Addressing gender disparity in the Physics National Qualifying Exam for the Australian Science Olympiads

by Kate Wilson, Nyree Kueter, Graham Dennis, Alix Nulsen and Matthew Verdon

While the number of girls attempting the Australian Science Olympiads Physics National Qualifying Exam (NQE) is representative of the percentage studying physics at secondary schools, they are under-represented at the top of the mark distribution and hence at the scholar school, which is the first stage in team selection. The disparity could be due to a number of causes including question type, question context and external effects.

We present some data on gender differences in the NQE and a discussion of possible causes. Changes to future NQEs are being made on the basis of the results, and this study is ongoing.

Introduction

Physics has always been a male-dominated discipline, and women are still under-represented at all levels in spite of the slow improvement in gender balance over the last few decades. Amongst undergraduate students around 15% to 30% of students studying physics are female (see, for example, Feteris, 2005), and at high school level the fraction is around 30% (Fullarton and Ainley, 2000). There have been many reasons suggested for this poor participation of females, including different socialisation of girls at home and at school, and teaching and assessment practices that discourage girls (see, for example, Robertson, 2006). There is evidence that changes in year 12 curricula, including broader assessment practices, are improving the participation rate and achievement of girls in physics (Hildebrand, 1996; Trimmer, 1995). Effects due to more recent syllabus changes have not yet been reported. Nonetheless, participation rates of girls are still significantly lower than those of boys in year 11 and 12 physics.

While it is not surprising that girls are under-represented in the Australian Science Olympiads physics scholar school, the degree to which they are under-represented is surprising. In the last 9 years only 24 girls have attended the scholar school out of approximately 200 students, or an average of 12% girls. However those girls that do participate in the scholar school generally perform well, and often finish in the top of the mark distribution and go on to perform very well at the international competitions, the Asian and International Physics Olympiads (APhO and IPhO). On average, the girls perform better than the boys compared to their initial ranking on the National Qualifying Exam (NQE).

Selection for the scholar school is based on performance in the NQE. Each year around 1000 students, mainly from year 11, sit the physics NQE. The top 26 to 30 students are invited to the January scholar school, from which each year we select the teams to represent Australia at the Asian Physics Olympiad and at the International Physics Olympiad. Typically 22 to 24 students attend the scholar school. No girls attended the 2004 scholar school, in 2005 there were three girls, and last year (2006) five girls were invited and four chose to attend. These are not large sample sizes, and no statistically strong conclusions can be drawn. However, the pattern of small numbers of girls in the top 24 to 30 is a long-standing one, and these numbers are typical. If the participation of girls at the January scholar school were representative of the participation of girls in the year eleven high school cohort, we would expect, on average, 7 or 8 girls each year.

As shown in Table 1, this is not because
of a lack of girls attempting the NQE. Around one third of candidates are
girls, which is representative of the high
school cohort. Hence, it seems that
either the exam is biased towards boys
or there is some difference in the male
and female students attempting the exam.

The main purpose of the NQE is to
select a group of students for further
training, with the eventual aim to select
a team to represent Australia at the Asian
and International Physics Olympiads.
As the ultimate aim for the students
is to compete at a Physics Olympiad,
we try to select students who will be
able to develop the skills required to
perform well at a Physics Olympiad.
We are concerned about all possible
sources of bias in this selection process,
including bias from students having
studied physics under different curricula,
and gender bias. In this paper, we are
focussing on possible gender bias, and
have set out to discover why
girls are under-represented at the
top end of the mark distribution,
and what can be done about
this if it is due to bias in the
NQE. We have reconsidered
the format of our exam and made
some modifications as described
below.

Gender and the NQE

Prior to 2004 no data on the
gender of NQE participants were
gathered. In 2004 and 2005
this information was collected
on the student data section of
the answer sheets. In 2004, 297
girls, 609 boys and 52 students who did
not indicate their gender undertook the
physics NQE. Hence, around one third
of the students undertaking the NQE in
2004 were girls, which is representative
of the year eleven high school cohort.
In 2005, a total of 1054 students sat the
physics NQE, of whom 708 were boys,
332 were girls, and the gender of the
remainder could not be ascertained. This
is summarised in Table 1. Again, this is
similar to the fraction of girls in the year
eleven high school cohort of physics
students, if not slightly higher. So is it
simply small number statistics at the top
end?

In the top 25% of students in 2004,
there were 50 girls and 169 boys, 30%
girls as compared to 33% in the total
cohort. Of the top 100 students, only
21 were girls and there were only
three girls in the top 30 – those that
are eligible for team selection. Hence,
while girls are not under represented
amongst those taking the exam, they are
under represented at the top of the mark
distribution. A careful look at the marks
awarded by gender for each question on
the 2004 exam indicated some patterns.

The 2004 NQE followed the long
standing format of 20 multiple choice
questions, each worth one mark, three
short written questions each worth six
marks and one long answer question
worth 12 marks. Examples of past NQEs
including the 2004 and 2005 papers
can be found at the Australian Science

On average, boys did better on the
multiple choice questions than girls
did, a pattern which has been observed
before in physics (Hazel et al., 1997)
and in a range of other areas of study
such as mathematics, English and other
languages (see, for example, Bolger and
Kellaghan, 1990). A statistical analysis
question was also not done as well by
girls than by boys, although the average
mark on this question was very low and
many candidates did not attempt it.

While the numbers are not statistically
significant, it was interesting to note that
the only question in the written section
of the 2004 exam on which girls scored
on average higher than boys was the
question which emphasised process
more than content and calculation and
required candidates to draw a graph and
use it to find a value.

Changes to the 2005 NQE

In response to this data, the format of
the 2005 NQE was modified in the
hope of decreasing gender bias, while
at the same time reducing the effect
of different students having studied
different curricula, and making it more
consistent with later examinations.

The number of multiple choice
questions was decreased to
10, again with each worth one
mark. The number of multiple
choice questions involving
numerical calculations,
as opposed to applying
conceptual understanding of
physics, was also decreased.
This is in line with assessment
at the school level and
afterwards, including the
Olympiad, where the majority
of questions involve algebra
and often some calculus, but
actual numerical calculations
are less common and generally
worth substantially less than
obtaining correct expressions.

The number of written questions was
increased to six, with students being
asked to attempt any four questions.
These long answer questions were worth
10 marks each, hence the entire exam
is marked out of 50 possible marks.
The written questions, as in previous years,
addressed specific syllabus points. By
giving students choice, we hope to
allow students to avoid questions that
may be biased against them, whether it
be due to gender or differences in state
syllabi.

Hence, the multiple choice weighting
was reduced from 40% in the 2004
NQE to 20% in 2005. This is appropriate
given that multiple choice is not used at
any stage in team selection or training,
or at the Olympiad competitions.
The purpose of the multiple choice section is
to provide students an (apparently) easy
start to the exam, and a section that
any student may at least attempt. It primarily
acts as a warm-up for the later problems. Since the Olympiad competitions test laboratory skills, we have begun including questions that specifically address experimental skills such as plotting of data and using graphs in the NQEs. The 2005 exam included a question in which students had to read and interpret a graph on a topic with which we hoped they would be unfamiliar.

**Results for 2005**

Statistical analysis was again carried out on the 2005 NQE multiple choice questions, and it was found that again the multiple choice section favoured boys overall. The overall average score in the multiple choice section was significantly higher for boys (mean = 5.31, standard deviation = 1.95) than girls (mean = 4.45, standard deviation = 1.77) \((t(703)=7.109, p<0.001\). It was again evident that some questions in particular were more difficult for girls than boys. For example, a question involving two dimensional kinematics (shown below) was generally answered correctly by boys, but girls chose the distractor that indicated that they understood the basic concepts in one dimension, but did not treat the problem as a combination of constant acceleration in the vertical direction and constant velocity in the horizontal direction. This could be indicative of difficulty dealing with problems in more than one dimension, which would also contribute to the difference in scores on the three dimensional resistor network question from the 2004 paper.

However a difference in spatial abilities is unlikely to have been the only contributing factor to the difference in scores on the resistor network question, as there was a general pattern on both 2004 and 2005 exams of girls having more difficulty than boys with all questions involving simple circuits of resistors. The students who achieved a reasonable mark seemed to be those with either prior experience of this question or of electronics in general. These students were almost exclusively male. This may be due to socialisation reasons, such as differences in the types of toys given to boys and girls, as well as different patterns of interest in topics within physics, such as the way electrical equipment works (Reid and Skyabinia, 2003).

There was also a substantial difference between the average scores of boys and girls for the written question section. The average mark was significantly higher for boys (mean = 6.40, standard deviation = 5.64) than for girls (mean = 4.47, standard deviation = 4.03) \((t(868)=6.266, p<0.001\).

The option of choosing which questions they would attempt did not seem to help the girls. Girls and boys tended to choose to answer the same questions in the same proportions, even though analysis revealed that some questions were more difficult for girls or more biased towards boys.

A plot of percentage of boys obtaining a given mark minus the percentage of girls gaining that mark as a function of mark is shown in Figure 1 for question 12 on the 2005 NQE. Bars above the line indicate boys were more likely to obtain that mark on this question than girls were, and bars below indicate girls more often obtained this mark than boys (normalised for number of students attempting the question). The general pattern seen for this question is consistent across all six written answer questions – boys more often achieved higher marks and girls more often achieved lower marks.

In general, where more than a few students are entered from a single school the bulk of the candidates from that school obtain extremely low marks, an effect that we call the ‘whole class effect’. Figure 2 shows a plot of percentage of boys obtaining a given mark minus the percentage of girls gaining that mark as a function of mark for the entire exam for the whole cohort of students, and Figure 3 shows a similar plot for that section of the cohort from schools that entered 5 or fewer students. Girls make up 34% of the 509 students in this section of the cohort, which is similar to the overall percentage of girls. However, the difference between boys and girls is larger for the entire cohort, which includes several large groups of 20 or more students from the same school. The small number of students forming an isolated cluster at the high end (marks greater than 40) is entirely made up of students who have had Olympiad program training in the previous year, either in the Physics program or one of the other science or maths programs.

The fractions of girls in the overall cohort, the top quartile, the top 100 students and top 30 students are shown in Table 1 for both the 2004 and 2005 Physics NQEs. The changes made did not increase the number of girls in the top section of the mark distribution, except in the top 30 students, where

**Question 1.**

A ball is thrown into the air and it moves in the path shown below. Ignore air resistance in this question.

![Diagram of a ball in motion](https://via.placeholder.com/150)

At position A the ball is at the highest point in its path, position B is just before it hits the ground. Which of the following statements is true?

A) The speed of the ball at A is zero and the acceleration of the ball at B is the same as at A.

B) The speed of the ball at A is the same as the speed at B and the acceleration at B is higher than at A.

C) The speed at A is lower than the speed at B and the acceleration at A is higher than the acceleration at B.

D) The speed at A is lower than the speed at B and the acceleration at A is the same as the acceleration at B.

E) The speed at A is higher than the speed at B and the acceleration at A is the same as the acceleration at B.

**Question 1 from the 2005 NQE. Boys tended to choose the correct answer (D) while girls tended to choose distractor (A).**
the numbers are small enough that it is unlikely that they are statistically significant.

Hence, the changes made to the format of the NQE have not significantly decreased the apparent gender bias of the exam. Nonetheless, the changes to the format of the exam including the reduction of multiple choice and the choice of long answer questions, were retained in 2006 as we believe that this format better selects for students with the ability to perform well at the international competitions because it more closely represents the style of exam they will encounter there. An exam of exactly the same format as the International Physics Olympiad exams would be inappropriate, as a five hour exam with three questions would be too unfamiliar and unappealing for many students to attempt. It is also not necessary to use an exam of this format for initial selection as appropriate exam technique can be taught during training.

Other factors
One factor not considered in the changes for the 2005 exam was question context. An analysis of question context of the two exams reveals that there is a general bias towards simple contexts (or even no context), with a few questions having 'fantastic' contexts (such as a superhero). Many of the questions have ‘typical physics exam’ contexts such as cars and trucks colliding, balls being thrown, skiers on frictionless snow, and so on. There was little change from 2004 to 2005 in the style of context used, although some attempt at including more context, particularly for written questions, was made.

There is evidence that question context can play an important role in students’ success, and some research has been done by McCullough (2004) to investigate the effects of context on standard physics diagnostic tools such as the Force Concept Inventory (Hestenes et al., 1992). McCullough’s work showed evidence of gender bias due to context in physics tests, where male oriented contexts such as firing cannon balls, throwing baseballs and cars colliding with trucks are common. The physics NQEs certainly do include these sorts of contexts, although there is no clear pattern of correlation between gender bias (in terms of poorer performance by girls) and apparent gender bias of the question context. This is of course difficult to assess, as contexts traditionally of interest to girls
Table 1. Gender distribution

<table>
<thead>
<tr>
<th></th>
<th>overall</th>
<th>top 25%</th>
<th>top 100 students</th>
<th>top 30 students</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004 NQE</td>
<td>girls</td>
<td>297 (33%)</td>
<td>50 (23%)</td>
<td>21 (21%)</td>
</tr>
<tr>
<td>(n = 958*)</td>
<td>boys</td>
<td>609 (67%)</td>
<td>169 (77%)</td>
<td>79 (79%)</td>
</tr>
<tr>
<td>2005 NQE</td>
<td>girls</td>
<td>332 (32%)</td>
<td>46 (18%)</td>
<td>15 (15%)</td>
</tr>
<tr>
<td>(n = 1054**</td>
<td>boys</td>
<td>708 (68%)</td>
<td>216 (82%)</td>
<td>85 (85%)</td>
</tr>
</tbody>
</table>

* 52 of the 958 were of unknown gender, and hence are ignored in subsequent columns.
** 14 of the 1054 were of unknown gender, and hence are ignored in subsequent columns.

and boys are steadily becoming less stereotyped.

Another important factor may be guessing behaviour, specifically boys being more likely to simply write down a guess where girls are more likely to leave blank spaces. Guessing behaviour is common on these difficult exams, and there are two distinct patterns when students run into difficulties after the first parts (usually the easiest parts) of a question. One is to simply stop, and try a different question, leaving subsequent parts unanswered. The other is to guess at answers, or at least attempt subsequent parts. This typically gains a student an extra mark or two out of the ten marks available, and hence different guessing behaviours could result in this observed spike if boys are more likely to write something and girls are more likely to leave a blank. Students are advised at the start of the exam: “Read each question completely. You may be able to do later parts of the questions even if you cannot do the early parts.” However, it may be that boys are more likely to take this advice than girls are. Our data indicate that on average girls attempt fewer questions than boys, and hence, it is not unlikely that they also attempt fewer parts of questions. The advice to read carefully and attempt all parts of the questions was more strongly emphasised in the 2006 NQE paper. While good exam technique should be rewarded, this type of guessing behaviour is easy to teach during team training. Hence we do not wish to discriminate against girls on this basis.

In general, we attempt to select students on the basis of potential ability ahead of prior training, and we spend a substantial time during training on exam practice and exam technique.

There are possible external effects, such as how teachers choose which students will register for the exams. It is possible that different criteria are applied to boys and girls when a teacher is making this decision. It is also possible that talented boys are more likely to ask their teacher to register them for the exam than talented girls. There is research which indicates (Stadler et al., 2000) that girls are likely to consider that they understand physics when they can relate it to the external world and put it into a broader context. In contrast boys do not need to be able to put physics itself into a broader context to feel they have understood it. Hence, it is likely that boys develop confidence in their understanding of physics earlier, and are more likely to either ask their teacher to nominate them or agree if their teacher suggests it. This could easily cause a skew towards boys at the top end of the group.

Another possible contributing factor is the interests of the students themselves. Anecdotally, students at the January scholar school typically are students that do extra reading around the physics they are learning at school, often from popular science books and magazines. In general, boys at high school level have a greater interest in science than girls (Skamp and Logan, 2005) and hence, we presume that they are more likely to read around the subject. While this specific extra knowledge is in general not likely to be of much help in answering the NQE questions, it may give a psychological advantage in that they are at least working on questions from areas in physics that they have heard of, and so may be more likely to attempt parts of a question other students may simply give up on.

**Implications and changes to the 2006 NQE**

There is still much work to be done in discovering why girls appear to be disadvantaged on our NQE. It would be interesting to investigate some of the possible contributing factors described above. Investigation of these factors will rely on collaboration with schools, to discover the factors affecting how a teacher chooses which students to nominate, and what preparation is provided and to whom. We have begun investigating these factors as part of the registration process for the 2006 NQE, and initial surveys which ask how teachers select the students that they nominate have been sent to schools. When analysis of these surveys is complete we intend to interview groups of teachers to gain greater insight into their nomination process and any training they do with their students. There are also implications for exam design. Giving the option of choice in the written questions did not appear to reduce the bias. However, with the variety of syllabi from state to state, the element of choice in the written answer section will be maintained in the foreseeable future to reduce possible bias towards or against students from any single state.

The fraction of multiple choice questions also remained the same in 2006, however the element of calculation was largely removed from these questions so that they could focus on conceptual understanding. The multiple choice questions themselves were ordered to be easier at the start of the section, becoming more difficult towards the end, so that some confidence can be built up in the early part of the exam. The main purpose of retaining the multiple choice section at all was to give students an easy introduction to the exam.

More diagrams were also included where visualisation of problems in two or more dimensions is required, although students will still be required to draw diagrams for many questions. Girls, on average, do not score as well on tests of spatial ability as boys (Gray, 1981). We hope that by providing more assistance on two and three dimensional problems, and not including problems as spatially complex as the three dimensional resistor network, it will reduce the bias towards boys due to this type of problem.

Questions which deal with process, such as how to make measurements and plot data, will also be included to a greater extent in future NQEs. These questions appear to be neutral or even favour girls slightly, and represent an important skill which is necessary for success in Olympiad training and competition where laboratory work is an important component. In the 2006 NQE two questions of this type were included, which required students to consider the process of taking measurements or analysing data from measurements. The first required students to plot an appropriate graph from supplied data and obtain a result from the gradient of teaching science
the graph. The second asked students how they would go about measuring rate of flow of water through a pipe, and also dealt with experimental uncertainties. We do not intend to produce an exam biased towards girls overall, but the inclusion of at least one or two questions biased towards girls amongst those which appear to favour boys should reduce the overall bias of the exam which we consider a desirable objective.

A broader range of contexts may also improve the performance of girls on the NQE. The questions will be constructed to take place in a variety of contexts, some familiar everyday situations and some outside the experience of all students, with relatively few context free questions.

Setting any selection exam is a balance between maximising benefits for the participants and successfully selecting students for the task required, in this case training for and competition in the International Physics Olympiads. We have a responsibility to ensure best possible representation at the International Physics Olympiads while at the same time encouraging all students in physics.

The 2006 NQE was written with these principles in mind. Analysis of the 2006 NQE data is currently underway to determine how successful these changes have been in improving the accessibility of the exam to all students, in particular girls.

Conclusions
While the number of girls attempting the Australian Science Olympiads Physics National Qualifying Exam (NQE) is representative of the percentage studying physics at secondary schools, they are under-represented at the top of the mark distribution in both the 2004 and 2005 NQEs. On the basis of data from the 2004 exam, the format of the exam was changed in 2005 in an attempt to decrease the bias in favour of boys, without compromising the purpose of the exam. However, the mark distribution was not significantly altered by the changes and the distribution of question choices by boys and girls were very similar and independent of the gender bias of the questions.

On average girls have lower marks in both the multiple choice section and the written answer section of the exam and almost all questions were found to be biased towards boys except for questions involving processes more than content. These differences between the marks of girls and boys could be due to a number of causes including question content, question context and external effects. The disparity between the marks of boys and girls is reduced when only entries of groups of five or fewer students, not whole classes, are considered.

Changes to the NQE for 2006 were made on the basis of the results of this study. Further work analysing the results from the 2006 NQE and attempting to improve future NQEs is ongoing.

Acknowledgements
We would like to acknowledge the assistance of Rhona McDowall, the Australian Science Innovations education officer and Colin Taylor, executive director of Australian Science Innovations. Jolyon Bloomfield assisted in preparing exam questions and made very helpful suggestions for presentation of data.

References


About the author(s):
Kate Wilson is program director of the Australian Science Olympiads physics program, a part time lecturer at the ANU and a part time student at the University of Canberra.
Nyree Kueter works as a statistician, researcher and survey designer at the ANU and is also a part time Masters student at the university.
Graham Dennis is a tutor for the Australian Science Olympiads physics program, and a PhD student at the Australian National University.
Alix Nulsen is a tutor for the Australian Science Olympiads physics program, and a Ph.D. student at the University of Sydney.
Matthew Verdon is the deputy director of the Australian Science Olympiads physics program, and a Ph.D. student at the University of Sydney.