QTAC | CHOOSE THE FUTURE

INTER-SUBJECT SCALING: AN INTRODUCTION

DECEMBER 2019

PAGE INTENTIONALLY BLANK

SETTING THE SCENE

The year is 1912. And the location is Stockholm, Sweden. The occasion? The Summer Olympics and the introduction of the modern pentathlon. It was comprised of pistol shooting, fencing, swimming, horse riding and running. As its sponsor Baron de Coubertin stated, this new event would test *"a man's moral qualities as much as his physical resources and skills, producing thereby a complete athlete."*

A prominent aspect of the modern pentathlon is the point scoring system. Each competitor is awarded a certain number of points based on their performance in each of the five events. The gold medallist is the competitor with the highest point total at the end of the five events. Because each competitor is competing in exactly the same five events, the raw scores from each event can simply be added together to provide a final score.

BUT...

What if the Olympic competitors were able to select five events to compete in from a list of 60 possible events – with some of the events more challenging than others? Would it be fair if each competitor's raw scores in each event were added together and the highest point total used to determine the gold medallist, regardless of whether the sport was fencing, shotput or swimming?

It may be **SIMPLE** to add five raw scores together to produce a final score, but it wouldn't be **FAIR** to do so, as not all sporting events are equal, with some events more challenging than others. Competitors would simply enter the five easiest sports to maximise their points tally. The same logic could be applied to senior school subjects. If subject scores were simply added together and used for tertiary entrance, it would make sense to only take what are perceived to be the 'easiest' subjects.

This is not how parents, schools, universities or the wider education sector wishes to see students achieve academically, which is why QTAC will use scaling as part of the ATAR calculation process in 2020.

MEET CHRIS AND MIA...

Chris is taking General Mathematics. He is planning on studying International Relations and is also taking English, Modern History, Chinese and Spanish. Chris has achieved a result at the end of Year 12 of 82/100 for General Mathematics.

Mia is taking Specialist Mathematics. She is also taking English, Mathematical Methods, Biology and Chemistry and wants to study Biomedicine at uni and work as a medical researcher. At the end of Year 12, Mia has secured a result of 82/100 for Specialist Mathematics.

THE QUESTION...

In the wider picture of progressing on with tertiary study, is Chris's mark of 82 in General Mathematics the same as Mia's mark of 82 in Specialist Mathematics?

We know that Specialist Mathematics is academically more challenging, so instinctively the answer is no. However, when it comes to calculating ATARs, how do we equitably compare Chris's result with Mia's result? The answer is through scaling.

SOME TECHNICAL DEFINITIONS

To get your head around scaling, there are a few technical terms that need to be expanded upon as they may arise during our discussion (and you'll sound like a real pro when you discuss this with others).

Iterative Process

A process which is repeated many times. In the context of scaling it is a process that refines scaled scores into more refined scaled scores. The objective of an iterative process is to bring you closer to the desired result, as you repeat a set of instructions many times. An example of an iterative process is when a jury examines evidence and scenarios several times to come closer to a resolution.

Mean and Variance

Quantitative measures indicating the shape of a set of points. A **mean** is the measure of the average, while the **variance** is a measure of the variability of the observations, defined as average squared deviation from the mean.

Polyrank

A combined result of subject scaled scores. In the context of scaling, it is the student's best five scaled subject results.

Quantile

Quantiles are a set of values dividing a frequency distribution into equal groups, each containing the same fraction of the total population. Quantiles are also called quartiles when you divide a distribution into four groups, or deciles when you divide the distribution in ten groups, or percentiles when you divide a distribution into a hundred groups. The most well known quantile is the median, which is the value that divides the population into two equal sized groups. The median is also equivalent to the 2nd quartile, the 5th decile, and the 50th percentile.

Raw Subject Result

The subject result achieved by a student. For General subjects, General Extension or Senior External Examination subjects, students will be provided with a result out of 100 and a level of achievement grade (A to E). For Applied subjects, students will be provided with a letter grade (A to E) only.

Scaled Subject Result

The subject result which has been transformed from the raw subject result via the scaling process. A student's raw subject result in English might be 50 out of 100, and the students eventual scaled result may be 52.16 out of 100.

Scaling

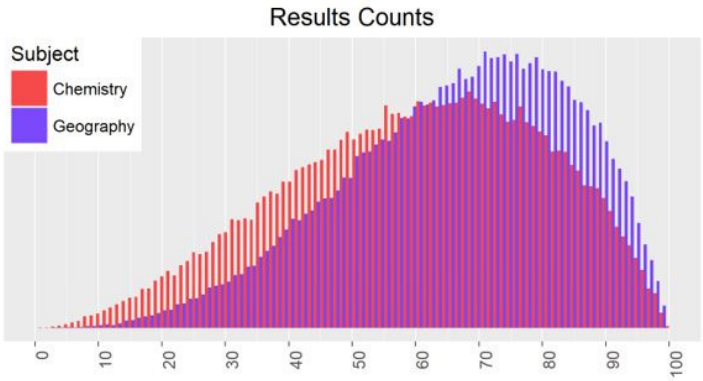
The process by which raw results for all subjects are adjusted to allow results from any subject to be fairly comparable, for the purpose of calculating an ATAR.

So why do we need to scale?

THE NEED TO SCALE

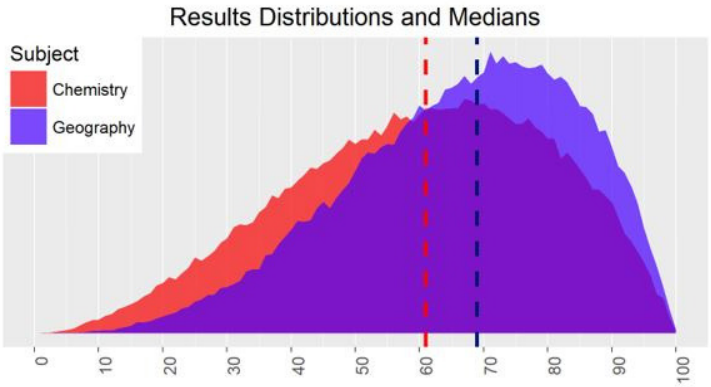
Every subject offered at senior level in Queensland schools is unique. Each one has its own syllabus objectives and assessment criteria. Each subject assesses different knowledge and skill sets. Students' performance varies across different subjects and the range of results achieved can be significantly different from the results that students achieved in another subject. This means that a straight comparison of raw subject scores can be misleading.

In this fictional example, Chemistry and Geography have different distributions of results, with different mean and variance. Geography appears to have a higher and more concentrated proportion of students achieving high results, whereas Chemistry has a more dispersed distribution with a lower proportion of students achieving high results. In practical terms this means that achieving 90 out of 100 in Geography is relatively easier than achieving 90 out of 100 in Chemistry. So, how can we translate a result of 90 in Geography into an equivalent mark in Chemistry?



A common way to compare the distributions of two populations is using the quantiles describing the two distributions.
Quantiles are a location index that can help you to map one distribution onto the other, by inferring what raw result in Chemistry would be equivalent to a certain raw result in Geography.

For instance, if you know that 50% of students achieved a score lower than 69 in Geography, and 50% of students achieved a score lower than 61 in Chemistry you would be able to conclude that a result of 69 in Geography is comparable with a result of 61 in Chemistry.



This correspondence can be used to form a hundred equally sized groups (rather than two). The idea is that the percentiles can be used to represent both subject results on a common scale. The logic behind the subject scaling is that each result within a subject can be translated into its relative percentile rank.

A score of 69 in Geography and 61 in Chemistry would be mapped to a "scaled result" of 50, and therefore be treated as equal.

MEET DOMINIC AND SASHA

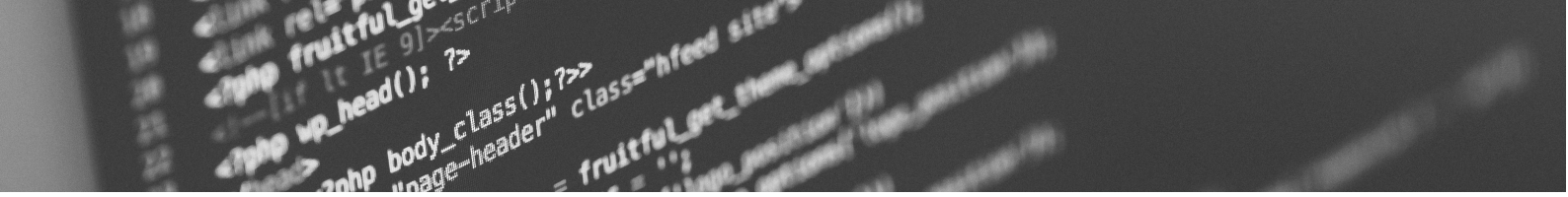
They are good friends and attend the same school. They are both studying (amongst other subjects), Geography and Chemistry. Here are their raw subject marks.

	DOMINIC	SASHA
GEOGRAPHY	90	85
CHEMISTRY	80	90

We can assume that Dominic’s raw Geography result of 90 is better than Sasha’s Geography raw result of 85. We can also assume that Sasha's result of 90 in Chemistry is better than Dominic's result of 80 in Chemistry.

However we cannot assume that subject results of 90 in different subjects are comparable. Similarly we cannot assume that Dominic's result of 90 in Geography is better than Sasha's result of 90 in Chemistry.

Sasha's 90 in Chemistry could have been achieved only by the top 5% of students, whilst Dominic's 90 in Geography could have been achieved by a larger proportion of students. Using the percentile ranks will enable the raw subject results to be fairly comparable across different subjects.



MEET DOMINIC AND SASHA

cont'd...

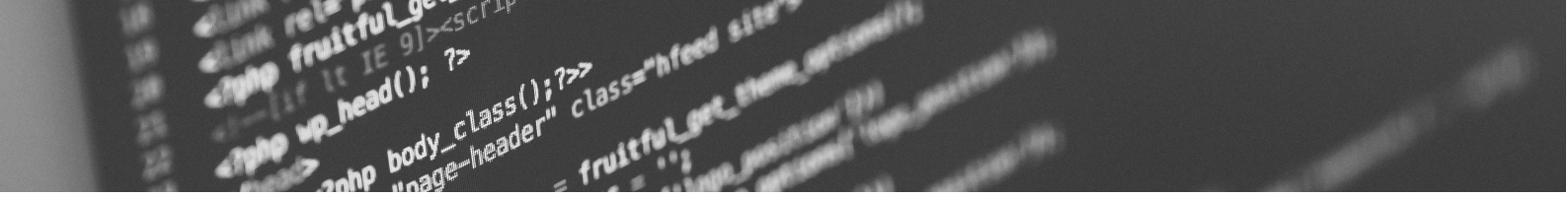
Using the subject results' distributions, these raw results could be transformed into the following scaled results:

	Scaled DOMINIC	Scaled SASHA
GEOGRAPHY	**92**	84
CHEMISTRY	82	**95**

Using scaled results makes immediately evident that Sasha's 90 in Chemistry is worth more than Dominic's 90 in Geography. What remains unaltered is that Dominic's Geography raw subject result is always higher than Sasha's, and Sasha's Chemistry raw subject result is always higher than Dominic's.

This idea to compare subject results can be extended to all subjects studied by students. As it will be described in the next section, it can also be applied to scaled subject results, obtaining a set of scaled results of scaled results, and then it can be done again, obtaining a set of scaled results of scaled results of scaled results, etc.

Percentile	Chemistry	Geography
1%	14	22
2%	18	27
3%	21	30
4%	23	32
5%	25	34
...	...	...
10%	32	42
20%	42	51
...	...	...
50%	61	69
..	...	...
82%	80	84
...	...	..
84%	81	85
...	...	...
92%	87	90
...	...	...
95%	90	92



THE 'HOW' OF SCALING

Since the introduction of the OP in 1992, a common scaling test – the Queensland Core Skills (QCS) Test has been used to scale Year 12 results. However, with the introduction of the ATAR as the standard measure for tertiary entrance in Queensland from 2020, the QCS Test will be discontinued. In the absence of this common scaling test, Queensland will use inter-subject scaling, a process which is used in other Australian states and territories.

1

The first step is to compare subject results to determine the relative position of one student against the relative position of another student. In order to see this in action, let's continue with Dominic and Sasha and their subjects.

Let's say Dominic has a raw subject mark of 81 in English. This tells us that Dominic got 81 marks out of a possible 100 and it also tells us where Dominic sits in relation to the other English students. But it doesn't directly let us compare Dominic's English result with results from other subjects.

To work out if Dominic's raw subject result of 81 in English is comparable to say, Sasha's raw subject result of 85 in Ancient History (a subject Dominic didn't take), we need to calculate the percentile ranks for all possible subject results, including Dominic's and Sasha's results. This can be done from the distributions of results, as shown in the previous example.

2

Once the first subject's scaled results are available, each student will have a set of provisional scaled results. These outputs will be able to tell us that Sasha's subject result of 85 in Ancient History is comparable to a provisional scaled result of 74, and that Dominic's subject result of 81 in English is comparable to a provisional scaled result of 83. This process is repeated for all subjects and students.

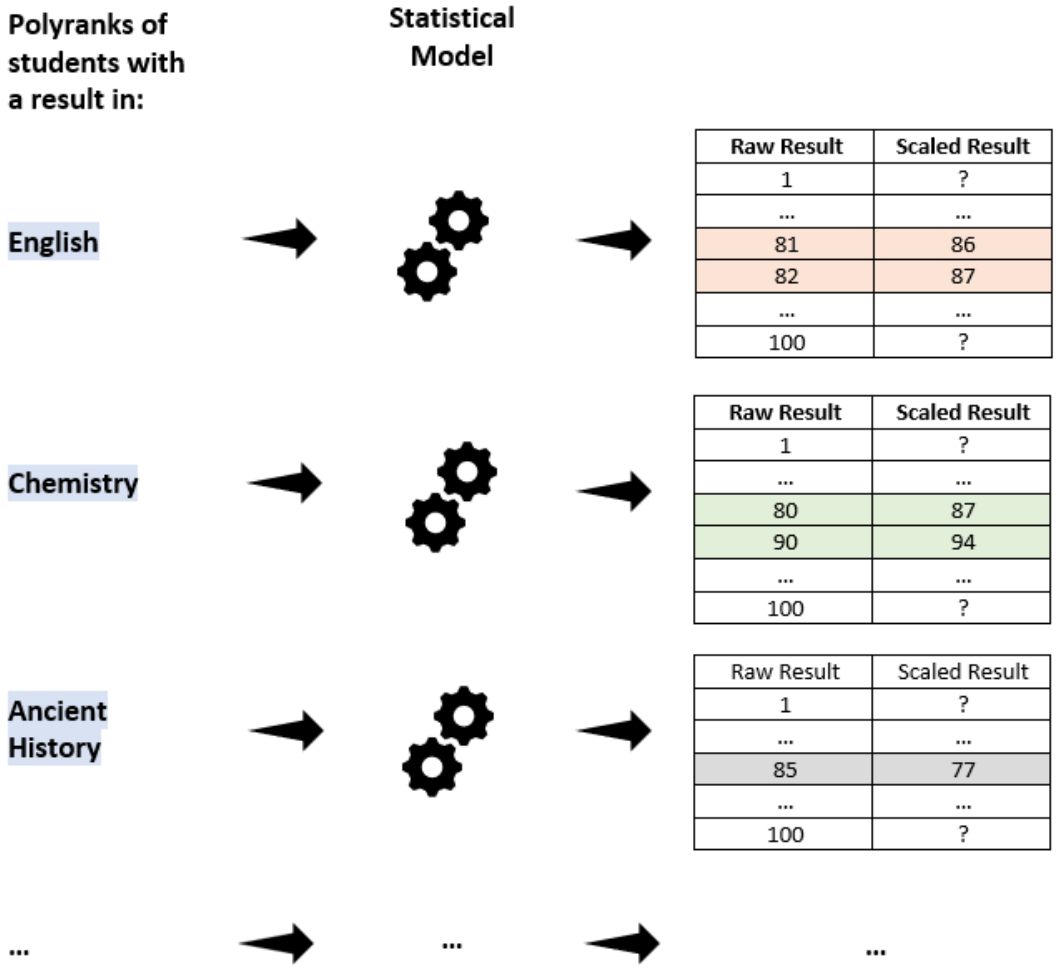
3

With the first set of scaled results, it is possible to infer the overall performance (or polyrank) of each student, by averaging the scaled results from each subject they took.

Subjects	Dominic		Sasha	
	Raw	Scaled	Raw	Scaled
Geography	90	92	85	84
Chemistry	80	82	90	95
Mathematical Methods	75	82	72	78
English	81	83	82	84
Ancient History			85	74
Biology	40	45	40	45
Legal Studies	65	63		
Average (Polyrank)		74.5		76.7

4

This first estimate of polyrank is used to refine the estimates for the subject scaled results. A statistical model is used to trace the relationship between raw results and polyranks of students in each subject. The values of the fitted model will "overwrite" the first set of scaled results.



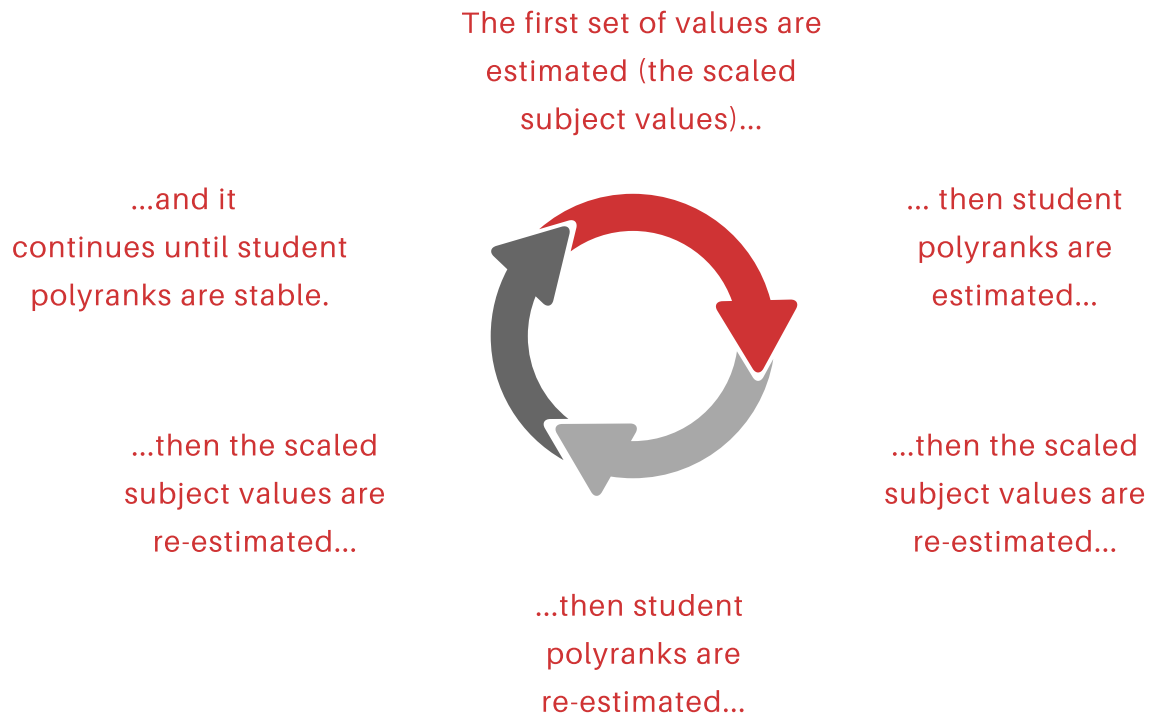
Achievement of students in their other subjects influences how a particular subject scales.

For example, if high results in English and English Extension are achieved by students who achieve high scores in other subjects, especially those where it is challenging to achieve a high score, then English and English Extension will tend to "scale well".

Notice that the polyranks used to scale subjects derive from all the available results of Dominic and Sasha. Their final ATAR however will be derived from their best five results.

5

New scaled results are used to re-calculate students' averages, which are then used to re-calculate subject scaled results, which are then used to re-calculate students' averages, and so on. Scaling continues iteratively (see, understanding the technical terms is paying off!) until the estimates stabilise.



This process does not change a student's raw subject result, nor does it change their ranking relative to other students in that subject. All scaling does is allow for the comparison of performances across all subjects, for the purpose of calculation of the ATARs.

So how did Dominic and Sasha end up after the scaling process?
You can see from the table on page eight that from the **first cycle of scaling**, that Dominic's raw subject result of 81 in English scaled to 83, and that Sasha's raw subject result of 85 in Ancient History scaled to 74.

But once the iterative process of scaling **over many cycles** has occurred, Dominic's raw subject result of 81 in English finally scaled to 84 and Sasha's raw subject result in Ancient History finally scaled to 76.

RECAP AND SOME FINAL POINTS ABOUT SCALING...

In the ATAR calculation process, scaling is the most difficult part to understand. It is however, a necessary part of the process to ensure fairness.

Scaling is a process by which raw subject results are adjusted for a given subject to allow the results for that subject to be fairly compared with the results from any other subject when calculating ATARs.

The scaling process will adjust the raw results in each subject to take account of how strong students are in their subjects and how difficult it is to achieve a result in the subject relative to achievements in other subjects.

Scaling outcomes for individual subjects are not predetermined and are expected to be different from one year to the next based on the performance of the student cohort for each year. Although trends will form, school, students and parents are advised not to use historical scaling data to predict future outcomes.

RECAP AND SOME FINAL POINTS ABOUT SCALING...

cont'd...

This document has focused on scaling of General subjects. Applied subjects and completed Vocational Education and Training (VET) qualifications at Certificate III level and higher may also be included as one of the five inputs into an ATAR. In the scaling process:

- Each Applied subject will have scaled scores based upon the letter grade awarded in the subject.
- Each VET level (Certificate III, Certificate IV, Diploma and Advanced Diploma) will have a single scaled score, and this will be independent of the area of study, duration, mode of assessment or study provider of the VET qualification.

QTAC will issue a scaling report in early 2021. The purpose of making the scaling reports available is to ensure openness, accountability and transparency of the ATAR calculation process. It is not for the purposes of creating predictive ATAR calculators or in any way to suggest that a subject (or group of subjects) is better than another.

In particular, the reader is reminded that the intent and rigour of any two General subjects is comparable even if the mode of learning and curriculum pedagogy is different and unique to that subject. Take for example these two subjects – Physics and Ancient History:

Physics: *…. "The purpose of senior Science subjects in Queensland is to introduce students to a scientific discipline. Students will be required to learn and apply aspects of the knowledge and skill of the discipline (thinking, experimentation, problem-solving and research skills), understand how it works and how it may impact society."*

Ancient History: *…."A course of study in Ancient History empowers students with multi-disciplinary skills in analysing textual and visual sources, constructing arguments, challenging assumptions, and thinking both creatively and critically. Ancient History students become knowledge creators, productive and discerning users of technology, and empathetic, open - minded global citizens."*

RECAP AND SOME FINAL POINTS ABOUT SCALING...

cont'd...

Across both Physics and Ancient History it is the range of results achieved in these subjects that can be significantly different, rather than the perception that one subject may be “easier” or “harder” than the other. This is especially borne out in a review of the syllabuses for these subjects which will show comparable intent and rigour.

Lastly, in this document QTAC has used subject terminology that will be applicable for students completing Year 12 and qualifying for ATARs in 2020. As part of Queensland’s move to adopt national curriculum, the following Mathematics subjects will be redesigned as follows:

- Mathematics A will become General Mathematics
- Mathematics B will become Mathematical Methods
- Mathematics C will become Specialist Mathematics

The Authors

Phillip Anthony, Chief Technical Officer

Phillip holds a range of postgraduate qualifications in international relations, Asian politics, commercial computing and administrative studies. Phillip has been with QTAC for over 25 years and has a wealth of technical knowledge unrivaled in the tertiary admissions sector. As Chief Technical Officer, Phillip ensures that QTAC systems deliver institutional admissions requirements with maximum efficiency and innovation.

Marco Lombardi, Statistician

Marco holds a master's degree in biostatistics and experimental statistics and a bachelors degree in statistics and information management. He has worked since 2012 as a statistical analyst in Italy and now Australia. Marco is responsible for the technical processing for the calculation of the ATAR for Queensland secondary students and will be managing the documentation of QTAC's ATAR processes.

www.qtac.edu.au/atar-my-path

Information accurate as at December 2019, please check website for updates.