An Analysis of Accounting Exam Answers: An inductive case study of response characteristics

Gail Ridley
School of Accounting and Corporate Governance
University of Tasmania, Private Bag 86, Hobart, Tasmania, Australia 7001. Gail.Ridley@utas.edu.au.

ABSTRACT

   This paper reports upon an exploratory study that aimed to identify and benchmark characteristics to account for the failure of students in short essay-style answers in the final examination of a second year core accounting unit. The purpose of doing so was to develop a benchmark against which future examination outcomes could be compared, to identify any differences in the outcome characteristics between international and domestic students and to inform the future redevelopment of the unit content and its delivery. An inductive approach was taken to the descriptive, quantitative analysis of 120 examination answers from 45 students, identifying themes using thematic analysis techniques. Although international students dominated both the group of students who had failed in the examination, and the failing answers, the same themes were found in the answers of both student groups, and the ranked order of the themes for both was similar. Modifications to the teaching and learning framework used for future delivery of the unit, and subsequent evaluations of similar assessment, will contribute to a better understanding of the causal links between the delivery and assessment outcomes. Future research will confirm whether the characteristics found apply to other accounting higher education contexts.

INTRODUCTION

   A very considerable body of literature has been published that is oriented towards enhancing student outcomes within a higher education context. This literature considers assessment, as well as associated learning frameworks. The studies reported in the literature have been influenced by a diverse range of disciplines and fields, including psychology, sociology, education and specific knowledge domains, as well as accounting education. The diversity of these publications, and the disciplines from which they are derived, give recognition to the complexity of teaching and learning. In an attempt to reduce complexity, many streams of research have examined higher education teaching and learning processes and outcomes from a particular perspective. Just as one influential example, a considerable body of literature exists that supports use of constructivist-based teaching (see for example, Bednarski 1997).

   Academics and others closely involved with higher education teaching and learning practice, in accounting and elsewhere, are faced with a range of complex questions. These questions relate to the nature of the practice problems encountered, causal links between outcomes and processes, possible interventions and implementation and delivery. In this setting, examination of quantitative forms of evidence have been popular, guided, *a priori,* by theory, using a deductive approach. Theory itself is designed to reduce complexity, through focus on a particular perspective to guide the analysis of evidence. However, deductively-oriented quantitative research, most often manifesting itself as "self-report inventories" in approaches to the learning framework in accounting education (Lucas & Mladenovic 2004, p.400) "can lead to confusion (and) difficulty in comparing findings" (p.400). The learning process is too complex to be measured through this kind of instrument (Lyons 2006). Inductive research instead seeks patterns from data without the imposition of theory, and is often used to assist in the development of theory. Inductive studies are appropriate for exploratory research where causal connections have not yet been hypothesised.

   The case-study research reported upon in this paper sought to identify the characteristics of short essay-style answers of students taking Accounting Information Systems, a second year core accounting unit from an accounting major in a business degree, where the students had failed the final examination. The study sought to develop a benchmark against which future examination outcomes could be compared, in order to inform the future redevelopment of the unit content and its delivery, and to identify any differences for the group in the outcome characteristics of international and

domestic students. However, identification of the causal links between teaching and learning processes and the examination outcomes was outside the scope of the current study, which was exploratory in nature.

BACKGROUND

As a data-driven approach guided this study it was considered that a detailed review for themes from the literature may have "…contaminat(ed)… the inductive derivation of key concepts" (Parker & Roffey 1997). Instead, data analysis would be followed by a detailed analysis of the literature, looking for consistency and inconsistency with the findings that emerged from previous studies. Some unusual circumstances made it possible for this process to be followed, as it would be expected that an accounting academic would already be familiar with the approaches to teaching and learning in accounting. However, the researcher had relatively recently started teaching within an accounting school, having been appointed from another relevant discipline.

Case Context

This study was prompted by the researcher's experiences in being involved in the delivery for the first time of a second year unit in accounting information systems from within an accounting major in an undergraduate degree program within an Australian university. In particular, the study was motivated by the relatively high proportion of students who had gained less that 50% in the examination component of the assessment. The unit was very largely delivered as it had been in the previous year, both with regard to its content and style, as developed by an experienced academic with an accounting background, who was still closely associated with the delivery of the unit. The researcher was familiar with both the university and faculty in which the accounting school was located, having worked as an academic in another school elsewhere in the faculty, teaching and researching in Information Systems.

The unit had 161 enrolments as at the end of the semester in which the examination was conducted. Four teaching staff were involved in its delivery across three geographic locations, two of which were within Australia while one was located in a South-East Asian nation. The unit was delivered in a traditional style, with a two-hour lecture and a one-hour-tutorial each week over a 13-week semester. Group activities were undertaken in the tutorials. Notes and other materials were made available on WebCT Vista to those enrolled in the unit. Answers to tutorial activities and weekly multiple-choice questions were also placed on Vista weekly, while completion of the multiple-choice exercises was an optional independent activity. Three pieces of assessment were used for the unit. The first assessment was through a mid-semester test of multiple choice questions, "fill-in-the-word" questions and short answer questions. The second piece of assessment due for submission in the second-half of the semester was a small-group practical assignment, while the final piece of assessment was a traditional closed-book, three-hour examination. The latter assessment item was the source of the data used for this study. The summative examination consisted of two sections and multiple questions with a different style, an overview of which is presented in Table 1.

Section	Question	Description	Total marks
A	1	Analysis of case study & data flow diagram to identify problems; six 2–3 line answers after marking the data flow diagram	18
	2	Interpretation of flowchart; half-page answer	10
	3	Analysis of a list of 10 procedures & requirements to identify internal controls; answers consisted of a few words each	10
	4	Three-quarter page case study with an organisational context; choice of 4 from 6 questions relating to the case, each requiring a half-page answer	32
B		15 multiple-choice questions, each with four options; answer involved recording a letter indicating the correct option for each	30
			100

Table 1 Overview of Examination Used for the Unit

Of the 161 students enrolled in the unit, 66 (41%) were regarded by the university as being international while 95 (59%) were Australian domestic students. Of the total, 45 gained less than 50% for the examination component of the assessment, which represented almost 28% of the students enrolled in the unit. However, this percentage did not consider the ongoing assessment component of

the unit, which formed 40% of the overall mark. Of the 45 examination papers analysed, 36 (80%) came from international students and nine (20%) were from local students. It could be anticipated however, that around 18 of 45 papers that gained less than 50% would come from international students, while around 27 were from domestic students. The relatively high proportion of examination papers that had scored less than 50%, and the over-representation of international students in this group, partly motivated this study.

The methodology used for the study is discussed next.

METHODOLOGY
Philosophy and research approach
Quantitative, thematic analysis was chosen, using a case study method. Case study research allows investigation of a "phenomenon in its natural setting (using) multiple methods of data collection to gather information from one or a few…people, groups or an organisation (Benbasat, Goldstein & Mead 2002, p.21). In the current study, data would be first collected through the "natural setting" of final examinations, before answers were selected for analysis, to reflect the aims of the study. As additional evidence of the natural setting for the study, the researcher was also the marker of all the exam questions used for analysis, from all three delivery locations, although another academic involved in delivery of the unit marked the first two questions from Section A. Data would also be collected from the broader education literature and the more narrow accounting education literature, to determine whether the findings were unique, or confirmed those of previous studies.

Deductive research commences with an "abstract, logical relationship among concepts" (Neuman 2000, p. 49), and then seeks to test theory using gathered data.   However, the deductive method may obscure important themes that may not be identified in data collection and analysis, because of preconceptions of the researcher (Thomas 2000). The inductive approach assists the researcher to understand meanings from the data, through "…developing insights and generalizations out of the data collected" (Neuman 2000, p.122). Inductive research starts with detailed observations and then moves towards abstract generalisations (Neuman 2000). The inductive approach was highly appropriate for this study as it enabled "…findings to emerge from …common…themes inherent in raw data… (Thomas 2000, p. 3).

Thematic analysis involves the identification of themes in data, before summarising the findings under themes, often in the form of tables (Dixon-Woods, Agarwal, Jones, Young & Sutton 2005). There are two main forms of thematic analysis, data-driven, where themes are derived from the data, or theory-driven, that is, where data are interrogated for particular themes based on the literature (Dixon-Woods *et al.* 2005). The thematic analysis used in this study was data-driven.

Study aim
The study sought to investigate all failing sub-questions in Question Four from Section A of the examination paper, for all students who had failed the exam.   An understanding of the characteristics of those papers that accounted for the failures were sought, rather than reasons behind the response characteristics. It was reasoned that identification of the characteristics of the failing responses in the examination for the weakest students would be the first phase towards establishing causality for the failures and addressing those causes. It was assumed that the causes that had brought about the characteristics would be complicated and multi-faceted, and likely to be connected to the teaching and learning framework used for the unit. Consequently, for that reason *a-priori*, deductive research was not considered appropriate, as it was also assumed that one theory or conceptual approach was unlikely to be able to explain the complexity found. Benchmarking the characteristics found in the failed answers was sought, as this would allow modifications to be made to the teaching and learning framework for future iterations of delivery of the unit, followed by subsequent revaluation of examination responses. Such a benchmark would allow measurement of the effectiveness or otherwise of future treatments to reduce the proportion of failing answers, and to reach an understanding of causation.

The proportion of students who were regarded by the university's student administration as international was over-represented in the students who had failed the examination, when compared against the proportion of international students in the entire unit. Therefore it was assumed that some attributes associated with being an international student might be responsible for some of the

3

characteristics found. For this reason, the failed answers from the international studies were analysed in the same way, but separately, to the answers from the domestic students.

Procedures

As the examination paper had multiple questions, the sections for analysis first needed to be determined. The first three questions from Section A of the paper were more technically oriented, and offered limited scope for a broader analysis of the response characteristics, other than the answer was correct or incorrect. Section B of the paper consisted of 15 multiple-choice questions. Again this section was not considered appropriate for analysis, for a similar reason. However, Question 4 consisted of a case study of around one-page, with six smaller questions where the students needed to refer to the case study to respond. The questions then not only required the student to recall factual information, but also to apply it in a meaningful way to a context. The students needed to answer four of the six questions, writing a short-essay style response of around half a page for each. The choice of questions meant that the students could avoid two questions on topics that they had not prepared or could not recall.

After developing a written marking scheme, the researcher had originally assessed all the questions in this section of the exam paper, marking all the answers for the same question together, rather than marking different questions of the paper for one student before starting to mark the paper of another student. The researcher had taken this approach to assessment to help ensure that the marking was reliable. For the same reason the papers had been marked in a limited period, rather than over an extended period. When the marking had been completed, the examiner had reviewed a selection of papers marked earlier in the period, in the middle and then later, to compare the consistency of the marking for each sub-question in Question Four. Where variations were found, the marks across the papers were examined and adjusted. Two administrative officers audited all the papers, checking that all the pages on each examination paper had been marked and the marks for each question of the paper had been entered, and the totals were correct. Some minor errors were found and these were corrected by the assessor.

All papers where students received less than 50% overall for the exam, from any of the three centres, were then identified by the same assessor. All answers for Question Four from this group of students were examined. Not all of the students who had gained less than 50% in the examination had failed all four sub-questions within Question 4. Where a student had failed one or more of these sub-questions (ie had gained less than 50% for a sub-question), then only these responses were analysed.

Data analysis

Thematic coding was used for data analysis to identify themes or concepts from the data" (Ezzy 2001). The researcher read through each answer several times to identify prominent themes, and then summarised the findings under thematic headings before collating these into summary tables (Dixon-Woods *et al.* 2005). Although thematic analysis can either "reflect the frequency with which particular themes are reported" or may be "…weighted towards themes…(that) have a high level of explanatory value" (Dixon-Woods *et al.* 2005, p.47), for the analysis in the current study, the frequency of occurrence of the themes was recorded using a spreadsheet. For example, where an answer was seven lines or less, or occupied more lines but the number of words per line was considerably reduced, a tally point was recorded under the category "*too brief*". As another example, it was noted early on in the analysis process that some of the answers were mismatched to the question. However, it was also realised that there were different types of mismatched answers. Some answers provided information mismatched to the question, but the answer displayed considerable preparation for the examination, but on another topic, while other mismatched answers were not meaningful. Considerable exploration took place before the themes were derived from the data, and then were used for analysis. On occasions this required collapsing categories into one, where it was found that a theme was replicated, but under a different name. The themes that emerged from the analysis procedure appear in the results and discussion of the findings.

The examination papers were "deidentified" for analysis, for example, in order to group the papers into whether they were written by international or domestic students, the names of the students were not used. This procedure, and the requirement that extracts of the papers not be used for publication purposes, were conditions imposed in order to obtain ethics clearance to undertake the study.

As quantitative analysis is usually required to be reproducible, and for reliability, another researcher trained in research methods analysed a sample of the answers using the developed themes, before those results were examined.

RESULTS

Of the 45 papers examined where a total of less than 50% had been awarded, 120 of the responses to Question Four gained less than 50%. Of these 120 responses, 19 (16%) were written by domestic students while 101 (84%) were written by international students. Again, international students were over-represented in the proportion of responses for this question that had gained less than 50%.

Many failed answers contained multiple characteristics that contributed to their failure, all of which were coded. It was found that the characteristics could be grouped into three broader categories, relating to the quality, quantity and timing/organisation of the answers.

Six characteristics relating to quality were found in the failed answers, namely *irrelevant responses but where knowledge was shown*, *irrelevant but where little knowledge was shown*, *responses that lacked meaning*, *incorrect responses that were however meaningful and relevant*, *incomplete but partially correct responses* and *those that lacked depth or specificity*. Quality issues were responsible for the great majority of the characteristics, with 212 of 252 characteristics, or 84%.

There were two quantity issues that accounted for some of the failing responses, that is where there was *no answer for one of the sub-questions*, and where answers were *too brief*. The quantity characteristics accounted for a much smaller proportion of the failing answers when compared against the quality characteristics, with 37 of 252 characteristics, or almost 15%.

A third group of characteristics contributed to the failure of some answers, timing and organisation issues. There were two characteristics in this group, namely *where students had answered all six sub-question choices instead of four*, and where *answers had been extensively edited or rewritten*. Both characteristics would leave the students less time to answer the sub-questions. The third group was the smallest when compared to the other two broad groupings of characteristics, with 3 of 252 characteristics, or approximately 1%.

Table 2 sets out the characteristics or themes found associated with the failed answers, grouped into these three broad categories, and the frequency of each characteristic, for both international and domestic students. Table 2 also gives the anticipated frequency calculated for each characteristic, along with totals and subtotals, for international and domestic students. The latter frequencies were derived from the proportion of each student group in the overall unit population, for the particular characteristic of interest. The calculated anticipated frequency is given to the nearest integer.

It can be see that in all cases the frequency found for the characteristics was higher for the frequency anticipated for international students, and lower than anticipated for domestic students, when the proportion of each group of students from the 45 who gained less than 50% was examined. In general terms the frequency found for the international students was around twice the number anticipated, while for the domestic students, the frequency found was around one-third or less than that anticipated for domestic students.

Characteristics	International Students		Domestic Students	
Quality issues	*Frequency Found*	*Frequency Anticipated*	*Frequency Found*	*Frequency Anticipated*
Irrelevant (but knowledge shown)	21	10	4	15
Irrelevant (little knowledge shown)	55	25	7	37
Lacked meaning	18	8	1	11
Incorrect (but meaningful/relevant)	23	9	0	14
Incomplete (partially correct)	7	4	2	5
Lacked depth/specificity	57	30	17	44
Subtotals	*(181)*	*(87)*	*(31)*	*(125)*
Quantity issues				
No answer	1	0	0	1
Too brief	32	15	4	21
Subtotals	*(33)*	*(15)*	*(4)*	*(22)*
Timing/organisational issues				
Answered all 6 questions	1	0	0	1
Extensively edited	2	1	0	1
Subtotals	*(3)*	*(1)*	*(0)*	*(2)*
Totals	***217***	103	***35***	149

Table 2 Frequencies found and anticipated for Question 4 failed answers for international and domestic students

Table 3 displays the ranking for the same characteristics for both the international students and the domestic students by frequency.

Characteristics	Ranking for International Students	Ranking for Domestic Students
Irrelevant (but knowledge shown)	5	3
Irrelevant (little knowledge shown)	2	2
Lacked meaning	6	6
Incorrect (but meaningful/relevant)	4	7
Incomplete (partially correct)	7	5
Lacked depth/specificity	1	1
No answer	9	7
Too brief	3	3
Answered all 6 questions	9	7
Extensively edited	8	7

Table 3 Ranked frequencies for Question 4 failed answers for international and domestic students

It can be seen that both the international and domestic students who gained less than 50% for the exam had similar rankings when the characteristics for the failed responses for Question Four were ranked by frequency. Both groups of students had the same ranking for *lack of depth or specificity* (first), *irrelevant with little knowledge shown* (second), *too brief* (third) and *lacked meaning* (sixth). The remaining characteristics were similarly ranked, with only the characteristic, *incorrect but meaningful/relevant,* being more than two ranked places away when the ranking for the international and domestic students were compared.

Table 4 sets out the differences between the found and anticipated frequencies for international and domestic students, for Question 4, which have also been ranked. As just the two groups of students were used for analysis, international and domestic, the combined differences for each group totalled zero. When the differences between the found and anticipated frequencies for international and domestic students were ranked by the size of the difference, disregarding the sign, the rankings differed from those displayed in Table 3. The rankings in Table 4 incorporated the anticipated frequencies when the proportion of both the international and domestic students was considered, and so reveal the areas where the international students were under-performing to the greatest degree in their answers to Question Four, and the reverse for the domestic students.

Characteristics	Difference in Frequencies for International Students	Difference in Frequencies for Domestic Students	Ranked Difference
Quality issues			
Irrelevant (but knowledge shown)	-11	+11	5
Irrelevant (little knowledge shown)	-30	+30	1
Lacked meaning	-10	+10	6
Incorrect (but meaningful/relevant)	-14	+14	4
Incomplete (partially correct)	-3	+3	7
Lacked depth/specificity	-27	+27	2
Subtotals	*-94*	*+94*	
Quantity issues			
No answer	-1	+1	9
Too brief	-17	+17	3
Subtotals	*-18*	*+18*	
Timing/organisational issues			
Answered all 6 questions	-1	+1	9
Extensively edited	-1	+1	9
Subtotals	*-2*	*+2*	*8*
Totals	***-114***	***+114***	

Table 4 Differences between found and anticipated frequencies for international and domestic students, with ranking

The sample of the short essays independently analysed by a second researcher using the same ste of characteristics yielded similar results, with a 83% match, which was interpreted as confirming that the coding was sufficiently reliable.

The results will be discussed in the next section, in conjunction with reference to prior studies, which were consulted after analysis was completed, for reasons explained earlier.

DISCUSSION

Of the 45 students who gained less than 50% on the examination paper, the 36 international students in this group had an average 2.8 of the Question Four sub-questions of the examination below 50%, while the nine domestic students had 2.1 failing answers in this question each on average. Consequently, not only were international students considerably over-represented in the group of the 45 failing students in terms of the number of students, they had, on average, 0.7 more failing answers to the sub-questions for Question Four, when compared to the domestic students. Moreover, international students were over-represented on every one of the ten characteristics considered to contribute to a failed response, while domestic students were under-represented for all, when the proportion of each student group in the unit was taken into account. Therefore, initial concern about the performance of the international students in the unit was confirmed from closer examination of the failing responses for Question Four.

The literature identified concern about assessment outcomes in higher education, and reflected on the role of essays in the assessment process. Accounting schools in universities were considered to "struggle with both philosophical and practical aspects of … assessment (Hindi & Miller 2000). Concern was expressed in the recent literature about the under preparation of students in higher education, and the high attrition rate, where students did not study sufficiently to achieve (Tinnesz, Ahuna & Kiener 2006). However, essay style examinations were seen as more suitable for monitoring learning and teaching than other forms of assessment (Stiggins 2002), and had higher face validity than multiple-choice exams (Burke 1992). Stout, Bordon, German and Monahan (2005) commented that the accounting education literature had limited descriptive publications of implementation results. Therefore, there was support from previous studies to validate the researcher's concern about assessment outcomes for students in an examination context. Moreover, the choice of essay-style

questions for evaluation was appropriate, while studies that provided descriptions of implementation results, like the current one, were sought.

In the current study, although quality issues made the major contribution to failed answers for Question Four for both international and domestic students, they were proportionally more important for domestic students. Quality characteristics were eight times more important than the next most important group of characteristics for domestic students, but they were around five times more important for international students. Characteristics relating to quantity were the second most important group in the failed responses for both student groups, while timing/organisational issues were the least important of the three groups for international students, and were not identified in the examined answers of the domestic students.

When individual characteristics were considered, the three characteristics ranked highest by frequency of occurrence not only had the same ranking for both international and domestic students, but together they also accounted for a high proportion of the failed answers. To illustrate, the characteristics, *lack of depth and specificity*, *irrelevant with little knowledge shown* and *answers too brief*, together accounted for 66% of the failed answers for international students and 80% of the answers for domestic students. If these characteristics had been eliminated or considerably reduced in the responses for both groups, then it is possible that far fewer students would have failed individual sub-questions for Question Four.

When the differences between the actual and anticipated frequencies for the characteristics from the failed responses to Question Four for international and domestic students were ranked in Table 4, different rankings were found when compared to the ranked frequencies of Table 3. It is argued that these former rankings were the most meaningful, as they indicated the areas where the international students under-performed the most, when the proportions of both the international and domestic students enrolled in the unit were considered. When the first three ranked characteristics were examined, the order of the first and the second was the reverse of that obtained when the characteristics were ranked by frequency alone, while the characteristic ranked third was the same. This suggests that the same three characteristics obtained from an analysis of frequency alone could also account for the failed responses to Question Four when the proportion of the international and national students was included in the analysis. This finding also seems to confirm that the characteristics, *lack of depth and specificity*, *irrelevant with little knowledge shown* and *answers too brief,* have the greatest scope to increase the performance of both the international and domestic students in questions similar to those of Question 4 in the examination for the unit.

It will be remembered that this study aimed to identify those characteristics that accounted for the failures in the answers examined, rather than the reasons behind them. An assumption was made that the reasons behind the characteristics would be multi-faceted. This seems likely from analysis, as characteristics associated with failing were identified 217 times in the sub-questions in Question 4, even though only 120 answers were analysed. In other words, on average, approximately 1.8 characteristics that may account for failure were found in each response examined. This figure tends to confirm the assumption that explanation for failure is multi-faceted.

Of the three most important characteristics accounting for failure, *irrelevant with little knowledge* could be explained in a number of ways, each of which could act independently or in association with other causes. One explanation (*Explanation 1*) could be that some students had undertaken little preparation for the examination, particularly when combined with little attendance at classes. Support for this explanation in an accounting context can be seen in the literature, as the final examination mark is positively correlated with the proportion of classes a student attends (Paisey & Paisey 2003). Another explanation (*Explanation 2*) could be linked to the limited language skills of some international students, in that they may have misunderstood the question. However the responses of some domestic students also displayed the same characteristic, as it was ranked the second highest for local students, who would be expected to be highly competent in English. A further explanation (*Explanation 3*) could be that students had prepared for the examination, but had prepared in an ineffective way.

Answers displaying a *lack of depth and specificity,* the characteristic with the second highest frequency linked to failure, may be explained simply by poor examination skills (*Explanation 4*). Where this is the case, provision of model answers to students when preparing for future examinations

may help. However, explanation 1, 2 or 3 could also contribute to this characteristic. *Explanation 2* appears to be even less likely in this case than for the characteristic, *irrelevant with little knowledge*, as 17 domestic students displayed a *lack of depth and specificity* in their responses, the highest number for this student group.

The third ranked characteristic accounting for failure, *too brief*, could be explained by one or more of *Explanation 1*, *Explanation 3* or *Explanation 4*. *Explanation 2,* students' limited language skills, may also be a possible cause for failure, particularly for international students. There were few domestic students who fell into this category, even though it also ranked third for this latter group. However, as 79 responses to other sub-questions from international students were not brief, with many well exceeding the half-page length suggested for these questions, *Explanation 2* seems unlikely. Another possible explanation for brief answers may have been limited handwriting fluency (Connelly, Dockrell & Barnett 2005) of the students. However, this is considered unlikely because of the longer length of some of the other Question Four responses.

Figure 5 summarises some possible causes for the most highly ranked failing characteristics, as outlined in the previous three paragraphs.

Most Highly Ranked Characteristic	Explanation 1	Explanation 2	Explanation 3	Explanation 4
lack of depth and specificity	X		X	X
irrelevant with little knowledge	X		X	X
too brief	X		X	X

Table 5 Possible Reasons for the 3 Most Highly Ranked Characteristics

Considerable discussion was found in the accounting education literature of "deep" and "surface" approaches to learning, influenced by the original studies from Marton and Saljo (1976; 1984).   In deep learning, students try to make sense of a subject against a framework of concepts, which enables them to link the material to their existing knowledge. However, in surface learning, the student focuses on unconnected facts to be memorised, and is not guided by patterns to integrate the material. As a result, when using surface learning, students find it difficult to apply their learning to a new setting (Davidson 2002). Earlier research by Volet and Chalmers (1992) suggested that a continuum exists from deep to surface learning. A surface approach to learning has been negatively correlated to marks in a commerce university unit (Eley 1992).

Undergraduate accounting students have been found to take higher surface approaches to learning, and lower deeper approaches (English, Luckett & Mladenovic 2004). Two groups of studies have been undertaken to evaluate the learning styles of undergraduate accounting students, aiming to bring about deeper learning. The first investigated the learning approaches used by accounting students at universities. This work has included comparing the student's approaches with those from other disciplines (for example, Booth, Luckett and Mladenovic 1999) and with international students, usually Asian (for example, Hassall & Joyce 2001). A second body of research has tried to encourage students to take a deep approach to learning through interventions to the teaching and learning framework, some of which have been successful (for example, English, Luckett & Mladenovic 2004).

Students' approach to learning was found to be affected by both their orientation to learning and the learning context (English, Luckett & Mladenovic 2004). The latter belief has motivated studies where interventions manipulate the learning context, including the nature of the course and the teaching (English, Luckett & Mladenovic 2004). Although it has been found that the curriculum, assessment, workload levels and student-centred learning styles can positively influence the adoption of a deeper learning approach by students, previous educational experiences may work against this (English, Luckett & Mladenovic 2004).

Although the findings of research to compare Australian and Asian students studying in Australia for their approach to learning has had mixed results, studies have not confirmed an initial perception that Asian students were mainly surface learners, who learnt by rote (see for example, Cooper 2004). However, Hassall and Joyce suggested that students may believe that a surface approach is warranted for assessment (2001). In the latter authors' evaluation of the learning approach of accounting students from the United Kingdom (UK) and from outside the UK, studying in the UK with an accounting professional body, international students had higher scores for surface

learning earlier on in their studies, which then decreased later in their studies (Hassall & Joyce 2001). The same authors claimed that a deep approach to learning can be encouraged through assessment, curriculum content and delivery style.

Explanation 3, the possibility that students had prepared for the examination, but in an ineffective way, appears consistent with many of the findings on deep and surface learning. Moreover, the potential of positive outcomes from intervention, through changes to the teaching and learning framework, are encouraged by previous findings from the literature. If students who failed the examination were surface learners, the essay-style questions of Question Four would be the most difficult, where knowledge needed to be applied to the case study.

CONCLUSION

This exploratory case study used an inductive approach to identify the characteristics that accounted for the failure of a group of low-achieving, mainly international students, who failed short-essay style questions on a second-year accounting unit at an Australian university. It was found that the international students were considerably over-represented in the failing examination answers examined. Ten characteristics were found that were associated with the failure of the answers, grouped into three categories; quality issues, quantity issues and timing/organisational issues. Of these the first group of characteristics had the highest frequency of occurrence, followed by quantity characteristics. The three individual characteristics with the highest frequencies were answers that were *irrelevant with little knowledge shown*, those that *lacked depth and specificity* and answers that were *too brief*. Although the frequencies of characteristics for domestic students were far smaller, and the number of domestic students in the failing group was limited, a surprising finding of the study was that the frequencies for the characteristics for the domestic students largely mirrored those for the international students, particularly for the most highly ranked characteristics. This was true whether the raw frequencies were considered or where the frequencies were adjusted for the proportion of both groups of students enrolled in the unit.

The significance of this finding is that it appears that interventions to improve the outcomes for international students on essay-style examination questions like those in Question Four may be likely to assist both international and domestic students in the unit. Furthermore, the findings of this study suggest that it may be possible to start to reduce the proportion of students failing essay-style questions in examinations by taking some simple measures with both international and domestic students. One is to communicate more effectively to the students the length of the responses required, as "too brief answers" was a highly ranked characteristic for both student groups. Another corrective measure that may assist students is to provide model answers to mock or past examination questions that demonstrate the degree of specificity and depth considered appropriate.

Now that this initial benchmark has been established after analysis, potential exists to make changes to the teaching and learning framework and measure outcomes on essay-style examination questions in the unit in future years. If effective interventions are identified, further research will need to be done in other contexts before it is possible to generalise beyond this case study setting. The literature points to the possibility that differences in the unit student population regarding deep and surface learning styles may contribute to the differences in examination outcomes for the students. The extensive accounting education literature available on deep and surface learning approaches may lead to possible future interventions.

The framework of ten characteristics developed from thematic analysis of the 120 examination essays may need refinement as a result of further iterations of examination evaluation in the unit, after this initial benchmark.

REFERENCES

Bednarski, M. (1997) *Constructivism and the Use of Performance Assessment in Science: A comparative study of beliefs among preservice and inservice teachers* (USA: The University of Connecticut).

Booth, P., Luckett, P. and Mladenovic, R. (1999) The Quality of Learning in Accounting Education: The impact of approaches to learning on academic performance, *Accounting Education: An international journal*, 8(4), pp. 277–300.

Burke, T. (1992) *An Investigation of Exam Type Preference, Performance Attribution and Exam Outcome*, Burke, T. Memphis (USA: Memphis State University).

Connelly, V., Dockrell, J. and Barnett, J. (2005) The Slow Handwriting of Undergraduate Students Constrains Overall Performance in Exam Essays, *Educational Psychology*, 25(1), pp. 99–110.

Cooper, B. (2004) The Enigma of the Chinese Learner, *Accounting Education*, 13(3), pp. 289–310.

Davidson, R. (2002) Relationship of Study Approach and Exam Performance, *Journal of Accounting Education*, 20(1), pp. 29–44.

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B. and & Sutton, A. (2005), Synthesising Qualitative and Quantitative Evidence: A review of possible methods, *Journal of Health Services Research and Policy*, 10(1), pp. 45–53.

Eley, M. (1992) Different Adoption of Study Approaches Within Individual Students, *Higher Education*, 23, pp. 231-254.

English, L., Luckett, P. and Mladenovic, R. (2004) Encouraging a Deep Approach to Learning Through Curriculum Design, *Accounting Education*, 13(4), pp. 461–488.

Ezzy, D. (2001). *Qualitative Analysis: practice and innovation* (Crows Nest, Sydney: Allen & Unwin).

Hassall, R. and Joyce, J. (2001) Approaches to Learning of Management Accounting Students, *Education + Training*, 43(3), pp. 145–152.

Hindi, N. and Miller, D. (2000) A Survey of Assessment Practices in Accounting Departments of Colleges and Universities, *Journal of Education for Business*, 75(5), pp. 286–290.

Lucas, U. and Mladenovic, R. (2004) Approaches to Learning in Accounting Education, *Accounting Education*, 13(4), pp. 399–407.

Lyons, J. (2006) An Exploration into Factors that Impact upon the Learning of Students from Non-Traditional Backgrounds, *Accounting Education: An international journal*, 15(3), pp. 325–334.

Maron, F. and Saljo, R. (1976) On qualitative Differences in Learning, *British Journal of Educational Psychology*, 46, pp. 4–11.

Maron, F. and Saljo, R. (1984) Approaches to Learning, *in Marton, F. et al.* (eds.), *The Experience of Learning* Edinburgh, (UK Scottish: Academic Press).

Neuman, W. (2000) *Social Research Methods: Qualitative and Quantitative Approaches*, 4th Edition. (Needham Heights: Allyn & Bacon).

Thomas, D.R. (2000) Qualitative data analysis: Using a general inductive approach [Internet]. University of Auckland. Available from www.auckland.ac.nz/mch/hrmas/qualdatanalysis.htm [Accessed 30 October 2006].

Paisey, C. and Paisey, N. (2003) Student Attendance in an Accounting Module – Reasons for non-attendance and the effect on academic performance at a Scottish University, *Accounting Education*, 13(Supplement 1), pp. 39–53.

Stout, D., Bordon, J., German, M. and Monahan, T. (2005) Designing and Implementing a Comprehensive Assessment Plan for a Graduate Accounting Programme, *Accounting Education*, 14(4), pp. 395–410.

Parker, L. and Roffey, B. (1997) Methodological Themes Back to the Drawing Board: Revisiting grounded theory and the everyday accountant's and manager's reality, *Accounting, Auditing & Accountability Journal*, 10(2), pp. 212–235.

Stiggins,R. (2002) Assessment Crisis: The absence of assessment for learning, *Phi Delta Kappan*, 83, pp. 758–765.

Tinnesz, C., Ahuna, K. and Kiener, K. (2006) Toward College Success: Internalizing active and dynamic strategies, *College Teaching*, 54(4), pp. 302–306.

Volet, S. and Chalmers, D. (1992) Investigation of Qualitative Difference in University Students' Learning Goals, Based on an Unfolding Model of Stage Development, *British Journal of Educational Psychology*, 62, pp. 17–34.