International conference
The Future of Education
2nd Edition

Gender Differences in Chemistry Performance: What is the Relationship between Gender, Question Type and Question Content

Ross D. Hudson
Australian Council for Educational Research (Australia)
ross.hudson@acer.edu.au

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 students' performances on the State University Entrance Examination for Chemistry (the Victorian Certificate of Education examinations; VCE). This examination is very high stakes as it forms the basis of selecting students for entry into tertiary courses. Students were performing differently on the two semester examinations. Female students achieved better grades in the second semester examination (November) whereas male students were much more successful in the first semester examination (June). This difference was demonstrated in the grade distribution data released by the examining authority see Figure 1 where the distribution of the A+ grade is significantly skewed in favour of male students in both semesters though less so in semester 2.

Why should this be occurring? Were any observed differences actually significant? A number of factors were relevant in considering this problem. Apart from the gender of the students the two other important factors were likely to influence the performance of the students. These were: type of question, short-answer or multiple-choice, and the content type of the question, recall or application.

The following research questions were addressed in the study:

1. Do multiple-choice or short-answer questions more positively emphasise student understanding?
2. Do students perform more effectively on recall type questions or on application questions?
3. Does students' gender and ability influence performance in chemistry?

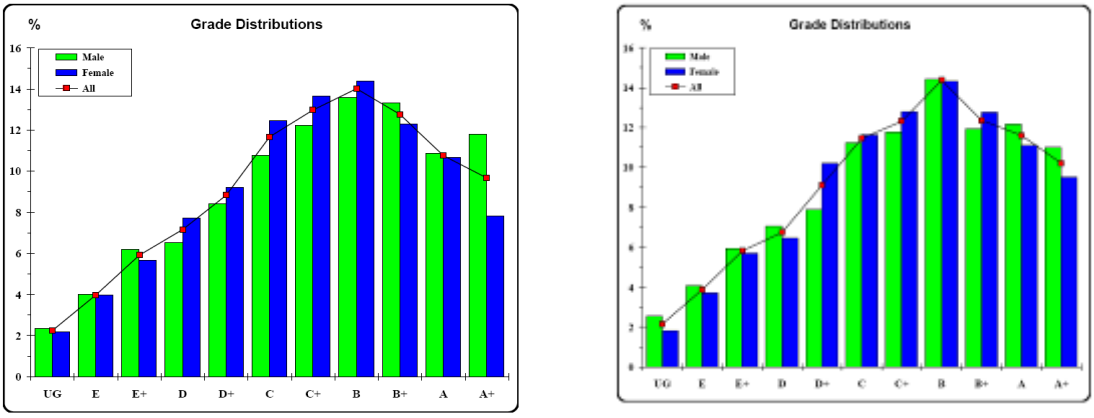

Figure 1: Grade distributions for the 2005 VCE Chemistry Examination1and 2 [17]

2. Literature Perspectives

2.1 Question type and content

Multiple-choice questions generally have less scope and complexity than short-answer questions and therefore are likely to be less difficult. This suggestion is supported in the literature [5] with the authors concluding that open ended type questions are superior in assessing student understanding of concepts because the solution methodology employed by the students in arriving at their answer can be examined whereas multiple-choice question answers give no indication of how students arrived at their answer [5]. Student perceptions about performance were also important. Students were generally more confident of their answers being correct when answering multiple-choice questions than short-answer questions regardless of the fact that the actual performance on the two types of questions was often almost the same. Students felt they would (or were going to) perform better on multiple-choice tests [14].

2.2 Gender Differences in Performance

The role of gender in chemistry performance, and other subject areas in general, has precipitated a variety of studies over time and will no doubt continue to do so. A study by Boli, Allen and Payne [4] explored the reasons behind the differences that were observed between the genders in undergraduate chemistry and mathematics courses. Their exploration sought reasons behind why male students were tending to outperform the female cohort, resulting in the suggestion that differences in mathematical ability were a very important consideration. The most important factor, through an analysis of previous studies, was that the male students' natural self-confidence and belief in the importance and need for mathematics had a positive influence on male performance. These findings with regard to mathematics can be fairly evenly transferred to the natural sciences [4]. Other than mathematics, there appeared to be no directly gender-related reasons for the male students outperforming the female students yet the evidence showed that this was the case. The study also showed that females were less likely to choose mathematics and science courses at the undergraduate level, often because of lesser preparation at the prior levels of schooling [3], [7], [16].

Many studies have agreed with the observation that male students usually outperform female students in assessments particularly in the areas of mathematics and science. The analysis of a number of large assessments has demonstrated that male students generally performed better than did female students [2], [11], [13].

3. Methodology

3.1 Data Sources

The evidence was provided by the test results conducted by the researcher with 192 year 11 students from four secondary colleges. The participation rate by the students was very good with few cases detected in the subsequent Rasch analysis suggesting instances of students not attempting the tests in a purposeful manner. The test results were analysed using both ANOVA and Rasch analysis. The Rasch analysis determined the validity and unidimensionality of the trial tests and the ANOVA analysis provided summative data about the tests.

3.2 Sample testing Procedure

The researcher constructed short tests that asked essentially the same question but in both multiple-choice and short-answer form. That is, pairs of questions were constructed so that the content loading of each was similar but one presented as a multiple-choice question and the other as a short answer question. The student's teachers participated in a process of checking that the question pairs were as much as possible of equal content loading. The equal loading of each question was later demonstrated by the excellent correlation found during the analysis of the trial tests. 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 [1], [15], have made some assessments and conclusions about the advantages and disadvantages of each type of question there has not been any study directed at examining the effectiveness of each type of question in how well they assess student understanding in chemistry, only a limited number have explored student performance where the questions are very similar in content but framed in the two question types [8].

3.3 Results

The results from the trial test analysis involved ANOVA and Rasch analysis based on the classification of the questions. An initial correlation analysis (plotting the item difficulty locations [logits]) importantly showed that the premise of the trial tests, that the tests would be testing the same skills but in different formats was supported (see Figure 2). This graph essentially showed that the easiest multiple-choice question was also the easiest short answer question and so on. The Pearson's r-correlation value of 0.72 supports this interpretation.

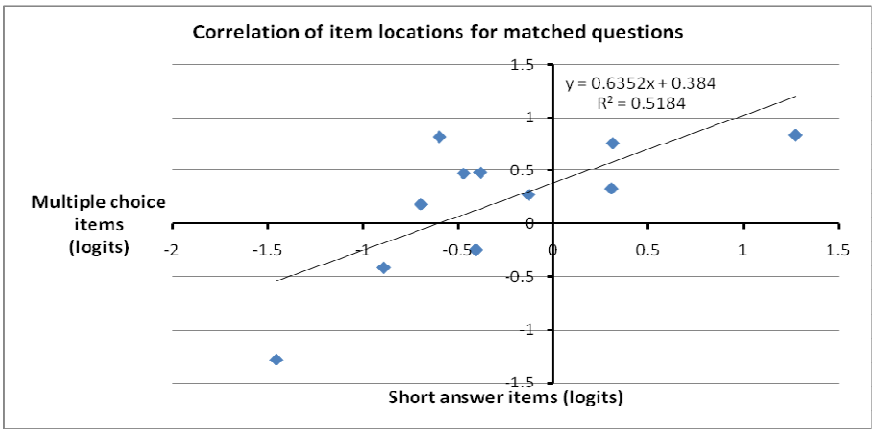

Figure 2: Correlation between the difficulties of the multiple choice items and short-answer items

The initial analysis of the trial papers was conducted using ANOVA. The result of this analysis is shown in Table 1. Summarising the trial test analysis, the differences in performance on different question types does occur even when the questions are matched for difficulty.

Table 1: ANOVA analysis of question performance by question type and classification Chemistry trial tests

			ANOVA results (N = 368, df = 1)		
Comparison variables	**Mean (%)**	**Std.deviation.**	**Sum of Squares**	**F**	**p (sig)**
Multiple-choice	71.0	20.8	333.84	0.72	0.39
Short-answer	69.1	22.3			
Application	74.8	25.3	6450.3	11.10	0.001
Recall	66.4	22.8			

Question Type
The comparison of means in performance between multiple choice and short answer questions overall was small (mean-multiple-choice=71.0, standard deviation = 20.8 and mean-short-answer = 69.1, standard deviation = 22.3). The difference was not statistically significant $F(1,369) = 0.72$; $p > 0.05$. It would thus appear that the multiple-choice questions do not offer any particular advantage to students over short-answer.

Question Content
However, the content of the questions does raise issues. The content of the question does affect the outcome of the response regardless of whether the question is presented as a multiple-choice or as short-answer. As presented in Table 1, students' performance on application questions (mean = 74.8 and standard deviation = 25.3) was surprisingly stronger than on recall questions (66.4 /22.8). The ANOVA results show a statistically significant difference $F(1,369) = 11.1$; $p < 0.001$ in performance on recall questions compared to application questions.

Gender Differences
Analysis showed that males performed significantly better in the examinations than did females, particularly at the A+ end of the score range (see Figure 1). An initial analysis of the student performance in the trial tests showed that the males again outperformed the females on the trial tests (Table 2).
The ANOVA results of the raw scores show that the males scored more highly (78.3%) than the females (68.9%) and that difference was significant ($F(1,181) = 15.9$; $p < 0.01$). The distribution of scores in the trial tests (Figure 3) closely matches that of the VCE examination distributions (Figure 1).

Table 2: Gender differences on the trial chemistry tests (means)

Groups	Count	Sum	Average	s.d.
Male %	92	7364.2	78.3	15.8
Female %	90	6174.4	68.6	17.3

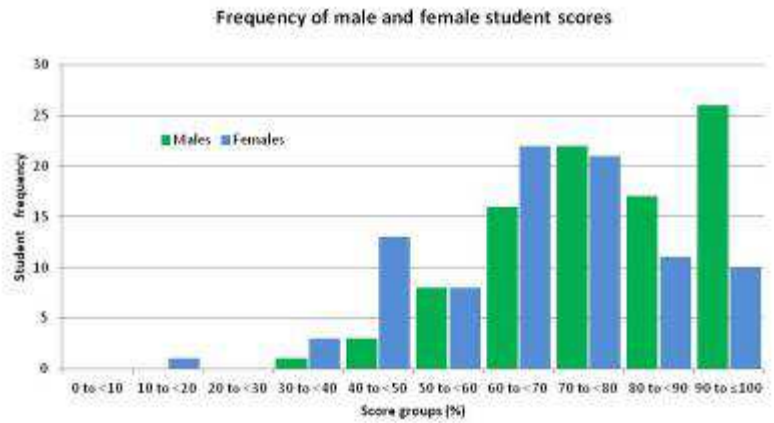

Figure 3: Distribution of male and female scores in the Chemistry trial tests

Whilst the distribution is less uniform than the VCE distribution (the smaller sample size in the trial tests needs would partly account for this) the significant aspects are the negative skewness and the peak in the performance of the males compared to females in the 90 to 100% score range. This distribution demonstrates that the trial tests are useful in that the tests appear to have reasonably mirrored the distribution of students' abilities even though the sample size was relatively small compared to that of the VCE examinations.
The trial tests, however, allowed a finer examination of student performance on the various category types of question asked that was not possible on the VCE examination and was able to shed some light on where the males were outperforming the female students. The following example analysis attempts to identify where, within the test structure, the males performed differently to the females.

Gender and Performance: Multiple-choice and gender
The ANOVA test results show that there is a significant statistical significance ($F(1,182) = 13.65$; $p < 0.001$) in performance on multiple-choice questions between the males and females. Male performance was better than female performance on the multiple-choice questions as supported by the means (males) = 76.4, standard deviation = 19.8 compared to mean (females) = 65.5/20.4. The Rasch graphical analysis difference shows, however, that the difference is less marked when the scores are adjusted for latent student ability as measured by RUMM2030.

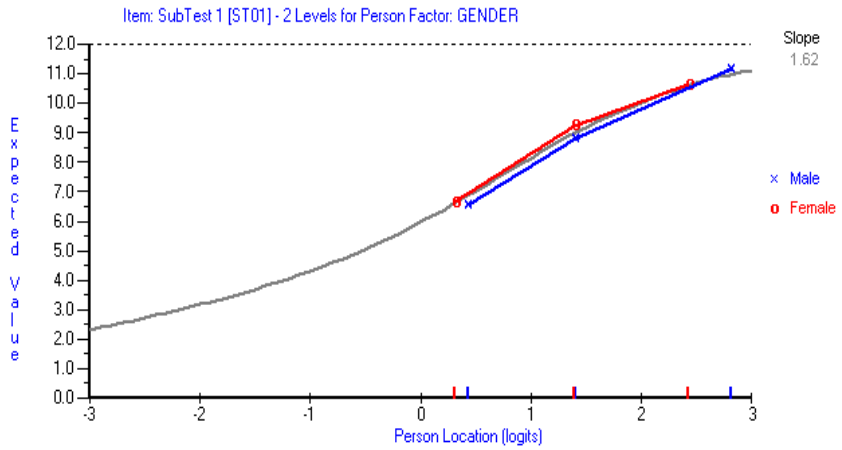

Figure 4: Multiple-choice questions showing gender difference against expected score and student ability

The graph (Figure 4) shows that the male students, within their ability ranges, find the multiple-choice questions slightly more difficult than do the female students. In other words, in spite of the higher raw scores obtained by the males they find them somewhat more challenging than expected. Put simply, the female students found multiple-choice questions easier than did the male students for students of equal ability.
A significant inference is that it would appear that there may be more high ability male students than female students in the trial test sample which could also account for the better performance of the males at the upper end of the performance scales in the VCE examinations (see Figure 1).

Comparisons between short answer questions, recall questions and application questions with gender all revealed similar finding to those demonstrated in the comparison between multiple-choice and gender above [10].

4. Conclusion

The trial test analysis showed similar patterns to the VCE examinations in terms of grade distribution. However, when gender difference analysis using RUMM2030 was performed the difference in performance was smaller. The Rasch analysis compared student performance according to gender but allowed for student ability as measured by the Rasch analysis. When this result was taken into account the performance by the students differed little by gender even though the mean scores were significantly in favour of male performance. In some instances, female performance (allowing for ability) was better than that of the male students even though the mean scores suggested otherwise. For example, in comparing multiple-choice question performance, the mean for males = 76.4 (s.d. = 19.8) was higher than that of the females, mean = 65.5 (s.d. = 20.4) (see Table 1). However, when analysed using Rasch, the gender difference analysis showed that females generally performed slightly better on multiple-choice than did males once student ability was allowed for (Figure 4).

Another important observation is implied in the trial test results. The initial findings of the trial test analysis support the observations of other researchers [2], [6], [9], [12], in that male students achieve higher scores than do female students (Table 1). When the performance allows for student ability however, the differences are quite small, suggesting that perhaps the ability of the male students is (at the top end) greater than that of the female students taking chemistry. It may well be that a greater proportion of high performing males are choosing chemistry than are high performing females. The reasonable assumption is that significant numbers of high achieving females are choosing to do other subjects and not chemistry. This would account for the skewed appearance of the results. This proposition certainly warrants further investigation.

Overall, however, the clear outcome from this section of the analysis is that male students achieve higher grades in chemistry than do female students, particularly at the top end of the grade scale.

The findings of this research show that there is not any underlying advantage of multiple-choice to short answer questions. Performance on both is similar when the question content is similar. Performance on multiple-choice questions was only slightly better than on short-answer even when allowances were made for the difficulty of the items and abilities of the students. Further research with a larger, more extensive cohort is suggested to clarify these findings.

References

[1] Anderson, G. (1998). Fundamentals of educational research (2nd. ed.). Bristol: PA:Falmer Press.
[2] Beller, M., & 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] Blickenstaff, J. C. (2005). Women and science careers: leaky pipeline or gender filter? Gender and Education, 17(4), 369-386.
[4] Boli, J., Allen, M. L., & Payne, A. (1985). High-ability women and men in undergraduate mathematics and chemistry courses. American Educational Research Journal, 22(4), 605-626.
[5] Bridgeman, B. (1992). A Comparison of quantitative questions in open-ended and multiple-choice formats. Journal of Educational Measurement, 29(3), 253-271.
[6] Bridgeman, B., & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. Journal of Educational Measurement, 31(1), 37-50.
[7] Buccheri, G., Gurber, N., & Bruhwiler, C. (2011). The impact of gender on interest in science topics and the choice of scientific and technical vocations. International Journal of Science Education, 33(1), 159-178.
[8] Chan, N., & 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.
[9] 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.
[10] Hudson, R. (2011). Multiple-choice questions compared to short-answer response: Which assesses understanding of chemistry more effectively? Perth: Curtin University of Technology.
[11] Korporshoek, H., Kuyper, H., van der Werf, G., & Bosker, R. (2011). Who succeeds in advanced mathematics and science courses? British Educational Research Journal, 37(3), 357-380.
[12] Lumsden, K. G., & Scott, A. (1987). The economics student reexamined: Male-female difference in comprehension. Journal of Economic Education, 18(4), 365-375.
[13] Neuschmidt, O., Barth, J., & Hastedt, D. (2008). Trends in gender differences in mathematics and science (TIMSS 1995-2003). Studies in Educational Evaluation, 34, 56-72.
[14] Pressley, M., Ghatala, E. S., Woloshyn, V., & 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.
[15] Simkin, M., & Kuechler, W. (2005). Multiple-choice tests and student understanding: What is the connection? Decision Sciences Journal of Innovative Education, 3(1), 73-98.
[16] Spelke, E. S. (2005). Sex Differences in Intrinsic Aptitude for Mathematics and Science?: A Critical Review. American Psychologist, 60(9), 950-958.
[17] VCAA. (2005). Grade distribution report for 2005. Retrieved 9th June 2008, from **http://www.vcaa.vic.edu.au/vce/statistics/2005/section3/vce_chemistry_ga05.pdf**