To:
Hon. Rosemary Menkins
Chair
Inquiry into assessment methods
for senior maths, chemistry and physics
in Queensland schools
eic@parliament.qld.gov.au
burdekin@parliament.qld.gov.au

From:
[redacted]



June 7, 2013

# Supplementary Written Submission

I am the author of submission 121 to the inquiry which has been published with my name supressed. I am employed as a senior science teacher and currently hold a middle-management position responsible for the analysis of senior school results and data. As a data analyst I provide advice and feedback to a number of schools on the usefulness and reliability of senior assessment practices and how to improve their QCS Test performance. I also have a good understanding of the assessment processes used interstate and their reliability.

I am concerned with the statements presented to the inquiry asserting the high reliability of current assessment and tertiary entrance processes for maths and science**. A lot of opinions have been presented but little hard data. It is my belief that the Inquiry has not obtained from the QSA data that could be compared with that interstate to provide a more definitive answer on the reliability and accuracy of current assessment practices**. As a consequence I have prepared the material contained in this document.

I am prepared and have applied to attend the final hearing on July 10 at my own cost (I live in north Queensland) and explain and discuss the data in this document in plain English terms (not educational jargon) so that members of the Inquiry will understand its implications.

The material in the documents provides hard evidence that:

1. **There are clear cases of school assessment in maths and science being very inaccurate in terms of maths-science skills measured by the QCS Test;**
2. **That students with only average ability in maths can gain inflated grades in maths and science and enter tertiary maths-orientated courses;**
3. **That for maths-science students a strong performance in literacy questions in the QCS Test can mask poor ability in maths providing an advantage for their OP and as a consequence making tertiary entrance selection procedures unreliable.**

The data has been rounded to the nearest whole number in the body of the submission to protect the identity of schools. **The original data is provided in Appendix 2 which is to be regarded as confidential and not for publication.**

In addition I have presented information from publicly available documents that report:

1. Only a modest reliability between school-based assessment and the QCS Test;
2. Assessment processes in individual maths and science subjects in West Australia (and other states) showing much higher assessment reliability than that occurring here;
3. A recent external audit of the QCS Test informed the QSA that current assessment processes are producing a bias which is giving girls higher than expected OPs based on QCS Test performance – the QSA have failed to mention this to the inquiry.

I thank you for considering the contents of this document.

[redacted]

**Assessment Accuracy**

The QSA uses a standard scale to measure student ability both in school-based assessment and the QCS Test. On this scale the mean (average) ability is defined as a score of 175.

From this scale cut-off scores indicate relative ability, for example OPs. An OP1 is about 50 points above the average (175) while an OP25 is about 50 points below. The precise numbers vary slightly from year to year. An example of the scale used recently by the QSA is provided below.

I am presenting this scale simply so that the members of the Inquiry will understand what follows when discussing student performance in terms of 'OP ability' as a useful comparison by referring to this table – **relative ability in terms OPs on a scale of 1 to 25 is something that everyone can understand**:

| OP Band Minimum Score | OP Number |
|---|---|
| 224.385 | 1 |
| 215.14 | 2 |
| 208.1073 | 3 |
| 202.1195 | 4 |
| 197.0231 | 5 |
| 192.1927 | 6 |
| 187.7555 | 7 |
| 183.5075 | 8 |
| 179.711 | 9 |
| 175.8189 | 10 |
| 171.8975 | 11 |
| 168.2553 | 12 |
| 164.4654 | 13 |
| 160.8723 | 14 |
| 157.209 | 15 |
| 153.7281 | 16 |
| 150.0425 | 17 |
| 146.4721 | 18 |
| 142.6748 | 19 |
| 138.6079 | 20 |
| 133.9212 | 21 |
| 129.1487 | 22 |
| 122.7309 | 23 |
| 111.4473 | 24 |
| 0 | 25 |

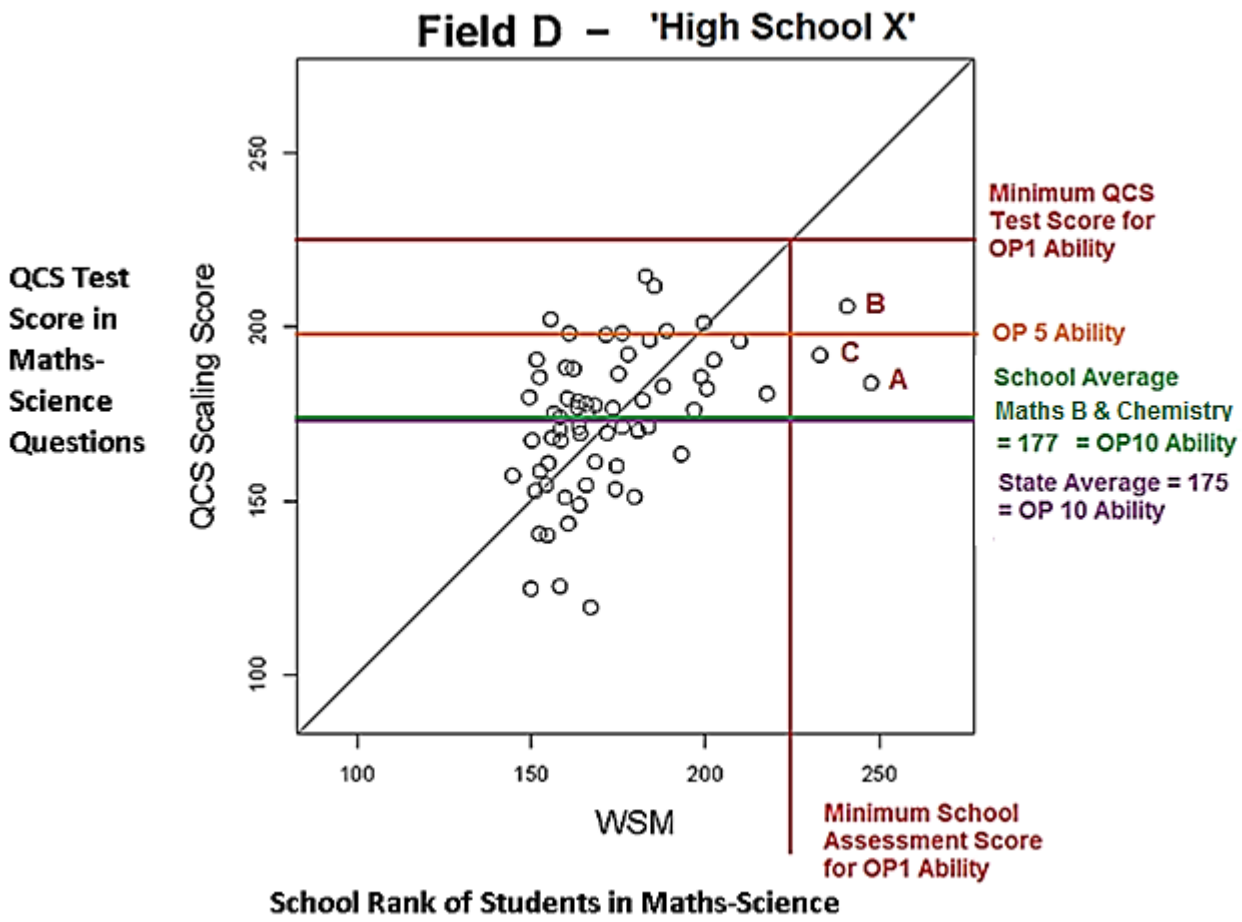
Each year the QSA provides data and graphs to schools comparing the rank of students in relative ability in a school (Within School Measure or WSM) to their QCS Test grades.

Below is a graph of the results by students in a school in Field D defined by the QSA as: *Solving complex problems involving mathematical symbols and abstractions.*

**A Field D score is largely determined by the subjects that concentrate on these skills – Maths A, B & C, Chemistry and Physics**. The QSA publish Field Scores (reported from 1 to 10) but OPs are what are used primarily for tertiary entrance. It has to be kept in mind that **the QCS Test is stated explicitly by the QSA to assess student skills at a Year 10 level of ability**.

I have drawn in the coloured lines on the graph to indicate the ability level based on the table on the previous page.

In this example school 'x' (a real school) and several panels agreed that 3 students (labelled A, B and C) had greater than OP1 ability in Field D skills based on the school assessment (compared to other students in the school) but an independent test (QCS) says otherwise! Also note the general poor fit of the results compared to the diagonal black line representing a perfect match between school and QCS Test results:



Key points:

- The top student in the school (A) has a relative ability at maths-science problem solving skills compared to students in the rest of the state equivalent to an OP of 8 yet the school, supported by maths and science panels, agreed that the student was a superb OP1 at these skills. Student B (about OP4 ability) and student C (about OP6 ability) are also way below the OP1 mark;
- Very few students reached an OP5 level of ability in the Field D maths-science questions in the QCS Test;
- The average score in the maths-science questions in the QCS Test for the Maths B and Chemistry students (177) is almost exactly the same as the state average (175);
- School and QSA records indicate that in senior maths and science subjects the students obtained grades above state averages;
- **External subject-specific exams as used in other states would have automatically corrected the anomalous results and adjusted student grades. In Queensland grades are only altered by panels and the process clearly failed in this example.**

Below is the school OP scaling data for the same year as the previous graph (the figures have been rounded to protect the identity of the school – the original data is provided in Appendix 2 and is not for publication):

| **Subject** | **OP Scaling (all QCS Skills)** | **Field A** (extended written expression involving complex analysis and synthesis of ideas) | **Field B** (short written communication involving reading, comprehension and expression in English or a foreign language) | **Field C** (basic numeracy involving simple calculations and graphical and tabular interpretation) | **Field D** (solving complex problems involving mathematical symbols and abstractions) | **Field E** (substantial practical performance involving physical or creative arts and expressive skills) |
|---|---|---|---|---|---|---|
| Maths B | **191** | 190 | 190 | 186 | **177** | 183 |
| Chemistry | **191** | 189 | 188 | 188 | **177** | 181 |

Keeping in mind that a figure of 175 represents an average performance compared to ALL students in the rest of the state several things are evident:

- The Field D averages for Maths B and Chemistry of 177 indicate that the students were barely above average compared to all other students in the state in mathematical problem solving. These are the skills in the syllabuses that to a large degree indicate 'A' and 'B' grade students in these subjects;

- The Maths B and Chemistry students are well above average in Field A and B skills (literary), above average in Field C skills (basic numeracy) and Field E (performing arts);

- **In Maths B and Chemistry the Field D mean of 177 for the average student indicates an OP 10 level of ability at these maths skills BUT the average student's scores contributing directly to their OPs will be scaled for tertiary entrance with a score of 191 indicating OP 7 ability.**

**Conclusions:**

1. **The results approved by the school and panels for maths skills that separate 'A' and 'B' grade students and ranking them ahead of other students in the school were very unreliable – at least two groups of subject teachers and panels agreed that student A had an OP1 level of ability in maths problem solving when an independent test at Year 10 level suggests OP8 ability. Serious errors are also evident in the grades of two other students, B and C;**

2. **Maths B and Chemistry students with average ability in solving difficult maths problems had their OPs inflated as a result of better than average literary skills – maths problem-solving ability OP10, general ability OP7 – highlighting that current tertiary entrance scores in maths and science are largely determined by literary ability rather than maths skills;**

3. **The average Maths B and Chemistry students in the school obtained above average OPs and would have gained places into tertiary courses requiring a high ability in maths problem solving skills ahead of other students in the state who were better at these skills;**

4. **Very few of the Maths B and Chemistry students had the maths problem solving skills required for maths-orientated tertiary courses with an entrance requirement of OP6 (but their school results suggest an <u>average</u> ability of OP7).**

**This is not a rare example** – in fact the same pattern occurred at this school over a number of years and the school sought the author's assistance to improve assessment practices. I have seen quite a few other examples like this and also the opposite scenario – well above average performance in maths skills in the QCS Test but lower than average grades and OP outcomes. **A fundamental flaw in the system is the scaling of maths and science tertiary entrance scores using a test comprising less than 1/3 maths-science skills – the QCS Test.**

From the evidence I have seen **<u>the type of discrepancies demonstrated have become more evident since the introduction of the new maths and science syllabuses</u>**. I could provide more examples if requested.

**The QSA performs statistical analyses to detect anomalous results in terms of whole school performance compared to that on the QCS Test. <u>The QSA does not employ any process to detect anomalous results in specific subjects in schools</u>**.

**In other states anomalous subject results in schools are generally corrected by external exams which moderate school assessment. They do so by:**

1. **Testing subject specific and not generic skills;**
2. **Automatically adjusting school grades;**
3. **Identifying anomalous subject results and informing schools.**
   - **The Queensland system has a lower level of accountability than other states.**

**Reliability**

Other states have published measures of reliability (correlation or $r^2$) on school-based assessment moderated by external exams. The $r^2$ is one of the classical statistical measures of reliability in all types of research. A score of 1.0 represents a perfect match between pairs of measurements, e.g., school results and an external exam such as the QCS Test. I won't go into mathematical detail on precise meaning but simply refer to this measure as 'reliability'.

West Australia for a time adopted the maths and science assessment practices of Queensland then abandoned them. They now use external exams to moderate the assessment conducted in schools and publish measures of state-wide reliability in all subjects:
(http://www.scsa.wa.edu.au/internet/Publications/Reports/Statistical_Reports/Secondary_Education_Statistics/Secondary_Education_Statistics_2012 ).

Presented below are the $r^2$ values indicating the correlation between school-based assessment and the student scores in the external scaling exams in key science and maths subjects in West Australia last year. It is apparent that the assessment practices show very high reliability in West Australia and similar results have from time to time been published in NSW and Victoria:

| Subject | Reliability ($r^2$) |
|---|---|
| Biology | 0.90 |
| Chemistry | 0.93 |
| Physics | O.92 |
| Maths 3AB | 0.88 |
| Maths 3CD | 0.93 |
| Specialist Maths 3AB | 0.91 |
| Specialist Maths 3CD | 0.94 |

**The QSA does not publish equivalent reliability measures on individual subjects in schools compared to QCS Test results – so at this point no one knows the level of reliability of assessment practices at the subject level as an $r^2$ value**. A recent external audit of the QCS Test (see Appendix 1) reported the reliability between the combined results of all subjects in school used to rank students (WSM) and QCS Test performance for a number of groups. Measurement of student ability with the WSM involves statistical processing by the QSA whereas West Australia used raw marks in assessing their reliability. **Despite the data processing used by the QSA the reliability is far lower than that reported in West Australia**:

| Group | Reliability ($r^2$) |
|---|---|
| Boys Co-Ed Schools | 0.637 |
| Girls Co-Ed Schools | 0.679 |
| Boys Single Sex Schools | O.679 |
| Girls Single Sex Schools | 0.717 |

**A significant finding of the audit was that it identified girls gaining higher than expected OPs than boys from QCS Test scores – it recommended that research be undertaken to identify the source of the identified gender-related bias in OPs (see Appendix 1).**

**Maths-Science assessment in West Australia shows high consistency and accuracy (typically above 0.90) between the assessment occurring in schools and the external scaling exams. The tertiary entrance process is therefore fair, accurate and reliable as indicated by the high reliability scores.** The QSA has never revealed the equivalent data comparing school grades in maths and science and scores in the maths questions (Fields C & D) in the QCS Test under the current syllabuses. The only other $r^2$ values reported for maths-science in Queensland were in a study in 1994 that recorded reliability of above 0.90 for Chemistry and Maths school assessment based on tests and marks (refer to submission 121).

On the evidence I have seen maths-science assessment reliability in Queensland is probably below 0.70 (the ACT uses similar practices to that in Queensland and also uses a test similar to the QCS Test to scale school subjects. For maths and science subjects $r^2$ values below 0.5 have been reported in the ACT – refer to submission 121).

**It is my professional opinion that the current inquiry has not requested from the QSA data that could have provided definitive answers on the reliability of current assessment practices in maths and science in Queensland.**

**If the inquiry desires to obtain an insight into the reliability of current assessment practices then:**

1. **Ask the QSA to provide graphs plotting raw subject grades (on the 50 point scale) in Maths A, B, C, Chemistry and Physics against raw QCS scores in Fields C and D (maths-science) questions in the QCS Test for every student in the state along with correlation coefficients and descriptive statistics;**
   - **This will provide a measure of the accuracy and reliability of current assessment practices in terms of purely maths-science skills – the use of raw data provides a valid comparison with that from other states.**

2. **Get an independent statistician to examine the graphs and data, compare them to data from West Australia (and possibly other states) and tender professional advice to the inquiry.**
   - **This will allow a comparison of totally school-based assessment with combined school-exam based assessment and may well provide strong evidence to help answer one of the terms of reference.**

**Is the Queensland Parliament and taxpayers satisfied with a system of assessment and tertiary entrance for maths and science students with an uncertain reliability (probably less than 0.70) compared to West Australia which is known to be generally higher than 0.90?**

**Assessment Validity**

Much has been presented in other submissions about the usefulness and validity of the current syllabuses in senior science in terms of the use of assignments such as EEIs and ERTs. I will provide this exemplar to bolster existing evidence that current practices are assessing the wrong skills and unfair. School 'x' sought the author's assistance in improving the outcomes of their senior science students. **My advice to the school management was**:

1. **Spend the minimum time on assignments;**
2. **No drafts for assignments in Year 12 Chemistry;**
3. **Focus on maths-science and visual problem solving skills in teaching & assessment.**

The results (year 'x' = the previous results discussed, year 'y' = another Year 12 group):

| Subject | OP Scaling (all QCS Skills) | Field A (extended written expression involving complex analysis and synthesis of ideas) | Field B (short written communication involving reading, comprehension and expression in English or a foreign language) | Field C (basic numeracy involving simple calculations and graphical and tabular interpretation) | Field D (solving complex problems involving mathematical symbols and abstractions) | Field E (substantial practical performance involving physical or creative arts and expressive skills) |
|---|---|---|---|---|---|---|
| Chemistry (year 'x') | **191** | 189 | 188 | 188 | **177** | 181 |
| Chemistry (year 'y') | **193** | 188 | 180 | 189 | **190** | 194 |

The results:

1. **The local Chemistry panel was very critical of the assessment and lowered the grade of the top student (they attempted to lower others too). The students obtained lower average and top grades than the previous group**;
2. A **dramatic improvement in the Field D performance in Chemistry** – in fact they performed better than in the literary fields (A & B) which is what you would expect of great Chemistry students (improvement = ½ standard deviation);
3. The Field C and Field D scores are similar – this indicates that the assessment testing basic and advanced maths-science skills was comparable – suggesting that the teaching and assessment has improved and is balanced;
4. The scaling score of 193 for the OP for the average Chemistry students is a record for the school in the subject as is the Field D score of 190;
5. **The school obtained a record number of OP1s**;
6. **The student ranked fifth in the school in Chemistry was only awarded a grade of B+. The student obtained an OP1. The same student sat an exam at a prestigious interstate university and won a large chemical engineering scholarship**.

**Reach your own conclusion on assessment fairness & validity in Chemistry in this state!**

**Appendix 1: Extracts from an External Audit of QCS Test**
http://www.qsa.qld.edu.au/downloads/publications/report_qcs_test_review_2012.pdf

## A Review of Aspects of the Queensland Core Skills Test

**A review of the Queensland Core Skills (QCS) Test to ascertain the ongoing relevance of the test and the capability of the test to act as a statistical scaling device in the calculation of Overall Positions (OPs) and Field Positions (FPs) for tertiary selection**


Produced by Enterprising Minds Pty Ltd
Tony McLeod and Michelle Davidson

in association with
John Bennett
George Cooney
Bob Edwards

and in consultation with
Jim Tognolini


February 2012

r2 values

**TABLE 4**
WSM and QCS Test mean scores by gender and type of school

| School type | School WSM | | QCS Test | | Correlation between WSM and QCS Test scores | |
|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male | Female |
| Co-educational | 170.8 | 175.5 | 177.4 | 170.4 | 0.637 | 0.679 |
| Males only | 181.7 | | 181.6 | | 0.679 | |
| Females only | | 183.5 | | 183.5 | | 0.717 |

### 3.6.2 Implications

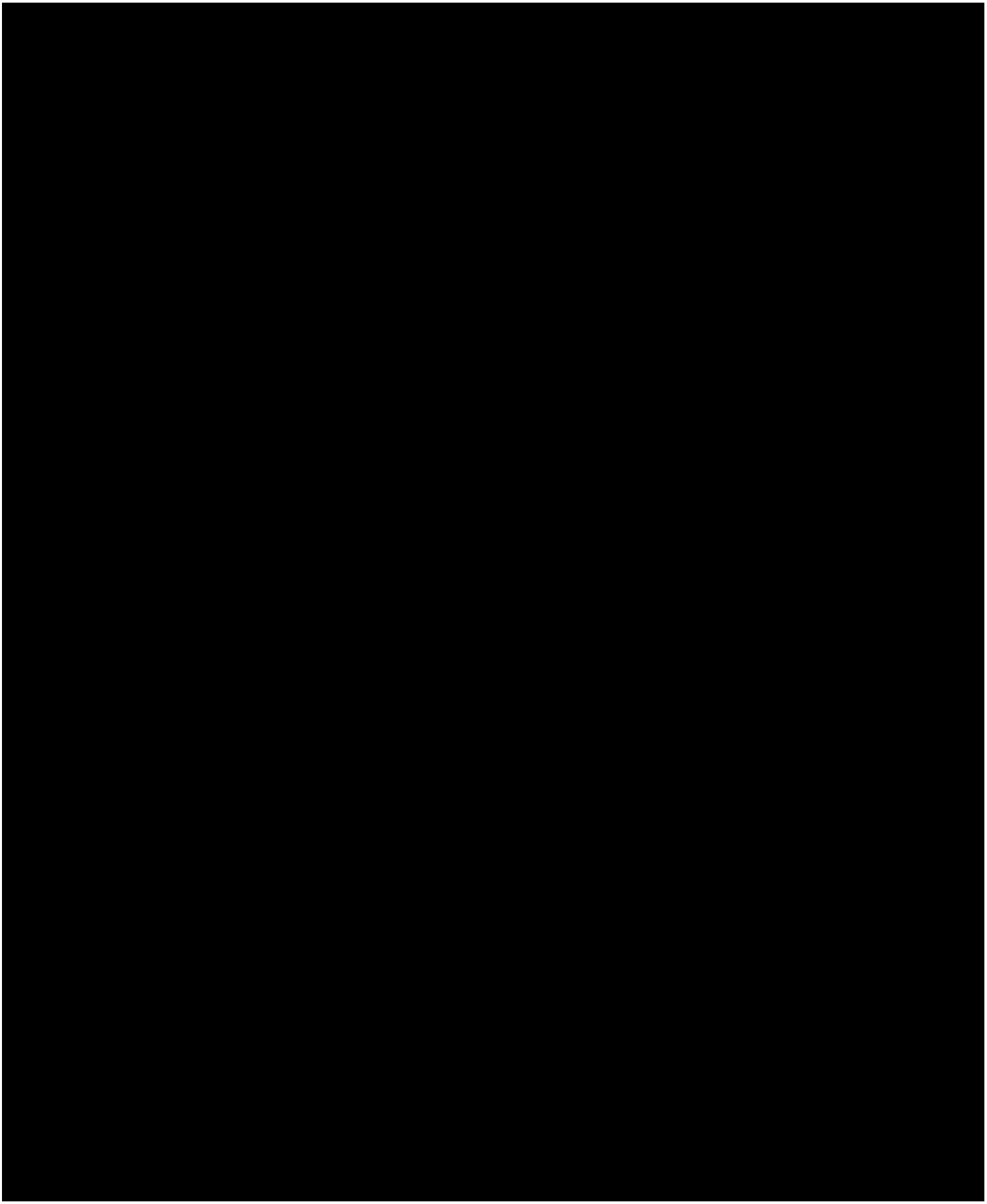
These analyses show that the relationship between WSM and QCS Test scores is not independent of gender in co-educational schools: males performed better than females in the QCS Test, but not in the school assessments; and on average, females received higher OPs than suggested by their QCS Test scores.

### Recommendation 4

*that the QSA conducts further research to determine the reasons for the gender-related difference in students' rankings based on QCS Test scores compared to their rankings based on WSM*

**Appendix 2: Original School Data used in the preparation of this document**.

# CONFIDENTIAL - NOT FOR PUBLICATION



**CONFIDENTIAL - NOT FOR PUBLICATION**

[redacted]

[redacted]