject, it might be assumed that students who are enrolled have a special interest in the topics and may be predisposed to using a deep approach to their learning. While this may be true for some students, other students might select the unit to fill gaps in their course, both in terms of timetabling and in terms of university requirements.
        When run in the second semester of 2006, 35 students, 22 males and 13 females, were enrolled with six of them studying towards a Master of Applied Statistics degree. More than half of the students come from countries in Asia (51%), and only 31% were local students.
        MIST812 – STAT820 has two parts - data mining (DM) and geographical information systems (GIS) – taught by two separate lecturers. A peer-assessment of oral presentations was trialled as part of a summative assessment of GIS oral presentations of assignment two. These presentations were one part of a group (maximum 4 students) assignment for which they later produced a written report. The assignment (Appendix 1) was worth ten percent, shared equally by a written report and the oral presentation. In past offerings, students did not participate in the assessment process (the GIS presentations were assessed by both lecturers and the average mark was given to students). One of the unit lecturers (Bilgin) felt that students were missing out on

learning from their peers and wanted to share the responsibility of assessing the project presentation with the students through peer assessment. She felt that this would increase their attention during the presentations, assist them to learn from their peers and encourage a deep approach to their learning, as well as serve to improve their critical thinking skills through the process of judging their peers' presentations.

Students approach their learning in qualitatively different ways (Marton & Saljo, 1976a, 1976b), they may take a surface approach in which the focus is on memorising material in order to pass an assessment task; a strategic approach characterised by students searching for 'cues and clues' (Ramsden 1979) in order to maximise their grade; or a deep approach in which students are interested in the search for knowledge and understanding for its own sake.

A presentation marking guide was made available to students through WebCT one week before their presentation (Appendix 2). Each student received a paper copy of the marking guide for each group presentation. They used this to give marks for the presentation in three aspects: topic (content); presentation skills and methods. After each ten-minute presentation, five minutes of question time enabled students to clarify what wasn't clear during the presentation. Two lecturers also used the same marking guide to assess the presentations. Student marking papers were anonymous.

A final mark for each presentation was calculated as the average of students' total marks (50%) and average of two lecturers' total marks (50%). The overall mark awarded for each group presentation was a percentage derived from the sum of marks awarded to the three components of the presentation (using the breakdown: topic contents 45%; presentation skills 45% and methods used to create geographical regions in the maps 10%; see Appendix 2). Each student in a group received the same presentation mark.

The presentations were scheduled just before the semester break. Of the nine groups, only seven presentations took place at the first sitting. Two groups were granted extra time to prepare their presentations. Group seven presented their project during the semester break, therefore there were no student assessors and is excluded from the analysis. The overall average mark of students for the previous seven groups was used as the peer assessment mark for this group. Group two's presentation was done straight after the term break where only two-thirds of the class were present.

ANALYSIS

The peer marks for each group by topic and presentation skills show that there were only five outliers out of 252 observations (Figure 1). Four of these outliers were topic-related and only one presentation-related. The median marks for presentation skills were lower than the median topic marks for five groups (left panel). It is also clear that topic and presentation marks have positive linear association (right panel). When the presentation skills marks are predicted by topic marks, 35.2 percent of the variability of presentation skills marks can be explained ($R^2 = .352$).

A one-sample t-test for the difference between the average of the marks given by peers and those given by the lecturers for the topic of all group presentations showed that there was no significant difference ($p > 0.05$). However the difference between the average marks given by peers and lecturers for presentation skills was statistically significantly different from zero ($p < 0.003$) with students' marks being lower than lecturers' marks for all groups.
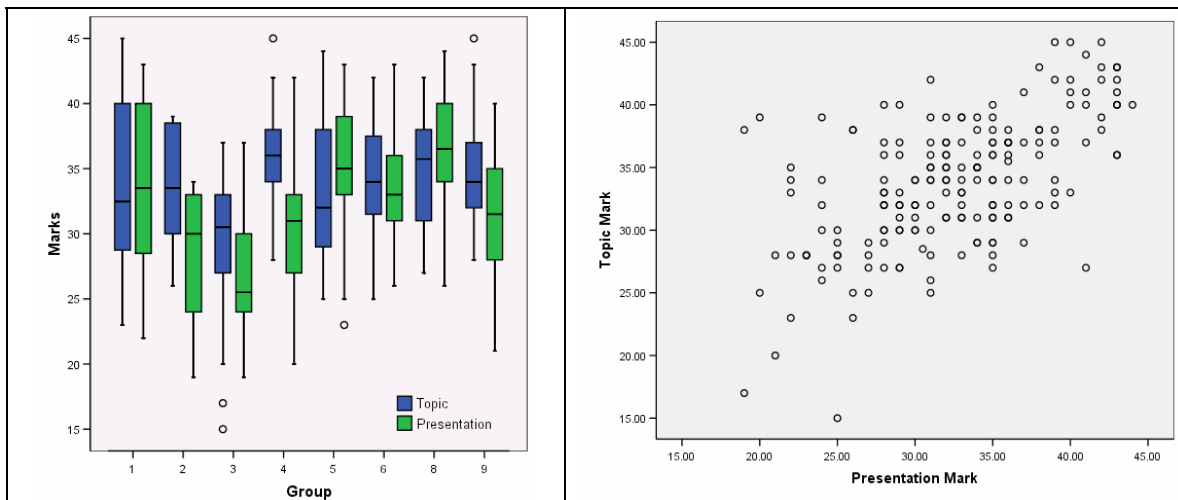
Figure 1: The distribution of marks for each group given by peers for topic and presentation skills

The average total mark out of 100 was 70.0 (SD = 1.1) given by peers and 72.6 (SD = 2.5) given by lecturers. The relationship between average total presentation marks given by lecturers and by peer markers (for each group) can be seen in Figure 2 (left panel). Students marked five groups lower than the lecturers and three groups higher. However, there was no significant difference between the lecturers' and students' average mark ($p > 0.05$).

As would be expected, group is a significant main effect for the total presentation mark ($F_{(8,242)}$ = 6.84, p = 0.000). However the type of the marker (peer or lecturer) is not significant ($F_{(1,242)}$ = 0.912, p = 0.34). The proportion of the variability accounted for by the group and marker type is 18.6% (R^2 = .186). The variation of the total presentation marks given by peer markers for each group is shown in Figure 2 (right panel).

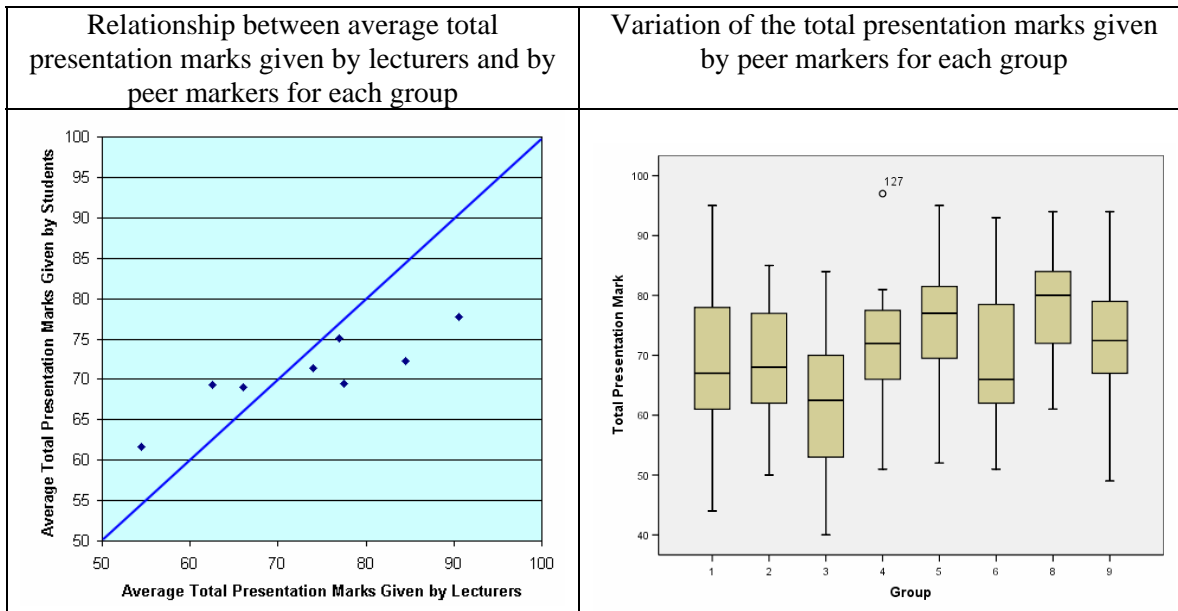| Relationship between average total presentation marks given by lecturers and by peer markers for each group	Variation of the total presentation marks given by peer markers for each group

Figure 2: The total presentation marks

The break down of the total marks for each criterion in the marking guide (Appendix 2) is presented in Figure 3. The only apparent point is that "methods used to create regions" represents the most variable category.
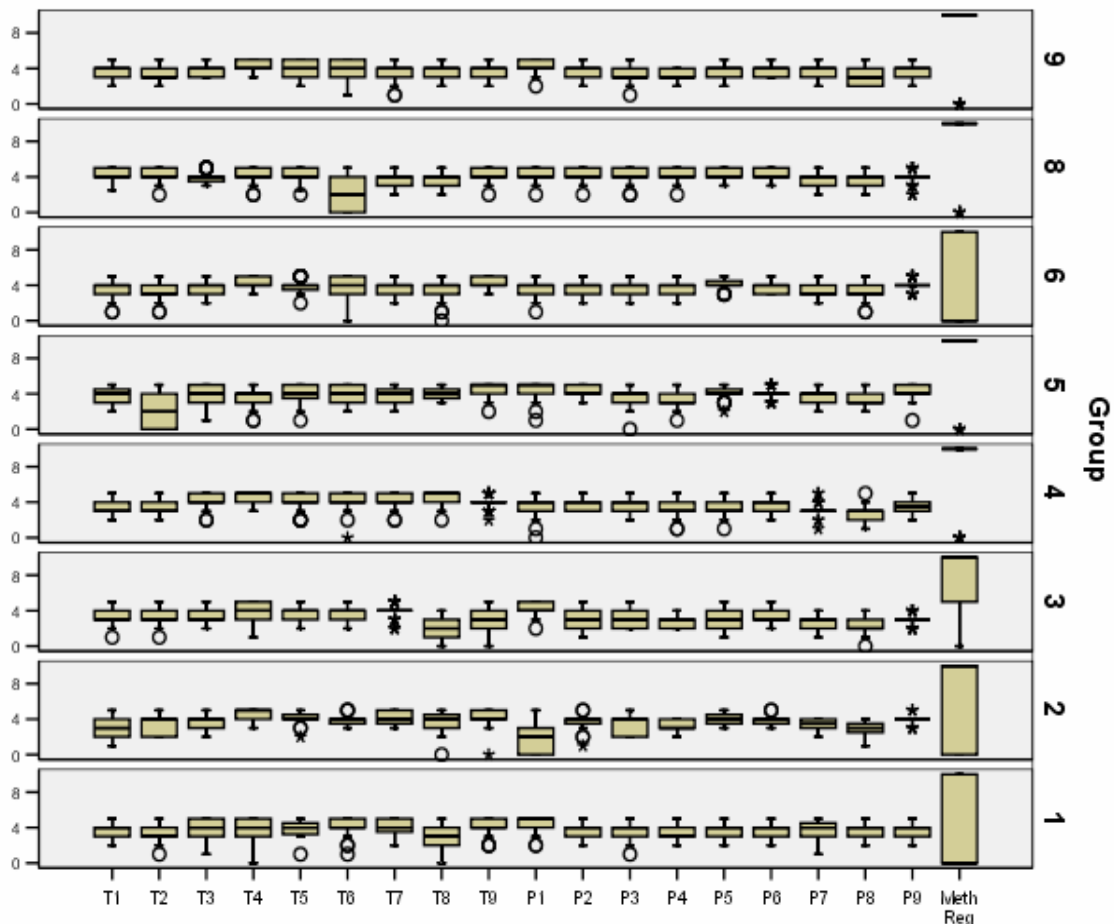
Figure 3: The category specific marks for each group

LIMITATIONS OF THE STUDY

A number of limitations of the study were identified that impacted on our ability to effectively analyse the data and draw conclusions about the assessment process or the factors contributing to, or limiting its effectiveness; these were:

(a) Marker details	Assessment outcomes were anonymous, and therefore information about the assessors (e.g. gender, international or local student, age, full-time or part-time student) was not collected.
(b) Marker consistency	It is not possible to know whether a marker was consistent (e.g. always marking lower or higher) because markers were anonymous. It is possible that there were biases (e.g. gender) and/or inconsistencies.
(c) Order of Presentations	The order of presentations was not recorded, and the literature indicates that this factor might have an effect on the marks given by students (Langan et al., 2005).
(d) Marking scheme	A marking scheme was *not provided* when students were *given the assignment*. This meant that for the most part, students prepared for the assessment task in the absence of criteria to guide their work.

(e) No Standards	The marking scheme showed the criteria but not the expected standards, limiting the information students could use to inform their judgements.
(f) Time	There wasn't enough time for students to write any comments on the marking paper which might have been valuable to presenting students.
(g) Verbal feedback	Students received both written and verbal feedback as recommended (Van Den Berg, Admiraal & Pilot, 2006), but the former was only given in *any detail* by the lecturer (students providing only box ticks), thereby limiting student engagement with the process.
(h) Summative assessment	The process was summative. Unknown factors may have influenced their marking, e.g. students may have been more lenient in giving higher marks to their peers (Farh, Cannella & Bedian, 1991).
(i) Norm-based grading	The understanding of norm-based grading by statistics students might have influenced the grading (e.g. lower marks given so that peers did not exceed their own marks).
(j) Poorly Trained Assessors	Reliability may have been undermined since there was no prior training for students with regard to the marking process.
(k) Evaluation Data	No data were collected from students about their perceptions of the process.

DISCUSSION AND CONCLUSION

Peer assessment was implemented into the postgraduate statistics unit with the aim of increasing the critical thinking skills of students and developing their ability as independent decision makers. We believe that simply by making the marking guide available to students before their presentation, they were more actively engaged with the assessment task. They had to think critically about what is expected and tailor their project around the criteria given to them. They also had to use the criteria to judge their peers' presentations and make decisions about the value. It could be argued that they had to use a deep approach (Marton, Hounsell & Entwhistle, 1984; Biggs 1987) to their learning so that they are capable of assessing their peers confidently. One lecturer observed that the students paid (more) attention to the quality and content of their peers' presentations (compared to previous years) as they were required to assess them. In previous years, students were still working on their presentations while waiting for their turn to come and not paying any attention to other presentations apart from listening to the questions at the end raised by the lecturers. In addition, the complaints regarding the grades for the presentations have dropped to zero which, might be due to the transparent nature of the peer assessment and the shared responsibility of the grading.

As the limitations outlined above suggest however, the peer assessment process in MIST812 – STAT820 could be improved. In particular, it is clear that students need to develop their skills in peer assessment, therefore the 'one off' experience of the assessment strategy, represented by our approach in statistics, is less than ideal. Involving students in the process of developing the criteria to be used in the peer assessment process has been found to enhance student learning (Sivan, 2000), as does an emphasis on the learning aspects associated with the task, rather than simply the allocation of marks (Cassidy, 2006). In the light of such recommendations, our next iteration of the peer assessment process will include:

(a) Student training

- ▪ Students will become familiar with the process of peer assessment by initially assessing their teacher using the marking guide;
- ▪ Their concerns will be discussed and related to the literature, highlighting the main areas for improvement;
- ▪ Students will be informed of the need for, and importance of peer assessment (e.g. necessary workplace skill);

(b) Improved marking sheet

- ▪ The 'training session' will enable the marking sheet to be improved;
- ▪ Coding to enable identification of the characteristics of students will be incorporated;

(c) Evaluation

- ▪ The students' experience of the process and its outcomes will be evaluated.

Our research in this area has also identified a gap in the literature regarding the attitudes of international students (especially students from Asian background) towards peer assessment. Since we haven't collected any data regarding the peer markers, we can not add to current knowledge however, we see this as a valuable area for future research, especially while the international student numbers are increasing in Australian universities. It is essential that we maintain a critically reflective approach to our implementation of the peer assessment process in statistics, to enable the continued improvement and contextualisation of the process, and to collect data about how well it is contributing to improving students' skills of critical thinking and independent decision making.

ACKNOWLEDGEMENTS

REFERENCES
Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment and Evaluation in Higher Education, 27*(5), 427–441.

Biggs, J. (2003). *Teaching for Quality Learning*. Buckingham: Open University Press.

Biggs, J. (1987). *Student Approaches to Learning and Studying*. Melbourne: Australian Council for Educational Research.

Bloxham, S. & West, A. (2004). Understanding the Rules of the Game: Marking Peer Assessment as a Medium for Developing Students' Conceptions of Assessment. *Assessment and Evaluation in Higher Education, 29*(6), 721-733.

Bostock, S. (2001). Student Peer Assessment. Website:
http://www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=422

Boud, D. and Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment and Evaluation in Higher Education, 31*(4), 399-413.

Brindley, C. & Scoffield, S. (1998). Peer assessment in undergraduate programmes. *Teaching in Higher Education, 3*(1), 79-89.

Brown, S. & Knight, P. (1994). *Assessing learners in higher education*. London: Kogan Page.

Brown, S., Rust, C. & Gibbs, G. (1994). *Strategies for Diversifying Assessment in Higher Education.*, Oxford: Oxford Centre for Staff Development

Cassidy, S. (2006). Developing Employability Skills: Peer Assessment in Higher Education. *Education & Training, 48*(7), 508-517.

Farh, J., Cannella & A. A., Bedian, A. G. (1991). Peer ratings: The impact of purpose on rating quality and user acceptance. *Group and Organisation Studies, 16*(4), 367-386.

Gibbs, G. (1981). *Teaching Students to Learn: A Student-Centred Approach*. Bristol, Pa.: The Open University Press.

Joliffe, F. (1997). Issues in Constructing Assessment Instruments for the Classroom. In I. Gal & J.B. Garfield (Eds.), *The Assessment Challenge in Statistical Education* (pp.191-204). IOS Press.

Langan, A.M., Wheater, C.P., Shaw, E.M., Haines, B.J., Cullen, W.R., Boyle, J.C., Penney, D., Oldekop, J.A., Ashcroft, C., Lockey, L. & Preziosi, F.P. (2005). Peer assessment of oral presentations: effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment and Evaluation in Higher Education, 30*(1), 21-34.

Marton, F., Hounsell, D. & Entwhistle, N. (1984). *The Experience of Learning*. Edinburgh: Scottish Academic Press.

Marton, F. & Saljo, R. (1976a). On qualitative differences in learning I. Outcome and process. *British Journal of Educational Psychology* 46, 4-11.

Marton, F. & Saljo, R. (1976b).On qualitative differences in learning II. Outcome as a function of learners' conception of task. *British Journal of Educational Psychology* 46, 115-27.

Murray-Harvey, Silins & Orrell, J. (2003). *Assessment for Learning: A Guide for Academics*. Adelaide: Flinders University, Flinders Press.

Ramsden, P. (2003). *Learning to Teach in Higher Education* (2nd Ed.). London: Routledge Falmer

Ramsden, P. (1979). Student learning and perceptions of the academic environment. *Higher Education*, 8(4), 411-427.

Sivan, A. (2000). The implementation of peer assessment: An action research approach. *Assessment in Education: Principles, Policy and Practice, 7*(2), 193-213.

Sluijsmans, D., Brand-Gruwel, S. & Van Merrienboer, J.J.G. (2002). Peer Assessment Training in Teacher Education: Effects on Performance and Perceptions. *Assessment and Evaluation in Higher Education, 27*(5), 443-454.

Stefani, L. (1994). Peer, self and tutor assessment: relative abilities. *Studies in Higher Education*, *19*(1), 69-75.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276.

Van Den Berg, I., Admiraal, W. & Pilot, A. (2006). Peer Assessment in University Teaching: Evaluating Seven Course Designs, *Assessment and Evaluation in Higher Education, 31*(1), 19-36.

APPENDIX 1: ASSIGNMENT

Macquarie University
MIST812 / Stat820: Decision Support Systems
GIS Assignment

This is a group assignment. The contribution to the assessment and the deadline are as stated in the course outline.

Each group should produce (a) a report in Word to be given to the lecturer, and (b) a set of files on the system in E4B308 to be demonstrated as a ten-minute presentation to an audience during the lab session in Week 7. The report should clearly state the name and contribution of each group member, and each group member should be present at the demonstration.

Use MapInfo and information available from the Internet to create an active earth map of a specified contiguous region on the earth's surface, such as southern Thailand or northern Sumatra, or a smaller area of a city. Your map should contain at least 24 components (suburbs or districts or provinces), geocoded to be linked to the map, with data (such as population density and ethnicity percent) for these components (and relations between them) graphed as thematic maps and/or Excel area charts. Your map should also contain separately addressable layers such as roads, walking or skiing tracks, caves, railway lines, rivers and landmarks of interest.

If a suitable map is not available from the Internet, create one from a hard copy using a scanner.

Your map should also have hyperlinks from which further details such as text and pictures appear when an appropriate point on the map is clicked.

APPENDIX 2: Presentation Marking Guide

Marking of GIS Presentation
Out of 100

Group Number: _____

Topic organisation

T1	0	1	2	3	4	5	The organisation of the topic work well
T2	0	1	2	3	4	5	The links between speakers work well
T3	0	1	2	3	4	5	The content is well researched, accurate and informative
T4	0	1	2	3	4	5	The map contains at least 24 components (suburbs or districts or provinces)
T5	0	1	2	3	4	5	Thematic maps and/or Excel area charts are (properly) used.
T6	0	1	2	3	4	5	Grid maps are created and explained.
T7	0	1	2	3	4	5	Colours in the maps are used properly.
T8	0	1	2	3	4	5	The maps contain separately addressable layers such as roads, walking or skiing tracks, caves, railway lines, rivers and landmarks of interest.
T9	0	1	2	3	4	5	The map/s also have hyperlinks from which further details such as text and pictures appear when an appropriate point on the map is clicked.

Presentation

P1	0	1	2	3	4	5	Your presentation is within the set time limits (10 minutes)
P2	0	1	2	3	4	5	Your presentation shows careful preparation
P3	0	1	2	3	4	5	Questions were answered correctly
P4	0	1	2	3	4	5	Good use of eye contact and body language
P5	0	1	2	3	4	5	Voice is audible
P6	0	1	2	3	4	5	Aids are readable
P7	0	1	2	3	4	5	Your presentation is effective
P8	0	1	2	3	4	5	Your presentation is entertaining
P9	0	1	2	3	4	5	Your use of layout, language, spelling and punctuation is correct

Methods used to create regions (tick one)
(First one is zero marks, and the other two are 10 marks)

☐ Uses existing MapInfo table (e.g. US counties)
☐ Uses existing MapInfo supplemented by students to obtain sufficiently many regions (24+)
☐ Created entirely by students

Any other comments: