MEETING THE CHALLENGE OF ASSESSING IN A STANDARDS BASED EDUCATION SYSTEM

Professor Jim Tognolini

6 December 2006

MEETING THE CHALLENGE OF ASSESSING IN A STANDARDS BASED EDUCATION SYSTEM

This report has been produced by Professor Jim Tognolini, formerly of the Australian Council for Educational Research (ACER) and currently Director of Research and Assessment at Pearson Research and Assessment. It was presented to the Curriculum Council at its 6 December meeting.

The Curriculum Council has endorsed recommendations 1, 2, 3, 4, 5, 7 and 8 and has referred recommendation 6 to an expert assessment measurement advisory group to be convened in 2007 to provide advice to Council.

1. THE PURPOSE OF THIS REPORT

The purpose of this report is to address a number of issues about assessment that have emerged as a consequence of the changes that have taken place since the inception of new Western Australian Certificate of Education (WACE) courses in Years 11 and 12 and to provide advice on how these issues might be resolved in the light of the recommendations of the Andrich report.

2. THE STRUCTURE OF THIS REPORT

This report outlines the background to senior secondary schooling reform in Western Australia, summarises the key issues identified by Professor Andrich, and explores ways in which the Council can meet the technical challenges identified without compromise to the spirit of the Curriculum Framework and the vision embodied in the *Our Youth, Our Future* report (2002).

Appendix 1 provides a summary of the Andrich report.

Appendix 2 reviews assessment, moderation and tertiary selection practices in other jurisdictions.

3. BACKGROUND TO THE SENIOR SECONDARY SCHOOL REFORM

In 2002, the *Our Youth, Our Future,* report a vision for the reform of senior secondary schooling in Western Australia, was published. It was the culmination of an extensive review of senior secondary schooling which found a strong consensus across the community that the time for significant reform was well and truly due.

The review and its report focused directly on meeting the needs of young people bound for a diversity of post-school destinations. It articulated a reform agenda that would see the introduction of a new WACE for students graduating from school between 2005 and 2009. *Our Youth, Our Future* describes a system that is tailor-made for Western Australian students and will bring Western Australia into line with other states and territories which have already overhauled their senior schooling courses.

Senior secondary schooling, as described in *Our Youth, Our Future,* presents all students with genuine options linked to their interests and abilities, regardless of socio-economic background or geographical location. The commitment to equity that permeates *Our Youth, Our Future* is both a philosophical and a research-based position that counters beliefs which emphasise the fixed nature of intellectual abilities. Its fundamental premise is based on research that supports the idea that cognitive abilities are not dependent solely upon innate ability but that they are developed through socially supported interactions. *(How People Learn: Brain, Mind, Experience, and School Committee on Developments in the Science of Learning. Bransford, Brown, and, Cocking, (Eds.) http://www.nap.edu/openbook/0309065577/html/index.html.)*

Acknowledging that one size does not fit all, the *Our Youth, Our Future* report delineates the imperative to meet the needs of all students as the fundamental driver of reform in senior schooling. It argues that schools must provide for students who are bound for work, trades and other vocations by 'selecting' the right students for such a pathway and providing them with appropriate educational and training foundation (in both school and workplace/community settings) to commence vocational learning. Similarly, schools must also provide for students who are university-bound and not only select the 'right' students, but provide them with appropriate academic foundations for university. In summary, the reform called for courses that would be relevant to the futures of all young people and which would give them opportunities to achieve improved education outcomes.

Defined by its legislative obligations, and guided by national agreements, contemporary research and advice received through the consultative processes, the Curriculum Council endorsed the following principles to underpin the new system.

- *Prepare young people to move into adult life, including work, higher education and vocational education.*
- *Learning outcomes for students should be maximised through a seamless focus on outcomes from kindergarten to Year 12, as expressed in the Curriculum Framework.*
- *Curriculum should be designed around learning outcomes that reflect general and vocational learning which can be achieved by all students to some extent.*
- *There should be flexibility to cater for the overall development of individuals; for different learning communities; and for recognition and valuing of diverse backgrounds and rates of learning.*
- *Opportunities for breadth and depth of learning should be embedded in a range of contexts and include the Curriculum Framework overarching learning outcomes.*
- *Standards should be made explicit to students, teachers, parents and post-school destinations through scales of achievement for each outcome that provide an effective basis for assessment, evaluation and reporting.*
- *Assessment should be valid, reliable, fair, educative, explicit and comprehensive, as described in the Curriculum Framework.*
- *The award of a Western Australian Certificate of Education (WACE) should reflect broad and significant achievement in senior secondary education.*
- *Structures should be designed to adapt to change.*

In relation to these principles, it should be noted that the Curriculum Council does not have a role in the way teaching and learning programs based on the courses of study are delivered in schools. (Our Youth, Our Future p.13, 14)

In other words, schools and school sectors are intended to have significant autonomy in deciding how they teach and how they assess student achievement in these courses.

3.1 The role of the Curriculum Council

Section 12 of the Curriculum Council Act 1997 specifies the functions of the Curriculum Council in relation to assessment. The Act specifies the Council's responsibility for establishing and implementing procedures for:

- accreditation of courses
- assessment of achievement of students undertaking senior secondary education, and the proper conduct of that assessment, including school and external assessment for the purposes of certification
- ensuring the comparability of assessments of student achievement.

This means the Curriculum Council is responsible for placing certain broad requirements on schools in relation to the processes associated with monitoring comparability, collection and submission of school-based assessment. It prescribes the framework, frequency and format for reporting but not the detail of assessment practice.

Schools are required to offer accredited courses and use their course standards; however, within the broad specifications of each course, schools have responsibility for more local decisions such as those associated with internal moderation, unit score calculation and the development of a school assessment policy.

4. STANDARDS-REFERENCED ASSESSMENT IN AN OUTCOMES AND STANDARDS EDUCATION SYSTEM

4.1 Introduction

Western Australia has a state-wide, mandated standards-referenced system, as defined by the Curriculum Framework and its progress maps. This system is characterised by a curriculum structured into eight learning areas, which in turn are described in terms of outcomes (composed of aspects) which are divided into eight levels that represent the 'path of learning' for that learning area. Courses are derived from the learning areas and are made up of semester units that have sequentially different degrees of cognitive demand. In most courses there are 6 units inclusive of standards from levels 4-8. However, in recognition of students with special needs, there are courses that have additional units targeting entry between foundation level to level 4. Units in each course have a syllabus that describes learning outcomes and content. The learning outcomes are the focus for teaching and learning. Student achievement is assessed against the outcomes using the course standards which are also referred to as scales of achievement.

One of the main advantages of a standards-referenced assessment system is that the results can indicate what it is students have achieved during the course. At the same time the scores that arise from the assessments can be used to locate the position of the students relative to the overall performance of the cohort. In other words, it is possible

to assess performance for certification and credentialing, and, at the same time, use the same scores for university entrance purposes (as is currently the case).

A standards-referenced system is not unique to Western Australia or to primary and secondary education. All states and territories in Australia have organised their curriculum around a standards-referenced system. Similarly, the International Baccalaureate (IB) Diploma is standards referenced, as is the Programme for International Student Assessment (PISA). Locally and nationally, tertiary education providers are implementing a standards approach. Educational outcomes clearly and unambiguously specified, have significant power and appeal in more globalised contexts as they contribute to a mechanism for tracking and comparing outcomes over time and across jurisdictions both nationally and internationally.

Standards-referenced assessment systems, like the one currently being adopted as part of the new senior secondary school curriculum reform in Western Australia, have the capacity to provide all students with a meaningful record of their achievements in all subjects throughout the senior secondary years. At the end of their schooling students are provided with a description of the types of skills that they have acquired in a subject. If used judiciously, this information facilitates smoother entry through different pathways into higher education and the workforce. In addition it enables schools, teachers and parents to see clearly their child's progress on a learning continuum defined by the outcomes and levels of achievement expressed in the Curriculum Framework.

While standards-referenced assessment systems offer valuable potential benefits, they also impose a number of requirements to ensure the integrity of the resulting measures.

4.2 *Requirements of standards-referenced systems*

If standards specify what students should be taught and what and how well they should be learning, and if assessment systems, be they external or school-based, measure what students know and can do, then the standards and assessments need to be closely synchronised. In the absence of strong synchronicity, assessments cannot provide accurate inferences about students' attainment of, and progress toward, standards.

One of these requirements therefore is that assessment tasks, whether constructed by the teacher or examiners, must link directly to the learning outcomes. These tasks provide the students with the opportunity to demonstrate what it is they know and can do. These expectations are captured by the course standards. There must be primacy of the learning outcomes in a standards-referenced assessment system.

The course standards are explicit statements of student performance that describe the levels of achievement for a course. These standards are derived from the outcomes in the Curriculum Framework and show development in relation to the construct being assessed. The performance of students as reflected through the assessment tasks is then assessed either directly or indirectly using these course standards.

A second requirement is that teachers must collect evidence to support their decisions regarding the performance of their students.

Another requirement is that a standards-referenced assessment system must have a reporting structure that is in harmony with the basic requirements of outcomes and standards education.

Still another requirement is imposed by the system when the information obtained from assessment tasks is to be used for different purposes eg. secondary certification (exit) and tertiary entrance (selection). For example, where students require a tertiary entrance score (TES), teachers should produce a school assessment score that reflects the relative differences in performance between the students at a fine enough level of precision to be useful for statistical moderation.

Recommendation 1
(Primacy of the Curriculum Framework and standards)

It is recommended that:

The Curriculum Council re-affirms the outcomes and standards base that under-pins the current senior secondary education reform and acknowledges that this approach is consistent with other Australian and international jurisdictions.

4.3 The Andrich Report

In July 2005 Professor David Andrich was commissioned to investigate issues around the comparability of standards for the new courses. The terms of reference specified that the report should investigate:

- the academic rigour of the course assessment processes
- the capacity of the assessment processes to provide sufficiently fine-grained information for university entrance
- whether the assessment processes provide reliable measurement that is susceptible to scaling and moderation.

The final report was released in September 2006 and is available from the Curriculum Council website: http://newwace.curriculum.wa.edu.au/pages/publications_report.asp .

Appendix 1 provides a summary of its argument and recommendations.

The requirement for fine grained assessment information has created tensions about coarseness of scale and use of marking rubrics and raised issues regarding the precision of measurement. These in turn created the need for the Andrich report that focused upon the role of measurement within the new system.

4.4 Challenges in a standards-referenced system

While the assessment process is quite straight forward, there are numerous points in the process that require judgement and interpretation and these present significant challenges to teachers, examiners and administrators. Learning outcomes are intended to describe what it means to progress through an area of learning. This path is not deterministic and hence there is scope for this developmental sequence to be challenged by data. Setting examination questions and assessment tasks that accurately assess the learning outcomes that are consistent with the requirements of the standards and are technically correct (particularly when there is no opportunity to pre-test the questions) is difficult. Setting marking keys that are fair, accurate and appropriate for the purpose for which the results are to be used and ensuring they are consistently applied is challenging. Accurately establishing the standards and presenting them to teachers, examiners and students in a manner in which they will all interpret them consistently is also a challenge, as is operationally defining the boundaries of the standards in the context of external and internal assessments.

4.5 Standards-referencing in Western Australia

Essentially, there has been a three-tiered system of subjects operating in Western Australian senior secondary schooling: Tertiary entrance examination (TEE) subjects and wholly school assessed (WSA) subjects and vocational education and training (VET) programs. All have operationalised standards-referenced assessment in some measure. TEE subjects have done so through grade related descriptors, and WSA subjects through performance criteria; however, they do not share a common standards framework. While each tier provides for particular groups of students, there has been no single achievement scale to monitor student performance irrespective of the tier. The new WACE provides standards against which student achievement can be measured. It provides a mechanism for articulating current subjects from across the three tiers.

Following feedback from teachers about implementation matters, particularly in relation to assessment, the Council resolved to conduct a course refinement process. In May 2006, the Curriculum Council announced two significant changes to the assessment process. The first of these gave the teachers more flexibility in their assessments by enabling them to use marks as part of the assessment process.

The second change dealt directly with the tertiary entrance issue. The decision was made to continue to use the external examination to statistically moderate the school assessment scores for tertiary entrance purposes. It also made the point that during the transition period there would be a scaling process to ensure that standards for new courses and existing TEE subjects were comparable. Universities have endorsed this approach.[1]

These changes have had an impact on the requirements associated with the new senior secondary school system. The situation has changed from one where there was going to be minimal use of statistical moderation to one where the school assessments that are to contribute to the tertiary entrance score (TES) will be statistically moderated before being combined with the examination score.

This means that the requirements for teachers have also changed. Teachers, for example, must now ensure that school assessments reflect the relative differences in

[1] Letter from the Curriculum Council to school principals, 10 May 2006.

performance between the students at a fine enough level of precision to be useful for statistical moderation. That is, the ranked order of the students is important and the relative difference in the marks is also important.

The Andrich report, with its focus on the need for high quality measurement, was largely instrumental in these changes being proposed and implemented. The report makes clear the need for reliable and valid assessments in all courses. It reinforces the need for the Curriculum Council to make clear to teachers that marks have meaning and that this meaning can vary according to the way in which the marks are to be used.

5 KEY RECOMMENDATION OF THE ANDRICH REPORT

The key recommendation from the Andrich report is

> " …that for both school based and external assessments, analytical marking of the traditional kind using marking keys that arise directly out of the assessment tasks, be used for student assessment for each unit of a course, and for each course as a whole at the end of Year 12.
>
> A related recommendation is that, simultaneously, a rating of student performance into one of eight generic levels of achievement that arises out of the outcome statements be used as part of the assessment. The former provides marks for the assessment and measurement of students at a relatively micro level suitable for feedback to students and for use in tertiary selection according to the policies of the Curriculum Council.
>
> The latter provides ratings for classification at a relatively macro level suitable for monitoring the general progress of students and the operation of a course and is commensurate with the generic nature of the level and outcome statements. The two assessment processes, distinguished by their level of precision and relevance, are compatible and can be combined and integrated. By taking advantage of this complementarity, the Curriculum Council can genuinely advance the communication of educational achievement in Western Australia"
>
> (Andrich, 2006, p. 2)

I have no hesitation in supporting this recommendation and the other recommendations in the report. They are focused upon improving assessment in the new system and providing direction about the professional development needs of teachers engaged in assessing students within such a system.

6. SPECIFIC ISSUES IN ASSESSMENT IN THE NEW WACE SYSTEM

6.1 The issue of prescription in the new system

Traditionally, the Curriculum Council and its predecessor the Secondary Education Authority (SEA) have not been prescriptive in dictating the types of marking keys that teachers have had to use and they have not been prescriptive in dictating the coarseness of the unit of measurement.

Historically, broad parameters and subsequent requirements for assessment have been prescribed. Teachers have been made aware of their responsibilities within these parameters to ensure the integrity of the outcomes. There has been professional development to support teachers in their role and there have been established procedures to verify and endorse the results.

When the new WACE system was first introduced, Western Australia was aiming to achieve a number of unique outcomes. One of these was to have school assessment contribute to the tertiary entrance score (TES) in the way that it had done in the past, without first statistically moderating it to ensure comparability across the state. While the intention had merit for a number of reasons, the requirement to produce a single rank order of merit for the purpose of tertiary entrance by combining it with an examination score, meant that the solutions nominated by the Curriculum Council were dubious from a statistical and measurement point of view. This necessitated the reintroduction of statistical moderation.

In the original scenario, teachers would be required to report the sum of the students' standard for each course outcome which would contribute directly to a TES without statistical moderation. There was undoubtedly a need to provide teachers with specific instructions as to how to make a judgement about the standard a student had achieved and the subsequent school assessment measure for inclusion in the TES. This led to the Curriculum Council being very prescriptive and providing specific instructions as to how teachers must assess and arrive at their final score for students. The prescription led to the requirement that all courses be treated the same. This led to some of the tensions about assessment within the system since its inception.

These tensions resulted in changes to the new system. The fine-grained comparability that is needed for the generation of the TES is now achieved by the familiar statistical moderation process that has been used since the introduction of school assessment into the tertiary entrance scoring process. Thus, the need for prescription has diminished.

The current situation means that the requirements for generating the school assessment component will be much the same as in the past. The most critical difference is that all of the assessments (school and external) will be derived from an assessment of the course outcomes and the marks will relate directly or indirectly to the course standards.

Teachers in current TEE and WSA subjects in Western Australia are experienced in assessing students at school within the assessment frameworks (tables) provided by the Curriculum Council. The requirement in the new system is similar in the sense that assessment frameworks, focussing on course outcomes will guide the assessment activities of teachers and examiners. In the traditional TEE system, teachers of tertiary

entrance subjects have also been experienced at producing a school assessment score that is generally reported on a scale between 0 and 100.

Recommendation 2
(Curriculum Council Prescription)

It is recommended that:

the Curriculum Council provides broad frameworks and guidelines for assessment in each course. These guidelines will address the balance of assessment types, weightings, the number and format of assessments required, and coverage of outcomes for each course.

It should be stressed that 'a scale between 0 and 100' carries with it the implication that marks take on any values in this range, with most being somewhere between 40 and 100. It is not intended that marks only take on values of, for example, 100, 90, 80, 70, 60 and 50. This does not convey the precision implied by "a scale between 0 and 100".

Traditionally, teachers have been aware that the school assessments used for the purpose of tertiary entrance must reflect the relative differences in performance between the students at a fine enough level of precision to be useful for moderation. That is, the ranked order of the students is important and the relative difference in the marks is also important.

Problems occur if either of these relativities is violated. Teachers should use parts of the scale to demonstrate 'how good students are relative to other students within the school cohort in the subject'. If the scale used by the teacher does not enable discrimination among the performances of the students then some students may be disadvantaged and others advantaged by the process. This is not fair.

Teachers of TEE subjects have traditionally been able to choose their own marking keys and numeric scales. It is important to note that teachers have had the freedom in the process of awarding marks in their schools with the provision that it meets the requirements summarised above.

Some use detailed marking keys of an extended task or performance to arrive at the marks for the various assessment tasks. Others use more holistic marking keys of small components of assessment; perhaps a short essay, a performance task. Some teachers use combinations of both depending on the kinds of tasks they set. In each case, however, the final mark built up for each student to be used in statistical moderation is fine enough so the students are separated meaningfully on the scale at a relatively finer degree.

In the TEE system the main requirement for teachers within a subject is to ensure comparability within the school. Statistical moderation carried out by the Curriculum Council then ensures comparability across schools within the subject.

At one level it does not matter that different teachers approach the collection of marks in different ways, providing they separate the students on a valid achievement scale

precisely enough for statistical moderation to take place. In reality, it matters that school assessment provides a ranked order of merit within a subject within a school and locates each student relative to others. The relative differences in the scores are critical as they are maintained after moderation.

The main point is that teachers have traditionally been made aware of the requirements of the system. They are informed of their assessment responsibilities within the system. They are provided with information regarding the measurement consequences of the assessment strategies that they choose and then empowered to take responsibility for their own marking.

I recommend that the Curriculum Council revert to its more traditional role in the area of assessment and provide a minimal level of prescription in the new system, with reference to the types of marking keys that teachers will use and the coarseness of the unit of measurement.

Recommendation 3
(Curriculum Council Prescription)

It is recommended that:

the Curriculum Council provides a minimal level of prescription in the new system, with reference to the:

- **types of marking keys that teachers will use**
- **coarseness of the unit of measurement**.

While this is not an explicit recommendation of the Andrich report, it is inferred consistently when it refers to traditional marking and providing professional development throughout the system so educational personnel are aware of the consequences of their assessment decisions. It specifically recommends provision of further professional development to:

> relevant educational personnel and to principals regarding the broader context of location of Year 11 and 12 study, the constraints imposed by competitive tertiary selection into particular courses, the advantages and disadvantages of the process implemented in responding to these constraints, and potential alternatives with their advantages and disadvantages as exemplified in other countries or other states.
>
> (Andrich, 2006, Recommendation 1)

It also refers to the need for the Curriculum Council to provide professional development to help teachers select the most appropriate unit of measurement for their situation:

> ...professional development should be provided by the Curriculum Council to principals, teachers and students regarding the

> arbitrariness of measurement units in educational assessment and the implications this has for placing the assessments on the same scale and ensuring that other policies of the Curriculum Council are applied correctly.
>
> (Andrich, 2006, Recommendation 4)

It is quite clear that Andrich is strongly advocating empowering teachers to take responsibility for choosing their own unit of measurement and their own marking keys in much the same way they have done in the past. He recommends that teachers and other educational personnel be provided with the knowledge and skill to do this effectively within the requirements of the standards-referenced system.

It should be reinforced that what is asked of teachers in the new system is much the same as what has been asked in the past, except that the assessment tasks in a standards-referenced system (and accompanying marking keys) must be more explicitly mapped to the learning outcomes and the content than they may have been in the past. The assessment tasks must provide students with the opportunity to demonstrate what it is they know and can do in relation to the curriculum and the course standards.

If the Andrich report's recommendation 1 is accepted by the Curriculum Council it will mean the Council has the basis to respond to issues such as the one related to the way teachers transform their marks into school assessments and achievement levels.

The next section considers this specific issue.

6.2. The issue of direct/indirect/composite marking

Each course outcome has a standard that is partitioned into five progressive levels of achievement, generally levels 4 - 8. There are 17 courses with levels Foundation-8. The higher levels correspond to higher standards of performance. These levels are explicit statements of student performance that describe progressive degrees of achievement. The course outcomes are derived from the Curriculum Framework outcomes and show development in relation to the construct being assessed. The performance of students as reflected through the assessment tasks is then referenced to these levels.

For each outcome, progress towards achievement of the next level of the course standards is recorded using three bands–first (F), middle (M) and high (H) band. Teachers refer the performance of the students, as determined through the school assessment, to the course standards for each outcome to make a professional judgement regarding, first, the level completely achieved and then the location of the student's performance within the next level. If the students are within the first band they are classified as having completely achieved the level and this is recorded as F; the second band level of achievement is recorded as M (middle band); while third band is recorded as H (high band). For example, a student who is demonstrating skills associated with the middle of level 5 is labelled 4M.

One of the most contentious assessment issues associated with the introduction of the new system has been the generation and use of marking keys. In the initial documentation from the Curriculum Council, teachers were required to develop tasks to collect information on the particular outcomes and then use the levels of achievement to identify the performance standard of students. The marking keys for the tasks are based on the achievement levels. There is a direct alignment of the task with the achievement

levels. (Please note that for the remainder of this report, the term 'achievement levels' will be used to refer to the bands associated with the levels.)

The method of arriving at a level for a task using the course standards directly for each task is referred to as "direct marking".

Where there is more than one task per outcome, the intention is that an 'on-balance' judgement is made to determine the achievement level for each outcome. The achievement level for each outcome is then aggregated to produce an overall achievement level for the unit. Teachers use a 25-point conversion table which maps the achievement level to a numerical score.

At the end of a unit within a course, teachers were requested to submit an achievement level for each course outcome to the Council.

In May 2006 the Curriculum Council provided teachers with more flexibility by enabling them to use marks as part of the assessment process. This means that teachers could submit the achievement level plus a numerical score for each course outcome to the Curriculum Council.

The intention of focusing the attention of teachers on the outcomes and the standards was to ensure that everything was aligned and teachers would be making judgements against the same 'ruler' (course standards). If teachers were using the 'same ruler' consistently across the state within the units for each of the outcomes then there would be no need to moderate the levels post-hoc to achieve comparability. This proposition was attractive because it meant that the final assessment awarded by the teachers would not need to be adjusted.

In the initial operationalisation of the new assessment model, teachers were asked to assess and collect evidence to support the achievement level that they believed best described the performance of their students. Under the existing grading system for TEE and WSA subjects, schools engage in consensus moderation to ensure the internal and external comparability of grading in each subject. Once schools submit their grades, these are inviolate. Statistical moderation adjusts only the relative position of whole school cohorts. It does not adjust the grades which relate to descriptions of performance. In the same way, the mechanism of statistical moderation has no capacity to adjust qualitative decisions about which achievement level best describes a student's performance.

With the decision to continue to use the external examination to statistically moderate the school assessment scores for tertiary entrance purposes, then the whole concept of marking as described above becomes problematic, because teachers have to also produce a school numerical score between 0 and 100 for those students wishing to have the mark contribute to their TES.

The initial classification into broad achievement levels may not enable teachers to reflect the relative differences in performance between the students which could lead to some students being disadvantaged and others advantaged in the construction of the TES. However, if teachers are aware of the requirement to differentiate within their school assessment marks between real differences in achievement within levels, then they may be able to do so within the level and submit this. Teachers have done this in the past using notations such as A^{++}, A^{+}, etc. and awarding different marks for this performance within the broad category (grade or level).

Andrich identifies the risks of using the direct method of marking where the units of measurement (achievement levels) are too broad when the results are to be used for precise comparisons.

An alternative method of marking enables teachers to use marks for assessment tasks and compile these marks in the way that they have traditionally done. Teachers are required to refer to the course outcomes and standards, and then compare student performance to align results to the achievement levels, obviating the need to do all assessment in levels. This process is most closely related to how most TEE teachers have marked and arrived at grades in the past. It also means that the information that is collected is more fined-grained and is more likely to meet the requirements for statistical moderation. At the same time, if the marking keys (analytical marking keys in the Andrich report) are constructed with the developmental sequence reflected in the course standards, then there is a better chance that the tasks and resulting marks will be directly related to the course outcomes–a primary requirement of standards-referenced assessment. This is the way that marking keys are developed for the Higher School Certificate (HSC) in New South Wales. This latter method of marking is referred to as 'indirect marking'.

A similarity between the direct and indirect methods is that they result in judgements about individual outcomes which are then combined to produce an overall level and band for the unit.

Another marking method has been described as the 'composite method'. The composite method was developed primarily to cater for those courses where it is difficult to disaggregate achievement information about each individual outcome for each assessment task. This method essentially requires teachers to develop a scheme of assessment to demonstrate how course outcomes will be covered in the tasks. A numerical scale developed by the teacher collects information on the task and the final numerical result is a summation of results for each task administered. This numerical result is then aligned to the course standards to produce a summative level and band for the unit. In many ways, this method is similar to current practice in TEE subjects. The composite method still requires reference to individual course outcomes in developing tasks and in making a final judgement using the numerical evidence–the judgement against the standards is, however, made post-hoc. It uses an analytical marking system with a degree of precision determined by the teacher, as is currently the case with TEE subjects.

There are of course similarities between the methods. They do, for example, all relate back directly to the outcomes and course standards. The difference relates to the crudeness or coarseness of the scale and the accuracy of measurement.

It must be remembered that teachers in different subjects traditionally use scales of different coarseness for different components of assessment, but make modifications within their marks to make sure that the school assessment mark reflects the relative differences in performance among the students. Some teachers choose to mark a task using fine-grained marking keys. In some instances they may also use analytical marking keys. In the latter case they generally make this choice to ensure that they can provide feedback to the students on the different criteria being assessed by the task.

Other teachers choose to mark a task in a more holistic manner and record the result as an A, B, C etc. In the new assessment context this may be seen as equivalent and analogous to making a level assessment as say a level 5H, 4M, etc. However, this latter method is still relatively crude, and the previous allocation of A, B, C etc, which would be converted to marks of 5, 4, 3, 2 etc, to build up a final mark is most likely to be within a narrow range of one of the outcome levels 5H or 4M and so on. Thus at the point of grading, the two processes have a similar principle, but the former allows for substantially greater precision than the latter.

Andrich strongly recommends that

> for both external and school-based assessments, analytical marking keys which arise from tasks set be used in conjunction with classification into one of only 8 levels. That the former and relatively precise marks be scaled as required to meet the tertiary selection policies and be used for tertiary selection, and that the levels be used for other educational purposes at the generic level at which they are described, for example, for monitoring and learning at a generic level.
>
> (Andrich, 2006, Recommendation 8)

I strongly endorse Andrich's recommendation 8 and believe that the decision as to the nature of the 'marking keys' for any particular task should be left to the teachers, as it has been in the past, rather than the Curriculum Council becoming prescriptive. The only requirement is that, to ensure precision, differences in performance of the task can be acknowledged in the marking. Marking that is generic and done in terms of the broad levels runs the risk of not recording valid differences and results in loss of precision.

It is with this in mind that I recommend that teachers be able to choose whichever method of marking they believe best suits their subject and their contexts, keeping in mind the primacy of the requirements promulgated by the Curriculum Council, and being aware of the measurement consequences of their decisions.

Recommendation 4
(Marking Methods)

It is recommended that:

teachers are encouraged to use whichever method of marking that suits them in their contexts; taking account of the nature of the subject, the requirements of the Curriculum Council, and good measurement advice regarding the consequences of their decisions.

This recommendation means, for example, that if teachers in a course in a school use the direct method to assess their students, and find that the marks that are emerging do not discriminate between their students when there are obvious differences between the performance of the students, then they must further refine their marking methods to ensure that the marks for the school assessment component that will contribute to the TES reflect the differences in performance between the students.

This recommendation highlights the need for teachers to be informed about the various marking methods and for them to be provided with enough information to make an informed judgement as to which method is most appropriate for them.

It is with this in mind that I reinforce the strong professional development theme in the Andrich report, by recommending that the Curriculum Council provide professional development to teachers and other relevant educational personnel regarding the requirements of the WACE assessment system; their responsibilities within that system and the advantages and disadvantages of using various marking methods to produce the final level and, where appropriate, the final school assessment for its inclusion, after statistical moderation, into the TES.

Recommendation 5
(Professional Development)

It is recommended that:

professional development be provided by the Curriculum Council to teachers and other relevant educational personnel regarding the requirements of the WACE assessment system; their responsibilities within that system; and the advantages and disadvantages of using various marking methods to produce the final level and band and where appropriate, the final school-based assessment for its inclusion, after statistical moderation, into the TES.

The Andrich report provides a rich source of information for workshops, brochures, information circulars and professional development activities.

Recommendation 2 requires the Curriculum Council to make policy decisions regarding the status of a number of communications that have provided directives to teachers regarding assessment arrangements for 2007 and beyond. The most urgent is to provide further information about the status of the conversion tables that are to be used to map standards to numerical values. (Memo to schools from the Curriculum Council dated 20 September 2006.)

I recommend, in accord with Recommendation 1, that teachers be advised of the advantages and disadvantages of using these tables as part of their assessment procedures, but that the decision as to whether they should be used or not, should be left to the informed professional judgement of the teachers responsible for preparing the assessments. This decision should be informed through information provided by the Curriculum Council as part of its program of professional development.

Recommendation 6
(Conversion tables)

It is recommended that:

- **the decision as to whether teachers use or do not use the conversion tables is left to the teachers, and**
- **to help them make their decision teachers should be advised as to the advantages and disadvantages associated with using conversion tables in arriving at the level and band and, for those units that contribute to tertiary entrance, for a school assessment.**

A question that has arisen in reference to marking methods is: 'does one marking method need to be used across a course?' That is: would all English teachers across the state, for example, need to agree to use the same marking method (i.e. direct, indirect, composite or other) or could teachers within the course choose different methods?

The statistical moderation procedure is used to make comparable the numerical marks obtained from school assessments: that is, to align them along the same scale. One of the requirements of this procedure is that the marks, within a subject and within a school, are already on the same scale. This means there is a requirement that the marks submitted to the Curriculum Council from the school are comparable within the school, and the statistical moderation procedure then accounts for differences between schools. With this in mind, I would recommend that teachers, within a school and within a subject, should use the same marking method for constructing their school assessments. However, there is no need to have the same method used in the same subject across schools or across subjects within the same school. A likely consequence of introducing a larger number of courses is that there will be an increase in small groups that will need to participate in small group moderation processes. For the effective

operation of these processes, partnering schools would also need to use a common method. Once again, this is the same as the procedure that exists in the current system.

Recommendation 7
(Consistency of marking methods)

It is recommended that:

within a school and within moderation groups within a subject the same marking method be used for constructing school assessments. However there is no need for teachers within a subject across schools to consistently use the same marking method nor for teachers within a school across subjects to use the same marking methods.

In selecting these specific questions to answer I am not implying that these are the only questions that need to be resolved. I have taken these to illustrate how I would address each of the specific issues that emerge regarding the construction of level assessments by referring back to my initial recommendation regarding the requirements of the assessment policy adopted by the Curriculum Council. I am sure there are other technical and policy issues that still need to be addressed in the assessment area.

There is no doubt that most of the issues that will arise regarding assessment in the new system will have as their genesis some technical aspect of assessment. I believe that the Curriculum Council does need sound technical advice regarding these issues. It also needs advice on preparing and conducting a research and development agenda that will mean that the Curriculum Council will be well positioned to respond to any future assessment issues or suggested policy options. A useful starting point for generating this agenda might be the list of challenges facing systems that adopt standards referenced assessment systems.

6.3. Technical advice on assessment issues

It is important that the Curriculum Council has access to expert technical advice regarding assessment in general and assessment in a standards-referenced system in particular. While it is important that the Council should employ staff with such skills so that it has constant access to this expertise, it is also important that the Council draws upon the wealth of assessment expertise within the state for advice and support. For this reason I recommend that the Curriculum Council establish an expert measurement and assessment advisory group to provide advice on the current assessment issues and research agenda that will position the Curriculum Council to identify potential issues in advance and evaluate the outcomes that are emerging from the current assessment decisions.

I further recommend that this group meet regularly to address and provide advice to the Curriculum Council on current issues.

While I do not wish to nominate the members of this group I would strongly recommend that the chief executive officer and the director of assessment and moderation from the Curriculum Council and Professor David Andrich are members of the group that may comprise up to six or seven permanent members.

Recommendation 8
(Expert measurement and assessment advisory group)

It is recommended that:

- **Curriculum Council establish an expert measurement and assessment advisory group to provide advice on the current assessment issues facing the Council and oversee a research agenda that will position the Curriculum Council to identify potential issues in advance and evaluate the outcomes that are emerging from the current assessment decisions**

- **the expert measurement and assessment advisory group meet on a regular basis for the first year and maybe less regularly after that period of time**

- **the chief executive officer and the director of assessment and moderation from the Curriculum Council and Professor David Andrich be members of the group that may comprise up to six or seven permanent members.**

Consistent with my recommendation that the Curriculum Council should provide support to teachers charged with assessing students in the new system, I recommend that the expert measurement and assessment advisory group commission a project that would lead to items from external examinations (in the first instance) being calibrated along the developmental continuum for a subject (eg. Drama, Physics, English) and then locate the results in an item bank that is available online to teachers. The intention would be that cut-scores for the levels corresponding to the course standards could be located on the scale.

One outcome expected from this calibration exercise would be the course standards being further informed by achievement data, and this in turn could lead to a fuller description of the various achievement levels. A second outcome would be teachers easily accessing the assessment items and using some of them in conjunction with their assessment tasks. This would provide an indication of where their students are located along the course standards. In this way, teachers will be supported in using and interpreting the standards which are so critical to the success of standards-referenced assessment.

APPENDIX 1

Summary of findings of the Andrich report

As defined by the terms of reference, the Andrich report focuses primarily on the capacity of assessment processes to provide adequate information for tertiary entrance; however, the recommendations are based on broader principles of good assessment practice and have relevance for the wider purposes of assessment processes which must always be consistent and fair.

Major Curriculum Council policies serve as the a priori assumptions of this report.

1. School–based and external assessments must be combined to provide a course assessment (which is essentially a ranking of students).
2. Course assessments must then be equated or scaled prior to being combined to form a Tertiary Entrance Ranking (TER).
3. These course assessments and the subsequent rankings must be sound and defensible.

Andrich explains that the data must meet certain conditions to be legitimately manipulated in this way. Both sources of information (the school and the external assessments) of a course must have the same level of precision to be legitimately combined to produce a course score, and different course scores also must have a similar level of precision if they are to be brought to a common scale (equated). The report explains that scaling processes require sufficiently large and homogenous data sets. Moderation procedures, including small group moderation procedures for schools with small cohorts of students, are identified as the key processes to ensuring optimum size and homogeneity of the data sets.

The report underlines the fact that scores derived from assessment processes do not constitute measurement. There is discussion of the misconceptions that are frequently held in relation to numerical ratings given in school assessments. It points out that level numbers, which are code for a set of skills and understandings, are mistakenly attributed the qualities of points on a calibrated scale and are then incorrectly interpreted and manipulated as if they had equivalent properties.

The report acknowledges that progress maps are useful, general and abstract frames of reference for organising teaching, learning and assessment. Andrich argues, however that, because of the inherent generality of level statements and the lack of empirical verification of the cognitive demand of levels within and across learning areas, the levels should not be used to determine assessment. To allow the organising framework to become the determinant of the form of assessment, he argues, leads to atomistic, highly prescriptive and burdensome assessment procedures. Furthermore, he provides evidence to indicate that it is difficult to make accurate a priori judgements about the levels that can be attributed to assessment tasks.

The report draws on recent research that clearly demonstrates that using the level-descriptions from the Curriculum Framework progress maps directly as a marking rubric results in:

- crude classifications, even though subtler distinctions are perceptible to teachers/ markers
- marking categories that do not match the particular assessment task

- halo effect on the marking categories resulting in redundancies in the data which further reduce its precision.

The report refers to the research findings that show how marking guides that have been developed from the task and describe the actuality of student response patterns produce finer-grained categories for marking that result in greater precision in assessment. These marking keys use language that describes specifically and concretely the actual features of student response to specific tasks or questions. These marking keys do not directly use levels or the abstract and generic language of the Curriculum Framework, although the specific descriptions map back to the Curriculum Framework.

The report addresses the issue of courses which have two quite poorly correlated components i.e. performance/product and written.

The format of examinations is addressed and alternative forms, such as 'open book' examinations, are canvassed. The need for continuation of existing policies and practices relating to scaling and weighting is explained.

Recommendations 1, 4 and 5 refer to the need for professional development for key education personnel to develop a deep understanding of the interplay between policy decisions and of measurement principles, specifically the arbitrariness of scale and scaling procedures.

Recommendations 2 and 3 refer to the need to delimit the prescribed number of assessments required per unit and the need to allow greater discretionary powers to schools in determining the final mark and level submitted to Council.

Recommendation 6 refers to the need for continued support of moderation processes.

Recommendation 7 refers to the need for Curriculum Council documentation to acknowledge the inherently probabilistic (as opposed to deterministic) definition of the level of assessments tasks.

Recommendations 8 and 9 refer to the need for assessments to be developed to operationalise the content of the course and the intent of the Curriculum Framework, and for them to be marked according to the finer-grained, specific expectations that arise from that particular task.

Recommendation 10 endorses the requirement that each course assessment intended to contribute to a TER have a 50% weighting to a written component. For particular programs (e.g., drama or fine arts), where performance/product achievement is considered to be more important than achievement on the written component, the selecting institution can be provided with disaggregated, scaled performance/product assessment scores.

Recommendations 11 and 12 refer to examination formats and the need to be explicit in identifying questions that relate to specific course units and the need for more innovative examination methods.

Recommendation 13 refers to the need for the continuation of scaling of both school and external assessments.

APPENDIX 2

Senior Secondary School Assessment Regimes

All jurisdictions across Australia recognise value in including school assessments of students' subject learning. Assessment is recognised as an integral part of good pedagogy and current arrangements draw on teachers' specialist expertise and professional judgment. In their arrangements for identifying, gathering and interpreting information about student achievement, the states/territories however place different emphases on external and internal assessments.

External assessment refers to the subject-specific examinations set by a body external to the school. All the questions refer to a syllabus that has been defined by a group of educators (teachers and/or examiners) and are devised to assess the attainment of knowledge and skill of students in a particular subject, whether by written, oral or practical questions.

Internal assessment refers to school assessment which is devised, constructed and implemented by schools. Internal assessment requires moderation. Teachers have to be trained to become consistent judges and there has to be a quality assurance process in place to guarantee comparability of results.

All jurisdictions use processes for describing school assessment and/or exam assessments that represent a shift to standards/criteria referenced reporting. They all value school assessment and all acknowledge through their assessment practices that many of the subjects offered are not validly assessed through paper and pen tests.

For school students applying to enter university, a process of establishing a rank–ordering is applied in all jurisdictions and all systems use some process or processes to support comparability of standards applied among teachers and across schools.

Consensus moderation
Social or consensus moderation is the preferred model across Australian jurisdictions. Exactly how social moderation is conducted varies between jurisdictions.

Statistical moderation
All jurisdictions use statistical moderation to remove the influence of harsh or lenient school assessments. The mechanism for this is an external examination in the subject, or/and (in the case of Victoria) a scholastic aptitude test not linked directly to syllabi or curricula. The General Achievement Test (GAT) is used by Victoria, the Australian Scaling Test (AST) is used by the ACT and Queensland uses the Core Skills (QCS) Test.

A detailed summary of the assessment, moderation and tertiary entrance selection processes used by the various jurisdictions follows.

Summary of Assessment, Moderation and Tertiary Entrance Selection Processes Used by the Various Jurisdictions

This information has been extracted from *Australian Certificate of Education: Exploring a way forward*, a report commissioned by Department Education Science and Training and compiled by ACER.

AUSTRALIAN CAPITAL TERRITORY (ACT)	
TERTIARY ENTRANCE	
Selection mechanism	University Admission Index (UAI), based on student performance in T courses, from 100.00 to 30.00. Provided on Tertiary Entrance Statement (TES).
Combining results of tertiary entrance	**Calculating the UAI** Colleges calculate a course score for each student completing a T course. Board of Senior Secondary Studies (BSSS) scales the college course scores where differences schools are provided by the AST results. This ensures that all T course scores can be meaningfully compared within and across colleges. Each student's aggregate score is the sum of the scaled scores in the best three T majors plus 0.6 of the next best T score, whether a major or minor. Aggregate scores for all eligible students are ranked (highest to lowest). Candidate rank is assigned to students, starting at the top of the list. Rank converted to a cohort rank via a lookup table supplied by the NSW Technical Committee on Scaling.
Standards setting/ maintenance	Unit grade descriptors are developed according to a set of principles and are stated in terms of outcomes. Generic criteria form the basis of unit grade decisions across all course frameworks.
ASSESSMENT ARRANGEMENTS	
Internal	Externally moderated, continuous school-based assessment Criterion-based; standards-referenced
External	There are no examinations set by a central authority for any subject.
Standardised testing	The ACT Scaling Test (AST) measures skills considered necessary for success at university. The AST provides group results for calculating the UAI. The test consists of a 2.5 hour multiple-choice test of 80 questions, a 2.5 hour writing test, and a 1.5 hour short response test.
Modes that contribute to high-stakes assessment	A range of task types (as outlined in course framework and course document).
MODERATION	
Type	Consensus and statistical
Purpose	To ensure consistency of teacher judgments and comparability of standards in reported grades. Statistical moderation ensures comparability of scores before aggregation to calculate UAIs.
Process	Structured peer review of standards and validation of unit grades assigned to student assessment portfolios Years 11 and 12 for all accredited courses; by matching student performance to criteria and standards outlined in the unit grade descriptors as stated in the course framework.

	Advice given to colleges to assist teachers with, and/or reassure them on, their judgments. The broad processes of moderation include: • establishment of system-wide assessment requirements, criteria and standards in board course frameworks • accreditation of colleges' programs of study (courses) from which student results may be recorded on board certificates • validation of portfolios of student assessment responses (Years 11 and 12) to establish standards and maintain comparability of assessment outcomes • feedback to colleges about consensus-based grade decisions • development of college action plans to address problems arising from the review process. All (1100 approx.) senior secondary teachers participate in the review process twice a year.

NEW SOUTH WALES	
TERTIARY ENTRANCE	
Selection mechanism	Index based on senior secondary school results, the UAI (Universities Admission Index), calculated by the universities in NSW via the Universities Admissions Centre (UAC)
Combining results for tertiary entrance	**Calculating the UAI** ***Step 1: Scaling HSC marks*** Quality of a subject's candidature defined in terms of their performances in other-subjects. Process modifies the mean, standard deviation and maximum mark in a course. Maximum mark in a course is related to the mean of the scaled marks in that course (to discourage students from taking easy courses in order to get high marks). ***Step 2: Combining scaled HSC marks*** Each student's scaled HSC marks are added together to produce that student's aggregate score (interim calculation not reported). ***Step 3: Ranking aggregate scores*** All students' aggregate scores placed in rank order. Individual student ranking expressed as position in the entire age cohort expressed as a percentile. This is the UAI.
Standards setting/ maintenance	Outcome statements written during development of new syllabuses. Along with course content, outcome statements guide teachers as to the knowledge, skills and understanding students are to develop through studying that course. Teams of experienced teachers considered student responses, statistical data and other materials from past HSC examinations, and prepared short statements (band descriptions) to summarise different levels of performance in the course. Mark of 90–100 corresponds to performance band 6; 80–89, band 5; 70–79, band 4; 60–69, band 3; 50–59, band 2; <50, band 1 (referred to as below minimum standard expected).

ASSESSMENT ARRANGEMENTS	
Internal	School-based assessments count for 50% of HSC. Expressed as a mark on a scale with ordinal and interval properties School determines timing and weighting of assessment tasks. Board recommends: • 3–5 tasks • weighting of each individual task at least 10% and up to 40% of total assessment • higher weightings for tasks towards end of the assessment program • outcomes and components assessed by more than one task. Schools submit students' marks in HSC board-endorsed courses to board.
External	External examinations count for 50% of HSC. The exceptions are VET courses, board endorsed courses and life skills courses.
Modes that contribute to high-stakes assessment	HSC examination may involve more than one component, such as written examination, submitted work or practical examination. Some courses require practical examinations or submission of works (eg. Dance, Industrial Technology).
MODERATION	
Type	Statistical
Purpose	To ensure that marks from internal assessment and external examination are aligned to the same standard
Process	For each course-group in a school, mean school assessment mark is set to be equal to mean examination mark, top school assessment mark to top examination mark and, where possible, bottom school assessment mark to bottom examination mark. Cut scores for each performance band are established through standards setting processes using subject experts (judges). Examination marks and school assessment marks expressed on a scale with anchors (70, 80, 90) to the boundaries between standards. Student's HSC mark in course is average of examination mark and moderated school assessment mark.

NORTHERN TERRITORY
TERTIARY ENTRANCE
Index based on senior secondary school results, the Tertiary Entrance Rank (TER), issued by SSABSA at the request of SATAC (see 'South Australia'). The TER is derived from the university aggregate, which is based on tertiary entrance points for best 5 scaled full-year (or equivalent) Stage 2 subjects. **Calculating the TER** • Tertiary entrance points calculated for all Stage 2 subjects using a scaling procedure. • University aggregate calculated as the total of the tertiary entrance points for best 3 full-year Stage 2 subjects plus half the tertiary entrance points for the 5th best subject, using the better of the subject achievement or scaled score for each subject. • TER derived from university aggregate and reported as a number between 0 and 99.95.
ASSESSMENT ARRANGEMENTS
In Stage 1, assessment of NTBS-accredited courses is school-based according to approved moderation schemes. Stage 2 courses are assessed and moderated under contractual arrangements with SSABSA.

QUEENSLAND	
TERTIARY ENTRANCE	
Selection mechanism	Student's position in statewide rank orders based on: overall achievement in Queensland Studies Authority (QSA) subjects, expressed as the OP (overall position); and up to five fields (areas of study that emphasise particular knowledge and skills), expressed as FPs (field positions). Calculated by QSA for the Queensland Tertiary Admissions Centre (QTAC) representing the universities in Queensland.
Combining results for tertiary entrance	Information used in the calculation of OPs (and FPs) comes from teacher assessment of student achievement in authority subjects and group scores on the QCS Test. The calculation of OPs involves two stages of scaling–between subjects within a school and between schools. Scaling aims to remove the bias that may be caused by differences in the competition in different subject-groups and school-groups. **Calculating the OP** First stage of scaling: The within-school stage is to make it possible to compare the achievement of students in one subject in the school with the achievement of students in other subjects in the school. For this a standard baseline of comparison is needed. This baseline of comparison is provided by the QCS test. To produce a single rank order of students within the school, an overall achievement indicator (OAI) is calculated. This is the average result across each student's best five subjects. The second stage of scaling: The between-school stage The first stage of scaling produces a single rank order within each school given by each student's OAI. The second (between-school) stage of scaling allows these rank orders to be compared across all schools. For schools with more than 19 students the average and spread of the OAIs for each school are re-set (scaled) to the average and spread of QCS Test scores for all the students in that school. Scaled OAIs place students in a single rank order across the whole State. Students are 'banded' so that students who have performed very similarly are not falsely reported as being very different. Banding also ensures that the results are relatively stable and not vulnerable to minor uncertainties in subject results. The cut-off for each OP is set each year so that there is approximate comparability with the standard of performance required to reach that OP in the previous year. This means that OPs are directly comparable from year to year. **Calculating FPs** Involves only one stage of scaling (between subjects). Unlike the OP where subjects are equally weighted, subjects are weighted differently for each of the five fields according to their emphasis on: extended written expression involving complex analysis and synthesis of ideas (Field A); short written communication involving reading, comprehension and expression in English or a foreign language (Field B); basic numeracy involving calculations and graphical and tabular interpretation (Field C); solving complex problems involving mathematical symbols and abstractions (Field D); substantial practical performance involving physical or creative arts or expressive skills (Field E).

Standards setting/maintenance	Level of achievement is an assessment provided by teachers of how well a student met the achievement criteria and standards for a particular subject. Achievement level statements (a) are couched in ordinary terms and refer (as far as possible) to familiar contexts (b) are specific in describing actual student achievements, and (c) indicate both those aspects. Achievement level statements consist of descriptions of typical achievements in a subject at each of the five levels, VHA through VLA.
ASSESSMENT ARRANGEMENTS	
Internal	Externally moderated school-based assessment Criteria-based Continuous Assessment criteria: information on the components for assessing the subject criteria and determining levels of achievement, including criteria-based principles of assessment, assessment techniques, standards descriptors and requirements for review folios.
External	No external examinations for students in full-time schooling.
Standardized testing	The QCS Test, administered to Year 12 students over 2 days in August/September, measures achievement in cross-curriculum skills. There are three modes of assessment – extended writing (2 hr), short response (2 hr), and multiple-choice (3 hr). The QCS Test provides individual results reported on the Senior Certificate as one of five grades (A–E). The QCS Test provides group results for calculating OPs and FPs, which are reported on the Tertiary Entrance Statement. OP-eligible students must sit the QCS Test. OP-ineligible students may choose to sit the QCS Test.
Modes that contribute to high-stakes assessment	The parameters for the assessment program used to award Levels of Achievement are stipulated in each syllabus.
MODERATION	
Type	Consensus
Purpose	To ensure that assessments given by schools meet minimum standards and are comparable across the State
Process	Based on a close partnership between QSA and schools. QSA contributes the design, operation and servicing of the structures that allow the system to operate. It accepts the responsibility for training the people who serve on review panels to review schools' work programs and student results. On their part, schools contribute the services of teachers as review panelists, and are responsible for developing and implementing work programs in line with syllabuses, and for assessing student work against statewide standards. They collect the student work samples and capture the data necessary for their students to receive Senior Certificates. The various stages in moderation are: work program and study plan approval; monitoring of Year 11 standards; verification and comparability of Year 12 standards; confirmation of Year 12 results; and random sampling (post hoc).

SOUTH AUSTRALIA	
TERTIARY ENTRANCE	
Selection mechanism	Index based on senior secondary school results, the Tertiary Entrance Rank (TER), calculated by SSABSA on behalf of the universities and TAFE
Combining results for tertiary entrance	The TER is derived from the university aggregate, which is based on tertiary entrance points for best 5 scaled full-year (or equivalent) Stage 2 subjects. **Calculating the TER** • Add scaled scores (or tertiary entrance points) for the student's best 4 subjects to half the tertiary entrance points for the 5th best subject. • Obtain total out of 90. This is the university aggregate score. • Calculate percentage of students at least each university aggregate score. • Obtain percentile distribution and corresponding percentile rank (0–100). • Student's percentile rank is her/his TER.
Standards setting/ maintenance	Assessment Standards Support Process • works through assessment standards support panels; • provides support and guidance to teachers on their assessment standards; • seeks to ensure a common understanding of the assessment criteria through inspection of marked student work, and approval of assessment plans, assessment tasks and teaching programs (some subjects only).
ASSESSMENT ARRANGEMENTS	
Internal	School marks count for 100% at Stage 1 and 50–100% at Stage 2.
External	No external examinations at Stage 1 of the SACE. At Stage 2, some subjects have external examinations; others are assessed using other tasks such as practicals and research work.
Modes that contribute to high-stakes assessment	A variety of writing-based, oral, practical and performance assessments with a recent increase in other methods (e.g. roundtable assessment and portfolio assessment). For majority of SACE subjects the required number of assessment components is prescribed. Assessment tasks within assessment components are negotiable; range of components not negotiable, except through special provisions in assessment in limited circumstances.
MODERATION	
Type	Statistical and non-statistical
Purpose	To validate marking standards
Process	1. Central moderation: Schools submit student materials and results sheets to a SSABSA central venue for validation of marks by a SSABSA moderation panel led by a Chief Assessor. 2. Group moderation: Teachers take student materials to meetings of a group of subject teachers. SSABSA moderator guides teachers in validating marking standards. Where necessary, SSABSA moderation panel led by a Chief Assessor conducts cross-group validations. 3. School-based moderation: SSABSA moderators visit the school to view student materials (e.g. practicals and performances), and validate the marking standards of the teachers. 4. Statistical moderation: External assessment results are taken as the standard for moderation of those school assessment results not

	included in central, group, or school-based moderation (i.e. non-statistical) processes. Statistical moderation only applies to those subjects that have an external examination component. Some subjects also have a school assessment component that is moderated using a non-statistical process before the components are added together. SSABSA provides in-Course Standards support moderation (selected subjects) and end-of-Course Standards validation (all subjects). Assessment Standards Support is designed to guide the teacher's marking standard during the course. All assessment standards support takes the form of non-statistical moderation. Assessment Standards Validation takes place at the end of the course and may involve changes to student results to ensure statewide comparability of marking standards.

TASMANIA	
TERTIARY ENTRANCE	
Selection mechanism	Selection of eligible Yr 12 students is based on tertiary entrance result. There are other pathways for non-Yr 12 students.
Combining results for tertiary entrance	Rasch Analysis is used to estimate the relative 'difficulty' of each award in each level 5 subject. The relative estimated difficulties are adjusted so that the weighted average values for the CA and the EA award remain the same from year to year. A (scaled) score is then calculated for each subject result. These range from at least 1 to 21+ approx. The TER is calculated by adding the three best (scaled) subject scores from level 5 subjects satisfactorily completed in Yr 12 (or a subsequent year), together with the next best two (equivalent) other subject scores taken from either the same year, or any other single year after Yr 10. The Tasmanian TER is determined from a ranking based on the tertiary entrance scores (using a method agreed to by all States) as a percentile ranking of students from the total age cohort.

Standards setting/maintenance	**TCE syllabuses** All TCE senior secondary syllabuses use criterion-based assessment. For each criterion (generic and subject specific) there are specific standards ranging over the levels of difficulty (i.e. levels 2–5) for which the syllabus has been provided. At each level there are three sub-sets of descriptors, distinguished by the ratings labeled *C*, *B* or *A*. The descriptors define the minimum requirement for achievement of the rating. A student's final award is determined from the profile of ratings. **TQA accredited course frameworks** The standards are defined in terms of a template that relates *required features of achievement* and the *awards* – EA, HA, CA, SA, PA. Each feature is a continuum. *Benchmarks* placed on each feature help to define the feature and to show the relationship of achievement on this feature and the final award. The final decision about an award is an on-balance decision, taking into account rules listed on the template. The template provides both a description of the standards and a tool for making and recording the assessment decisions.
ASSESSMENT ARRANGEMENTS	
Internal	The TQA approves students' internal assessments if schools ensure that: • the course of study complies with the syllabus statement, that all criteria are addressed, that specified content is covered, and that the broad objectives of the syllabus are reflected in the teaching and assessment; • each student's performance is assessed on the assessment criteria stated in the syllabus against the standards provided by the TQA for that syllabus; • each student's achievement on each criterion is given a rating of A, B, C at the end of the course of study; • the school complies with all moderation requirements for the syllabus.
External	All TCE level 5 syllabuses include an external assessment component, where students are assessed on half of the criteria stated in the syllabus. Students' performances on these externally assessed criteria are summarized as a rating of *A*, *B* and *C*. Final awards are determined from the combined set of internal rating and external rating, using the award rules that are stated in the syllabus. TQA accredited course frameworks at levels 3 and 5 include external quality assurance of the assessment and standards.
Modes that contribute to high-stakes assessment	TQA level 5 subjects, UTAS HAP subjects
MODERATION	
Type	There are two components of TCE syllabus moderation. Major emphasis is placed on consensus moderation: the process of attaining comparability in the assessment of student achievement. The second moderation component is statistical monitoring, where the

	TQA makes determinations about consistency in awards and takes actions to ensure comparability in assessments where appropriate. TQA accredited course frameworks specify an external TQA panel review of learning designs and assessment standards as shown in the evidence of student work.
Purpose	Statewide comparability of standards and consistency with syllabus/Course Standards minimum requirements.
Process	Consensus (as of 2005): Particular criteria (usually one or two), and tasks appropriate for assessing these criteria, are selected for moderation each year. One meeting of at least one teacher from each school offering the syllabus is held in March to decide tasks. A second meeting in September is held to examine examples of assessments to the particular descriptors of the selected criteria. Internal school moderation meetings are held to ensure that all teachers of the syllabus are fully informed of the requirements of and results from the moderation meetings. The TQA may reject a school's final ratings or adjust them if there is evidence to justify such action, for example, if: • assessment procedures have not been followed; or • moderation consensus recommendations have been rejected. Analysis: TCE level 5 syllabuses have an external assessment component. Half of the assessment criteria as assessed both by internal process and by one or more external instruments. The two assessments against the same criteria are analyzed. Class and school variations greater than those commonly observed are identified and discussed with school leaders. Strategies for rectification are identified by schools. Monitoring the following year is undertaken to determine the effectiveness of the strategy.

VICTORIA	
TERTIARY ENTRANCE	
Selection mechanism	ENTER score (a number between 0 and 99.95 in intervals of 0.05).
Combining results for tertiary entrance	**Calculating the TER** Assign study scores: Student gets a Study Score on a scale 0–50 (a measure of performance relative to others who took the study). Distribution of study scores (50 max, 0 min) cluster around 30 (for a given study approx. 70% of students get a study scores 23–37). Scale study scores to obtain ENTER subject scores for each study: For each VCE study, study scores are scaled according to the strength of the competition in that study (strength of competition in a particular study is gauged by comparing students' performance in all their other VCE studies with their performance in the particular study). This scaled study score is the ENTER subject score. Aggregate subject scores to obtain the ENTER aggregate: Use maximum of 6 results (up to 3 for VET sequences) in the aggregate.

	Where > 6 results exist, use the 6 legitimate results yielding the highest aggregate. Add ENTER subject scores according to the following sequence: • best subject score for an English study • next best 3 ENTER subject scores (of an allowable combination) • 10% of any fifth and sixth ENTER subject score as/if available Up to 2 scored VET sequences may be included in the primary four; a third may count as an increment; unscorable VET sequences may count as the fifth and/or sixth increment by adding 10% of the average of the primary four; the increment for the sixth study may be for an approved university study as part of the VCE extension study program. ENTER aggregate is between 0 and 210+. Rank all eligible students according to their ENTER aggregates. Assign a percentile rank that (as far as possible) distributes the students evenly (although ties might result in an increase in the number of students assigned a certain percentile rank). Convert the percentage rank to an ENTER, using a method agreed to by all States (except Qld). ENTER, a number between 0 and 99.95 in intervals of 0.05, is thus an estimate of a student's relative position in her/his age-group, having taken account of students who have moved or left school before Year 12.
Standards setting/ maintenance	According to *Curriculum and assessment principles and standards for VCE studies.* Also, examination panels report on distribution of grades for examination assessment and statistical moderation of coursework.
ASSESSMENT ARRANGEMENTS	
Internal	There are two forms of graded school assessment - Coursework and School-assessed Tasks. The form/s of school assessment and their weighting are specified for each study and are to be found in the Study Design. For each coursework component, the Study Design specifies a range of assessment tasks for assessing the achievement of the unit outcomes. School-assessed Tasks occur in studies where products and models are assessed (Art, Media etc.).
External	External examinations (written, oral, performance and electronic) are set and marked by VCAA.
Standardized testing	General Achievement Test (GAT): a test of general knowledge and skills in: • written communication • mathematics, science and technology • humanities, the arts and social sciences. Used for statistical moderation (see below); this is quality assurance of a predictive nature.
MODERATION	
Type	Statistical
Purpose	To ensure that schools' coursework assessments are comparable across the State and fair to all students
Process	The level and spread of each school's assessments of its students in each study is compared with the level and spread of the same students' scores in the external examinations. School scores are adjusted if necessary. In some studies, students' GAT scores (as well as their examination

	scores) are used for comparison purposes; specifically where GAT is a better match with schools' coursework assessments throughout the State. External examination scores, however, are the major influence in statistical moderation. For course work (7 studies), the GAT is used to check each school's assessments for School-assessed Tasks in Art, Design & Technology, Food & Technology, Media, Studio Arts, Systems & Technology, and Visual Communication & Design. In the case of a school's assessments for a particular School-assessed Task being significantly higher or lower than predicted by students' GAT scores, reviewers from VCAA visit the school to look at the corresponding work.

WESTERN AUSTRALIA	
TERTIARY ENTRANCE	
Selection mechanism	Tertiary Entrance Rank (TER) based on achievement standards in school assessments and WACE examinations
Combining results for tertiary entrance	**Calculating the TER (tentative process)** For each course combine the internal and external outcome levels of achievement on a 50:50 basis to produce a final course level of achievement (CSLA). Combine the highest 4 final CSLA's taking into account any unacceptable course combinations to produce a tertiary entrance aggregate (TEA). Convert TEA into a TER taking into account the number of students with a TEA and the total Yr 12 school leaving age population in WA.
Standards setting/ maintenance	Standards identified in the scales of achievement are derived from the K–12 Curriculum Framework progress maps. Typically, each course has 4 outcomes (e.g. reading, writing, speaking/listening and viewing for English). For each outcome, student achievement is to be assessed against 5 clearly defined levels (4–8). The higher levels show higher standards of achievement at increasing degrees of complexity.
ASSESSMENT ARRANGEMENTS	
Internal	Internal assessment counts for 50% of final result where students undertake the external assessment. If not, then it counts for 100% of the final result.
External	All courses have an external examination, the WACE examination. Only those students intending to seek selection for university are required to sit for the WACE exams. External assessment counts for 50% of final result.
Modes that contribute to high-stakes assessment	School-managed assessment is to be comprehensive, including evidence of achievement that cannot readily be obtained through an external assessment process (e.g. practical investigations, laboratory activities, and enriching tasks such as research projects, work projects and work performance).

MODERATION	
Type	1) Consensus 2) Statistical
Purpose	1) To ensure that the outcomes-focused standards are being applied consistently. 2) To ensure that judgments of student achievement from external and internal assessments are comparable.
Process	(1) For each course, an assessment and moderation panel has responsibility for managing external and school assessment requirements and ensuring that judgments about achievement in both contexts are comparable. Each year, consensus meetings are conducted in a sample of courses (as per negotiation with sectors and systems); anticipated to be essential in first year of full implementation and once again during the 5-year accreditation period. Each year, the Curriculum Council collects samples of student work from selected schools. Assessment and moderation panel views these with the aim of verifying teachers' judgments (this is within-school comparability). Adjustments to teachers' ratings will be made if necessary. (2) Results from external assessments will enable the construction of statistical models for the investigation of any systematic bias in school assessments. It is expected that results from school and external assessment will be closely correlated, as they are both assessments of course outcomes. The scales of achievement for course outcomes will provide the external measures for moderation.