

Report on scaling of stage 2 and stage 3 results

M.T. Partis

§1. Background

The system of scaling currently used in Western Australia for tertiary entrance purposes was introduced in 1998 and was modelled on the system used in New South Wales. It is normally referred to as average mark scaling (AMS)¹. Previously the Australian Scaling Test (AST) had been used as the anchor variable in the scaling process.

The Curriculum Council has been implementing changes in the structure of years 11 and 12. Three of the most significant changes are that

- the syllabuses in all courses offered in years 11 and 12 will conform to a three-tier structure with a lower level consisting of semester units 1A and 1B; an intermediate level consisting of semester units 2A and 2B; and a higher level consisting of semester units 3A and 3B,
- all intermediate and higher level courses will have tertiary entrance subject status, and
- from 2009 most year 12 students will be required to sit stage 2 or stage 3 examinations.

This report considers what modifications to the existing scaling model might be required to accommodate these new arrangements whilst preserving the integrity of the university entrance system.

There is broad agreement on the principles that should apply in any new arrangements which might be introduced:

- students should be encouraged to take courses which are appropriate for their abilities and aspirations,
- students should not be in the position of having to choose between maximising their tertiary entrance aggregates and taking the subjects best suited to their needs²,
- taking the more demanding stage 3 units should not disadvantage students,
- the public credibility and integrity of the university entrance system must be preserved, and
- the requirements of tertiary admission should not have a distorting effect on the courses of those students who are not university bound.

§2. Timetable

Most of the new courses will be introduced at year 11 level in 2008, and at year 12 level in 2009. Some new courses have already been introduced. It is intended that separate stage 2 and stage 3 examinations will be held for the first time at the end of 2009. For the 2007 and 2008 tertiary entrance examinations those courses which are being offered at two levels in year 12 will not have separate examinations. As an interim measure there will be composite³ examinations which cover both the lower and higher levels.

This means that any significant modification to the scaling process would be applicable to the 2009 year 12 cohort of students. This group of students will be in year 11 in 2008. In order for students, parents and schools to make informed choices it is essential that decisions are made by the relevant bodies and disseminated to schools by June 2007 at the latest.

A cautionary note should be inserted here. Although the details set out above reflect current Curriculum Council policy, the possibility of further changes cannot be ruled out. There may, for example, be

¹ More detailed information about the current system is available from the Curriculum Council and the Tertiary Institutions Service Centre (TISC).

² There is a widely held perception that, under the present system, some mathematically able students take TEE *Discrete Mathematics* rather than the more appropriate *Applicable Mathematics* simply to improve their tertiary entrance aggregates.

³ In Curriculum Council documentation the term 'generic' is used to describe these proposed composite examinations.

a decision to hold composite rather than separate examinations in some subject areas, there may be special arrangements needed for the language subjects, there may be pressure from schools to legitimise a year 12 pairing of units other than 2A + 2B or 3A + 3B, there may be ministerial directives to make changes. All of which means that any public pronouncements on the processing of the 2009 TEE results, and the requirements for university admission in 2010, should be accompanied by cautionary caveats. Indeed, any policies now put into place should be sufficiently flexible to allow for change – without disadvantaging students who have already selected their courses.

§3. The conservative response

The four public universities have been considering how their entry requirements need to be modified to reflect the new structure. In some subject areas prerequisites have been stipulated as an appropriate level of achievement, usually a scaled mark of 50% or more, in stage 3 units. This is really a continuation of the current policy of stating prerequisites in terms of tertiary entrance examination (TEE) subjects. For most practical purposes, however, this only applies to Mathematics, Physics, Chemistry and the LOTE subjects⁴.

It is fair to say that there is some nervousness within the universities about admitting students who have taken mainly stage 2 courses. Until the new system is fully operational there is no way of knowing what the standard of the stage 2 courses and examinations will be. This has caused some apprehension that students admitted on this basis might be ill-prepared for tertiary studies, and should not be accepted ahead of students who have taken the more demanding stage 3 courses.

A possible response to this dilemma might be to require a minimum number of stage 3 courses for the calculation of a tertiary entrance rank. This might vary from university to university, or indeed from course to course. So it might be a minimum of four stage three courses for a highly competitive degree course, but less for others. The reality is that there is variable demand for the many degree courses offered by the four universities.

There is a further difficulty with this approach. The Curriculum Council reforms have been motivated by the desire to open up the system and allow talented students access to higher education. This equity thrust is reflected in the decision to allow stage 2 courses to count for university entry. There would be particular problems in schools that were unable to offer a full range of subjects at both stage 2 and stage 3 levels. To put it bluntly a requirement that prospective university students would need to take at least four stage 3 courses⁵ would undermine an important part of the reform agenda.

Nevertheless, the specification of a minimum number of stage 3 courses, at least for the first two years⁶ of the new structure, remains an option unless a more acceptable solution is found.

§4 Introduction of a bonus

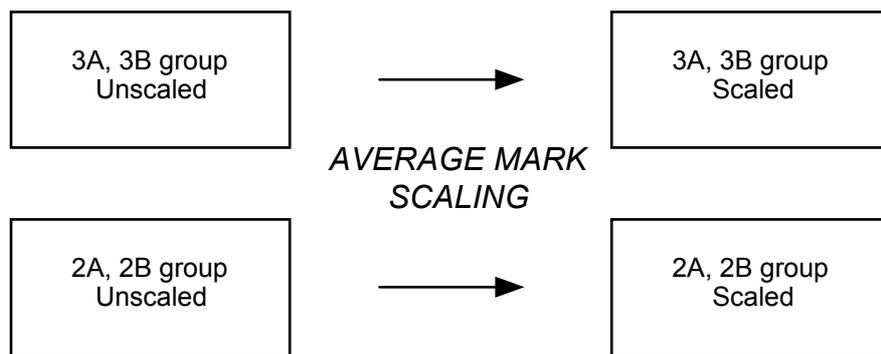
A different approach would be to introduce a numerical differential as an incentive for students to take the stage 3 courses. This would really be a device for separating the scores of the stage 2 students from those of the stage 3 students. Although this could be construed as a penalty on the stage 2 students, it would be more generally acceptable to describe it as a bonus for the stage 3 students. This adjustment could be made at a number of different points in the process of calculating a student's tertiary entrance rank.

Suppose that the current processes of standardisation and moderation are carried out to produce a combined subject mark for the group of students taking stage 2, and similarly for stage 3. The two groups, together with all the other upper and lower groups in each subject, could then be scaled in the usual way. In diagrammatic form the process would look something like this:

⁴ The requirement for English language competence is of a different nature and will be discussed in a later section.

⁵ It should be pointed out, however, that year 12 students normally take five or six courses, allowing the inclusion of one or two stage 2 courses.

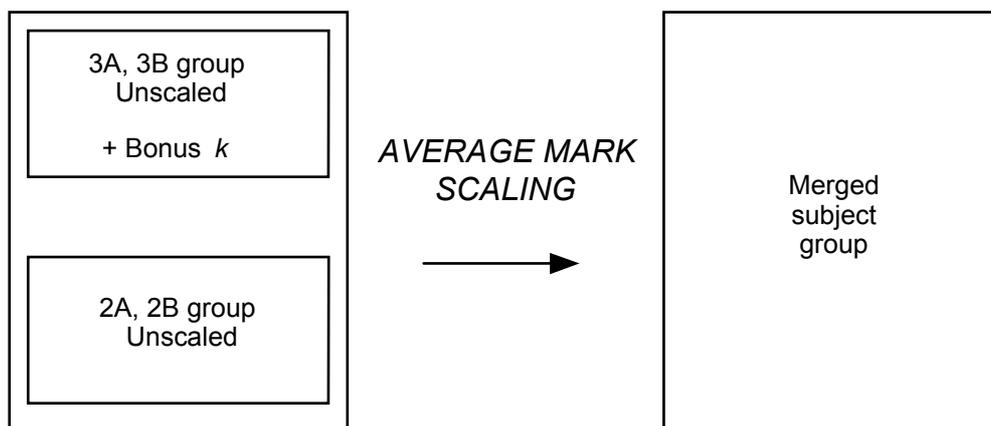
⁶ It is tempting to suggest that this option should apply for only one year, namely, for the 2009 year 12 cohort. Data would then be available to allow a different approach to be adopted for the 2010 year 12 cohort. However, this group of students would already be undertaking their year 12 courses, and any significant change to the rules of the game could not be countenanced.



As a consequence of the standardisation process the two groups on the left would have the same mean. The average mark scaling process uses performance across all subjects as an ‘ability measure’. It’s reasonable to conjecture that the 3A, 3B group will consist of abler students than the 2A, 2B group, and so will end up with a higher scaled mean. But this can only be a conjecture until there is some empirical data, and there is no way of quantifying the difference between the two groups at this time. Furthermore, the difference in scaled means between the upper and lower groups is certain to vary across different subjects.

It is possible that the average mark scaling process in itself would produce sufficiently higher marks for the stage 3 students, and that this would discourage students from opting for the lower level courses. However, there is a risk that a substantial number of students would opt for the lower level courses in the belief that this would generate a higher tertiary entrance aggregate. Indeed, there might be a particular problem with students who had taken units 2A and 2B in year 11, and then opted to repeat 2A and 2B in year 12. This could well be a good strategy if the only object were to maximise one’s tertiary entrance aggregate.

One way of applying a bonus would be to add k to the scaled marks of the 3A, 3B group, that is to say, add k to all the marks in the top right box. One problem with this is that the scaling process will already have introduced a differential between the lower and upper groups, and adding a further k might produce too big a gap⁷. A better approach would be to add the bonus to the 3A, 3B group before scaling. The marks of the 3A, 3B group (+ the bonus k) could then be amalgamated with the marks of the 2A, 2B group to create a single distribution for the subject. Again a diagram might help to clarify the concept:



There are two advantages in applying a bonus in this way. First, the bonus is carried through the scaling process and will be reflected in the final scaled marks. Second, the average mark scaling process will be applied to approximately 50 subjects, rather than about 100 subjects if the stage 2 and stage 3 groups are scaled separately.

A qualification needs to be made here. In a particular subject the number of students taking the stage 3 courses is unlikely to be the same as the number taking the stage 2 courses. This means that if, for example, there are many more students taking the lower level option their results will dominate the merged

⁷ It would be possible, of course, to only add a bonus if the differential produced by average mark scaling were less than some predetermined amount.

group as the scaling is applied. This could pull down the scaled marks of the stage 3 students, notwithstanding that a bonus had been applied.

The point made in the previous paragraph highlights the difficulty of making any decisions in the absence of empirical data. Until the first scaling exercise has been carried out on the 2009 results much of this analysis is speculative. Although many stakeholders might support the notion of adding a bonus k in the way indicated above, there are different opinions on what numerical value should be assigned to k . In the absence of hard data it's difficult to get a handle on this. Assigning a value to k is explored below.

§5 2006 Western Australian results

A pointer to a possible value for k is provided by the scaled means of the subjects examined in the 2006 tertiary entrance examinations⁸.

<i>Subject</i>	<i>Scaled mean</i>	<i>Subject</i>	<i>Scaled mean</i>
Chinese: Second Language	51.66	Geography	54.90
English	56.02	History	57.08
English Literature	64.39	Political and Legal Studies	60.04
French	65.26	Biology	56.13
German	63.10	Chemistry	62.77
Indonesian: Second Language	60.57	Geology	52.91
Italian	57.09	Human Biology	55.27
Japanese: Second Language	63.42	Physical Science	53.61
Chinese: Advanced	44.08	Physics	62.37
Indonesian: Advanced	44.19	Discrete Mathematics	54.87
Malay: Advanced	51.24	Applicable Mathematics	62.64
Aviation	50.02	Calculus	64.70
Accounting	55.94	Art	54.27
Information Systems	50.50	Music	61.58
Ancient History	59.90	Drama Studies	55.08
Economics	59.12		

If one considers cognate subject areas in the above table, more specifically,

English/English Literature,

Physical Science/Physics/Chemistry, and

Discrete Mathematics/Applicable Mathematics/Calculus,

it is clear that there is a difference in the scaled means between what are normally regarded as the easier and harder subjects of 8 to 10 marks. One needs to be careful about leaning too heavily on this data in assigning a value to k . The point must be emphasised that all the subjects in the above table have TEE status, and so are pitched at a standard more comparable to the stage 3 courses in the new structure than to stage 2. Nevertheless, it would be safe to conclude that an appropriate value for k would need to be greater than 10.

§6 New South Wales

In the New South Wales Higher School Certificate (HSC) similar problems of comparison have arisen with different solutions attempted. Years ago there were three levels of Mathematics in the HSC, with the separate courses having different maximum marks to reflect the difficulty of the course. This proved unpopular and was quickly abandoned. New South Wales has also attempted to align marks in related subject areas by setting up expert panels of teachers and examiners to make standards-based comparisons of examination papers. However, such an approach is labour-intensive and logistically complicated.

It is useful, though, to look at the way English and Mathematics are handled in the HSC. In English there are two separate year 12 courses: *English Standard* and *English Advanced*. Students take two examination papers in each of these courses with one paper being common. This allows a direct

⁸ The scaled means for 2005 were much the same except for a few small enrolment subjects such as *Chinese: Second Language*.

comparison of performance in the two subjects. There are also two additional courses: *English Extension 1* and *English Extension 2*, which can best be regarded as project work associated with the *English Advanced* course.

The Mathematics structure is similar, with separate courses in *General Mathematics* and *Mathematics* plus *Mathematics Extension 1* and *Mathematics Extension 2*. However, there is only one examination in each course so, unlike English, there is no overlapping content to allow a direct comparison to be made.

In the 2005 HSC the scaled means and candidature for these various courses were as follows⁹:

Course	Enrolment	Scaled mean
English Standard	30,140	35.6
English Advanced	27,542	62.4
English Extension 1	6,282	72.6
English Extension 2	2,608	73.2
General Mathematics	28,673	41.2
Mathematics	19,006	60.0
Mathematics Extension 1	9,359	79.2
Mathematics Extension 2	3,240	87.6

The large gap between the scaled means of the two main English courses is surprising – and this is based on a common paper taken by all students! It would be facile to assume that the New South Wales data would be replicated in Western Australia, but it does suggest that the bonus k should be much higher than 10.

§7 Victoria

Victoria has a different course structure, and there is little information to be gleaned which would shed light on scaling in Western Australia. There is only one English course, and so the problem of differential scaling doesn't arise. There are three courses in Mathematics: *Specialist Mathematics*, *Mathematical Methods* and *Further Mathematics*. Significant groups of students take both *Specialist Mathematics* and *Mathematical Methods*, making alignment of means fairly straightforward. A similar situation exists with *Mathematical Methods* and *Further Mathematics*, allowing all three Mathematics subjects to be aligned. In 2006 the scaled means in these three subjects were as follows¹⁰:

Course	Scaled mean
Further Mathematics	55.10
Mathematical Methods	71.26
Specialist Mathematics	81.26

An interesting aspect of the Victorian system is the existence of a LOTE bonus. This was introduced by ministerial directive some years ago. The bonus is applied by adding 5 marks¹¹ to the scaled mean of all LOTE subjects. There appears to be no evidence that the existence of this bonus has produced any increase in enrolments in the LOTE subjects.

§8 Back to k

If a bonus of k were to be introduced, a strong case could be made for assigning a different value of k for each subject. However, there is an insurmountable difficulty in trying to do this in the absence of any statistical data. The situation will be different after the 2009 examinations. By then it will be possible to make informed decisions about the size of bonus which would be appropriate in each subject area. The only safe policy is to have the same bonus in all subjects for the 2009 and 2010 examinations, and to explore the desirability of having subject-specific bonuses in early 2010.

⁹ In New South Wales and Victoria marks are reported out of a maximum of 50; the scaled means from those two states have been doubled to make them directly comparable with the Western Australian data.

¹⁰ See the previous footnote.

¹¹ The LOTE bonus in Victoria would be equivalent to 10 additional marks in the Western Australian system.

If the bonus is to have the desired effect it has to be announced in advance, that is to say, by mid-2007. If the bonus is too small it is unlikely to persuade border-line students to opt for the stage 3 courses. On the other hand, if the bonus is too large students taking the stage 2 courses will find it hard to achieve marks which would make them competitive for the purposes of tertiary entry.

Well, enough prevarication! There has to be a value assigned to the bonus k if that approach is to be adopted. The recommendation in the final section of this report is that k be set at 15 for 2009 and 2010.

§9 Other options

Earlier in this report three possible approaches were canvassed, namely,

- allow the average mark scaling to operate on all the stage 2 and stage 3 courses,
- apply a bonus to encourage students to take stage 3 courses, or
- require students to take stage 3 courses to qualify for university entrance.

Two other options should be considered.

The New South Wales English marks are aligned by requiring students to take a common paper. It would be logistically impossible, and educationally undesirable, to introduce common papers across the whole range of subjects here, but it might be possible to have common content in the stage 2 and stage 3 examinations. The overlap would need to be about 20% to 25% of each paper in order to make a statistically reliable comparison. But the common content would have to be accessible to both cohorts and so would need to examine material in the 2A, 2B courses. However, the stage 3 students would have taken the 2A, 2B courses in year 11 and would have had an extra year's exposure to the discipline. In the other words the marks on the common content would not be a simple measure of the relative ability of the two groups, but would give a significant advantage to the students who had greater acquired knowledge in the subject. Furthermore, the requirement to set papers with overlapping material is likely to have a distorting effect on the examination as a whole.

Another possibility would be to use the General Achievement Test (GAT) to compare the ability levels of the groups of students taking the stage 2 and stage 3 courses, and then make appropriate adjustments to the marks. The GAT was not set up for this purpose and there are, in any case, methodological problems with this type of test. Not the least of the difficulties is that there is no incentive for students to take the test seriously and so it is questionable how much reliance can be placed on the results. In any case the GAT tests, or purports to test, inherent academic ability, rather than demonstrated academic achievement¹². There is the further point that use of the GAT for scaling purposes would be unlikely to provide the necessary incentive for abler students to take the stage 3 courses.

§10 English language competence

The universities require all entrants to demonstrate English language competence. For local students the benchmark is currently a 50% scaled score in TEE *English* or *English Literature*¹³, though there are safety-net provisions that allow students to satisfy the English language competence requirement in other ways. There are interim arrangements in place for 2007 and 2008, but from 2009 onwards there will be separate stage 2 and stage 3 courses and examinations in English. In effect the stage 2 and stage 3 courses will absorb the student cohorts who previously took *Senior English* or TEE *English*. Hence, virtually all year 12 students seeking university entrance will take English or the new Literature course at either stage 2 or stage 3 levels.

The question that arises is how to specify the English language competence requirement in the new structure. The conservative answer is to require students who seek university admission to obtain a scaled mark of 50% in stage 3 English. This would put pressure on all schools to offer English at the higher level, and would mean that any student who wished to keep open the option of tertiary study would be coerced into the 3A, 3B course. In these circumstances the stage 2 English course would become a poor relation. Furthermore, if a bonus were applied in the manner indicated in §4, there would not be a separate scaling of

¹² Similar criticisms were made of the Australian Scaling Test when it was used as the anchor variable in Western Australia. This was part of the rationale for switching to Average Mark Scaling.

¹³ Different arrangements are in place for students whose first language is not English.

stage 3 English independently of stage 2 English. Whilst an algorithm could be devised to address this problem, this would be an unnecessary and confusing complication.

In any case the historic use of a scaled mark of 50% in TEE *English* sets a norm-referenced rather than a standards-referenced hurdle. An alternative approach to that canvassed in the previous paragraph would be to require a scaled mark of 50% in the merged English group, that is to say, the stage 3 cohort (with a bonus of k) amalgamated with the stage 2 cohort. Good students, encouraged by the bonus, would almost certainly opt for the stage 3 course, whilst able students who performed well in the stage 2 course could still satisfy the English language competence requirement.

Since this would still be a norm-referenced criterion it is worth making the point that average mark scaling produces a global scaled mean across all subjects and all students of approximately 58. Since most students are likely to take the new English course at one of the two levels, the scaled mean of the merged English group will inevitably be close to 58. In fact the 2006 scaled mean for TEE *English* was 56.02. This was almost certainly because a cohort of able students was 'siphoned off' into TEE *English Literature*. The same phenomenon may well occur under the new structure with some of the more able students opting to take the new Literature course instead of English.

§11 Mathematics in 2009

Under present plans all courses will be fully operational at stage 2 and stage 3 levels by 2009 except for the new Mathematics courses. These are scheduled for introduction at year 12 in 2010, a year later than the other courses. This presents a one-off problem with the scaling of the Mathematics units in 2009. In that year the subjects available will be TEE *Applicable Mathematics*, *Discrete Mathematics* and *Calculus*. In the new terminology these are all effectively stage 3 courses, not underpinned by stage 2 courses. As a consequence the average mark scaling process would push the scaled marks for all three subjects up significantly by comparison with previous years. This distortion would give students taking Mathematics an unfair advantage over their peers who were not taking Mathematics.

One way of handling this problem would be to restrict any movement of the scaled means in the three Mathematics units to not more than, say, 5 marks above or below the 2008 scaled means for those subjects. This might be called a blunt instrument solution, and would be difficult to justify in a public arena. In any case it could still give the Mathematics students an unfair advantage.

Another option would be to fix the scaled means of the Mathematics units in 2009 at the average for the three previous years. Although, at first sight, this would provide some sort of consistency, it ignores the fact that there are likely to be major perturbations of the scaled means in all other subject areas. This option would not guarantee fairness across all courses.

A possible solution might be to scale the Mathematics units against just the stage 3 units in all other disciplines. This would involve more computational complexity¹⁴ but might be easier to defend. More investigation is needed to come up with an acceptable solution.

§12 Scaling population

Under the current system the Curriculum Council has procedures in place to identify anomalous performers in the TEE. These are students whose examination results differ markedly from their school assessments. The anomalous performers are removed from the school moderation process, and then slotted back in afterwards for average mark scaling.

When the new system is in operation a much higher proportion of the year 12 cohort will be taking external examinations. It is likely that there will be a significant increase in the 'non-trier' group, that is to

¹⁴ I hesitate to get too far into the detail of how this might be done, which is why I have buried this comment in a footnote. The process would probably have to look something like this:

1. Apply the bonus in the manner indicated in §4 to all courses except the three Mathematics units.
2. Carry out the average mark scaling on all courses except the Mathematics units.
3. Identify the results of the stage 3 students in the merged scaled groups, and create a new subgroup for these students.
4. Carry out a new average mark scaling on all the stage 3 subgroups, this time incorporating the Mathematics units.
5. Use the results obtained in step 2 as the scaled marks for all subjects except Mathematics.
6. Use the results obtained in step 4 as the scaled marks for the Mathematics subjects.

say, students who take an examination because they are required to do so, but do not make a serious attempt at the examination paper. It is safe to predict that this will be more of a problem in the stage 2 courses, and that the pattern across all subjects will be uneven. The presence of significant numbers of 'non-triers' in any course could distort the scaling process and disadvantage other students taking that course. Once the 2009 data is available it will be possible to see whether there is a real problem with 'non-triers', and whether some modification of the scaling process is needed to counteract it.

Consideration also needs to be given to the composition of the scaling population, that is to say, those students whose results are used in the average mark scaling process. The most straight-forward solution is to define the scaling population as those students who have completed at least four year 12 courses.

§13 Conclusion and recommendations

The purpose of this project was to identify ways in which the scaling process used might be modified to take account of the structural changes in years 11 and 12. A particular concern for the universities is to preserve the integrity of the tertiary entrance system. In the final analysis this is based on the principle of selection on academic merit, and the problem is how to interpret this for the stage 2 and stage 3 courses now being introduced.

The core difficulty in this exercise is the lack of empirical data which would allow proper statistical analysis. Although there is historical data available for the Western Australian TEE and from other jurisdictions, there is a real risk that this data will mislead rather than inform.

It should be emphasised that the lack of data is only an interim problem. Once the 2009 results have been processed and analysed it will be possible to put in place a system that has more rigorous foundations.

Inevitably, then, some of the suggestions made in this report will appear arbitrary and contentious. The recommendations below are designed to carry the tertiary entrance system through the transitional years of 2009 and 2010.

Recommendation 1

That, for the 2009 tertiary entrance examinations, in all subjects except Mathematics,

- (a) a bonus of 15 marks be added to the unscaled stage 3 marks,**
- (b) in each subject the stage 3 marks, incorporating the bonus, be merged with the stage 2 marks in the same subject to form a common distribution,**
- (c) average mark scaling be then applied to the merged subject groups.**

Recommendation 2

That, for the 2010 tertiary entrance examinations,

- (a) a bonus of 15 marks be added to the unscaled stage 3 marks,**
- (b) in each subject the stage 3 marks, incorporating the bonus, be merged with the stage 2 marks in the same subject to form a common distribution,**
- (c) average mark scaling be then applied to the merged subject groups.**

Recommendation 3

That a review of scaling methods based on a detailed analysis of the 2009 results be carried out early in 2010.

Recommendation 4

That, for the 2009 and 2010 tertiary admission exercises, a scaled mark of 50% in either the merged English subject or the merged Literature subject be accepted as satisfying the requirement for English language competence¹⁵.

¹⁵ This assumes that special arrangements would continue for those students whose first language is not English.

Recommendation 5

That, for the 2009 tertiary entrance examinations, the Mathematics units be scaled against the stage 3 courses in all other subjects¹⁶.

Recommendation 6

That the scaling population be defined as those students who have completed at least four courses in year 12.

¹⁶ The mechanics of how this should be done require further investigation.

Appendix

In the preparation of this report a number of individuals were consulted. Their collective input proved invaluable, though, of course, the contents of the report are the direct responsibility of the author. The full list of people interviewed is as follows:

Bill Louden	Curriculum Council
David Wood	Curriculum Council
Jenny Morup	Curriculum Council
Alan Honeyman	Curriculum Council
Murray Jackson	Curriculum Council
George Leung	Curriculum Council
Steve Hoath	TISC
Judy Cresp	TISC
Chris Cook	Department of Education and Training
David Axworthy	Department of Education and Training
Lois Joll	Department of Education and Training
Steve Donatti	Department of Education and Training
Valerie Gould	Association of Independent Schools
Mark Newhouse	Association of Independent Schools
John Nelson	Catholic Education Office
George Cooney ¹⁷	Macquarie University
Pat Garnett	Edith Cowan University
Ken Harrison	Murdoch University
Bruce Shortland-Jones	Curtin University
Alan Rendell	Curtin University
Jane Long	University of Western Australia
Wayne Betts	University of Western Australia
Rob McCormack	University of Western Australia
Greg Marie	University of Western Australia
Mary Carroll	University of Western Australia

Many thanks to all the above¹⁷ for their helpful and constructive advice

¹⁷ Professor Cooney is the long-serving chairman of the scaling committee in New South Wales.