

# What Students Think and How They Really Perform in Chemistry

Ross Hudson

*Australian Council for Educational Research (ACER), Australia*

## Abstract

*This research was part of a larger study into student performance in senior chemistry with regard to question type and content. This paper examines student perceptions about question type and context and compares these perceptions to actual performance. How students perceive different types of questions and how it influences their self-belief and motivation were the focus of this study. Student responses to different styles or types of questions have been well researched over time. In this study Year 11 chemistry students were quizzed about their preferences to Multiple-Choice questions and Open Response question types and how the presence of each type was likely to influence their test performance. Student's perceptions were then correlated to their actual performance on sample chemistry tests. Students generally preferred MCQ questions and believed they were likely to perform better on these questions regardless of the topic. Test results did not always support this confidence. Suggestions for further research are also made.*

## 1. Introduction

This study initially developed from the researcher's observations, along with those of a number of teaching colleagues, that there were apparent differences between the observed student performances in the State University Entrance (VCE) examination in Chemistry. The assessment comprised of both multiple-and short-answer sections in two examinations. The effectiveness of student responses to each form of question, involving recall or application categories, was noted in anecdotal observations of the student performance in the examinations along with stated student preferences for the different examinations. Of importance to this paper was the issue of student perceptions with regard to self-belief about performance and how students believed that their performance would be influenced by the type and style of question asked. Students were asked in interviews about their preferences in chemistry tests with regard to question content and

type. Students were tested to measure actual performance and the two sets of data compared.

The research questions addressed in this study were:

1. Do students have a preference for the type of question style in terms of:
  - a. Multiple-choice or short-answer in general terms.
  - b. With respect to whether the question is assessing recall or application.
  - c. Gender
2. Does student performance correlate with student perceptions about chemistry testing?

## 2. Gender differences in performance

Whilst there have been numerous studies examining various aspects of the gender participation and performance by males and females in mathematics and the sciences, these reports have focused on either science or mathematics in general terms. For example, a study that examined Australia's participation in the Trends in International Mathematics and Science Study (TIMSS) in 2007 showed that Australian female students outperformed their male counterparts in Science at Years 4 and 8 decisively [26]. More pertinent to this research was a study that found females formed the larger proportion of all students studying all the sciences except physics, and, in terms of study score rankings, the females' outperformed males in all subjects except Chemistry [8]. A study of the GCSE examination in England over a three-year period demonstrated significant gender differences in both participation and performance. Generally, females performed better in most subjects except some of the sciences (chemistry and physics) and mathematics. Using English and Mathematics as benchmarks the researchers found that females significantly outperformed boys in obtaining A to C grades (54.6% of females obtaining these grades whereas only 41.5% of boys achieved A to C grades). In mathematics the

situation was reversed, 38.9% of males obtaining A to C grades and only 34.6 % of females [25]. Similar patterns were found in each individual science subject though the differences tended to be a little less dramatic due to the more selective entry to these [25].

A contributing reason for the lower participation rate and performance of female students was partly attributable to an apparent lesser interest by females in pursuing careers in the science field [3]. For example, in a study conducted in the USA researchers found that interest in science subjects waned along with career aspirations as female students progressed through secondary school [28].

Beller and Gafni [2] examined the performance of males and females in mathematics on the same examination over a three-year interval. The findings from this research demonstrated that the performances of females and males were not consistent. In the first study, correlations between performance and gender demonstrated that male students outperformed female students most significantly in multiple-choice questions but there was less difference on open-ended questions. This supported earlier findings that suggested that the open-ended questions tended to be more effectively done than multiple-choice by female students. The follow up study in 1991 demonstrated that in this instance the open-ended questions most significantly favoured male students' performances, essentially contradicting the first set of findings. What was consistent between the two studies was the finding that the harder the questions were (regardless of format) more likely to be done correctly by male students [2].

The difference between the genders in performance has been most studied in mathematics, where males have tended to outperform and also outnumber females in the higher levels of education. The number males studying physics and specialist mathematics are significantly larger than females. The numbers of males and females studying chemistry and biology have declined slightly over time as a percentage of the Year 12 cohort but have generally remained even in terms of males and females. To some extent this is probably due to growth in the numbers of students undertaking psychology and biology where the number of females is much greater than male numbers [6, 8].

## 2.1. Motivation

Motivation is a key factor in determining success in education. The level of students' commitment to learning is seen as a key factor in determining students' levels of motivation and success in any course of study. The factors that affect motivation, both positively and negatively, are therefore important [5]. Success is important to motivation. Students who believe that their success is in their own hands are described as internals whereas students who believe

their success in school is out of their control are described as externals. Success and motivation are more strongly linked to internal students than externals [5].

Difference in gender attributions to success or failure were demonstrated in a study that found male students, for example, were less likely to attribute success or failure to the level of help from their teacher, female students on the other hand were more likely to attribute failure to a lack of teacher support and a general belief that they were not likely to do well in the first place [16]. This attribution of success to gender was also demonstrated in a study of sociology students that found considerable unintended gender bias amongst the students. Students attributed success in their course more often to the instructor if that instructor was a male as compared to the instructor being female. Students were noted also to have a belief that the male instructors were more likely to be more highly qualified than the female instructors. These attitudes were even exhibited in the way the students referred to the instructors. Male instructors were referred to as the professors whereas the female instructors as teachers [19].

A study by Ryan and Patrick [22] found that motivation and engagement by students was strongly determined by the motivation and engagement that had been developed in prior years. The most important factors in determining motivation were found to be teacher support of the students and teacher encouragement of mutual respect and interaction. Teacher promotion of performance goals was seen to have a negative impact on motivation and engagement [18, 22]. The relationship between motivation and educational successes was examined in a review of studies conducted by Becker [1]. Differences in motivation and attitude towards science again favored male students but the difference was less than the difference in outcome achievement [1]. Further to this Becker was also able to support other findings [2, 12, 29] that have indicated that the subject areas most favoring males over females are the more traditional pure sciences of physics and chemistry. Jones and Kirk [14] explored the differences of males and females in their attitudes towards choosing sciences. Their study showed that when the issue came down to choice, females tended towards choosing the life sciences of biology and psychology rather than chemistry and physics because females were generally more interested studying a science they saw as a helping science, a people oriented science or a nurturing science. Subsequent studies [11, 29] have all demonstrated similar patterns in enrolments and participation in higher school sciences, however the differences are not as great as they once were.

### 3. Methodology

The main source of data for the student opinions was obtained from student interviews that were conducted with the sample of 59 students who had participated in the trial testing.

In this study between-method triangulation was employed as the primary source of triangulating the data sets [7, 9]. Triangulation can enhance the outcomes of research by demonstrating that several different mutually supportive sources of data are providing evidence for a particular conclusion and this gives much greater applicability and relevance to the findings [7]. In spite of the apparent good sense of triangulation it is not without its critics. Patton and Fielding and Fielding have suggested that having multiple data sources does not necessarily increase the likelihood of reliability or replication of the results and may even compromise the reaching of a conclusion as the researcher attempts to artificially find a conclusion that is supported by the different sources of data, thus compromising their objectivity [10, 20]. In spite of these critics it is difficult to dispute the inherent value of triangulation.

The particular aspect of triangulation that is relevant to this study is the mixed methodology approach of using both qualitative and quantitative paradigms that increase the scope for the data meeting triangulation requirements. In spite of the apparent advantages of the mixed methodology approach the case for it in the literature (as mentioned earlier) has, at times, been limited. Several reasons are offered for the preference of purely qualitative or quantitative research. By involving both qualitative and quantitative approaches it is believed that the length of the investigation will be extended beyond what a purely qualitative or quantitative enquiry may require and that by following two methods of research the clarity of purpose of the research may be compromised. The costs associated with mixed methods enquiries may also be unnecessarily prohibitive [17, 20].

#### 3.1. Interviews

Since the time of Piaget interviewing children to find their views and beliefs has been an integral part of developing understanding of how children learn and understand their learning environment [7]. Of particular concern in this research was gaining an understanding of students' views about the methods used to assess understanding in chemistry classrooms.

Interviews may be conducted in a number of different ways depending on the circumstances required. Largely the process chosen depends significantly on how predetermined the interview will be [12]. The conversational interview is informal in nature and the nature of the questions flows from the answers and responses of the both the interviewer and

interviewee. This approach is particularly common in ethnographic inquiries where it is part of the overall process of observation and fieldwork. This type of interview has the advantage of being able to match the interview to the particular respondent and the respondent's circumstances. However, because of the informal nature it is theoretically possible that no two interviews will proceed in the same way and therefore may produce a series of apparently unconnected and therefore difficult to summarize results [7]. Other more structured techniques included guided interview or semi-structured interview [7]; closed quantitative interview and standardized open-ended interview [7]. Of these techniques the open-ended structured interview is the most suitable for this particular research project. The targeted response areas are clearly defined thus suiting a formalized sequence of questions. Notwithstanding this, the interview process still allows some deviation from the expected course should student responses warrant it. With respect to reliability and validity, there is some contention as to whether both are equally achievable. The reliability of the interview is increased by greater control of the interview through a less open-ended response format (thus increasing the consistency of the responses) but it does so at the expense of validity [7].

Cross checking or validating the interview results is necessary to ensure the validity of the interview results, thereby contributing to the triangulation of the results obtained [7].

### 4. Summary of practical methodology

Whilst the main focus and data source for this paper was the student interview process, the other data source that served as a comparison for the student perceptions gained from the interviews was the trial testing. This data source was drawn from a sample of students from a local group of schools.

All willing student participants (59) were interviewed using an open ended standardized interview [15, 20]. Interviewees were each asked the same series of questions (refer to Appendix A for interview question details), which were recorded by the interviewer. The interviewees were asked to give their views about the following aspects of chemistry assessment

- Types of assessment they typically have experienced
- Preferences regarding that assessment
- Specific questions about multiple-choice and short-answer questions
- Preferences about these types of questions
- Perceived advantages and disadvantages about these question types

The participants, by responding to the questions, had the opportunity to reflect on their own practices and evaluate their own approaches to the learning environment that they were experiencing on a day-to-day basis. The interviewees were given the opportunity to reflect and clarify their answers and the interviewer sought clarification when this was warranted. To help facilitate the validity of the process, a peer review took place at a later date.

A few weeks after the initial interviews were conducted, a random sample of the original participants were re-interviewed and the interviewees given the opportunity to reflect upon and change/clarify their initial responses. If this process had revealed substantial differences between the initial and review interviews then further clarification was sought from the interviewees. Fortunately this was not a significant issue. This re-interview process enhanced the validity of the initial interview process in that it was possible to show that the responses initially made by the interviewees were of a considered nature, therefore enabling both credibility and trustworthiness in the interview process.

#### 4.1. Data sources-student groups

The methods of research were both quantitative (statistical analysis of the sample student tests) and qualitative (interpretation of the interview responses). The data was gathered at four local secondary colleges in a largely affluent middle class suburb. The applicability or transferability [7] of the results was therefore limited in the extent to which any findings could be applied to the wider school community.

#### 4.2. Student interviews

All willing student participants (59) were interviewed using an open ended standardized interview [15]. Interviewees were each asked the same series of questions which were recorded by the interviewer. A few weeks after the initial interviews were conducted, a random sample of the original participants were re-interviewed and the interviewees given the opportunity to reflect upon and change/clarify their initial responses. This re-interview process enhanced the validity of the initial interview process therefore enabling both credibility and trustworthiness in the interview process. The interview schedule is described in Appendix 1.

#### 4.3. Sample testing

The researcher constructed short tests that asked essentially the same question but in both multiple-choice and short-answer form. The testing is a crucial part of this study as it seeks to examine an area of research that has not been extensively studied. Whilst some researchers [24] have made some assessments

and conclusions about the advantages and disadvantages of each type of question, there have been few studies directed at examining the effectiveness of each type of question in how well they assess student understanding in chemistry [6]. The following test structure was adopted.

- Each student participated in a series of short (approximately 10 - 15 minutes) tests.
- At the conclusion of the tests the interview process took place.

### 5. Results

Several sources provided data from which the results have been drawn. The main data source were the student interviews. The second source of data were the student trial tests, however this is not covered in any great detail in this paper [13]. The student interviews were coded to allow a statistical evaluations of student opinions. Summaries of student responses and some actual responses are provided along with the post member checking.

#### 5.1. Results from interviews

Conducting the interviews provided a first person opportunity to determine whether or not the students' preferences on question types matched the actual performance on the tests. To gain a larger number of results a number of students were interviewed by their teachers who recorded their responses; this sometimes took the form of the students recording their opinions rather than the teacher recording the responses.

The results from the student responses were inconclusive in terms of determining any particular favoritism for either short-answer or multiple-choice questions. Responses to Question 3 (Does the content of the question influence your decision?) and Question 8 (If you had the choice would you prefer chemistry tests to be multiple-choice only or short-answer only or a combination of both?), that were the most likely to give some information to this issue, the results were quite uniform.

#### 5.2. Coding responses

To aid in this interpretation, some student responses were coded according to whether they were positive or negative with respect to each question. This method enabled an assessment of the general trends of each gender and their preferences for particular types of questions. Whilst students responded either briefly or with some elaboration it was possible to deduce that they either preferred (or not) one type of question over the other. Students who generally favored multiple-choice questions were coded 1 and students who preferred short-answer were coded 2. This process was

applied to question 3 from the interview schedule. Similar coding was then applied to preferences for application content and recall content questions.

**Table 1. Responses to Research Question 1 (n = 59)**

Question preference combination	Males	Females	Totals
Recall: as MC and Application: as MC	6	7	13
Recall: as MC and Application: as SA	19	22	41
Recall: as SA and Application: as MC	1	0	1
Recall: as SA and Application: as SA	1	3	4
<b>Totals</b>	27	29	59

The majority of students indicated a preference for multiple-choice for responding to recall type content and short-answer for responding to application questions. Responses from 59 students were obtained with the distribution shown in Table 1.

Of the students involved, 70% preferred recall questions to be multiple-choice and the application questions to be short-answer. There were no discernible differences between the responses of the males and the females. Further description and analysis of the data for Question 3 follows later. The evenness of the responses shown in Table 1 was a little surprising, as it might have been expected that there would have been some gender differences evident in the responses. The important observation from this limited sample is that the number of students who preferred to answer recall questions as multiple-choice questions was very strong. Only one student out of the 59 showed a preference for recall short-answer questions. The number wanting only recall questions was also very small, 3 out of the 59. This compares with the number of students who preferred multiple-choice questions only, which was 11 out of 59, nearly three times the number. Responses to Question 8 provided little useful additional information with almost all students indicating a preference for the status quo of a combination of multiple-choice and short-answer. This response possibly reflects the experiences they already have in chemistry assessment rather than the intended purpose of the question. The actual intention of Question 8 was to determine the students' preference for one type of question over the other if such a choice had to be taken.

To rectify this issue the participating teachers were asked to put the following question to their classes: "If a test was to be either all multiple-choice or all short-answer which would you generally prefer to do

regardless of the test topic? That is, which type of question do you like best?" The entire class was asked this question and the teachers forwarded the simple poll results to the researcher. Table 2 shows the outcome of this poll. It included virtually all the students who took part in the trial tests depending on the individual attendance on that day.

Chi-squared analysis of this data showed a value of about 4 with a significance of  $p < 0.05$ . The obvious differences between the data for males and females support results found through the observations of other researchers [3], though not strongly.

**Table 2. Preferences of Students for Question Type (n=100)**

Group	Percentage who favored multiple-choice	Percentage who favored short-answer
<b>Males</b>	66	34
<b>Females</b>	54	46

These results (Table 2) confirm the results demonstrated in Table 1. A strong preference was demonstrated for multiple-choice questions with males being the stronger supporters of the multiple-choice format.

### 5.3. Sample test results

The performance summary of the sample tests is shown in Table 3. Detailed results of the sample testing can be found in the paper by Hudson and Treagust [13].

**Table 3. Gender differences on the trial chemistry tests (means)**

Groups	Count	Sum	Mean	s.d.
<b>Male %</b>	92	7364.2	78.3	15.8
<b>Female %</b>	90	6174.4	68.6	17.3

This table clearly demonstrates that the male students scored more highly than their female counterparts and this difference was significant ( $F(1,181) = 15.9; p < 0.01$ ). Of interest to this study was whether the attitudes of students to the various types of questions were reflected in the subsequent performance.

#### 5.4. Student responses by interview question

The first two questions aimed to clarify the experiences of the students with respect to the styles of questions they had experienced. The students' responses were quite uniform. All students indicated that they had commonly experienced tests that contained both multiple-choice and short-answer questions. The proportion of multiple-choice to short-answer questions in the tests experienced by the students was reported at between 30-50% multiple-choice. The most common response indicated a roughly one-third multiple-choice proportion in the tests. This was to be expected as all the students were taught by experienced teachers who were more than aware of the need to adequately prepare their students for the examination regime they would encounter in the following year which had this distribution of item types.

Question 3 elicited a wide variety of responses. Students were asked if the type of question in terms of content influenced their choice of question style. The majority of students, around 90%, interviewed preferred content or recall questions to be multiple-choice. The most common reason given seemed to be based on the idea that multiple-choice gave the students an opportunity to be prompted by the options. A number of students indicated the possibility of eliminating incorrect options as an advantage with this type of question. Students who preferred short-answer for recall type questions did so mainly from the perspective of the confusion that sometimes occurred if the answer they thought was going to be the answer was not in the options. Typical responses representative of these views were: Female student B8: School-A: "multiple-choice as you have options to give you an idea but in short-answer you don't." Male student B18: School-D "Multiple-choice – if the question cannot be instantly answered, looking at available answers may remind you of the correct choice".

Question 3b, (If the question is an application question [e.g. Calculate the number of mole of something] is it better for the question to be multiple-choice or short answer? Please explain your response) focused on application or calculation type questions. The views of the students were considerably different to the attitude towards multiple-choice questions. The students were generally of the view that short-answer responses gave them the best opportunity to perform well. There were, however, a number of students who preferred multiple-choice for both types of question. The percentage of students who opted for the short-answer application questions was about 70%. Typically the students indicated that the short-answer approach allowed them the opportunity to gain marks for correct working even if they could not provide the final correct answer. Typical examples of these responses were: Female student B4: School-A: "short-answer, because I

can get marks for showing working if I were to get the answer wrong"-There were a number of responses similar to this, Male student A9: School-D "short-answer – writing down my process of working out helps in checking my answers – less chance of making mistakes".

A view held by the small number of students who preferred multiple-choice for application questions focused on the advantage of having the possible answer presented in the options and that having worked out an answer that was not an option meant that the students knew they had made an error and could then have a second chance at trying to calculate an answer that was presented in the options. A typical response was: Male student B2: School-C "multiple-choice because that way when you work out the answer, it has to be one of the answers given, so that way you can work out if you have made a mistake or not." Of some significance was the observation that of the 59 interviewed students only one responded with short-answer for recall questions coupled with multiple-choice application questions.

Questions 5, 6, 7 and 8 (see Appendix 1) from the interview schedule appear to ask the same questions. This was apparent on one occasion where a student exclaimed that I was trying to trick him by asking the same question again. This strategy was employed with the knowledge that some students would respond in a nearly identical manner to questions 5 and 8 and, 6 and 7. However, the method was designed to ensure that students did have the opportunity to consider each type of question in the focus of their thoughts from both a positive and negative viewpoint.

**Table 4. Advantages and Disadvantages of Multiple-choice questions**

Advantages	Disadvantages
Possible to eliminate or narrow down response.	Can be confusing or tricky with good options
Can work backwards from the answers	Can appear to have more than one correct answer
Prompting of answer from given options.	Penalized if you make a silly mistake or small error
Checking answer against options	Can't show your working out
Can guess the answer if it can't be worked out	Questions seemed designed to trick students
Usually easier questions	
Quicker to do	

This tactic appears to have paid dividends with students giving new information in subsequent questions that was not in the previous questions. For example, whilst question 5 seeks advantages of multiple-choice over short-answer, question 8 seeks the opposite, that is, the disadvantages of short-answer over multiple-choice. (Questions 6 and 7 reinforce the students' ability to express their ideas about the two types of questions).

By asking the questions, e.g. question 5 and question 8, about the same idea but from the reverse logic, that is one from a positive and the other a negative mindset, the questions encouraged the students to consider each perspective before answering. Nevertheless, this was not always the case with a number of students giving responses in the style of "I can't add to what I said in the last question". The student responses from questions 5 to 8 are summarized in Tables 4 and 5.

**Table 5. Advantages and Disadvantages of Short-answer questions**

Advantages	Disadvantages
Proves that the student really knows the work	Don't have an opportunity to check answer against any given options.
Can get marks for partially correct answers	If you don't know what to do you can't attempt the question at all.
Worth more marks	Usually worth a lot of marks each
Can see where you went wrong	Usually harder questions
Teacher bias can influence marking	No prompts from the question like multiple-choice

As could be expected from different students, opposing viewpoints were often expressed. For example, some students thought that the larger number of marks attributed to a short-answer question was an advantage whilst others thought it was a disadvantage.

Some of the comments from students that illustrated the results in Tables 4 and 5 were:

### 5.5. Example student responses

**5.5.1. Multiple-choice advantage.** Female student B4: School-A *"Sometimes because the answer is on the page you just need to select it"*

Female student D10: School-A: *"Multiple-choice only. It is much easier and reassuring. If you don't come up with the right answer it forces you to look*

*over your working and then try to work out where you have gone wrong. This cannot happen in short-answer questions, as you can't be sure you are right"*.

Female student B4: School-A *"Sometimes because the answer is on the page you just need to select it"*

**5.5.2. Multiple-choice disadvantage.** Female student B4: School-A *"You don't get any marks if you do the working out method partially correct but get the wrong answer."*

**5.5.3. Short-answer advantage.** Male Student A3: School-D *"Short-answer, helps students show their working out step by step for future exams, and rewards points for each step, rather than lose all of the marks. Also once you finished the test, you can look at what went wrong in your steps thus correcting the mistakes. Also you can correct your method of approach. Also can help your vocabulary by writing."*

**5.5.4. Short-answer disadvantage.** Male student A17: School-D *"Yes, if you don't know how to do the question all you can do is leave it blank"*

Male student A4: School-D Q6: *"If you know your stuff there are more marks available and generally short-answer questions are more straight forward."* Q7: *"Often it is easy to drop one mark (3/4 or 2/3) on short-answer questions with a minor mistake"*

In the last two quotes from the same student the advantage of asking the double question is indicated as the student was able to reinforce his opinion as given in the first response with a second supporting argument.

The general opinion gained from the responses was that student views of the advantages and disadvantages of multiple-choice and short-answer were somewhat mixed but in keeping with the observations of other researchers [4, 21, 24]. Generally, the male students were more positive than were the female students about multiple-choice and females more positive about short-answer.

An interesting observation is that the students who offered guessing as an advantage to multiple-choice questions were amongst the lower scoring students on the trial tests. The students who appreciated the advantage of being able to get marks for showing correct working tended to be amongst the higher scoring students in the trial tests. For example two students who did not perform particularly well on the tests made the following similar observations about multiple-choice questions.

Male student C22: School-D: *"Helps you work through process of getting the right answer, and a chance of getting it right if you're clueless"*. This student scored an average of 65.3% and was ranked 127<sup>th</sup> amongst all the students.

Female student A23: School-B: *"If you are unsure of the answer you have a 1 in 4 chance even if you guess and you have something to work to"*. This

student scored an average of 48.3% and was ranked 166<sup>th</sup> amongst all the students

This contrasted to the better scoring students.

Male student A3: School-D: For Q5 multiple-choice disadvantages, *“It allows guessing most of the time rather than attempting having a go. Also some students can guess and get it right. It doesn’t show their strengths or weaknesses. Also, skipping steps rather than showing how you got there. Also, if you were taught the wrong method of working, it doesn’t let the teacher know or you know. Thus not allowing you to correct an error.”*

and,

Q6, short-answer advantages, *“helps students show their working out step by step for future exams, and rewards points for each step, rather than lose all of the marks. Also once you finished the test, you can look at what went wrong in your steps thus correcting the mistakes. Also you can correct your method of approach. Also can help your vocabulary by writing.”* This student scored an average of 93.1% and was ranked 18<sup>th</sup> amongst all the students.

### 5.6. Post interview member checking

Several weeks after conducting the interviews a sample of six students from the original interview groups were asked to reflect upon the transcripts of their interviews to see if their viewpoints had changed. This process enhanced the triangulation validity of the data [7]. In each case the students indicated an affirmation of what they had originally said. One student said that she may have answered some of the questions a little differently having experienced the interview but it was unlikely to actually alter the way she viewed chemistry questions.

## 6. Implications for student motivation

The relationship between expectation and actual performance in chemistry is likely to impact on student motivation. It has been well recorded that male students both prefer the “harder” sciences like chemistry and have generally been shown to outperform female students [2, 8, 12, 27]. These findings are unlikely to encourage female participation in these subjects. The findings in this research generally support the findings that male students have a clear domination in the awarding of the higher grades. Were this information to be taken at face value then motivation of female students would be disadvantaged. However, it appears that there may be an explanation offered from the analysis of the gender performance highlighted through Research Question 3. Initial analysis showed that the males outperformed the females as may have been expected in terms of previous research. When the Rasch gender differential

analysis [23] was included it showed that if latent student abilities are allowed for then there was little between the performance of the male students and female students. That is, male students and female students of equal ability perform very similarly in chemistry. This suggests that the male students choosing chemistry include a greater proportion of high ability males compared to the proportion of high ability females choosing chemistry. Whilst this proposition will need more testing there is some support when comparing the VCE Biology results to the VCE Chemistry results grade distributions (see Figures 4.5 and 4.6) where the females have achieved a greater proportion of the higher grades than the males in biology, implying that more high ability females choose biology than do high ability males.

## 7. Conclusions

A sample (59) of the 192 students was interviewed. With respect to the preferred type of question the students indicated a significant ( $p < 0.05$ ) preference for multiple-choice (59%) over short-answer (Tables 1 and 2). The results support those previously found in the literature [24]. Generally the strongest features reported by students in favor of multiple-choice were that the options offered prompted them towards the correct answers, allowed the possibility of cross checking results from calculations and finally offered the possibility of at least being able to make an informed guess rather than leaving a blank space. Students who favored short-answer questions most often offered the response that it gave them the opportunity to gain partial credit for incomplete responses and also the opportunity to “show what they knew”. Overall the results of the interviews were indicative rather than conclusive but generally supported the findings of previous research.

## 8. Future Work

The results of this research provide the framework for further analysis, particularly into the performance of males and females in different VCE examinations. The most interesting outcome was the possibility that the differences in performance of the male students compared to female students may be due more to the abilities of the students actually choosing the subject than it has to do with either the nature of the assessment (test structure issues) or any latent ability advantage that male students have over female students. That notwithstanding, further deeper analysis of the actual VCE data may provide some useful insights into this issue.

The limited size of the trial sample and the low stakes nature of the trial tests placed some limitations on the transferability of the results. However, sufficient information was gained to suggest the need for a wider scale test program, which analyzed



performance on matched multiple-choice and short-answer items to clearly determine whether the item type influences success or demonstration of chemistry understanding. Whilst there were substantial similarities between the VCE examination analysis and the trial test analysis the variations mentioned on the previous page suggest that a more detailed analysis and larger trials may give more certainty to the observation that students perform better on recall questions. Such a test program would also help provide insights into differing performances of the male and female students. This line of examination may lead to findings that will ultimately suggest an examination structure that more evenly assesses the performances of male and female students.

## 9. References

- [1] Becker, W. E., and Johnson, C. (1999). The relationship between multiple-choice and essay response questions in assessing economics understanding. *Economic Record*, 75(231), 348-357.
- [2] Beller, M., and Gafni, N. (2000). Can item format (multiple choice vs. open-ended) account for gender differences in mathematics achievement? *Sex Roles*, 42(1-2), 1-21.
- [3] Boli, J., Allen, M. L., and Payne, A. (1985). High-ability women and men in undergraduate mathematics and chemistry courses. *American Educational Research Journal*, 22(4), 605-626.
- [4] Bridgeman, B. (1992). A Comparison of quantitative questions in open-ended and multiple-choice formats. *Journal of Educational Measurement*, 29(3), 253-271.
- [5] Buehl, M. M., and Alexander, P. A. (2005). Motivation and performance differences in students' domain-specific epistemological belief profiles. *American Educational Research Journal*, 42(4), 697-726.
- [6] Chan, N., and Kennedy, P. E. (2002). Are multiple choice exams easier for economics students? A comparison of multiple choice and equivalent constructed response exam questions. *Southern Economic Journal*, 68(4), 957-971.
- [7] Cohen, L., Manion, L., and Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge Falmer.
- [8] Cox, J., Leder, J., and Forgasz, H. (2004). Victorian Certificate of Education: Mathematics, science and gender. *Australian Journal of Education*, 48(1), 27-46.
- [9] Denzin, N., & Lincoln, Y. S. (2005). *The SAGE handbook of qualitative research*. Thousand Oaks, CA: SAGE.
- [10] Fielding, N. G., & Fielding, J. L. (1986). *Linking data*. Beverly Hills: SAGE.
- [11] Francis, B., Hutchings, M., Archer, L., and Melling, L. (2003). Subject choice and occupational aspirations among pupils at girls' schools. *Pedagogy, Culture and Society*, 11(3), 425-442.
- [12] Hamilton, L. S. (1998). Gender differences on high school science achievement tests: Do format and content matter? *Educational Evaluation and Policy Analysis*, 20(3), 179-195.
- [13] Hudson, R.D. and Treagust, D.F. (2013): Which form of assessment provides the best information about student performance in chemistry examinations? *Research in Science and Technological Education*, 31(1), 49-65.
- [14] Jones, A. T. and Kirk, C. M. (1990). Gender differences in students' interests in applications of school physics. *Physics Education*, 25, 308-313.
- [15] Kvale, S. (1996). *Interviews*. London: Sage publications.
- [16] Lloyd, J. E. V., Walsh, J., & Yailagh, M. S. (2005). Sex Differences in Performance Attributions, Self-Efficacy, and Achievement in Mathematics: If I'm so Smart, Why Don't I Know It? *Canadian Journal of Education / Revue canadienne de l'education*, 28(3), 384-408.
- [17] Maor, D., & Fraser, B. (1996). Use of classroom environment perceptions in evaluating inquiry based computer assisted learning. *International Journal of Science Education*, 18(4), 401-421.
- [18] Middleton, J. A., and Spanias, P. A. (2005). Motivation for achievement in mathematics: Findings, generalizations, and criticisms of the research. *Journal for Research in Mathematics Education*, 30(1), 65-88.
- [19] Miller, J., & Chamberlin, M. (2000). Women Are Teachers, Men Are Professors: A Study of Student Perceptions. *Teaching Sociology*, 28(4), 283-298.
- [20] Patton, M. (1990). *Qualitative evaluation and research methods*. Newbury Park: SAGE.
- [21] Pressley, M., Ghatala, E. S., Woloshyn, V., and Pirie, J. (1990). Sometimes adults miss ideas and do not realize it: Confidence in responses to short answer and multiple choice comprehension questions. *Reading Research Quarterly*, 25(3), 232-249.
- [22] Ryan, A. M., and Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, 38(2), 437-460.
- [23] RUMM\_laboratory\_P/L. (2009). RUMM2030.
- [24] Simkin, M., and Kuechler, W. (2005). Multiple-choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education*, 3(1), 73-98.

[25] Stobart, G., Elwood, J., & Quinlan, M. (1992). Gender bias in examinations: How equal are the opportunities? *British Educational Research Journal*, 18(3), 261-276

[26] Thomson, S., Wernert, N., Underwood, C., and Nicholas, M. (2007). *Highlights from TIMSS 2007 from Australia's perspective*. June 4th, 2009, from [http://www.acer.edu.au/documents/TIMSS\\_2007-AustraliaHighlights.pdf](http://www.acer.edu.au/documents/TIMSS_2007-AustraliaHighlights.pdf)

[27] VCAA. (2009). *Grade distribution report for Chemistry 2008*. Retrieved 28 February, 2010, from [http://www.vcaa.vic.edu.au/vcaa/vce/statistics/2008/section3/vce\\_chemistry\\_ga08.pdf](http://www.vcaa.vic.edu.au/vcaa/vce/statistics/2008/section3/vce_chemistry_ga08.pdf).

[28] Watson, C., Quatman, T., and Edler, E. (2002). Career aspirations of adolescent girls: Effects of achievement level, grade and the single sex environment. *Sex Roles: A Journal of Research*, 46(9), 857-871.

[29] Wolleat, P. L., Pedro, J. D., Becker, A. D., and Fennema, E. (1980). Sex differences in high school students' causal attributions of performance in mathematics. *Journal for Research in Mathematics Education*, 11( 5), 356-366.

## 10. Appendix 1

1. a. What type of questions have you experienced in your chemistry tests this year, multiple-choice, short-answer or both?  
  
b. If yes have the test been all MC all SA or a mixture of both in the one test?
2. If you have responded to both what proportion of the test is MC and what proportion is SA? E.g. 50% MC 30% SA etc.
3. Does the content of the question influence your decision?
  - a. For example if the question tests recall (e.g. name a strong acid) then is it better for the question to be MC or SA? Please explain your response.
  - b. If the question is an application question (e.g. Calculate the number of mole of something) is it better for the question to be MC or SA? Please explain your response.
4. Do MC have any advantages compared to SA questions? Please explain your response.
5. Do MC questions have any disadvantages compared to SA questions? Please explain your response.
6. Do SA have any advantages compared to MC questions? Please explain your response.
7. Do SA questions have any disadvantages compared to MC questions? Please explain your response.
8. If you had the choice would you prefer chemistry tests to be multiple-choice only or short-answer only or a combination of both? Please explain your response.