

Assessment method difference: comparisons between international and domestic students within a first year undergraduate management course

E John McElvaney*, Jan Fermelis** and Helene Strawbridge***

This paper investigates the assessment methods and student results within a first year undergraduate management course offered within the business faculty of an Australian university. This course is compulsory for those studying for a commerce or management degree. The assessment results of full fee paying international students were compared with those of domestic students, during four teaching semesters in 2009 and 2010. Analysis compares 2,682 students' numerical results for two constructed response assignments to their results for an examination comprising both multiple choice questions and constructed response questions. It also compares the results of international and domestic students across metropolitan, regional and rural campuses. However due to little comparison data for multiple campuses, findings are consolidated by domestic and international students, university-wide. International students were found to achieve lower results than domestic students for constructed response assessment tasks, but higher results than domestic students for multiple choice question assessments. These findings have implications for instructors eager to provide a level assessment playing field for both domestic and international students, enabling both groupings to take advantage of existing strengths but also to improve their weaknesses. This research led to a restructuring and rescheduling of assessment tasks for the 2012 academic year.

Field of Research: Management discipline assessment methods

Key Terms: Higher education, management education, assessment, multiple choice questions, constructive response questions

1. Introduction

Within many Bachelor of Commerce or Bachelor of Business undergraduate degrees, Management is a core and foundation course. Generally taught during the first or second term of the student's university life, a management course is also a popular elective subject for students from other faculties and partner institutions. The course has a wide relevance as no matter what occupation the student later chooses, it is expected that the majority of university graduates will eventually find themselves managing people and resources (McElvaney 2009; Deakin University 2009).

* Eugene John McElvaney, Faculty of Business and Law, Deakin University, Australia.

Email: john.mcelvaney@deakin.edu.au

** Jan Fermelis, Faculty of Business and Law, Deakin University, Australia.

Email: jan.fermelis@deakin.edu.au

*** Helene Strawbridge, Faculty of Business and Law, Deakin University, Australia.

Email: helene.strawbridge@deakin.edu.au

McElvaney, Fermelis & Strawbridge

However, first-year management courses face some serious problems. In many universities these units have a high rate of incompleteness (McElvaney 2009). This student attrition rate exacts a high price in dollars, missed opportunities, and human lives (Erickson, Peters & Strommer 2006). For example, prior to 2008, 13% of students enrolled in the management unit investigated in this study withdrew prior to taking examinations. During the same period, 20% of students who did complete the unit failed to achieve a satisfactory final score. Therefore, in excess of 30% of students who had been enrolled in the unit at the beginning of the teaching period were unable to successfully complete.

The nature of assessment is also frequently problematic in such courses. To achieve consistency across such a diverse range of teachers and teaching locations there is a tendency for academics to apply Fredrick Taylor's management theory of choosing the one best way to do the job. This is often the case for those in charge of the teaching of large units (Waddell et al. 2007). Yet teaching, like management, is all about human beings. Few service industries in the world would be as people-oriented as education. When McElvaney (2009) surveyed students about their likes and dislikes about the management course prior to 2008, 80% indicated that they were less than happy about the then purely writing-based, or constructive response (CR) methods used for assessment. International students in particular expressed a strong preference for including Multiple-Choice (MC) tests within the total assessment mix.

These twin problems led to this university's management course team choosing in 2008 to trial use of students' preferred assessment method of MC testing. The trial aimed to increase the percentage of students passing the course and in 2008 the assessment mix within the course became:

Assignments

- 1. 10% - CR individual literature review research essay; and
- 2. 30% - CR group analytical research case study report

Examination

- 1. 30% - 60 x MC questions; and
- 2. 30% - 5 x CR theory and analytical questions.

Once these changes had been introduced, the management course team analysed students' assignment, examination and final results to determine the success of the new assessment mix. Those results form the basis of this paper.

2. Literature Review

Designing appropriate assessment is an integral task within the planning and organisation of any course, as it is regarded as of vital importance in order to accurately and fairly evaluate a student's knowledge. As part of the unit planning stage academics need to ask themselves 'what should our students be able to do intellectually, physically or emotionally as a result of the learning in that unit?' (Biggs 1999, p.49). First year students in particular frequently struggle to come to terms with the standard university requirements of research, and of the analytical and applied requirements of many assessment tasks.

McElvaney, Fermelis & Strawbridge

A great deal has been written about the strengths and weaknesses of assessment by multiple choice questions (MC) compared to written or constructed response (CR) questions, in relation to research and analytical assignments as well as written examinations. It has been noted that 'Constructed response items are preferred over multiple-choice by many in the education community because the former are believed to measure more important skills, be more relevant to applied decision making, better reflect changing social values, and have more positive social consequences' (Kuechler & Simkin 2010, p. 57). Although CR formats may often be viewed as a more suitable tool for accurately evaluating student knowledge, they do possess certain disadvantages. These include the subjectivity that can exist or be seen to exist in the marking process, the requirement for markers to possess subject mastery and the increased time to mark the CR component. Most universities do not reward or penalise faculties for the quality of their examinations (Bible, Simkin & Kuechler 2008, p. S56), so CR components effectively punish academics. Feedback takes longer, and the format may favour students with better English and handwriting skills (Kuechler & Simkin 2010, p. 57). As one investigator noted, on the basis of several studies, 'all ethnic groups have a relative disadvantage compared with Europeans in CR' Hickson (2010, p. 275). For all groups other than those from an Asian background the disadvantages do not occur when controls for student quality are introduced. It was concluded that 'rather than an ethnicity issue, this is likely to be a language issue' (Hickson 2010, pp. 275-6).

There is empirical evidence that MC questions may well be used with undergraduate classes as it has been found difficult to design MC questions to assess beyond the first three cognitive levels of Bloom's taxonomy (Simkin & Kuechler 2005, p.90). Later research by these authors in the computer programming discipline concluded that CR questions are not homogenous at the upper levels of knowledge in Bloom's taxonomy (Kuechler & Simkin 2010, p.68). From another perspective, new university students have expressed a preference for MC in order to smooth the transition from high school to university learning in ways that are both familiar and gradual (Erickson, Peters, & Strommer 2006; Bacon 2003).

Research differs about the intellectual value of MC assessment over CR assessment. Some evidence suggests that it is theoretically possible to construct MC items that measure many of the same areas as CR questions (Kuechler & Simkin 2010, pp. 56-57). Wainer and Thissen (1993) challenge the theory and posit that 'the multiple choice items may not be measuring the same thing, and so validity is served better by having both methods of assessment' (cited in Kennedy & Walstad 1997, p. 360).

Wallace and Williams discovered in their study of undergraduates that approximately two-thirds of their MC questions required the use of some level of critical reasoning (cited in Hautau et al. 2006, p. 260). On the other hand, Turner found that when writing activities were used during the semester, performance on MC exams improved (cited in Hautau & Turner 2006, p. 260).

Other concerns have been raised by a number of authors relating to whether MC tests evaluate the same level of understanding as that measured by CR tests. 'Test reliability as well as item difficulty and discrimination need to be considered to gain useful insights as to the validity of any instrument for measuring desired learning outcomes' (Moncada & Moncada 2010, p. 27). Nevertheless, indications are that MC questions are being frequently utilised for assessment purposes within at least some

McElvaney, Fermelis & Strawbridge

Australian universities, in response to tight budgetary constraints and increased teaching workloads (McElvaney 2009).

Consequently, especially when dealing with a large student cohort for compulsory first-year units, and students from diverse ethnic backgrounds, such as is currently the situation in most Australian universities, it is extremely important that adequate time be allocated to analysing, testing and refining MC questions, so that assessment outcomes are produced are not only cost-efficient, but also valid and reliable in accordance with course objectives, as well as reasonable and appropriate for diverse student populations.

In this report international students (IS) are those who, according to Organization of Economic Cooperation and Development (OECD), who travel to a country different from their own for the purpose of tertiary study. For Australia, 'International Students' are defined as those studying onshore only with visa subclasses 570 to 575, excluding students on Australian-funded scholarships or sponsorship or students undertaking study while in possession of other temporary visas. New Zealand citizens do not require a visa to study in Australia, so are not classed as international students. Domestic students (DS) are those who are Australian citizens, New Zealand citizens or holders of a permanent visa.

3. Methodology and Research Design

This investigation focuses on the assessment methods used in a first year management unit, compulsory for student enrolled in an undergraduate degree course, within a business and law faculty at an Australian university. The aim is to investigate a large data set to discover whether IS and DS receive similar or different final results, and similar or different scores for CR questions compared with MC questions. The total sample of 2,682 students comprises the 1995 domestic and 687 full fee-paying international students who completed and received a final score for that management unit, across the metropolitan, regional and rural and on line campuses for two semesters in each of years 2009 and 2010. Analysis focuses on comparing the results obtained by IS with those of DS.

Within the management unit, the assessment regimen consists of 40% progressive assessment and an examination worth 60%. The progressive assessment is divided into two assignment tasks: first, a 1,000 word, individual literature review and research essay worth ten percent of the student's final score; and second, a 3,000 word case study report assignment worth thirty percent, which requires teams of two students to research, analyse and compare various companies.

For the time period of this study, the final examination covered the entire semester's work, was a three-hour, independently administered session divided into two sections, each of which is worth 30 possible marks. The first section included 60 MC questions worth half a mark each. Each question referred to a separate aspect of management theory and provided four possible answers, labelled A to D.

Students answered this section of the examination by blackening a square on a Scantron scoring sheet for each question. Marking was conducted by scanning answer sheets electronically, with points awarded for a correct response, but without any point deduction for an incorrect response. Even though students were free to complete the examination tasks in any order, most students in practice appeared to complete this MC section first.

McElvaney, Fermelis & Strawbridge

The second section of the examination contained five short answer CR questions, each worth 6 marks. Each student wrote their answers into an examination script booklet, which was then individually read and graded by an examiner. As the number of students examined within each semester was so large, it was not feasible for a single examiner to grade all CR answers in all papers within the allocated period of time. Therefore student CR responses were graded by experienced examiners who were pre-trained in how to assess the specific CR questions on the examination paper and whose scoring was also moderated to maximise consistency. Because up to ten different examiners were working on the examination papers at the one time, any potential for inconsistency was also minimised by three markers being used for each student exam paper, spot-checking of individual examiners' scores by the Unit Chair, and blind double-marking of the papers of all students who had achieved a final examination score below 50%.

Assessment data from all students' assignments and examination results was downloaded into an Excel spreadsheet with all marks converted to a percentage to enable consistent comparisons across the data as some results related to a total possible score of 30 and others to a total possible score of 60 marks. There were five separate scores for each student: the first CR assignment, the second CR assignment, the exam MC, the examination CR plus the student's final total score for the course. Data was then imported into statistical software SPSS for in depth statistical analysis with the view of proving or disproving the hypotheses below:

Hypotheses:

H1: That international students will on average achieve lower scores than domestic students for constructive response assessment tasks; OR

H1 alternative: That international students will on average achieve higher scores than domestic students for constructive response assessment tasks;

H2: That international students will on average achieve higher results for multiple choice assessment tasks; OR

H2 alternative: That international students will on average achieve lower scores than domestic students for multiple choice response assessment tasks;

H3: That international students will on average achieve lower scores for exam constructive response tasks than for assignment constructive response tasks; OR

H3 alternative: That international students will on average achieve higher scores for exam assignment constructive response tasks than for assignment constructive response tasks

4. Findings and Discussion

The study found that the average student final result score for the unit was 61.58% with domestic students (DS) receiving an average overall mark of 63.23% compared to international students (IS) who received an average overall mark of 57.07%. These final results are included in Table 1 below, where students are categorized according to their citizenship (international student or domestic student), and according to their campus of enrolment (metropolitan, regional or rural/online).

McElvaney, Fermelis & Strawbridge

Table 1. Students' final results according to citizenship status and campus

All Cohorts of Management students	N	Mean	Std. Deviation	Std. Error Mean	Average Variances %
International Students	687	57.07	14.82	0.595	-7.3
Domestic Students	1995	63.23	13.24	0.465	+2.7
Average all Students	2682	61.58	13.64	0.498	
Metropolitan Campus	N	Mean	Std. Deviation	Std. Error Mean	Average Variances %
International Students	660	58.01	14.367	.559	-5.8
Domestic Students	1286	64.59	13.304	.371	+4.8
Regional campus	N	Mean	Std. Deviation	Std. Error Mean	Average Variances %
International Students	20	62.23	17.786	3.977	+1
Domestic Students	445	63.84	13.622	.646	+3.66
Rural campus & On Line cohort	N	Mean	Std. Deviation	Std. Error Mean	Average Variances %
International Students	7	65.42	9.285	3.509	+6.23
Domestic Students	264	63.36	12.522	.771	+2.9

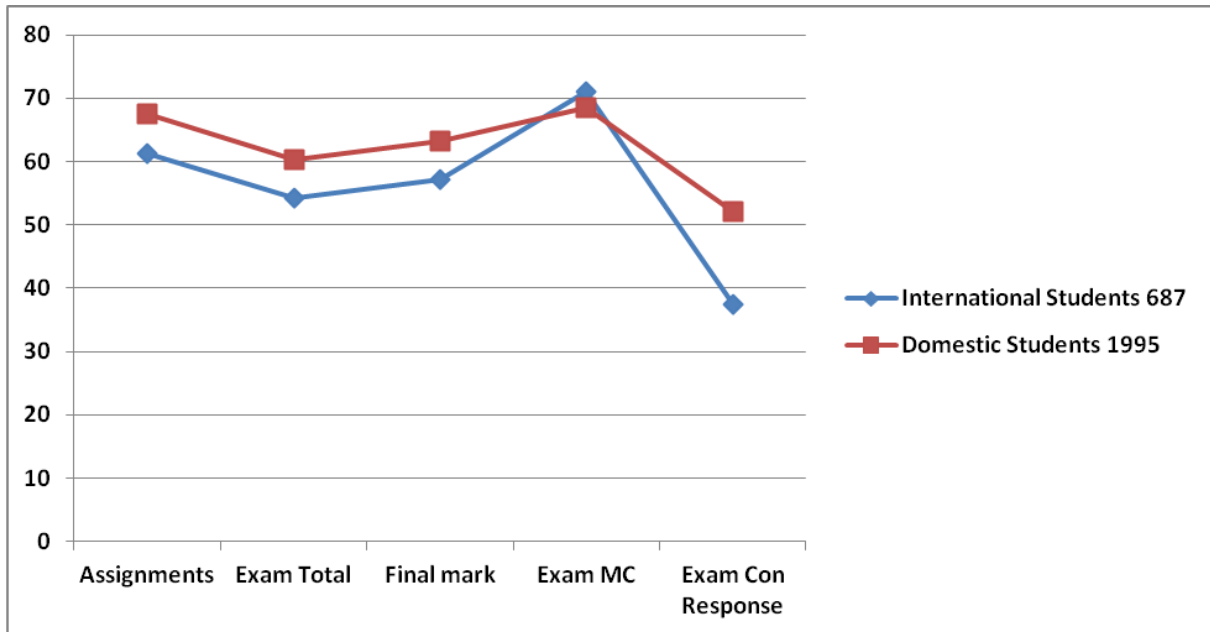
These results may appear satisfactory for a first year university unit considering the many language and other difficulties faced by international students. However, a deeper analysis of the data, as shown by Figure 1 below, shows that whilst the ISs achieved much lower scores than DSs for the Constructed Response (CR) question assessment tasks, they actually achieved higher scores than the DSs for the examination MC questions. In other words, the IS may have been disadvantaged by CR questions, but they actually outperformed DS in the MC questions in the exam.

Due to the similarity of data and low number of international students on non metropolitan campuses. All data will now be presented based on the comparison of International and Domestic students assessment results university wide.

As shown in the figure 1 below, IS scored lower results than DS for the two CR assignment tasks, and substantially lower results than DS for the CR examination questions. This finding agrees with earlier research published by Hickson (2010). Both CR and MC assessment methods are further examined to determine the reasons for the above findings.

McElvaney, Fermelis & Strawbridge

Figure 1. Comparison of International and Domestic students' assessment results



4.1 Assignments

Individual Literature Review and Research (10 marks) and group analytical research case study report (30 marks) with the results combined.

On average DS received a statistically significant difference in the average mark with DS clearly attaining a 10.1% higher result on average than IS ($p\text{-value} = 0.000$). These results could be anticipated for a first year university unit considering the many difficulties faced by international students experiencing their initial year in a country with a different language and culture.

Table 2. Analysis of Assignment Performances by Citizenship Status

Assignments 1 & 2	Mean % N mark	Std. Deviation	Std. Error Mean	Variances
International Students 687	61.32	15.54	0.565	-10.1
Domestic Students 1995	67.58	15.41	0.425	+10.1
All management Students 2682	65.84	15.44	0.48	

Based on the above, the hypothesis **H1**: That international students will on average achieve lower scores than domestic students for constructive response assessment tasks, should be accepted.

4.2 Exam-Overall result

Total 60 marks made up of 5 short constructed response answer questions worth 6 marks each totaling 30 marks. Plus 60 multiple choice questions worth 0.5 marks

McElvaney, Fermelis & Strawbridge

each totaling 30 marks. Exam conducted under strict examination supervision in week 14 after a 12 week teaching semester.

Table 3 below demonstrates a statistically significant difference in the average exam mark with the DS clearly attaining a higher result on average than IS ($p\text{-value} = 0.000$). Readers can be 95% confident that on average DS attain an exam mark anywhere from 5.35% to 7.98% higher than IS. As stated earlier this finding agrees with earlier research published by Hickson (2010), but with only a 10% average difference between the cohorts these statistics by themselves don't ring any alarm bells, because it's in line with the findings in the written assignments. It's when we analyze the different components of the exam that we can clearly notice the problem associated with articulation in a stress environment.

Table 3. Analysis of Performances by Citizenship status Examination Combined Components

Overall Exam Results	N	Mean % mark	Std. Deviation	Std. Error Mean	Variances
International Students	687	54.24	13.81	0.57	-10.4
Domestic Students	1995	60.31	13.15	0.45	+10.4
All management Students	2682	58.73	13.31	0.48	

4.3 Exam- Multiple Choice (MCs)

60 multiple choice questions worth 0.5 mark each totaling 30 marks. Exam conducted under strict examination supervision in week 14 after a 12 week teaching semester.

For Table 4 In addition a statistically significant difference shows in the average mark with IS clearly attaining a significantly higher result on average than the domestic student ($p\text{-value} = 0.002$). This is opposite to the CR SAQs data. In fact readers can be 95% confident that on average IS attain 0.8% to 3.7% more marks than DS in a multiple choice examination.

Table 4. Analysis of Performances by Citizenship status Examination Multiple Choice Questions

Exam multiple choice	N	Mean % mark	Std. Deviation	Std. Error Mean	Variances
International Students	687	71.12	15.82	0.62	3.9
Domestic Students	1995	68.48	13.2	0.385	-3.9
All management Students	2682	69.15	13.87	0.445	

Based on the above, the hypothesis **H2**: That international students will on average achieve higher results for multiple choice assessment tasks should be accepted.

McElvaney, Fermelis & Strawbridge

4.4 Examination- Constructed Response Short Answer Questions (SAQs)

5 short answer questions of 250 words worth 6 marks each totaling 30 marks. Exam conducted under strict examination supervision in week 14 after a 12 week teaching semester. Table 5 below firstly demonstrates a significant difference in the variation of scores attained by DS and IS in the SAQs .IS show greater variation in the SAQ marks. However a statistically significant difference shows in the average mark with DS clearly attaining a significantly higher result on average than IS ($p\text{-value} = 0.000$). In fact readers can be 95% confident that on average DS attain 13.8% to 17.5% more in their mark than IS.

**Table 5. Analysis of Performances by Citizenship status
Examination Constructed Response Short Answer Questions**

Exam Construction response	N	Mean % mark	Std. Deviation	Std. Error Mean	Variances
International Students	687	37.35	19.75	0.62	-39.7
Domestic Students	1995	52.18	18.05	0.385	+39.7
All management Students	2682	48.38	18.48	0.445	

Based on the above, the hypothesis **H3**: That international students will on average achieve lower scores for exam constructive response tasks than for assignment constructive response tasks should be accepted.

4.5 Summary of Findings

For both assignments, the examination and the unit overall, DS outperform IS by a little more than 10%. These data provide a satisfactory outcome for a first year university unit, considering the many difficulties faced by international students experiencing their initial year in a country and having to cope with a different language, culture and methods of teaching.

However, it would appear that under the pressure of an examination IS' ability to express themselves in Constructed Response Short Answer Questions is a major problem. Constructed Response exam (CR) data shows that the mean average score for DS in this assessment is 39.7% better than IS.

It would be expected that given the mean average CR exam mark of 37% for IS and then applying a standard multiple choice logic factor of 1.25, the average MC exam marks for IS should be 52%. However, the mean average score for IS in the M/C section of the exam was over 71%. Thus IS improved their CR exam marks by over 90% for MC exam marks.

This data has major implications on the fair assessment methods that we apply to first year international students as discussed below.

5. Conclusion and Implications

The results of this study align with earlier research in indicating that International Students may know or be able to memorise the content of the course, but are less successful when they express their course knowledge in writing. However, business graduates need to be able to have excellent communication skills as well as

McElvaney, Fermelis & Strawbridge

specialised subject knowledge, which is more the focus of other higher level courses within the students' tertiary degree. This research has led to changes to the submission dates of the assessments for 2012. Assessment tasks early within the academic program using non-constructive response formats are introduced now, to allow for the quicker identification of 'at risk' students and provide faculty time to assist international student to make the transition to meet the academic standards expected by Australian universities. The changing demographics of students at Australian universities lead to the need for further research into assessment methods.

References

- Bacon, DR 2003, 'Assessing learning outcomes: a comparison of multiple-choice and short-answer questions in a marketing context', *Journal of Marketing Education*, vol. 25: no. 1, pp. 31-6.
- Bible, L, Simkin, MG & Kuechler, WL 2008, 'Using multiple-choice tests to evaluate students' understanding of accounting', *Accounting Education*, vol. 17, Supplement, pp. S55-S68.
- Biggs, J 1999, *Teaching for quality learning at university*, SRHE and Open University Press. Buckingham, England.
- Deakin University, Institute of Teaching and Learning 2009, *Teaching and Learning Modules*, retrieved 10 August 2009, <http://www.deakin.edu.au/itl/pd/tl-modules/index.php>.
- Erickson, BL, Peters, CB & Strommer, DW 2006, *Teaching first-year college students*, John Wiley & Sons Inc., San Francisco, Ca.
- Hautau, B, Turner, HC, Carroll, E, Jaspers, K, Parker, M, Krohn, K & Williams, RL 2006, 'Differential daily writing contingencies and performance on major multiple-choice exams', *Journal of Behavioral Education*, vol. 15, pp.259–276.
- Hickson S 2010, 'The impact of question format in principles of economic classes: evidence from New Zealand' *New Zealand Economic Papers*, vol. 44, no. 3, pp. 269-287.
- Kennedy, P & Walstad, WB 1997, 'Combining multiple-choice and constructed-response test scores: an economist's view', *Applied Measurement in Education*, vol. 10, no. 4, p.359.
- Kuechler, WL & Simkin, MG 2010, 'Why is performance on multiple-choice tests and constructed-response tests not more closely related? Theory and an empirical test', *Decision Sciences Journal of Innovative Education*, vol. 8, no. 1, pp. 55–73.
- McElvaney, J 2009, 'Do teachers of management "practice what they preach"?' *International Review of Business Research Papers*, vol. 5, no. 1, pp. 174-183, World Business Institute, Melbourne, Vic.
- Moncada S & Moncada, T 2010, 'Assessing student learning with conventional multiple-choice exams: design and implementation considerations for business faculty', *International Journal of Education Research*, vol. 5, no. 2, pp. 15–29.
- Simkin, MG & Kuechler, WL 2005, 'Multiple-choice tests and student understanding: what is the connection?', *Decision Sciences Journal of Innovative Education*, vol. 3, no. 1, pp. 73–97.
- Waddell, D, Devine, J, Jones, GR & George J 2007, *Contemporary management*, McGraw-Hill Australia, Sydney, NSW.