

Fair and meaningful measures?

A review of examination procedures
in the NSW Higher School Certificate

Published 2002 by the Australian Council for Educational Research Ltd
19 Prospect Hill Road, Camberwell Victoria 3124 Australia

Copyright © 2002 HSC Examination Program Review
Suite 1203, Level 12, 31 Market Street, Sydney, New South Wales

All rights reserved. Except as provided for by Australian copyright law,
no part of this book may be reproduced without written permission
from the publisher.

ISBN 0-86431-600-3

Printed in Australia

Preface

This report presents the results of a review of examination procedures introduced for the new Higher School Certificate (HSC) in New South Wales.

The adoption of new assessment and reporting procedures was a key element of the Government's reform of the Higher School Certificate following the McGaw Review. Changes made to examinations and to the reporting of HSC results were designed to provide more explicit assessment against standards and clearer and more informative reports of student achievement.

An independent review of the 2001 examinations was initiated by the Board of Studies as the first cohort of students to complete the new Higher School Certificate was preparing to sit HSC examinations. The Board asked me to review and provide advice on the implementation of the new standards-referenced examination and marking processes and, in particular, to consider the quality and effectiveness of:

- ◆ the processes for setting the examinations and developing marking guidelines in the new standards environment;
- ◆ the procedures for marking the examinations and determining the standard of achievement students had demonstrated; and
- ◆ the Board's quality assurance procedures for developing the examinations, marking examination papers and validating results.

I was asked as part of the review to consult with a number of organisations and to examine and report on issues raised by students, teachers and interest groups concerning the examination and marking processes for 2001.

A public call for submissions to the review was made in January 2002 and consultations with peak bodies were undertaken in February. A number of the written submissions I received were informed by substantial surveys that organisations had conducted of their members. The organisations involved in the consultations and the written submissions received by the review are listed in the Appendix.

Submissions to the review raised a variety of matters relating to the content of the 2001 examination papers, the guidelines for markers, the Board's marking processes, the new standards-setting process, and the new HSC reports provided to students.

Many of the submissions raised concerns about the Universities Admission Index (UAI). Although the construction and use of the UAI were not within the review's terms of reference, and are outside the control of the Board of Studies, in view of the level of concern about the relationship between HSC results and the UAI in 2001, I have included a brief discussion of this issue.

In undertaking this review I was greatly assisted by Andrew Rolfe and Andrew Goodyer. Their experience and deep understanding of HSC syllabuses, examinations and Board processes were essential to my own learning and thinking. They provided invaluable support to the review's analysis of issues, in the sourcing of additional information I required, and in listening to and responding to my thinking about matters as they emerged. I greatly appreciated their enthusiasm, counsel and friendship during the review.

I also wish to thank Carol Taylor, Rob Speers and John Bennett who provided necessary details of the Board's processes and clarified how the Board had responded to issues as they arose in 2001.



Geoff N Masters
Australian Council for Educational Research
March 2002

Table of contents

Preface	
1 Introduction	3
2 Designing new procedures: 1998–2001	6
3 Developing examinations	13
4 Marking examinations	31
5 Standards setting	42
6 Reporting HSC results	60
7 Using HSC results	70
8 Conclusion	74
Appendix	80

1 Introduction

In September 1997 the New South Wales Government introduced a number of significant reforms to the Higher School Certificate. These reforms, which were the most far-reaching since the introduction of the Higher School Certificate in 1962–67, followed an extensive review of the HSC by Professor Barry McGaw¹. The Government's proposed changes to the Higher School Certificate were outlined in the 1997 White Paper *Securing Their Future*².

The broad objectives of the 'new' Higher School Certificate, which was implemented for the first time in 2000–01, are to:

- ◆ increase the rigour and quality of the HSC curriculum;
- ◆ ensure HSC marks fairly reflect the standards achieved by students;
- ◆ provide more informative reports of students' HSC achievements; and
- ◆ enhance the chances for more equitable educational outcomes.

Although the new Higher School Certificate incorporates reforms of both HSC curricula and methods of assessing and reporting student achievement, the focus of this review is on new HSC assessment and reporting procedures and, in particular, on the procedures used in 2001 to set examinations, mark student responses and report student results.

Submissions to the McGaw review identified a range of concerns about existing assessment and reporting practices. Among those concerns was a view that the way in which HSC marks were awarded no longer accorded with community expectations. HSC reports had become too complex for employers, parents and the community to understand, and there was a need for simpler and more informative forms of reporting. Other concerns were that insufficient recognition was being given to students successfully completing advanced studies, school-based assessments often were based on assessment methods resembling the pen-and-paper examinations, and methods of reporting provided an inadequate basis for education systems to monitor levels of student achievement over time.

There was strong support at that time for an assessment system that would report on the substance of student learning rather than reporting only students' relative performances. HSC examinations were seen as a potential vehicle for the more explicit identification of standards and for reporting student achievement in relation to those standards.

The McGaw review made a number of recommendations for the reform of HSC examinations. At the heart of these reforms was the proposition that there should be a tighter and more explicit relationship between HSC curricula and HSC examinations, with student results indicating the standards (knowledge, skills and understandings) they had achieved. The review recommended that the existing practice of reporting students' results only in terms of the performances of other students taking the same course ('cohort-referencing') be replaced by a system of 'standards-referencing' that would show how students had performed in relation to a hierarchy of described levels of achievement.

¹ McGaw, B (1996). *Their Future: Options for Reform of the Higher School Certificate*. Sydney: Department of Training and Education Co-ordination.

McGaw, B (1997). *Shaping Their Future: Recommendations for Reform of the Higher School Certificate*. Sydney: Department of Training and Education Co-ordination.

² NSW Government (1997). *Securing Their Future: The New South Wales Government's Reforms for the Higher School Certificate*. Sydney: NSW Government Printer.

The Government's 1997 decision to adopt a standards-referenced approach to assessment and reporting in the Higher School Certificate has had implications for all stages in the development and marking of examinations and for the reporting and use of HSC results. In the period 1998 to 2001, the Board of Studies designed and field-tested significant modifications to its examination procedures to support the introduction of a standards-referenced approach to the assessment and reporting of HSC achievement.

Developing examinations

A standards-referenced system has implications for how examinations are developed. In introducing new procedures, the Board of Studies has sought to build a tighter connection between HSC curricula and examinations and, in particular, to require greater explicitness about the learning outcomes that individual examination questions are designed to assess. The rationale for this change is that the ability to interpret and report students' achievements in terms of standards depends on clarity about the knowledge, skills and understandings that examination questions address.

The Board also has introduced new procedures to make the development of marking guidelines an integral part of the examination development process. This change is intended to ensure that the guidelines, and thus the marking process itself, yield evidence about the syllabus outcomes addressed by examination questions.

Marking examinations

In a standards-referenced system, marking schemes are designed to reward evidence of the outcomes that examination questions are designed to address. Under the new Higher School Certificate, marking schemes are based on marking guidelines developed by the course Examination Committee. Samples of student work serve as 'benchmarks' to illustrate guidelines and marking schemes. The Board also uses pilot marking and marker briefings in an effort to ensure student work is marked fairly and consistently by different markers.

Standards setting

A key feature of a standards-referenced approach to assessment and reporting is the interpretation and reporting of students' performances in terms of the knowledge, skills and understandings they typically have demonstrated. In 2001 the Board introduced a 'standards-setting' process, the purpose of which is to allow performances on particular HSC examinations to be aligned with, and interpreted in terms of, a hierarchy of described levels of achievement in each course.

Reporting HSC results

For the first time in 2001, each student taking an HSC examination was given a Course Report showing their examination mark, school assessment mark, and HSC result in that course against a hierarchy of described standards of achievement. These standards describe the knowledge, skills and understandings typical of students within particular mark ranges.

Using HSC results

A claim sometimes made for standards-referencing is that it provides users with more informative descriptions of student achievement and so is likely to be of more assistance in decision making. The proposal to use standards-referencing to add 'meaning to marks' in the HSC was seen as a way of providing employers and tertiary institutions with richer information about students' school achievements, and of differentiating results in the new HSC from scores and ranks that show only how students have performed in relation to one another.

Sections 3 to 7 of this report consider each of these major phases in the development and marking of HSC examinations and in the reporting and use of students' HSC results. Given that the terms of reference for the review included an evaluation of Board processes, and given that these processes have not previously been documented in one place, each section provides a detailed description of the steps in the Board's processes. Each section includes:

- ◆ a description of the Board's current processes;
- ◆ a summary of issues raised with the review; and
- ◆ a reflective comment on these issues and on possible ways of addressing them.

HSC examinations 2001

- ◆ 62 751 students (33 031 females and 29 720 males) enrolled as candidates in the HSC.
- ◆ Students presented for 74 subjects organised into 224 courses, including seven new industry-accredited Vocational Education and Training frameworks.
- ◆ The Board of Studies established examination committees comprising almost 400 people to set 154 different HSC examination papers for 108 courses.
- ◆ HSC examination papers comprised more than 12 million printed pages, with students providing answers in 1.7 million writing booklets.
- ◆ More than 5000 people supervised the HSC examination sessions in some 750 examination centres.
- ◆ More than 7000 markers marked HSC examinations at 16 venues across the state.

2 Designing new procedures: 1998–2001

Following the release of the Government's White Paper *Securing Their Future*, the Board of Studies undertook two parallel but linked processes of curriculum and assessment reform in preparation for the introduction of the new Higher School Certificate from 2000.

Reforming curricula

The White Paper endorsed the development of a 2-unit curriculum model across all HSC subjects and called for a clear definition in each course syllabus of the content (knowledge, skills and understandings) students are expected to learn.

Commencing in 1998, the Board of Studies developed single 2-unit courses in most school subjects. The development of these courses involved discontinuing 3-unit courses and incorporating the learning outcomes from these courses into 2-unit courses; discontinuing 'general' courses in English, mathematics and other subjects and incorporating some of the outcomes and content from these courses into new 2-unit courses; and developing one or two 'Extension' courses for each of a small number of subjects. In undertaking the development of new courses, the Board followed a set of criteria for new and existing HSC courses provided in the White Paper.

Criteria for HSC courses (excerpt)³

For any new course developed or endorsed by the Board of Studies, there must be satisfactory evidence of the need for the course and an assurance of its quality through:

- ◆ a detailed explanation of the nature of the subject content (knowledge, skills and understandings);
- ◆ a clear rationale and statement of purpose for the course;
- ◆ a review of national and international research and practice;
- ◆ the learning outcomes students are expected to achieve;
- ◆ evidence that the learning outcomes for students are set at an appropriate standard;
- ◆ prior knowledge assumed in students enrolling in the course, and the manner in which that knowledge will be built upon;
- ◆ an identification of the subsequent uses students might make of the learning from the course; and
- ◆ internal and external assessment procedures to be used.

Drafting standards of achievement

As part of the introduction of new HSC courses and syllabuses, the Board of Studies initiated the development of a set of described levels of achievement for each course. The Board's intention for this hierarchy of described levels was that it would provide a set of 'standards' against which students' examination performances and school assessments could be mapped and reported. These levels also would make explicit the nature of increasing achievement within each course.

³ NSW Government (1997), *op. cit.*, p 8.

The first step in the development of the levels of achievement in each course was a detailed consideration of the questions/tasks in the 1996 and 1997 examinations. Within each course, teachers and academics with experience in HSC examining and marking – including Examination Committee Chairs and Supervisors of Marking – worked with Board officers to identify the area of the syllabus addressed by each question and the knowledge and skills required to answer that question. The course Examination Committee also judged how students at different levels of attainment in the course were likely to perform on each question.

Examination results were then used to establish and describe how students at different levels of HSC marks in a course had performed on each question. These descriptions were compared with the Examination Committee's expectations. The outcome of this process was a draft description of five levels of achievement ('bands') in each course.

In the third step, a group of experienced teacher-markers worked as 'judges' to consider the questions one at a time and, by referring to the descriptions of the five achievement levels, to judge the mark a student at the boundary between any two bands was likely to achieve on each question. After making independent judgements, they met to discuss and reach consensus on their judgements. These question-by-question judgements were then added to obtain the overall examination mark expected of a student at each boundary.

Finally, some student scripts that had achieved marks around these 'cut-off' marks between bands were retrieved and given to the judges to allow them to check the implications of their judgements. Was the work of students achieving these marks consistent with the draft description? If it was not, then cut-off marks between bands were changed and/or band descriptions were revised.

Consistency across courses

Once draft band descriptions had been developed for each course, it was possible to make comparisons of descriptions across courses. These comparisons considered the kinds of knowledge and skills expected at different levels of achievement, and the degree to which descriptions were course specific.

By highlighting common features of band descriptions across courses it was possible to identify general kinds of skills typical of each band across HSC courses. This process was not intended to produce a consistent definition of bands across subjects, but to achieve greater consistency in band wording.

The skills identified through this process were compared with other frameworks such as the Key Competencies and the common curriculum elements in the Queensland Core Skills test. An expert advisory team of academics and curriculum advisers was formed to assist in reducing the number of elements and refining band descriptions.

Further work was done to clarify distinctions between the bands and to identify patterns of knowledge and skill development. Terms were chosen to describe identified sequences of development through these five levels of achievement.

A mapping exercise was then undertaken by curriculum and assessment officers to map the descriptions to the new HSC outcomes in each course. As a result either the syllabus outcomes or the descriptions were modified.

Consultation

Consultation on the draft standards for each course occurred at the same time as consultation on the draft syllabus. Consultation strategies included surveys of teachers and meetings with focus groups.

Following surveys of teachers and meetings with focus groups, a team consisting of Board officers, academics with expertise in the subject area and experienced teachers was formed to amend the band descriptions on the basis of feedback from the consultations.

After the publication of the syllabus packages, including the described bands of achievement, further consultations were undertaken with parent groups, principals and groups of employers.

Developing examination specifications

Beginning in 1998, examination specifications were developed for each new HSC syllabus. The specifications, which were informed by previous examinations and the Board's syllabus development handbook, indicated the structure to be followed by the examination.

A revised syllabus and an Examination, Assessment and Reporting Supplement (EARS) were distributed to schools in June-August 1999. This material included:

- ◆ a sample examination paper that outlined the general structure of the new HSC examination and the question types;
- ◆ sample marking guidelines; and
- ◆ draft band descriptions.

Consultations on these materials were conducted in late 1999. Teacher surveys and focus groups were used to determine whether the sample papers provided students at different levels of attainment in a course with opportunities to demonstrate what they had learned, and whether the questions were appropriate for the outcomes being addressed.

Glossary of key terms

To provide greater consistency in the language used to describe expectations of students in statements of syllabus outcomes, band descriptions and examination questions, the Board developed a glossary of key terms (see page 9). It was intended that the glossary would assist teachers and students to understand what was expected in responses to examination questions and assessment tasks.

Principles for setting examinations

To provide a common framework for the development of HSC examinations, the Board produced a set of principles for setting examination papers (see page 10) and an accompanying set of principles for developing marking guidelines (see page 11).

The principles for setting papers addressed such issues as the coverage of syllabus content, the level of difficulty of the examination, the format and length of the paper and the question structure. The Board's principles for marking guidelines included requirements that they elicit evidence relevant to the outcomes addressed by each question, permit flexible interpretation, be clear and unambiguous, and allow different levels of response to a question to be recognised and rewarded.

Glossary of terms

Account	Account for: state reasons for, report on. Give an account of: narrate a series of events or transactions
Analyse	Identify components and the relationship between them; draw out and relate implications
Apply	Use, utilise, employ in a particular situation
Appreciate	Make a judgement about the value of
Assess	Make a judgment of value, quality, outcomes, results or size
Calculate	Ascertain/determine from given facts, figures or information
Clarify	Make clear or plain
Classify	Arrange or include in classes/categories
Compare	Show how things are similar or different
Construct	Make; build; put together items or arguments
Contrast	Show how things are different or opposite
Critically analyse/evaluate	Add a degree or level of accuracy, depth, knowledge and understanding, logic, questioning, reflection and quality to analysis/evaluation
Deduce	Draw conclusions
Define	State meaning and identify essential qualities
Demonstrate	Show by example
Describe	Provide characteristics and features
Discuss	Identify issues and provide points for and/or against
Distinguish	Recognise or note/indicate as being distinct or different from; to note differences between
Evaluate	Make a judgement based on criteria; determine the value of
Examine	Inquire into
Explain	Relate cause and effect; make the relationships between things evident; provide why and/or how
Extract	Choose relevant and/or appropriate details
Extrapolate	Infer from what is known
Identify	Recognise and name
Interpret	Draw meaning from
Investigate	Plan, inquire into and draw conclusions about
Justify	Support an argument or conclusion
Outline	Sketch in general terms; indicate the main features of
Predict	Suggest what may happen based on available information
Propose	Put forward (for example a point of view, idea, argument, suggestion) for consideration or action
Recall	Present remembered ideas, facts or experiences
Recommend	Provide reasons in favour
Recount	Retell a series of events
Summarise	Express concisely the relevant details
Synthesise	Put together various elements to make a whole

Principles for setting examinations

Content

- 1 The examination will test a representative sample of the knowledge, understanding and skills outcomes in any given year. The intention of the examination in its formulation is to avoid predictability and encourage students to prepare for all syllabus outcomes. Over a number of years, it is expected that the full range of syllabus outcomes that are appropriately measured by an examination will be covered.
- 2 The examination as a whole will be constructed in such a way that it provides a representative sampling of a range of syllabus outcomes and questions that allow demonstration of performance across all levels in the performance scale.
- 3 The coverage of syllabus outcomes and content in the examination must allow students to demonstrate the levels of performance that are described in the bands on the performance scale. In preparation of a paper, each question should be mapped against syllabus outcomes, content and performance descriptions that students may demonstrate in answering the question. These will be addressed in the table of specifications, constructed by the examination committee each year.
- 4 Values and attitudes outcomes will not be included in the examination.

Level of difficulty

- 5 The examination paper as a whole will provide the range of candidates with the opportunity to demonstrate what they know, understand and are able to do and will allow for appropriate differentiation of student performance at each band on the performance scale, including demonstration of higher order skills.
- 6 The level of difficulty of a paper should be maintained consistently from year to year.

Paper format, length and layout

- 7 In accordance with the examination specifications, the examinations should include a range and balance of question types, including multiple-choice questions, short-answer free response questions, open-ended questions and extended responses including essays.
- 8 The demands of the examination in terms of the number and length of student responses required, the amount of reading time provided and the complexity of the questions will be appropriate for the time allocated for the examination.
- 9 Examination layout will assist students in working through the paper and instructions will be clear and concise.
- 10 Questions will be set simultaneously with marking guidelines and will allow for marks to be awarded commensurate with performance.
- 11 The mark allocations and space provided to answer questions will be appropriate for the anticipated range of responses.
- 12 The marks allocated for each question or part question will be clearly indicated.
- 13 Wherever appropriate, explanatory information will be placed at the top of a section or page, rather than written within a question.

Question structure and language

- 14 The language used in questions will be accessible to candidates. It is preferable to use the simplest and clearest language in the wording of questions so that it is clear to all students what they are expected to do.
- 15 Questions will require minimal reading time except where reading and comprehension are being specifically examined.
- 16 Stimulus material will only be provided when it is essential to answering the question.
- 17 Questions must be free of culture or gender bias, stereotyping or tokenism.
- 18 The requirements of the question will be clear to all adequately prepared students while encouraging flexibility in their responses.
- 19 Free response questions will have simple structures with a minimal number of parts and sub-parts. The parts will be sequenced in order of difficulty and allow the candidates to demonstrate what they know, understand and are able to do.
- 20 Where definitions such as 'describe', 'analyse', 'synthesise' and 'evaluate' are used they will be used consistently and appropriately.

Comparability and moderation

- 21 To assist in achieving comparability, optional questions within a section of the paper must be marked using similar marking criteria. Choices within questions should have a comparable degree of difficulty.
- 22 To assist moderation in papers where there is a core and options there will be no internal choice within questions in the core section of the paper.

Principles for developing marking guidelines

Content

Marking guidelines will be developed in the context of relevant syllabus outcomes and content. Marks will be awarded for demonstrating achievement of aspects of the syllabus outcomes addressed by the question.

Marking guidelines will reflect the nature and intention of the question and will be expressed in terms of the knowledge and skills demanded by the task.

Specificity of marking guidelines

Marking guidelines will indicate the initial criteria that will be used to award marks.

Marking guidelines will allow for less predictable and less defined responses, for example, characteristics such as flair, originality and creativity, or the provision of alternative solutions where appropriate.

Marking guidelines for extended responses will, as far as possible, use language that is consistent with the outcomes and the band descriptions for the subject.

Marking guidelines are to incorporate the generic rubric provided in the examination paper as well as aspects specifically related to the question.

Language of marking guidelines

The language of marking guidelines will be clear, unambiguous and accessible to ensure consistency in marking.

Discrimination between different levels of performance

Where a question is designed to test higher-order outcomes, the marking guidelines will allow for differentiation between responses, with more marks being awarded for the demonstration of higher-order outcomes.

Marking guidelines will indicate the quality of response required to gain a mark or a sub-range of marks.

High achievement will not be defined solely in terms of the quantity of information provided.

Optional questions

Optional questions within a paper will be marked using comparable marking criteria.

Generic type questions

Marking guidelines for questions that can be answered using a range of contexts and/or content will have a common marking guideline exemplified using appropriate contexts and/or content.

Providing specimen papers

Following consultations with schools on the 1999 sample papers, the Board of Studies produced a set of specimen papers and accompanying marking guidelines for all HSC courses. Schools were told that:

- ◆ the specimen examination paper would be a good guide to the first new HSC examination;
- ◆ the specimen papers provided examples of the types of examinations that could be prepared under the examination specifications in the syllabus;
- ◆ future examinations would be based on the syllabus and would test a representative sample of syllabus outcomes;
- ◆ the range and balance of outcomes tested in HSC examinations in 2001 and subsequent years might be different from those in the specimen papers; and

- ◆ the provided specimen papers highlighted important points about the paper, such as where question types were different from those in the past, and sometimes indicated the ways in which HSC examinations could vary from year to year⁴.

Communicating new procedures

In the course of developing procedures for the new Higher School Certificate and during the implementation of new syllabuses in 2000–01, the Board of Studies prepared and disseminated a variety of information packages and Board Bulletin articles explaining aspects of the new HSC examination and assessment procedures. For example, the principles for setting examinations and developing marking guidelines were published in the August 1999 and May 2000 Board Bulletins and information about the stages in the standards-setting process for each course was published in the September 2000 Bulletin.

Communication with schools included advice on the Board's expectations in relation to school-based assessments. *The New Higher School Certificate Assessment Support Document* (Term 3 1999) provided advice under the headings:

- ◆ Board of Studies Requirements for Internal Assessment;
- ◆ Moving to a Standards-Referenced Approach;
- ◆ Developing an Assessment Program for a Standards-Referenced Approach (an outline of steps towards implementing a standards approach to school-based assessment over the period 2001–03);
- ◆ Designing Assessment Tasks;
- ◆ Using the Examination and Assessment Supplement; and
- ◆ A Glossary of Key Words.

In the Board's communication with schools, emphasis was placed on the way in which the new HSC brought together intended learning outcomes specified in course syllabuses, achievement standards in each course, examinations and school assessments. Figure 2.1 was developed in 1999 and widely used by the Board in information and training sessions to illustrate the relationships between syllabus standards, performance standards and assessment and examinations.

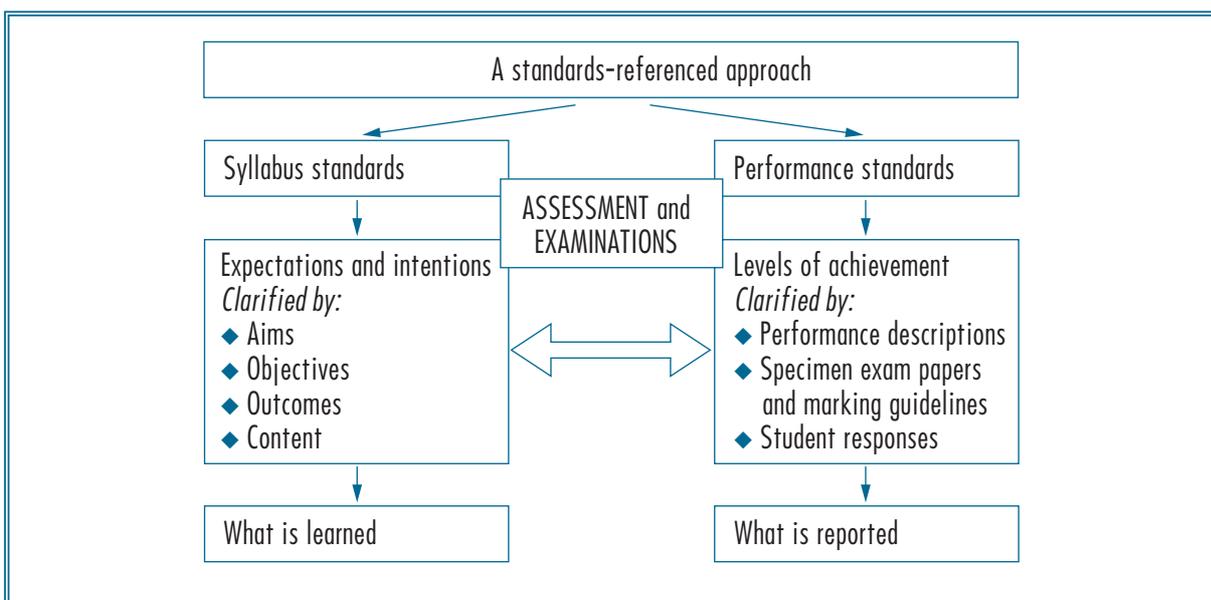


Figure 2.1 Diagram used by the Board of Studies in communications about new HSC arrangements

⁴ *Preparing students for the New HSC examinations*, Board Bulletin, June 2001.

3 Developing examinations

Each year the Board of Studies develops Higher School Certificate examinations in all Board Developed courses. The purpose of these examinations is to provide reliable and fair measures of students' achievements in the courses they have taken. Each examination is designed to address the syllabus objectives, outcomes and content of the course and to provide users of HSC results with meaningful indications of what students know, understand and can do at the time of the examination.

Paper-and-pen examinations are developed for all Board Developed courses. In some courses, the paper-and-pen examination is supplemented by other forms of examination: oral and aural tests in Languages other than English and Music; performances in Music, Dance and Drama; research projects in Agriculture and Society and Culture; and products in Design and Technology and Visual Arts. In 2001, the Board developed 154 examination papers, 56 practical examinations and 7 project-based examinations under the supervision of the Manager of the Board's Assessment Branch.

A significant change introduced as part of the new Higher School Certificate was to include in the examination development process not only the writing of questions, but also the development of guidelines for the marking of students' responses to those questions. In previous years, the development of marking guidelines had been the responsibility of the course's Supervisor of Marking, working with senior markers.

The decision to make the development of marking guidelines an integral part of HSC examination development was consistent with usual test development practice. When professionally developed tests are constructed, careful consideration is given to the purpose of each test item, to the range of possible responses to that item, and to the way in which those responses are to be scored. In psychometric testing, a test item is considered to exist only when both the test question and the key to be used in scoring responses to that question have been developed. This practice has not always been followed in public examination systems, where one group of persons may be responsible for setting the paper and a quite different group may decide what is to be rewarded in students' responses to that paper.

In 2001, a specially appointed course Examination Committee was responsible for developing both the HSC examination paper and its associated marking guidelines. The intention in integrating these two tasks was to ensure that, as questions were written, explicit consideration was given to the learning outcomes (knowledge, skills and understandings) each question was designed to address, and to ensure that the marking of students' responses provided information relevant to those intended learning outcomes.

The Examination Committee for each course followed the Board's general principles for developing examinations and marking guidelines (pages 10–11) and worked to an examination 'brief' that identified the learning outcomes to be addressed, the content to be covered and the types of questions to be used. Once completed, the draft paper and the draft marking guidelines were subjected to a variety of reviews designed to ensure that each paper met quality standards established by the Board.

The process

As part of the introduction of the new Higher School Certificate, the Board of Studies approved a revised process for the development of HSC examinations. This revised process is designed to enhance Board procedures and to support the development of standards-referenced examinations.

Key features of the new process include the use of a project-based approach to examination development, with the assignment of a Board assessment officer to manage each Examination Committee; the inclusion of quality controls and sign-offs at key stages in the examination development process; and the involvement of key individuals such as the Board Inspector and the Supervisor of Marking at important steps in the process.

The major stages in developing HSC examinations are common to all courses. They include: appointing a person to chair the Examination Committee, developing an examination brief, appointing and training committee members, writing the examination questions and marking guidelines, and reviewing, evaluating and ratifying the final paper.

Chair of examination committee

For each HSC course, a person is appointed to chair a course Examination Committee.

The process by which Chairs are appointed includes a call for expressions of interest circulated to current chairs, members of examination committees and academics. Nominations of appropriate people also are sought from Board Inspectors. Senior Assessment Officers of the Board evaluate applicants against a set of criteria, and the appointment is made on the recommendation of a selection panel chaired by the President of the Board and including representatives of school systems, the university sector and senior Board officers.

The Chair provides leadership to the committee in the preparation of the examination paper and the guidelines for marking the examination. It is the Chair's responsibility to ensure that the final examination is accurate and error-free, conforms to the examination specifications, and addresses an appropriate range of content and intended learning outcomes in the course syllabus. The Chair also responds to any issues raised about the examination and is a spokesperson, when required, for the Examination Committee.

Following the sitting of the examination, the Chair participates in the standards-setting process for the examination and leads the team reporting to the Board's Consultative Committee.

Examination planning and review group

For each course, an Examination Planning and Review Group is established to develop an examination brief specifying the structure of the examination questions and the content and outcomes to be addressed. It is this group's responsibility to ensure that the Examination Committee has the necessary expertise to develop the examination and to confirm that the final paper meets all requirements (including accessibility to the full range of candidates).

This group consists of the Examination Committee Chair, the Board Inspector, the Senior Assessment Officer and the Supervisor of Marking for the course.

The examination brief is based on the examination specifications in the course syllabus and indicates the outcomes to be addressed, the content to be covered, the types of questions to be used and the approximate performance bands to be targeted by each question.

Based on the examination brief, the Assessment Officer who project manages the development of the examination prepares an outline of the experience and expertise required of the committee members who are to prepare the examination.

Examination committee

The next step in the process is the appointment of a course Examination Committee to develop the paper in accordance with the examination brief. Each Examination Committee is appointed for two years, although, for a variety of reasons, there usually are annual changes in membership. The appointment of replacement members also takes account of the requirements of the examination brief.

Examination committees are structured to have, as far as possible, equal numbers of tertiary and secondary members, representation from non-metropolitan areas, gender balance and representation from both government and non-government schools. Examination Committee members are nominated by a selection panel that includes the Examination Committee Chair, the Senior Assessment Officer, the Inspector, the Supervisor of Marking, and a professional association representative. Selections are approved by the General Manager on behalf of the Board.

Examination Committee members are provided with training to support them in their work to develop examinations and marking guidelines. The Chair and Supervisor of Marking, as well as being given opportunities to familiarise themselves with the course syllabus and previous examination papers, are provided with training in: the role of the Chair; the nature of assessment in a standards-referenced framework; the standards-setting process in the Higher School Certificate; the examination-setting process and the roles and responsibilities of the people involved; and elements of examination setting (the examination brief, marking guidelines, rubrics, the Board's principles, achievement bands, glossary of terms).

All members of the Examination Committee, including the Chair, are provided with a booklet of guidelines on the writing of examination questions and take part in training sessions covering: the outcomes-based approach of HSC syllabuses; principles and processes for setting examinations and developing marking guidelines; roles and responsibilities; and item design.

Examination development

The Examination Committee meets over a number of days between October and April. The Chair leads the committee in the drafting of questions and guidelines consistent with the examination brief and the Board's operational requirements.

The Board Assessment Officer, as project manager, works with the Chair to plan meetings of the committee to ensure that the draft paper is prepared by the due date and provides guidance to committee members in the writing of questions and marking guidelines.

Board Inspectors or, where appropriate, the Curriculum Officer review drafts at various stages of development to ensure that questions are interpreting the syllabus correctly and that marking guidelines describe standards of performance consistent with the intentions of the syllabus.

Assessment Branch officers ensure that the draft paper is consistent with the specimen paper, complies with the Board's principles, uses key terms in accordance with the Board's glossary, and is formatted in accordance with Board style. Marking guidelines are monitored to ensure that they are consistent with the sample marking guidelines published for the course.

Review and evaluation

Once the draft examination paper and accompanying marking guidelines have been developed, they are subjected to a variety of reviews.

Multiple-Choice Review – An independent expert reviews all multiple-choice questions on the examination and provides feedback and suggestions to the Examination Committee.

Assessor Review – A practicing teacher (in some cases more than one) provides answers/worked solutions to all questions except extended-response questions. For extended-response questions, the assessor is asked to indicate the characteristics they would expect a high-quality response to have. They also are asked to comment on such features as the wording of questions, clarity of instructions, accuracy of content, level of difficulty, space provided for answers, and the extent to which the paper is likely to be accessible to candidates at all levels of ability.

Curriculum Review – The Board Inspector or delegate reviews each examination paper to verify its appropriateness to the syllabus, including whether the examination tests a representative sampling of content and outcomes, whether the level of difficulty is such that it allows students to demonstrate performance at all levels on the achievement scale, and to verify that all questions have been correctly mapped against syllabus outcomes, content and band descriptions.

In addition to these three major reviews, the examination paper is subjected to a number of other reviews including:

Copy Editor Review – An independent professional copy editor checks the paper against the Test Development Style manual, checking spelling, punctuation, grammar and clarity of expression.

Vision/Hearing Impairment Review – Specialists in visual and hearing impairment provide advice on amendments to questions to assist candidates with visual and/or hearing loss.

Examinations Branch Review – The Examinations Branch reviews instructions and layout for clarity and ease of use for candidates, presiding officers and markers.

Information Services Branch Review – Information Services reviews the paper and develops sheets for capturing and recording students' marks.

The Supervisor of Marking reviews the draft marking guidelines for consistency with the Board's principles for marking guidelines and to ensure that they reflect the intentions of each question and are appropriate for the marking operation. The Supervisor of Marking signs-off that the final guidelines are consistent with the Board's requirements for effective and efficient marking.

The course Examination Committee considers and responds to issues raised through this set of reviews, making changes to questions and marking guidelines where necessary. The Examination Planning and Review Group checks that all issues raised in the review phase have been adequately addressed by the Examination Committee. Any unresolved issues at this stage are referred to the Manager Assessment for negotiation and resolution. The Assessment Officer ensures that all agreed changes are made.

The examination paper is then proofread by the Examination Committee Chair and the Assessment Officer. The final examination is signed off by the Examination Committee Chair and countersigned by one other committee member. The marking guidelines are signed off by the Chair and countersigned by the Supervisor of Marking.

These stages in the development of examinations are described in the flow chart on page 17. The final stage is to evaluate each year's paper following the examination and to forward all feedback on the paper to the Chair of the Examination Committee for consideration in developing the following year's paper.

These stages are illustrated on pages 18–22. This illustration relates to a particular item (Question 27) on the Personal Development, Health and Physical Education (PDHPE) paper. Relevant sections of the course syllabus and examination specifications are shown on pages 18–19. The examination brief, the question itself, and the marking guidelines developed for this question are shown on pages 20–22.

Development of examinations

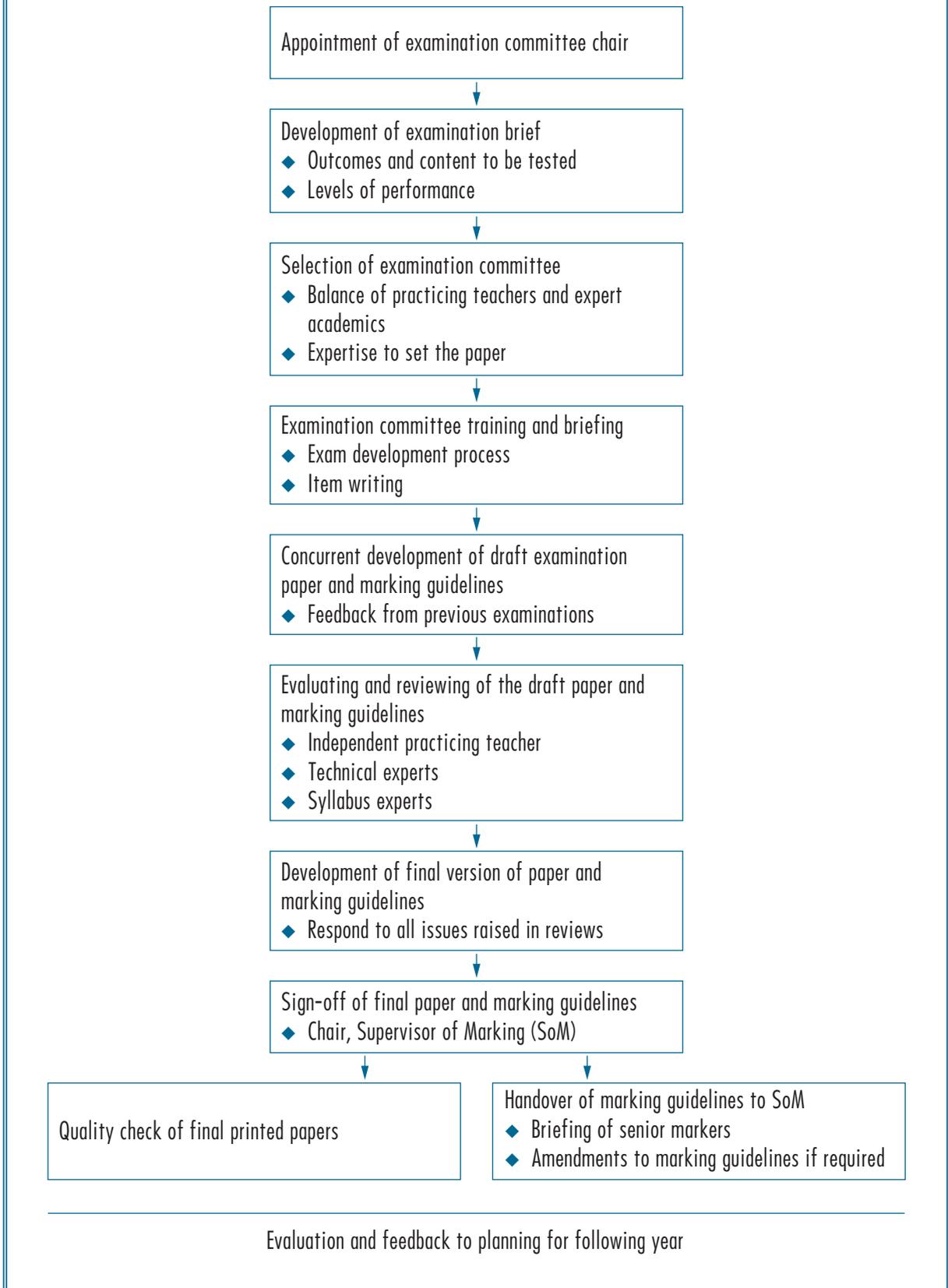


Figure 3.1 Steps in the Board's annual development of examinations

COURSE SYLLABUS (excerpt)

Personal Development, Health and Physical Education

Option 5: Equity & Health Module

Module description

This option module is concerned with the achievement of health for all and the actions necessary to realise this goal. In this module, students build upon their understanding of equity and social justice introduced in the core module Health Priorities in Australia. They examine the populations that experience health inequities and critically analyse social, cultural, economic and political factors that impact on the health status of these populations. Students think critically in order to discern actions that work towards reducing the gap in health status between populations.

As the major focus of this module, students conduct an examination of two populations experiencing significant health inequities: Aboriginal and Torres Strait Islander people and people living in rural and remote communities.

In this module, students investigate the following critical questions:

- ◆ Why do inequities exist in the health of Australians?
- ◆ How may the gap in health status of populations be bridged?
- ◆ What inequities are experienced by Aboriginal and Torres Strait Islander people and rural and remote communities?

Outcomes

A student:

- H1 describes the nature, and justifies the choice, of Australia's health priorities
- H2 analyses and explains the health status of Australians in terms of current trends and groups most at risk
- H3 analyses the determinants of health and health inequities
- H4 argues the case for the new public health approach to health promotion
- H5 explains the different roles and responsibilities of individuals, communities and governments in addressing Australia's health priorities
- H14 argues the benefits of health-promoting actions and choices that promote social justice
- H15 critically analyses key issues affecting the health of Australians and proposes ways of working towards better health for all
- H16 devises methods of gathering, interpreting and communicating information about health and physical activity concepts.

Suggested assessment strategies

- ◆ Prepare a critique of the major factors contributing to a particular health inequity.
- ◆ Conduct a comparative study of the inequities experienced by two different groups.
- ◆ Analyse the appropriateness of a current health promotion strategy targeting a particular health inequity.
- ◆ Formulate a set of recommendations for future directions in addressing the health inequities of a particular group.
- ◆ Describe how a commitment to social justice impacts on decision-making related to health and the allocation of health resources.

Figure 3.2 Example of HSC course syllabus (excerpt from PDHPE)

EXAMINATION SPECIFICATIONS**Personal Development, Health and Physical Education**

Time allowed: Three hours (plus 5 minutes reading time).

Section I Core (60 marks)**Part A (20 marks)**

- ◆ There will be TWENTY multiple-choice questions.
- ◆ All questions will be compulsory.
- ◆ All questions will be of equal value.
- ◆ Questions will be based on the two HSC core modules.

Part B (40 marks)

- ◆ There will be TWO questions: one on each of the HSC core modules. Each question may consist of several parts.
- ◆ Both questions will be of equal value.
- ◆ Both questions will be compulsory.

Section II Options (40 marks)

- ◆ There will be FIVE questions: one on each of the five HSC options. Each question may consist of several parts.
- ◆ All questions will be of equal value.
- ◆ Candidates must attempt TWO questions.

Figure 3.3 Example of examination specifications (excerpt from PDHPE)

EXAMINATION BRIEF
Personal Development, Health and Physical Education

Section II (40 marks)

- ◆ There will be FIVE questions: one on each of the five HSC options. Each question may consist of several parts.
- ◆ All questions will be of equal value
- ◆ Candidates must attempt TWO questions

Examination section and/or question number	Marks	Syllabus/course outcomes	Content area	Question structure	Targeted performance bands
Section II Question 27 Equity and Health	20	H3 H3, H4, H5, H14 H4, H5, H14, H15, H16	<ul style="list-style-type: none"> ◆ Actions that improve health ◆ Factors that create health inequities in relation to Aboriginal T.S.I. people ◆ Evaluation question that focuses on the characteristics of effective strategies in relation to improving the health of people living in rural and remote locations 	a 3 marks b 5 marks c 12 marks	2–3 2–4 2–6

Figure 3.4 Example of examination brief (excerpt from PDHPE)

EXAMINATION QUESTION**Personal Development, Health and Physical Education****Question 27 – Equity and Health (20 marks)**

- a Outline how 'advocating' can be used to enhance the health of a disadvantaged group. 3 marks
- b Discuss how socioeconomic status AND access to health care influence the health status of Aboriginal and Torres Strait Islander peoples. 5 marks
- c Evaluate the characteristics of effective health promotion strategies for improving the health of people living in rural and remote locations. 12 marks

Figure 3.5 Example of examination question (excerpt from PDHPE)

EXAMPLE

MARKING GUIDELINES

Personal Development, Health and Physical Education

Question 27 (c) (12 marks) Outcomes assessed: H4, H5, H14, H15, H16

Criteria	Marks
<ul style="list-style-type: none">◆ Demonstrates an understanding of the characteristics of effective health promotion strategies◆ Demonstrates a clear understanding of the factors which influence the health status of people living in rural and remote locations◆ Makes judgements about the appropriateness of different health promotion strategies in improving the health of people living in rural and remote locations◆ Provides a range of relevant examples	10–12
<ul style="list-style-type: none">◆ Demonstrates an understanding of a range of characteristics of effective health promotion strategies◆ Discusses how these strategies can improve the health of people living in rural and remote locations◆ Provides relevant examples	7–9
<ul style="list-style-type: none">◆ Describes characteristics of health promotion strategies with some link to the health of people living in rural and remote locations <p>or</p> <ul style="list-style-type: none">◆ Thorough description of the characteristics of effective strategies of health promotion <p>or</p> <ul style="list-style-type: none">◆ Thorough description of the factors influencing the health of people living in rural and remote locations <p>or</p> <ul style="list-style-type: none">◆ Thorough description of the health status of people living in rural and remote locations	4–6
<ul style="list-style-type: none">◆ Identifies a range of effective health promotion strategies <p>or</p> <ul style="list-style-type: none">◆ Identifies a range of factors influencing the health of people living in rural and remote locations <p>or</p> <ul style="list-style-type: none">◆ General statement about health promotion strategies or people living in rural and remote locations	1–3

Answers could include:

Effective strategies: eg

- move towards equity
- working with target groups in program design and implementation
- ensuring cultural relevance and appropriateness
- ensuring credibility of those delivering the study

Factors influencing the health of people living in rural and remote locations:

- geographic and social isolation
- access to services
- lack of infrastructure
- structural factors (economic resources, education)
- occupational hazards
- unemployment

Figure 3.6 Example of marking guidelines (excerpt from PDHPE)

Issues

A number of submissions to the Review raised issues concerning the content of the 2001 examinations.

Some of these issues were not specifically related to the new HSC arrangements. Following most public examinations, concerns are raised about aspects of papers (for example, that a particular multiple-choice question had two correct answers, or that students were disadvantaged by the wording of a question or by the choices offered). Some submissions to the review raised issues relating to the wording, layout and reproduction of the 2001 examinations. Particular questions were claimed to be poorly-worded, ambiguous/misleading, too difficult, too easy, to have more than one correct answer, or to contain grammatical or typographical errors. Other concerns were raised about examination instructions, reading times and response spaces (particularly answer booklets). Among these concerns, the most frequent claims were that there had been:

- ◆ grammatical, spelling and punctuation errors in the Japanese Continuers paper;
- ◆ inconsistencies and poor reproduction in the Geography broadsheet; and
- ◆ a complicated Ancient History paper that made it difficult for students to follow instructions and resulted in students making incorrect choices.

There is no reason to believe that the number of issues of this kind was greater than the number that would have been raised in relation to past HSC papers or in any large, complex examination system.

Other issues appear to have arisen as a direct consequence of changes to syllabuses and examinations under the new Higher School Certificate. These issues can be grouped into a few broad areas of concern.

Syllabus coverage

The most commonly raised concern about the content of 2001 HSC examinations was a concern about the extent to which some papers adequately reflected the course syllabus. A small number of submissions claimed that there were examination questions on content not included in the relevant HSC syllabus, but a far more common claim was that examination papers provided narrow and unrepresentative samples of syllabus content and outcomes.

Submissions raising this concern often claimed that examinations had ignored important areas of the syllabus and had placed a disproportionate emphasis on relatively minor aspects of the course, leaving students and teachers feeling that much of what had been taught and learned had remained untested. Some submissions speculated that particular 2001 examinations would have an impact on pedagogy as teachers endeavoured to cover all aspects of a syllabus rather than concentrating on broad themes and higher-order understandings.

Concerns about syllabus coverage were raised most often in relation to examinations in history, Visual Arts, PDHPE, Business Studies and a number of science and technology courses. In some cases, the issue was presented as a problem of too much syllabus content, creating problems in achieving adequate coverage in the examination. A number of submissions claimed that this was a particular issue in science courses. One submission referred to students 'learning vast amounts of information that they were unable to use'.

In other cases, the issue was a perceived disparity between the importance given to topics in the syllabus and their relative contributions to the examination. For example, one submission claimed that minor concepts in the Modern History syllabus had become the focus of major questions in the examination. That submission argued that, if each 'dot point' in the Modern History syllabus could become the basis of an extended examination question, then the result would be 'a focus on minutiae and detail, trivialising learning and leading to poor teaching practice'.

Related to the issue of syllabus coverage were occasional concerns that the marks awarded to examination questions did not always reflect their relative demands.

Specimen papers

With the introduction of the new Higher School Certificate, schools looked to the specimen papers provided by the Board, rather than to past examination papers, for guidance on the likely content of the 2001 examinations. Differences between the specimen papers and the 2001 examination papers were the subject of considerable comment, particularly in relation to English Extension 1.

Submissions argued that the specimen paper in English Extension 1 had established an expectation about the structure of the paper. In the specimen paper, there were two different types of questions for each elective and module: one requiring students to answer in a critical fashion and the other providing an opportunity for an imaginative response. The specimen paper had applied this pattern consistently, leading to 'a reasonable assumption of teachers that the HSC exam would follow this pattern'. The 2001 examination was considered not to have followed this pattern. Submissions argued that both questions on the examination required an analytical response, that this had confused students, and that a narrower range of learning outcomes had been addressed in the 2001 examination than in the specimen paper.

Language demands of examinations

A substantial number of submissions raised concerns about terminology and language. Many of these concerns related to the glossary of key words (see page 9) published by the Board to provide a common language and consistent meaning and to assist teachers and students in understanding what is expected in responses to examination questions and assessment tasks.

A number of submissions argued that the use of the glossary had resulted in too much emphasis on the decoding and application of verbs, resulting in an over-emphasis on students' literacy skills and placing students who were less familiar with the glossary at a disadvantage. Some submissions commented that examinations appeared to be composed of formulaic questions: a verb from the glossary followed by a dot-point outcome from the syllabus.

There were also concerns that a rigid adherence to the glossary would lead to the loss of subject integrity as the distinctive vocabulary of disciplines was discarded in favour of generic terms. Some claimed that the result would be less precise and potentially misleading examination questions within a subject; others, that the terms in the glossary could more readily be applied to some topics than others, resulting in distortions and a loss of meaning when the same verb was applied in each of several optional questions.

Further concerns relating to the glossary were that:

- ◆ there was sometimes an inconsistency between the verb used in the syllabus outcome and the higher-order behaviour required by the examination (eg, 'evaluate' as opposed to 'discuss');
- ◆ the mark values of questions did not always reflect the demands inherent in the verbs (eg, questions requiring students to 'describe' could be worth more than questions requiring students to 'analyse'); and
- ◆ some examinations used verbs that were outside the glossary, or used verbs incorrectly.

A somewhat different concern raised in a number of submissions was that there had been an overall increase in the language demands of some examinations. This concern was most often raised in relation to science and technology examinations. It was argued that the number of extended-response questions had increased in these examinations at the expense of questions requiring quantitative analysis. In some courses, particularly in the sciences, this shift in balance was seen as being associated with an increasing emphasis on historical and social issues in syllabuses. There were concerns that the increasing literacy demands of these courses and examinations would disadvantage particular groups of students, such as boys and students from non-English speaking backgrounds.

Parity of options

Many submissions raised concerns about apparent disparities in the difficulty levels of optional components of examination papers. While a number of examinations were mentioned, this concern was raised most frequently in relation to English and history subjects. There was reference in particular to the perceived difficulty of the King Lear question in English Advanced and to the 'narrowness and obscurity' of questions on Germany and Indochina in the Modern History paper.

In a few instances, the concern about lack of parity related not only to the relative difficulties of options but also to differences in the opportunities options gave students to demonstrate higher levels of attainment in a course.

Examination development

A number of submissions addressed the examination development process, particularly the appointment and roles of examination committees and assessors. There were calls for increased transparency in the criteria and processes for selecting committee members and assessors. The importance of recent experience teaching the relevant syllabus was raised, and a number of submissions noted the particular difficulties created by excluding current Year 12 teachers from the examination development process in the first year of the new Higher School Certificate.

The most frequent request was for increased representation of practicing schoolteachers on examination committees. There were several requests for the number of assessors to be increased to provide broader coverage of the syllabus and to allow consideration of the draft marking guidelines together with the paper. There were also concerns about feedback mechanisms from assessors to examination committees, and about changes made to papers after they left examination committees, leading to a loss of 'ownership' by the committee. Some respondents raised questions about the Board's processes for responding to concerns and for incorporating feedback into future examination and marking procedures.

Marking guidelines

References to marking guidelines were made in a number of submissions and in relation to a range of courses.

A common concern was that marking guidelines were not always able to encompass the diversity of students' responses to examination questions. They were sometimes considered too inflexible to be able to recognise and reward unanticipated responses; to be narrow and written with a specific answer in mind. In some instances, there was a view that students had been penalised for omitting relatively trivial details in otherwise high-level responses, and that answers that would have satisfied more general guidelines were undervalued.

The perceived deficiencies in marking guidelines were commonly attributed to shortcomings in the process used to develop them. Examination committees were sometimes described as lacking the teaching and marking experience that provides exposure to the breadth of likely student responses, and calls were made for the increased involvement of experienced senior markers in the development of marking guidelines. Respondents who were members of examination committees sometimes referred to difficulties encountered in developing marking guidelines, and to their concerns about the time and expertise available for this task. Some submissions also suggested improvements to the processes for reviewing draft guidelines, including the use of independent assessors to appraise not only the examination, but also the marking guidelines.

Comment

Public examinations must satisfy a number of criteria. Among these criteria are the requirements that they meet acceptable standards of validity and reliability and do not unfairly disadvantage particular categories of candidates.

Higher School Certificate examinations provide *valid* measures of achievement to the extent that students' marks accurately reflect their levels of achievement of the intended learning outcomes in each course syllabus. The validity of examination marks as measures of achievement in a course depends on an appropriate and adequate sampling of the course syllabus by the examination. If an examination does not adequately sample the syllabus, then not only are marks less valid indicators of students' course achievements, but there is a risk of future teaching and learning being distorted to address only those outcomes that have been examined.

HSC examinations also must provide measures of achievement that are *fair* to different groups of students. It is well known that the choice of contexts and subject matter for test and examination questions can affect the relative performances of males and females. Care must be taken in developing examinations to avoid unintended dependence on local or cultural knowledge that may be available to some students but not to others. And the way in which questions are worded can make them overly dependent on English language competence and unfair to students from non-English speaking backgrounds.

Many of the issues raised in submissions to the review related either to the validity or fairness of the 2001 HSC examinations.

Ensuring validity

The examination procedures introduced for the new Higher School Certificate have as their primary purpose the measurement of students' levels of achievement in the area of learning described by each course syllabus. Students' examination marks are intended to reflect their levels of achievement of learning outcomes (knowledge, skills and understandings) identified in the syllabus. Examination marks complement school assessments which in general are expected to address a somewhat broader range of outcomes.

This emphasis on marks as measures of students' levels of achievement in a course is subtly different from the emphasis in past HSC examinations. The McGaw review noted that, at that time, assessment procedures in the Higher School Certificate were primarily concerned with ordering students by results and did not address directly the question of the levels of achievement students had reached⁵. The primary purpose of examinations under the new Higher School Certificate is to assess and report students' levels of achievement of intended learning outcomes. This subtle shift in emphasis has implications for the way in which examinations are developed and validity is evaluated.

Syllabus coverage

The focus of teaching and learning in the senior secondary school is strongly influenced by the content of HSC examinations. In an ideal world, examinations might be less influential in shaping what happens in schools, but given that HSC results are a key determinant of students' chances of being admitted to higher education and of securing future employment, HSC examinations continue to exert a powerful influence on teaching and learning. For this reason alone it is essential that examinations focus attention on valued learning outcomes in a course and do not distort syllabus intentions by addressing only conventional or easily measured outcomes.

⁵ McGaw (1996), *op. cit.*, p 64.

Public examinations sometimes narrow the focus of teaching and learning because of the predictability of their content. Relatively open questions on major content areas appear year after year and come to be expected by candidates and their teachers. Students often enter examinations, if not with prepared answers, then at least with the expectation that they will be able to find questions that will allow them to use the material they have memorised. In the course of the review consultations, one person referred to this approach to answering examination questions as a 'brain dump'. Examinations of this kind can encourage a narrowness of learning: a focus on preparing in-depth responses on a limited number of predictable topics.

The intention of the new Higher School Certificate is to provide a more explicit specification of the knowledge, skills and understandings that students are expected to develop through a course and to use examinations and school assessments to measure students' levels of achievement in relation to those outcomes. Under this approach, the point of reference for the design of each year's examination is the set of intended learning outcomes in the relevant course syllabus.

Importantly, examinations will be designed to *sample* each course syllabus. Examination questions will test students' knowledge, skills and understandings, not in an attempt to cover everything in the syllabus, but to achieve an adequate sample of the syllabus content and outcomes. The result of this approach is likely to be more variability and less predictability in the content of examinations from year to year. The validity of each year's examination will depend not on whether students have been given an opportunity to demonstrate everything they have learned and prepared, but on the adequacy (amount and representativeness) of the examination's sampling of the course syllabus.

Some submissions to the review expressed concern that examinations in 2001 ignored important areas of the syllabus and left students and teachers feeling that much of what had been taught and learned had remained untested. Some of these concerns may reflect a lack of familiarity with the intention under the new HSC to sample syllabus content and outcomes in examinations. The decision to address representative samples of syllabuses means that everything within a course is available to be assessed in an examination, but no one examination will ever assess an entire syllabus.

Other submissions expressed concern about the level of detail tested in some 2001 examinations. In the move towards greater explicitness about intended learning outcomes in course syllabuses, and to develop examinations that more directly address and provide information about those outcomes, it will be important that the Board monitors the level of detail addressed by examination questions. In general, examinations should endeavour to test for fundamental rather than peripheral or superficial knowledge, and to assess for conceptual understanding rather than the memorisation and formulaic application of knowledge.

Related to this concern was a belief that some syllabuses, particularly in the sciences, currently contain too much content, and that attempts by teachers and students to cover this content in preparation for examinations are likely to lead to superficial treatment and learning. While the consideration of syllabus content is outside the terms of reference for this review, the frequency with which concerns about overcrowded science curricula were raised in submissions suggests that this issue may warrant further attention.

In light of questions raised about syllabus coverage following the 2001 examinations, the Board should reinforce with schools the reasons for sampling in HSC examinations. Guidance on what can be done to ensure a representative sample of syllabus content and outcomes needs to be given to those planning examinations. There also needs to be inclusion in the Board's examination development process of explicit checks on the adequacy of sampling and on the level of detail assessed in each examination.

Specimen papers

Some concerns about the specimen papers prepared and distributed by the Board prior to the 2001 examinations may be a further reflection of a lack of understanding that, under the new HSC, different examinations in the same course will sample the course syllabus differently. It will be less possible than in the past to predict the content of an examination from the content of past examinations. Some 2001 examination papers were not as similar to the specimen papers as some may have anticipated, and the 2002 papers are likely to be different again.

Nevertheless, there does appear to have been an issue in relation to the English Extension 1 specimen paper. Although the specimen paper indicated that in modules A, B and C, questions may require a critical, interpretive or imaginative response or any combination of the three, the fact that the specimen paper consistently required both an analytical and an imaginative response seems to have led many teachers and students to assume that this structure would be repeated in the 2001 examination. Presumably this will not be an issue from 2002 because it will be understood that the modules in the English Extension 1 paper may require any combination of writing genres.

Ensuring fairness

Submissions included a number of concerns about the fairness of some questions or parts of papers to particular groups of students.

Clear instructions

It obviously is important that examination instructions make clear to students what is expected of them. Only one significant concern was raised about the clarity of examination instructions, and that related to the Ancient History paper. There was a belief that a number of students were confused by the complex instructions for that paper, and that some students were disadvantaged as a result.

The complexity of HSC history papers is widely recognised. The McGaw review concluded that in Modern History 'there was not really one History examination; there were many history examinations within the same paper'⁶. Given the complex structure of the history papers, complex instructions are required to navigate through them. The instructions for the 2001 Ancient History paper appear to have been particularly complex, and the Board should examine ways to clarify instructions and layout in papers of this kind.

Accessible language

Although proficiency in English is an expectation of all students undertaking the Higher School Certificate, it is important that HSC examinations do not unintentionally disadvantage students from non-English speaking backgrounds because of their dependence on high levels of competence in English. Some submissions questioned whether students had been disadvantaged in 2001 by the English language requirements of some examinations.

The glossary of terms has been developed by the Board in an attempt to ensure a consistent understanding of what is expected of students, both in syllabus statements of intended learning outcomes and in examination questions. It was clear that, despite the availability of the glossary prior to the 2001 examinations, and discussion of the intent of the glossary through the professional development program, some students and teachers were not familiar with the terms in the glossary.

An issue that the Board will need to monitor is the possibility of overly-rigid interpretations of terms in the glossary. The glossary appears to have value as a way of communicating what is expected of students. But the rigid application of the terms in the glossary may result in formulaic approaches to question development and an unintended level of inflexibility in the marking of students' responses. Particular attention should be given to the possibility that the terms in the glossary may

⁶ McGaw (1997), *op. cit.*, p 136.

require slightly different interpretations in different disciplines. It probably will be necessary for the Board to clarify further its reasons for introducing the glossary and its intentions regarding the glossary's use in developing examination questions and marking student work.

A more general concern was that increases in the language content of some examinations may have disadvantaged particular groups of students (usually boys and students from non-English speaking backgrounds). One person commented that some science examinations were now in part tests of 'literacy'; another, that the physics examination had become less a test of physics and more a test 'about' physics.

It is inevitable that an attempt to place less emphasis on testing students' abilities to recall and apply mathematical formulae, and more emphasis on students' conceptual understanding, ability to apply knowledge and skills to problems, and understanding of the contexts in which learning can be applied, will require greater use of language. Concerns about the increased language content of some examinations often appear to have been concerns about the changed nature of the syllabus. It has not been possible or appropriate for this review to form a judgement about changes to syllabuses that may have resulted in increased language demands in some examinations. The challenge no doubt will be to maintain an appropriate balance in syllabus content in courses such as Physics.

Parity of options

The use of optional questions in examinations is designed to allow a level of student choice and to cater for differences in the way in which HSC courses are taught. The intention is that students should be able to develop their knowledge, skills and understandings in different contexts (eg, in relation to different periods of history) and to choose to answer different questions without being penalised for these choices. A number of submissions felt that students were disadvantaged by the difficulties of some optional questions.

The issue here is not so much whether optional questions are equally difficult, as whether students choosing different options are treated fairly when those options are not of the same difficulty. In general, it is desirable that an attempt is made to write optional questions so that they are of approximately equal difficulty. However, this is not always possible, and when options differ in difficulty, these differences should be identified and statistical adjustments made to remove the effects of unintended differences.

The Board currently has a routine 'question scaling' process that looks for unintended differences in the difficulties of optional questions and makes appropriate adjustments. However, this process does not seem to be widely known – even by markers – and it may be useful for the Board to include the intention to adjust for differences in the difficulties of optional questions in its examination development principles.

Examination development processes

As is evident from the description of the Board's process for developing HSC examinations (pages 13–22), a thorough set of procedures has been established for appointing key personnel to undertake the development of the examination in each course, for planning and developing the content of each examination, and for checking and reviewing examination papers. Most of these procedures have evolved and been modified over a number of years and appear well suited to ensuring high quality HSC examination papers. A small suggestion for improving these procedures would be to consider having more assessors and/or giving them a larger role in some courses.

The process for developing public examinations such as the HSC differs from processes used to develop other kinds of educational tests, largely because of the need to ensure strict security of examination materials. In the development of standardised achievement tests it is usual to begin by developing several times the number of questions required for the final test, to field test those questions on several hundred students, to analyse statistically responses to those questions, and

then to select questions for the final test to provide the desired test properties (eg, range of question difficulties). The marking key for each question is seen as an integral part of the test question and is developed and trial tested together with the question. This process provides an opportunity to see how students respond to each question, to identify unanticipated responses and to modify marking keys where necessary. The finished test consists of a set of questions and a finalised marking key for those questions.

These standard test development procedures are not available to the development of examinations. Trial testing within the same state is impossible because of risks to test security. Trial testing in other states or countries introduces questions about the likely usefulness of the trial data given differences in curricula. And even the common practice of including in secure high-stakes tests some unscored trial items for possible use in future tests is not possible because each year's paper becomes public immediately following the examination.

Examination systems have been able to function without the usual trial testing in part because marking schemes are not finalised at the same time as examination papers. It is not uncommon for questions to be written, papers printed and examinations taken before markers turn their attention to how students' responses are to be marked. The development of schemes for marking papers is informed by an inspection of how students responded to each question in the examination.

The Board's decision to make the development of marking guidelines part of the examination development process is consistent with usual test development practice. It also is consistent with the intention to move to standards-referenced examinations and marking procedures. This move requires greater clarity about the outcomes that questions are designed to address, and marking schemes that are based on these intended outcomes. However, the inability to trial HSC examinations means that marking guidelines and marking schemes must remain tentative and can be finalised only on the basis of information about actual student responses.

The new HSC procedures for the development of marking guidelines sit between standard test development practice and past HSC practice under which the development of marking guidelines usually was the responsibility of senior markers rather than the examination committee and occurred after the finalising of papers. As will become clear in the next section, there was a lack of clarity on the part of some markers about the extent to which marking guidelines should be treated as given, and the extent to which they could be modified. This is a matter that should be clarified for all markers from 2002.

Recommendations

It is recommended that the Board:

- ◆ ensure better dissemination of the reasons for, and nature of, sampling of content and outcomes in HSC examinations;
- ◆ include in the Board's examination development process explicit checks on the adequacy of sampling and on the level of detail assessed in each examination;
- ◆ further clarify its rationale for the Glossary of Terms and its intentions and processes for the use of the Glossary in developing examination questions and marking student work; and
- ◆ provide an explanation of the 'question scaling' process designed to adjust for unintended differences in the difficulties of optional questions.

4 Marking examinations

Higher School Certificate examinations are marked each year by several thousand experienced teachers in marking centres across country New South Wales and the Sydney metropolitan area.

Supervisors of Marking (SoMs) for each course or component of a course are appointed to manage the process which is designed to ensure that each student's performance, project or written response is marked accurately and fairly. The Board has developed a range of quality assurance processes, including the training and briefing of markers, pilot (practice) marking and the monitoring and review of the marking process and its outcomes.

In 2001, 7000 markers were involved in the marking of HSC examinations. Their duties included marking performances or projects in 12 courses, speaking skills in 51 language courses, and students' responses to 154 different written examination papers. Approximately 440 markers visited schools across New South Wales to mark student performances and projects. The marking of written examinations was undertaken at 13 venues, mainly at major centres in Sydney.

The process

The major stages in the Board's marking process include the selection and training of markers; the hand-over of the marking guidelines; the development and pilot testing of a marking scheme and benchmark responses; the briefing of markers; and the marking process itself.

Appointment of markers

The Board appoints various categories of markers who play different roles in the marking process. In any given course, the Supervisor of Marking (SoM) has responsibility for overseeing marking in that course. Under the direction of the SoM, senior markers lead and supervise groups of markers. The appointment of markers generally is based on their level of experience in teaching the course, the recency of that experience and their general teaching experience.

Supervisors of Marking (SoMs) have administrative responsibility for the marking operation for a course or course component. They are responsible for recommending the appointment of markers and senior markers, for developing and implementing a marking plan, for managing the operation of the marking centre, and for ensuring the accurate and timely marking of all examination responses.

Supervisors of Marking are appointed following interview by a cross-sectoral panel and usually have extensive experience both as a marker and senior marker.

Assistant Supervisors of Marking assist the Supervisor of Marking in courses with very large candidatures or in multi-discipline subjects such as Visual Arts and Mathematics.

Senior Markers are appointed on the basis of their marking experience and demonstrated leadership skills. Their role is to lead and supervise groups of up to ten markers in the marking of one section or question on a paper, depending on the size of the candidature and the nature of the examination. Senior markers are responsible for the briefing and monitoring of markers, check marking, the analysis of group results and the resolution of discrepant markings.

Pilot Markers begin marking earlier than other markers and assist senior markers to establish initial understanding of the marking guidelines and to assess the calibre and range of responses received from students.

Preparation for marking

Following their appointment, Supervisors of Marking take part in a range of training and preparation activities. Briefings are conducted in broad subject groups (eg, all the Sciences; all English courses).

In 2001 the briefing sessions for Supervisors of Marking focused on the changed procedures to support standards-referenced assessment. Input was provided by staff of the Board's Examinations and Certification Branch, Assessment Branch, Information Services Branch and Curriculum Branch. It covered:

- ◆ new HSC context;
- ◆ developing examinations and marking guidelines;
- ◆ the marking operation/management issues;
- ◆ quality control processes in marking;
- ◆ marker reliability operation;
- ◆ standards setting;
- ◆ marking logistics;
- ◆ marking resource plans;
- ◆ marker appointments; and
- ◆ special cases.

Most Supervisors of Marking are directly involved in the development of the course marking guidelines as part of the prior development of the examination. All are responsible for signing-off an approval of the marking guidelines as capable of being implemented in the marking centres.

The formal 'hand-over' of the marking guidelines occurs at the beginning of the marking process. Ideally, a Board officer who has been directly involved in the development of the examination and guidelines, the Chair of the Examination Committee, and the Supervisor of Marking are involved in briefing senior markers to explain the process by which the examination and guidelines were developed.

Prior to marking, Supervisors of Marking submit a range of marking management plans for approval. These include:

- ◆ a quality control plan;
- ◆ a resource plan;
- ◆ a selection of markers and senior markers;
- ◆ senior marker briefing notes;
- ◆ marker briefing notes; and
- ◆ mark sheet sign off.

Marking schemes and benchmarks

The next step in the process is the development of a more detailed 'marking scheme' that is later used by marking teams. The marking scheme, based on the Examination Committee's marking guidelines, is developed by the senior markers working with the Supervisor of Marking. While the marking guidelines identify and describe performances at the various mark levels, the marking scheme adds detail and elaborates particular components of anticipated student responses.

As an initial step, senior markers inspect a range of student responses to test the guidelines, to select illustrative 'benchmark' samples of student responses, and to develop annotated marking schemes if required. In some cases, a marking scheme takes the form of additional notes to accompany the marking guidelines; in other cases the marking schemes take the form of documented 'benchmark' performances, especially in practical examinations.

For example, in Visual Arts, each expressive form has a marking scheme that includes photographs of student work at each mark range, and a detailed description of the elements that have contributed to the award of those marks.

In written examinations, benchmark scripts are sometimes provided as examples. In other courses, such as Software Design and Development, benchmark scripts are incorporated into a booklet of information with notes on the elements of the marking guidelines demonstrated in each of the benchmark examples. This booklet forms the marking scheme provided to markers.

Pilot testing of marking guidelines

Pilot testing occurs concurrently with the finalisation of guidelines, marking schemes and associated materials. In this process senior markers test the guidelines and scheme against a representative range of responses. This pilot testing provides an initial test and assists in refining the selection of benchmark scripts. In some courses senior markers are assisted by particularly experienced markers brought in specifically for this process.

Alterations to guidelines

It is anticipated that the pilot marking process will lead to elaboration of the marking scheme and the identification of a broader range of examples to illustrate the marking scheme. Pilot marking also may result in proposed changes to the marking guidelines, although advice from the Board of Studies anticipates limited alterations to marking guidelines at this stage, particularly in relation to the outcomes and content being examined.

Alterations to marking guidelines proposed by the Supervisor of Marking as a result of pilot marking are endorsed by the Chair of the Examination Committee and/or the Assessment Officer and are referred to the Director, Examinations and Certification for approval. Changes are made to the master copy submitted to the Office of the Board following marking.

The marking process

Following development of the marking scheme and pilot testing of the marking guidelines and related materials, the marking process itself commences. Markers are first briefed and various procedures are then used to monitor the progress of the marking.

Marker briefing

During the briefing process, markers are introduced to the marking guidelines, detailed marking schemes (where they are necessary) and 'benchmark' responses at a range of performance levels described in the marking guidelines. The materials are discussed at length to ensure that all markers have a shared understanding of the requirements of each performance level.

In 2001 the Board prepared briefing notes for all markers and senior markers covering all aspects of the setting and marking process. The purpose was to ensure that the information provided was consistent across subjects and to reassure markers that, while the process for the development of marking guidelines had changed (in that the marking guidelines were now developed as part of the examination and not by senior markers), the marking process itself was largely a confirmation of previous 'best practice'.

Pilot (practice) marking

In this process, markers apply the guidelines and other materials to a range of responses drawn from a number of representative examination centres. Typically, markers discuss in small groups the marks they have awarded. This discussion provides a further opportunity to achieve consensus and to consider responses that test the breadth of the marking guidelines. Pilot marking continues until senior markers are confident that the guidelines are being applied accurately and consistently. This activity sometimes is referred to as 'pencil marking' because the marks awarded are subject to check marking, and responses may be returned to the pool for marking at a later time.

As a result of pilot marking, changes may be made to marking guidelines to ensure that they adequately capture evidence of the outcomes and content that individual questions are designed to assess. Changes are recommended through the Board facilitator and the Chair of the Examination Committee to the Director, Examinations and Certification, and are approved if justified by the provided evidence. In 2001, the majority of Supervisors of Marking requested alterations to some aspects of the marking guidelines.

Check marking

Once marking has commenced, a proportion of the scripts marked by each marker is passed to a senior marker to check mark. The senior marker marks each student's response and checks that the mark awarded by the marker is accurate. If a pattern emerges showing that marking guidelines are not being applied correctly by a particular marker or group of markers, then re-briefing occurs to redress this anomaly.

Check marking generally begins with a high proportion of scripts being check marked. As consistency and confidence develop, the proportion of check marking is reduced.

Statistical reports also may identify a need to focus check marking on a particular marker. Check marking is subject to audit by the Supervisor of Marking.

Marking common (control) scripts

Supervisors of Marking and senior markers also use common (control) scripts to identify markers who are not applying the marking guidelines or scheme consistently.

The use of common scripts can vary from one subject to another. Generally a script from each question being marked within the centre is photocopied and distributed to all markers of that question at least once per marking session.

Results on common scripts are compared with previous sessions and group data to identify patterns in the marking process and to either confirm satisfactory marking patterns or to identify inconsistencies. Markers showing inconsistency or lack of appropriate application of the marking guidelines receive individual attention to redress the problem.

In subjects with high proportions of objectively-marked questions, common scripts are less relevant, and marking guidelines, marking schemes and check marking are used instead to ensure comparability.

Statistical reports

Statistical checks form part of the Board's quality assurance processes for marking. Marks awarded by each marker are tallied and processed at regular intervals. A report identifying markers who are marking significantly above or below the average for the marker group is provided to the Supervisor of Marking. While this report may simply reflect the standard of a particular group of responses, it also can point to a marker who is unusually lenient or unusually harsh. Early identification ensures that re-briefing can be given where necessary.

Statistical reports also identify markers using an unusually narrow range of marks. Again, this may simply indicate a group of average students' responses, or it may indicate conservatism or lack of confidence on the part of a particular marker.

A report on the rate of marking for each member of the marker group is used to adjust the allocation of resources within a marking centre, and allows the Supervisor of Marking to monitor the overall progress of the marking operation.

Supervisors of Marking use statistical reports to highlight perceived anomalies between the anticipated target range of marks (as indicated by the Examination Committee's mapping grid) and the spread of marks resulting from the application of the marking guidelines.

These various statistical reports can be used by the Supervisor of Marking to identify areas that need finetuning. Their use throughout marking provides a monitoring mechanism for the Supervisor of Marking that contributes to the quality assurance of the marking process.

Double marking

In subjects in which students provide responses that must be judged subjectively, the Board uses a system of double marking in which two markers make independent judgements of a student's response. Each marker allocates a mark in accordance with the approved marking guidelines, and is unaware of the other marker's judgement.

Double marking usually is applied to questions requiring an extended response (eg, essays, creative writing, projects and performances) and not to short answer questions of the kind commonly used in mathematics and the sciences.

When the marks assigned to a double-marked question differ by more than the maximum acceptable difference set by the Board, this difference is considered a 'discrepancy' and a third or possibly fourth marking of the student's response is undertaken. The Supervisor of Marking or another senior marker then takes all these independent markings into consideration and uses professional judgement to determine the most appropriate mark for the student's response.

Resolution of atypical responses

Atypical responses include answers that are outside the range of answers described by the marking guidelines, but which appear to be valid alternative interpretations of a question. They also include non-serious and offensive responses. Supervisors of Marking are required to have strategies in place to deal with atypical responses, all of which must be brought to the attention of a senior marker.

Valid alternative responses are discussed with the senior marker (and Supervisor of Marking if required). Senior markers and the Supervisor of Marking together determine and allocate a mark based on the quality of the response.

In the case of a non-serious attempt, senior markers may allocate a mark of zero, or may refer the matter to the Supervisor of Marking. Significant concerns are referred to the Director, Examinations and Certification for consideration. The Board's Examination Rules Committee considers cases of alleged breach of examination rules and malpractice.

Responses that exceed stated parameters

In some examinations, particularly projects and performances, there are limits placed on students' work. This limit may be a limit on the size, time or word length of the work. Where a student's response exceeds the stated parameters, a mark penalty may be imposed.

In written examinations, some questions indicate a limit on the length of the response. Overly long responses generally are not penalised because the student is considered to have already imposed a limit on the time available to answer other questions.

Revision marking

In subjects in which single marking applies (due to the objective nature of the questions), revision marking is used to provide further quality assurance.

At the end of the marking process, a report is run to identify students whose total examination marks are significantly different from their school assessment marks. The responses of these students are extracted and a group of senior or other experienced markers is recalled to re-mark these students' papers. This process confirms or alters the marks received by students.

Review and evaluation

Following the completion of each year's marking, an evaluation is undertaken of various aspects of the marking process, and reports are provided to the Director, Examinations and Certification. Evaluations are conducted of:

- ◆ the training of markers;
- ◆ the management of the marking;
- ◆ examination paper issues;
- ◆ marking guideline issues;
- ◆ quality control processes in marking;
- ◆ the marker reliability operation;
- ◆ marking logistics;
- ◆ marking resource plans;
- ◆ marker appointments; and
- ◆ the handling of special cases.

Confidential reports are prepared by all Supervisors of Marking.

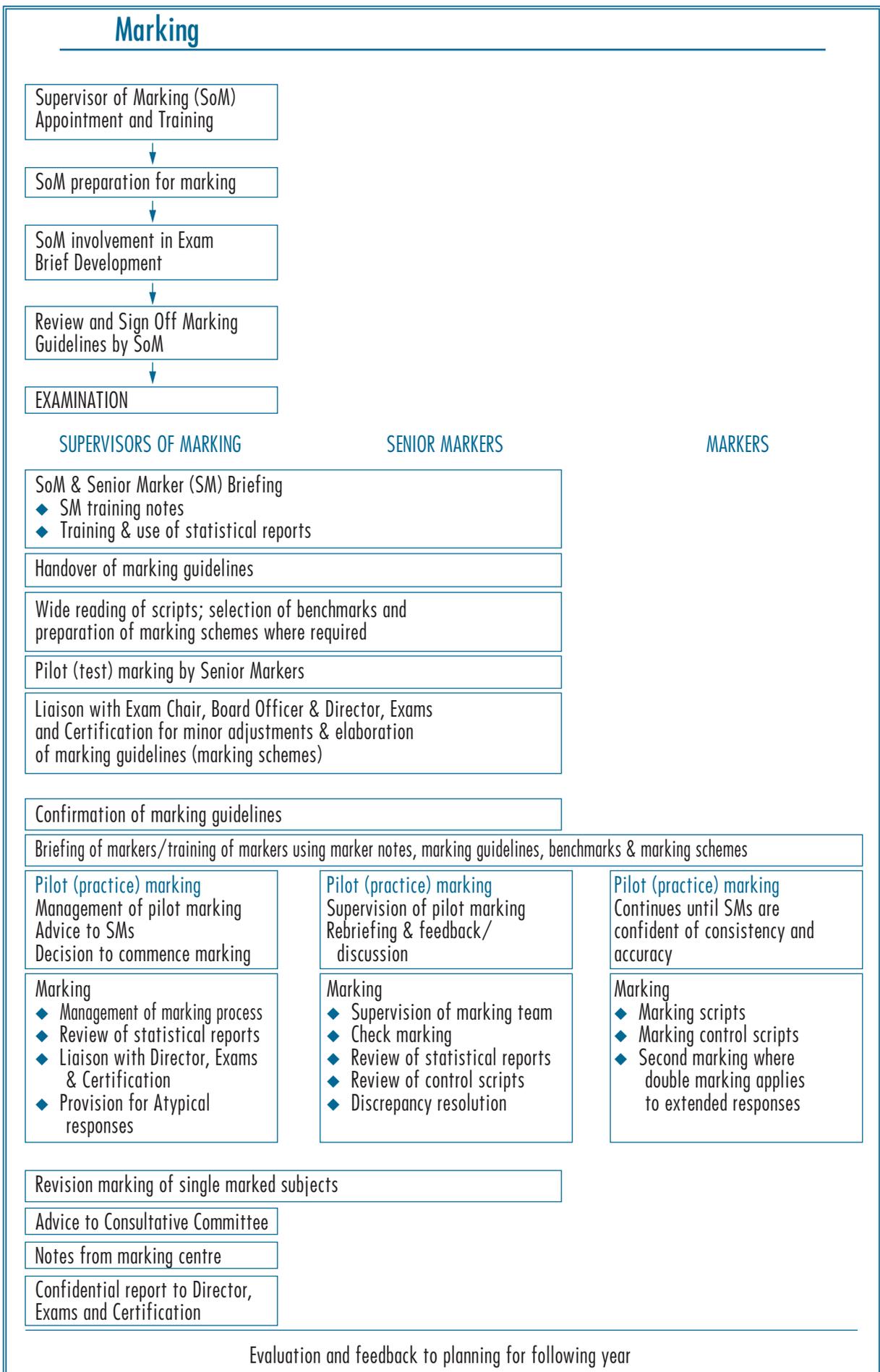


Figure 4.1 Steps in the Board's marking of examination papers

Issues

Many submissions to the review raised matters relating to the marking of students' HSC papers in 2001. Two general and related issues dominated these comments: the nature of the marking guidelines for some papers and the consistency with which marking guidelines were applied.

Marking guidelines

A common claim was that some marking criteria were too narrow. There was a belief that narrow criteria sometimes were enforced rigidly, denying markers the opportunity to use an appropriate level of professional judgement in applying the guidelines. The result, it was claimed, was that students who provided unanticipated responses or who answered outside the guidelines were penalised.

Some submissions expressed concern that marking criteria occasionally included 'key words' that students were required to use in their responses. According to these submissions, marks sometimes were influenced by the presence or absence of key words, rather than being based on the quality of a student's response. One submission claimed that 'poor responses which hit the key words got good marks; good responses which didn't mention them scored poorly'.

Other submissions took a more moderate view, although still expressing concern that some students 'did not receive the marks they deserved'. This concern often was expressed as a tension between instructions to mark against provided criteria, and markers' desires to reward what they saw as quality responses – even if those responses did not fulfil the criteria. One submission called for 'more flexibility in the marking process so that insight, sophistication and higher order thinking and expression can be appropriately rewarded even where there is a failure to meet all the demands of a question'. Failure to address the question was frequently cited as the reason for failure to satisfy the marking criteria, leading to calls for increased efforts to ensure that students were in no doubt about what questions were asking them to do. One submission claimed that: 'students who demonstrated knowledge of the course content and outcomes but who failed to correctly interpret the question failed to gain marks'.

Consistency in applying marking guidelines

In addition to concerns that some marking guidelines were too narrow and had been applied inflexibly were claims that guidelines had not been applied consistently. It was claimed that marking guidelines had been applied differently in different courses, between marking centres (in one instance), between questions or options within the same paper, and between individuals or groups of markers marking the same question. There were claims that marking guidelines had been 'thrown out', and other claims that guidelines had been applied rigidly despite concerns about their inflexibility.

The review was told that, in at least one instance, disquiet at apparent difficulties in the application of marking criteria led to the remarking of one section of a paper, while in another case the overly-rigid interpretation of criteria in one marking centre became apparent later in the process and had to be addressed by statistical adjustment.

A number of suggestions were made for improving consistency in the application of marking guidelines. These included:

- ◆ ensuring that Supervisors of Marking and senior markers have sufficient and timely input into the development and review of the guidelines;
- ◆ providing adequate processes for consultation with the Examination Committee in the implementation of marking guidelines and in any subsequent alterations;

- ◆ increasing the time allocated to markers and senior markers to develop an understanding of marking guidelines, prepare marking schemes, select appropriate scripts, and participate in pilot marking;
- ◆ ensuring that markers are thoroughly briefed on relevant aspects of standards-referenced approaches in the new Higher School Certificate, including the terms used in rubrics and marking guidelines; and
- ◆ establishing overarching coordination positions (sometimes termed Chief Examiner) to oversee examination and marking operations across a course or subject.

Comment

The introduction of a standards-referenced approach to Higher School Certificate examinations involves:

- ◆ a more explicit identification of the learning outcomes (knowledge, skills and understandings) that each examination question or task is designed to address; and
- ◆ the design of marking guidelines and schemes to provide information about students' achievements in relation to those outcomes.

Under past arrangements, there was a much weaker and less explicit connection between the writing of questions for a paper and the development of guidelines for marking responses to those questions. Many of the issues raised about the marking of HSC papers in 2001 originate in these changed arrangements under a standards-referenced approach.

General questions underlying many of the comments made in submissions relate to when and how marking guidelines are developed and finalised, how much flexibility there should be in the interpretation and modification of marking guidelines and schemes, and who should be involved in making these decisions. In 2001, there was some uncertainty around each of these questions.

Within the Office of the Board of Studies, the development of examinations and the marking of student scripts are the responsibilities of two different branches. The Assessment Branch is responsible for developing examinations, including marking guidelines. The Examinations and Certification Branch is responsible for marking student scripts, including revising marking guidelines and developing benchmark scripts and marking schemes. Although this division of responsibilities in general makes good sense, under the new HSC it has resulted in questions about the level of involvement Supervisors of Marking should have in the drafting of guidelines, and the level of involvement examination committees should have in the revision and operational interpretation of those guidelines.

To ensure quality guidelines and marking schemes consistent with syllabus outcomes and the intentions of examination questions, the Board has developed processes through which most Supervisors of Marking have some direct involvement in the development of marking guidelines. Supervisors of Marking also sign off an approval of the marking guidelines and, with senior markers, take part in a hand-over meeting with the Chair of the Examination Committee and a Board officer.

Following the 2001 marking operation, officers of the Examinations and Certification Branch identified a number of steps that could be taken to improve the hand-over and the relationship between the examination development and marking processes. These steps include:

- ◆ ensuring that all Supervisors of Marking attend the meeting with the Exam Committee/Chair and Assessment Officer at the time of sign-off;
- ◆ allowing sufficient time to discuss issues and to finalise guidelines prior to sign-off;
- ◆ scheduling the hand-over of marking guidelines after senior markers have had an opportunity to read the guidelines and to begin reading student responses;
- ◆ ensuring that all senior markers are present at the hand-over meeting;
- ◆ where possible, appointing the Chair of the Examination Committee to the marking team; and

- ◆ ensuring that the Chair of the Examination Committee or nominee is available after the hand-over of guidelines while senior markers are reading responses, selecting benchmark scripts and developing marking schemes.

Each of these suggestions should clarify the intentions of the marking guidelines and ensure that revisions to the guidelines prior to sign-off are informed by a consideration of students' responses.

Given the concerns that were raised in relation to some aspects of the marking in 2001, it is worth considering whether more could be done to integrate the processes of developing guidelines, revising guidelines, developing marking schemes and selecting benchmark scripts. The general challenge is to combine the intentions of the examination developers for each question and its associated marking criteria with evidence of how students actually responded to that question. The obvious solution is to ensure that those drafting the marking guidelines have marking experience, and that the Examination Committee's basic intentions for each question are not lost in the revision and implementation of the marking guidelines.

Better integration also might be achieved by making one person (a 'Chief Examiner' for a course) responsible for overseeing both the development and marking of an examination. This person could chair the Examination Committee as well as overseeing the marking operation for the paper.

Marking guidelines

A valid examination is one in which questions and tasks provide an adequate sampling of the intended learning outcomes for a course and so do not distort the focus of teaching and learning. A valid marking guideline is one that provides evidence about the outcome/s each question is designed to address. A number of submissions to the review raised issues about the marking guidelines used for some questions and some sections of papers in 2001.

In the past, marking schemes usually were developed by senior markers under the direction of the Supervisor of Marking. These schemes were different from the marking guidelines developed in 2001 to provide a more explicit identification of the syllabus outcome/s that each question was designed to address.

In the new Higher School Certificate, marking guidelines are developed as part of the examination and are provided to senior markers. Markers sometimes felt that provided guidelines were too narrow and restrictive and took insufficient account of the responses students actually gave to questions. It was clear that some markers felt that they were less able in 2001 to use their professional judgement to give credit to 'quality' responses – even though those responses may not have addressed the question or provided evidence of achievement of the outcomes targeted by the question.

Some concerns of this kind probably could be anticipated in the move to a marking system that is less impressionistic and more criterion-based. Nevertheless, reports of markers applying guidelines rigidly, including awarding credit for the mention of 'key words', suggest that some markers, in encountering the new approach for the first time, may have adopted a particularly inflexible interpretation of guidelines.

During and following HSC marking, claims were made in the media of particular instances where marking guidelines had been so unusable that they had been 'abandoned'. This claim is not consistent with the findings of telephone interviews conducted by Board of Studies officers with all Supervisors of Marking in November 2001. Those interviews revealed that no marking guidelines had been abandoned. In sixty per cent of marking centres, as a result of the pilot marking process, guidelines were expanded or enhanced to accommodate unanticipated student responses.

It clearly is important that marking guidelines are not interpreted as fixed and immutable or applied rigidly by marking teams. The challenge will be to achieve an appropriate balance between maintaining a focus on the outcomes that questions are designed to address, and retaining sufficient flexibility to allow professional judgement in the case of responses that are not well

covered by the guidelines. In most courses and most marking centres in 2001, this balance appears to have been achieved. It needs to be achieved in all courses and centres.

It may be useful for the Board to provide Supervisors of Marking and senior markers with further training and advice on principles and procedures for modifying marking guidelines. Particular attention could be given to advice on the application of marking guidelines and the place of marking schemes and benchmarks in supporting professional judgement and the flexible interpretation of guidelines. Where appropriate, training could include further advice on the interpretation and application of language (eg, the Glossary of Terms) in the marking process.

The Board also may wish to give consideration to ways of enhancing the work of senior markers in the processes of pilot testing and adjusting marking guidelines, developing more detailed marking schemes (where appropriate) and identifying benchmark scripts. The more thorough these steps in the process, the more useful the revised guidelines, marking schemes and benchmarks are likely to be in guiding the subsequent marking process.

Consistency in applying marking guidelines

The fairness of the marking operation to students depends on consistency in the application of marking criteria by the same marker over time, between markers, and across marking teams and centres. As the description of the Board's marking processes (pages 31-37) shows, a variety of steps are taken to ensure consistency of marking. These steps have been carefully designed and implemented and reflect international best practice in ensuring consistency of examination marking.

Although various claims of inconsistent marking were made, there was very little evidence of significantly inconsistent marking in practice. For example, the discrepancy rate between markers in 2001 does not appear to have been higher than in the past.

There was, however, some evidence of inconsistency. In the case of one marking centre, the Board's internal processes identified such a rigid interpretation of marking criteria that a statistical adjustment to students' marks was made to remove its effect. However, even in this centre, markers appear to have been consistent in applying their overly-rigid interpretation of criteria.

Finally, several submissions raised questions about the Board's confidentiality requirements, arguing that they acted as an impediment to markers expressing their concerns. Although it clearly is important that confidentiality requirements are in place for those involved in the development of examinations and marking guidelines throughout the year, and also for markers during marking, it is important that senior markers and markers are able to share lessons learned through the marking process. The Board's current practice of placing no restrictions on the ability of markers to present at conferences or to speak to groups of teachers should assist in this regard.

Recommendations

It is recommended that the Board:

- ◆ explore the feasibility of appointing a 'Chief Examiner' responsible for both the setting and oversight of the marking of the examination paper in each course;
- ◆ develop still closer integration of the examination development process and the process for revising marking guidelines, developing marking schemes and identifying benchmark scripts; and
- ◆ continue to explore ways of enhancing the processes of pilot testing and adjusting marking guidelines, developing more detailed marking schemes (where appropriate) and identifying benchmark scripts.

5 Standards setting

An important element of the examination procedures under the new Higher School Certificate is the process of interpreting students' examination performances in terms of a set of pre-specified course achievement 'standards'. The achievement standards in each course take the form of a hierarchy of six described levels referred to as 'bands'. Band 1 represents a level of achievement below the minimum standard expected in the course. Bands 2, 3, 4, 5 and 6 represent increasing levels of course achievement and are described in terms of the kinds of knowledge, skills and understandings typically displayed by students achieving at those levels.

Each year, the standards-setting process is used to determine the mark on that year's examination corresponding to each of the course achievement standards. For example, the standards-setting process determines the minimum examination mark a student must achieve to be assigned to Band 6. This raw examination mark will not necessarily be the same in different years because examinations differ slightly in difficulty from year to year.

The standards-setting process is conducted for each examination by a group of judges drawn from the examination markers for that course. They make judgements about how students at the borderlines between the course achievement standards are likely to perform on each of the examination questions.

The process

The standards-setting process in each course is undertaken by a group of markers who also are appointed as judges for the standards-setting exercise. The selection of judges is made from recommendations provided by the Supervisor of Marking based on applicants' experience in teaching and marking the course, and on their experience in making judgements against standards.

The outcome of the judging process is a set of five cut-off marks. These are the minimum marks students must achieve on a particular examination to be assigned to Bands 2, 3, 4, 5 and 6 in the course. Once these cut-off marks are established, each student with an examination mark is assigned to one of the six achievement bands.

This procedure has been used since 1998 to report against standards in the Year 10 School Certificate tests and was pilot tested with a number of HSC courses in 1999 and 2000.

Making judgements

The judging process involves three stages, giving judges several opportunities to review and refine their earlier decisions.

To inform their decisions, the judges review statistical data showing how students at varying levels of attainment in the course typically performed on each examination question. They also review samples of students' responses to each question.

Judges are trained for this task and are given a copy of the course band descriptions, a copy of the examination paper and specially prepared recording sheets.

Stage 1

Working independently, the judges study the band descriptions and develop an 'image' of the kinds of knowledge and skills characterising students in each band. Having done this, they then develop an image of students at the borderlines between bands.

Still working independently, each judge considers the examination questions/tasks one at a time and judges and records the mark that they believe a student at the borderline between Bands 5 and

6 is likely to receive on that question/task. Each judge's marks are then added across all examination questions to give a total examination mark corresponding to the borderline (or cut-off) between Band 5 and Band 6 for that judge.

The cut-off marks between Bands 5 and 6 determined by all the judges in the team are then averaged to provide the first estimate of the minimum examination mark required by a student to be assigned to Band 6.

The judges follow these same procedures to determine the cut-offs for the other bands.

Stage 2

The judges then meet to discuss the individual judgements they have made. To assist their discussion they are given a statistical report showing how students at different levels of attainment in the course performed on each question in the examination. The judges work through and discuss this information. These statistical reports are referred to as 'item-student scales'. Judges also receive a report showing how students responded to the various alternatives of each multiple-choice question. During this process each judge has the opportunity to modify any decisions they made during Stage 1. In this way the team starts to develop a shared image of students at the borderlines between bands.

The judges' recording sheets are again collected and processed as in Stage 1, resulting in a new set of band cut-off marks.

Stage 3

A sample of student examination scripts with total marks equal to or near each cut-off mark is then retrieved and judges review and discuss these scripts.

Where a student at the borderline has been awarded the mark on a question that judges expected of students at that borderline, the student's response is shown to judges who are asked to confirm that it is typical of what they would expect of students at that level. During this process judges have the opportunity to further refine their band cut-off marks.

At the completion of this process, judges recommend to the HSC Consultative Committee a set of examination marks corresponding to the cut-offs between bands.

Role of consultative committee

The HSC Consultative Committee is appointed by the Board of Studies to review the distribution of marks in each course. The committee, which has operated since the early 1980s, is an expert technical committee made up of leading authorities in educational measurement in NSW.

The role of the Consultative Committee in the new Higher School Certificate is to ensure the integrity of the standards-setting process and the integrity of the final decisions. The Committee seeks advice from judges and, where necessary, makes adjustments within the natural variability of judges' decisions. On behalf of the Board, the Consultative Committee also determines appropriate action if the standards-setting process has not been applied appropriately or if there are other anomalies or problems in the process.

Following the standards-setting process, the Consultative Committee meets with the judges for each course, the chair of the Examination Committee and the Supervisor of Marking. The role of the Committee is to:

- ◆ ensure through discussion with the judges that the standards-setting process followed by each team of judges was in accordance with the Board's requirements;
- ◆ identify any issues that may have impacted on the effectiveness of the standards-setting process;
- ◆ receive advice from the subject representatives as to what adjustments, if any, should be made to the distributions of marks for optional questions; and
- ◆ receive from the judges their recommended band cut-off marks.

The Consultative Committee then:

- ◆ determines whether the recommendations submitted by each team of judges are appropriate or whether minor amendments are necessary; and
- ◆ approves the final band cut-off marks for each course.

Once the cut-off marks are approved by the HSC Consultative Committee, the cut-off examination mark between Band 5 and Band 6 is assigned a value of 90. The cut-off mark between Band 4 and Band 5 is assigned a value of 80; between Band 3 and Band 4, a value of 70; and so on. Examination marks between the cut-off marks are adjusted in a linear manner.

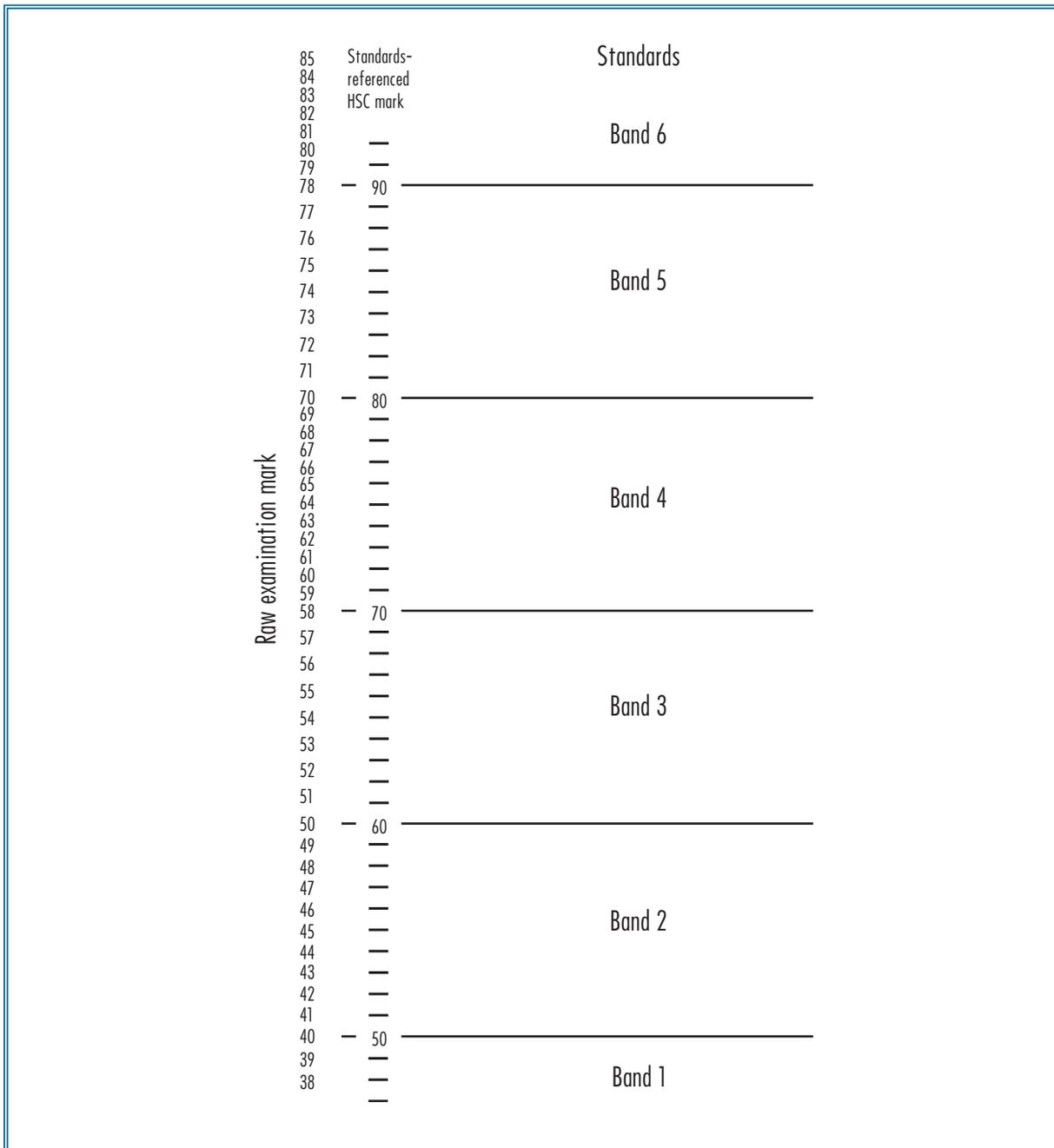


Figure 5.1 Illustration of the result of the standards-setting process (hypothetical HSC course).

Standards setting

Office and/or Board activity

Judges' activity

Nomination, selection and appointment of judges

Training of judges prior to the examinations

Briefing of judges prior to commencement of standards-setting

Stage 1:

Each judge independently records the mark, for each examination question, that corresponds to a performance on the borderline of each pair of adjacent bands. The sum of these question marks provides each judge's recommended cut-off marks.

The cut-off marks proposed by each judge are averaged. These average marks are the first estimate of the examination marks that represent the borderline between each pair of bands.

Stage 2:

Judges meet to review and discuss the decisions they arrived at individually, and consider special statistical reports. Each judge may modify any of the decisions recorded during Stage 1.

The new cut-off marks proposed by each judge are averaged. These averages are the new set of examination marks that represent the borderline between each pair of bands.

A range of student responses that correspond to each borderline mark is selected, together with works above and below the proposed cut-off mark.

Stage 3:

Judges consider the samples of student work and have the opportunity to further refine their band cut-off marks.

The new cut-off marks proposed by each judge are averaged. These averages are the new set of examination marks that represent the borderline between each pair of bands.

Judge representatives meet with the Consultative Committee. The final cut-off marks are recommended.

The consultative committee approves the final cut-off marks.

The Board's computer aligns the raw examination marks to the performance scale by mapping the cut-off marks to the examination marks (5/6 mark → 90, 4/5 mark → 80 and so on).

Standards package:

Judges select samples of student work that best represent borderline performances. A CD-ROM is produced for each course and includes the examination paper, band descriptions, marking guidelines and 3 samples of student work at each band cut-off.

Evaluation and feedback to planning for following year

Figure 5.2 Steps in the Board's standards-setting process

Finalising standards

The band descriptions, the examination paper, samples of student responses at each borderline between bands, and selected statistical information are collected and incorporated into a 'standards package' that is made available to schools on CD-ROM. The aim of the standards package is to assist teachers and students to develop a clear understanding of the standards for each course.

The standards packages also play an important role in future standards-setting exercises, assisting judges to form a mental image of students at the borderlines between bands and hence providing comparability of standards from year to year.

Issues

Most comments about the standards-setting process focused on the outcomes of the process. The most commonly raised issues related to the different percentages of students achieving the highest bands (5 and 6) in different courses, the concentration of results in a limited range of bands in some courses, and the relatively low percentage of students achieving the highest bands in English.

Differences across courses

More than 60 per cent of submissions referred to differences in the percentages of students achieving the highest bands in different courses. Concerns most frequently focused on differences in the percentages achieving Band 6, with frequent references to English (Advanced and Standard), courses in science, Studies of Religion, Legal Studies, courses in computing, Engineering Studies, VET curriculum frameworks, PDHPE and Textiles and Design.

Submissions often compared the percentage in Band 6 in a course with the percentage in other courses that were believed to have similar or even less able candidatures. In support of this concern, references were made to historical patterns of achievement in HSC courses, the similarity of syllabus content and of band descriptions in different courses, and the relative abilities of candidatures as indicated by the UAI scaling process. Submissions from schools often compared results across courses within a school to illustrate apparent discrepancies.

There was concern that differences in the percentages of students achieving the highest bands could have a range of consequences, including:

- ◆ placing students who had taken courses with low percentages in Band 6 at a disadvantage in competitive situations such as employment-seeking;
- ◆ causing student, teacher and parent disappointment and concern at apparently poor results;
- ◆ lowering the relative standing of some courses;
- ◆ causing changes in course enrolments, with students moving from courses in which it appeared to be harder to achieve in the top bands (there were claims in submissions that such changes were already occurring);
- ◆ leading to inequities in the number of students placed on First in Course and Distinguished Achievers lists, and in turn reducing the chances of students attaining awards for All-Round Excellence;
- ◆ producing distorted school 'league tables' based on the percentages of students achieving Band 6 in a school;
- ◆ affecting school, faculty and teacher accountability procedures based on HSC results; and
- ◆ leading to confusion over the possible influence of these differences on students' UAIs.

Distribution across bands

Two other features of the 2001 results attracted comment in submissions to the review: the relatively small number of students in Band 1, and the concentration of students in just a couple of middle bands.

It was noted that the great majority of students reached the minimum standard expected in courses, with generally fewer than 10 per cent of students being placed in Band 1. Some submissions saw this result as 'over-rewarding those who haven't done well' or 'artificially inflating the value of some very poor performances'. Some were concerned that, with such a high percentage of students being awarded at least Band 2, the achievements of high performers had not been adequately valued, particularly in courses where it was felt to be difficult to attain the higher bands.

Allied to concerns that there were too few students in the lowest and highest bands, a number of submissions pointed to a concentration of students in the middle bands and a perceived lack of discrimination between students. It was pointed out that in many courses, the majority of students were placed within two bands, most commonly Bands 3 and 4. In Advanced English over 85 per cent of students were in Bands 4 and 5 and in Standard English nearly 80 per cent were in Bands 2 and 3. In these two courses, more than 50 per cent of students were awarded the same band (Band 4 in Advanced; Band 3 in Standard). This 'clumping' in bands was seen as reducing discrimination by placing a wide range of students within the same achievement level.

English

Many submissions referred to the 2001 results in Standard and Advanced English.

English is a special case because it is the only subject that all students must study and that must be included in UAI calculations, and because it is the only subject in which there are two 2-unit courses (Standard and Advanced) reported on a common scale.

Standard and Advanced English were identified in submissions as courses in which the numbers of students achieving Bands 5 and 6 were particularly low. The percentages in these bands were seen as a significant problem in Standard English which had a candidature of over 36 500, but had no students achieving Band 6 and fewer than 130 achieving Band 5. Some believed that a 'ceiling' had been placed on Standard English, making it impossible for students ever to achieve Band 6. Many expressed concern that the 2001 results would compel students to take Advanced English, even where Standard English would be more appropriate to their needs and abilities or for reasons of workload.

A related issue was a belief that there had been inconsistencies in the results obtained in different English courses. The most frequent concern related to English Extension courses and to the fact that some students had achieved the highest band (E4) in the Extension courses but had not achieved Band 6 in Advanced English. Submissions from schools also identified apparent discrepancies in the Standard and Advanced English results of their students. One peak body claimed that 'many schools found that their poorest student in Advanced English achieved higher results than their best Standard student', an outcome that schools saw as 'inexplicable' in the light of previous school results and teachers' knowledge of their students. It was also noted that the percentages of students in the upper bands were higher in English as a Second Language (ESL) than in Standard English. There was a belief that some students obtained better results in ESL than they would have had they attempted Standard English.

Submissions also contrasted results in English with results in mathematics. The most common observations related to:

- ◆ the fact that English 2-unit courses were reported on a common scale but mathematics 2-unit courses were not;
- ◆ the greater percentage of Band 5 and 6 results in General Mathematics than in Standard English;
- ◆ the requirement for English Extension 2 students to undertake the English Advanced examination papers, while Mathematics Extension 2 students did not sit the 2-unit mathematics paper; and
- ◆ differences in mark ranges used in reporting results for Extension 2 students.

Extension courses

A number of submissions raised questions about standards and reporting for Extension courses. Some questioned the relationship between the achievement bands for Extension courses and for 2-unit courses and wondered whether these different sets of bands were intended to be part of a single hierarchy. Others questioned whether the numbering system for the four Extension bands (E1, E2, E3, E4) was creating unnecessary confusion. At least one submission suggested that these bands be renumbered from 3 to 6.

Band descriptions

The fact that relatively low percentages of students were placed in Band 6 in some courses often was interpreted as evidence that the standard for Band 6 in those courses had been 'set too high'. Some submissions concluded that the band descriptions needed to be revised in particular courses. There was also a view that the bands in some courses were difficult to interpret and use, and that the development of some band descriptions had been rushed or based on inadequate material. A number of submissions expected that the Board would be revising band descriptions in light of the 2001 results.

Some submissions raised questions about the relationship between band descriptions and school assessments. There were concerns about whether a single set of bands could adequately describe achievements in courses that use a range of assessment methods including paper-and-pen examinations, student performances, practical tasks and projects. It was also argued that, in courses such as Software Design and Development, it was not possible in a 3-hour examination to demonstrate achievement in relation to all outcomes included in the current band descriptions.

Judging process

A relatively small number of submissions raised questions about the judging process itself. Concerns were most frequently related to specific aspects of the process, such as the selection, training and monitoring of judges or the number of judges and their qualifications. Other issues included the reliability of judgements between judges and over time, potential difficulties encountered as a result of judges also being markers (such as adoption by judges of normative frames of reference based on patterns evident in marking) and the perception that judging may have commenced too early in marking, when marking schemes and marker accuracy were seen as less well-developed.

A number of submissions made positive and supportive comments, particularly in the few submissions from people directly involved as judges. Some suggested that the experience of the judging exercise, and the understandings developed through that process, could provide a useful basis for professional development activities with teachers.

Some submissions suggested changes that could be made to the judging process. The most common suggestion was for the inclusion of normative data to assist judges to monitor and moderate the consequences of their judgements. There were also suggestions for a return to a norm-referenced approach to defining bands, with the same percentage of students being assigned to a particular band in all courses, and a description of that band being developed for each course based on those students' responses.

Comment

Achievement standards

A key feature of the new Higher School Certificate is the set of 'standards' that describe increasing levels of achievement within each course. Each course syllabus identifies intended learning outcomes (knowledge, skills and understandings). The hierarchy of standards for each course provides descriptions of the knowledge, skills and understandings typically demonstrated by students at different levels of achievement in the course.

For 2-unit HSC courses, six described standards ('bands') have been defined. The boundaries between these six standards have been assigned the values 50, 60, 70, 80 and 90.

The development of described standards of achievement, and the interpretation of students' performances in terms of such standards (known in the literature as 'standards-referencing') is now routine practice in many large-scale tests and examinations. International testing programs such as the Third International Mathematics and Science Study (TIMSS)⁷ and the Programme for International Student Assessment (PISA)⁸ use standards-referenced interpretations of students' test performances, as do large national educational surveys such as the US National Assessment of Educational Progress (NAEP) and the Australian National School English Literacy Survey (NSELS)⁹.

Standards referencing allows students' performances to be interpreted in terms of what they know, understand and can do, rather than being interpreted only in terms of the performances of other students ('norm-' or 'cohort-referencing'). The decision to interpret students' performances in the Higher School Certificate in terms of standards is consistent with international best practice in educational measurement. This decision does not mean that a student's HSC course performances cannot be compared with, and interpreted in terms of, the performances of other students, but it does mean that each student's performances also can be interpreted in terms of explicit standards of achievement.

Judging

In general, standards of achievement in an area of learning are intended to describe increasing attainment in that area of learning and to be independent of any particular test, examination or other method of assessment. Standards are generalised descriptions of levels of achievement. Particular tests, examinations and other assessments are conducted to infer students' levels of achievement (ie, the standards they have reached) from their performances. This distinction between 'standards' intended as generalised levels of achievement in a course and observed 'performances' on particular assessment tasks is an important distinction in standards-referenced assessment.

A fundamental task in standards referencing is to interpret performances on particular assessment tasks in terms of a general framework of achievement standards: for example, to infer the standard of achievement indicated by a student's portfolio of artwork, a Drama performance, or a raw mark of 65 on a German language examination. The construction of frameworks of described standards and the mapping of performances on particular assessment tasks on to these frameworks are central activities in modern educational measurement.

The Board's judging process is designed to map performances on each year's examination on to general standards of achievement for each course. This process also is referred to as 'standards-setting', although because the process is one of interpreting examination performances in terms of standards that have already been set, it may be less confusing to refer to this process simply as 'judging'.

The judging process used by the Board is a variant of internationally used procedures developed and refined over the past thirty years¹⁰. These procedures often are applied to determine a single cut-off mark on a test or examination: for example, the minimum examination mark required to graduate or to be allowed to practice in a profession, or the test score corresponding to a minimally acceptable standard of literacy or numeracy achievement.

⁷ Mullis IVS, Martin MO, Beaton AE, Gonzalez EJ, Kelly D & Smith TA (1998). *Mathematics and Science Achievement in the Final Year of Secondary School*. Boston: International Association for the Evaluation of Educational Achievement.

⁸ OECD (2001). *Knowledge and Skills for Life: First Results from PISA 2000*. Paris: Organisation for Economic Co-operation and Development.

⁹ Masters, GN & Forster, M (1997). *Mapping Literacy Achievement: Results of the 1996 National School English Literacy Survey*. Canberra: Department of Employment, Education, Training and Youth Affairs.

¹⁰ The method used by the Board is a form of the widely-used Angoff method (see Angoff, W (1971) Scales, norms and equivalent scores. In RL Thorndike (Ed) *Educational Measurement*. Washington DC: American Council on Education).

In the HSC context, the judging process considers an examination in detail and estimates the mark on that examination corresponding to each of the boundaries between standards. That raw mark is then given the relevant boundary value (50, 60, 70, 80 or 90). This approach to assigning marks to HSC examination performances is very different from the past practice of assigning marks so that, in each course, the median mark was 60 and approximately the same percentage of students achieved the same marks (eg, between 1% and 2% of students were assigned marks above 90 in each course).

Percentages achieving standards

An obvious consequence of this changed basis for assigning marks is that there is no longer a reason for the percentage of students achieving particular marks to be the same from course to course. For example, the percentage of students scoring 90 or better in a course depends on the number of students who demonstrate achievements of the kind described by the highest standard (Band 6) for the course.

Differences across courses in the percentages of students achieving Band 6 are not in themselves a problem. In fact, in a standards-referenced system, these differences are expected because described achievement standards are developed independently for each course and reflect what subject matter experts consider to be desirable standards of achievement in that course.

While this feature of the new Higher School Certificate was reasonably well understood in submissions to the review, there was surprise at the magnitude of differences across courses. The reasons for these differences were the subject of extensive discussion and speculation:

- ◆ Some submissions believed that differences in the candidatures might have been part of the explanation. For example, it was suggested that it was not surprising, given the nature of the candidatures, that relatively high percentages of students achieved Band 6 in French and German.
- ◆ The standards themselves were sometimes seen a possible source of variability. It was suggested that, in some courses, the standards for the upper bands had been set too high. Some submissions linked this to the belief that the band descriptions for some courses included outcomes that could not be demonstrated in a written examination.
- ◆ Others looked to syllabuses and teaching for the explanation. It was claimed that, in some courses, teachers were better able to prepare students and had better resources and more accurate expectations of what was required.
- ◆ The examinations also were seen as a possible explanation. Some submissions argued that papers differed in the extent to which they allowed students to demonstrate performance at the upper levels; others, that the marking criteria and their implementation made it less likely that students would achieve high levels in some courses.
- ◆ Other submissions pointed to judges' interpretations of the standards and to the harshness or leniency of their expectations of student performance as possible influences on examination cut-off marks. And still others wondered about the intervention of the Consultative Committee and its role in changing cut-off marks between bands.

In practice, each of these suggested influences could have had some influence on the percentage of students achieving the upper bands in a course.

Course candidatures

The suggestion that differences in the percentages of students achieving Bands 5 and 6 might be related in some way to differences in course candidatures can be investigated only by accepting the notion that it is possible to compare candidatures across courses. There is a question about the validity of this assumption: given the very different kinds of content and learning outcomes in different HSC courses, it probably is not possible to make meaningful comparisons of course 'candidatures'. Nevertheless, this assumption is made each year in the scaling of HSC results by universities for the construction of the Universities Admission Index, and so – despite the

questionable validity of these comparisons – the results of the UAI scaling can be used to explore the ‘candidature’ explanation for between-course differences.

Figure 5.2 shows the cut-off points between bands for a selection of eight HSC courses after between-subject scaling¹¹. The eight courses are ordered by average scaled mark. The course with the ‘most able’ candidature in this picture is English Advanced.

The dotted line shows how the scaled cut-off mark between Bands 4 and 5 varies across courses. For the eight courses shown here, the lowest scaled mark required for Band 5 was in Drama; the highest, was in Studies of Religion II. It is not suggested that all courses should require the same scaled mark to achieve Band 5, but if they did, the dotted line would be horizontal.

An inspection of Figure 5.2 (and of similar plots for other HSC courses) shows that variations in the percentages of students achieving the upper bands are not well explained by variations in course candidatures. In fact, there is some tendency for courses with more able candidatures to set higher standards for each band than courses with less able candidatures.

Table 5.1 shows HSC courses with relatively high percentages of students in Bands 5 and 6, and courses with relatively low percentages of students in Bands 5 and 6, given their average scaled mark. If the Board wished to look more closely at courses with unusually low or unusually high percentages of students in Bands 5 and 6, then the courses at the top of these lists would be possible starting points.

High	Low
Arabic Continuers	Studies of Religion II
Music (Extension)	Chemistry
Modern Greek Continuers	Physics
History (Extension)	Studies of Religion I
Spanish Continuers	Engineering Studies
English (Extension 2)	Software Design & Dev
English (Extension 1)	Legal Studies
Dance	English (Advanced)
Food Technology	

Table 5.1 Courses with high/low percentages in Bands 5 and 6 given average scaled mark

¹¹ The between-subject scaling used here was the scaling conducted by the universities for the construction of UAIs in 2001. The main purpose of Figure 5.2 is to show the mark ranges corresponding to the bands. Some attempt also has been made to show the percentages in bands, but it should be noted that the percentage in each band is indicated by the *width* of each rectangle, not by its area. For example, in Drama, the percentages of students in Bands 4 and 5 were almost identical, with Band 5 extending over a wider range of scaled marks.

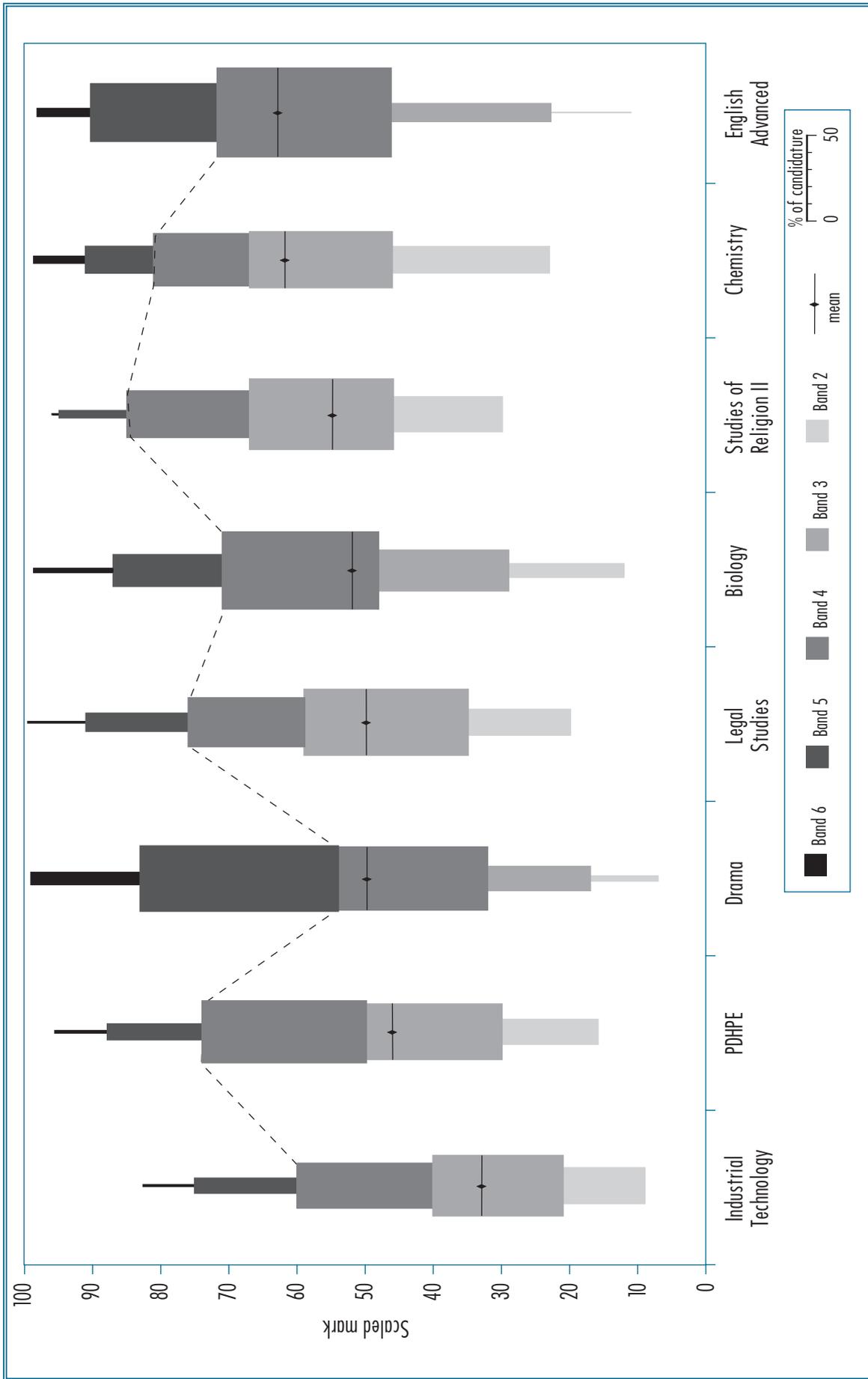


Figure 5.3 Bands for some selected HSC courses (after between-course scaling).

Band descriptions

Explanations offered for differences in the percentages of students achieving Bands 5 and 6 included the suggestion that some courses required higher levels of knowledge and skill than others.

In an attempt to ensure a degree of consistency, at least in the language used to describe achievement standards, the Board made comparisons of the kinds of skills and understandings described by the same band in different courses. Although not designed to provide a constant definition of a band across courses, this exercise should have ensured some consistency in the levels of knowledge and skill required by Bands 5 and 6 in different courses.

An exercise that the Board might find useful would be to identify some courses in which the percentages of students in the upper bands were significantly different, and to compare both the band descriptions and the work of students achieving those bands. For example, Modern History and Studies of Religion had almost identical candidatures (mean scaled scores) in 2001, but very different percentages of students in Bands 5 and 6. It may be useful to compare the band descriptions for these two courses and the examination responses of students achieving Bands 5 and 6 in these two courses. The relevant questions would be: To what extent do the band descriptions identify the same kinds of achievement? To what extent do the examination responses of students achieving the same band appear to represent the same general level of achievement?

Syllabuses and teaching

It is possible that some differences in the percentages of students achieving Bands 5 and 6 in 2001 were due to differences in teachers' and students' understandings of new syllabuses. Greater familiarity with course content and objectives, increased experience with new assessment procedures, and increasingly rich resources may result in increased levels of achievement in some courses and thus reduced differences between courses. It was suggested, for example, that while the percentage of students achieving Band 6 in some particular courses was low, this was not an indication that the standards expected of students were too high, but that the newness of the courses meant that students were not yet achieving high standards.

The release of the Board's standards packages should assist schools in their understanding of the standards expected of students in examinations.

Examinations

There were some suggestions that the percentages of students achieving Bands 5 and 6 may have been relatively low in some courses because examination papers and the ways in which papers were marked made it difficult for students to demonstrate high levels of achievement and to gain high marks. It has been difficult to evaluate general claims of this kind. Nevertheless, this possibility may warrant closer exploration by the Board. Presumably, efforts to ensure adequate syllabus coverage, clear language and expectations of examination questions, closer integration of examination development and marking procedures, and thorough pilot marking would all assist in this regard.

The judging process

A final suggestion was that the judging process also may have contributed to differences in the percentages of students achieving the upper bands in different courses. As noted above, the Board of Studies went to considerable lengths to ensure that judges were trained in the same way and applied the same procedures in setting standards in different courses. Nevertheless, the nature of the judging process introduces the possibility of differences of interpretation across courses. Differences in judges' interpretations of standards could result in different expectations of students' performances in different courses. This possibility could be explored by examining students' responses to examination questions in similar courses to see whether judges were systematically harsher or more lenient in their interpretation of standards in one course than the other.

Summary

The different percentages of students achieving the highest standards (Bands 5 and 6) in different HSC courses were clearly a concern in many of the submissions to the review. There is no doubt that there were courses with atypically high, and others with atypically low, percentages in the highest bands. It is also clear that there are many possible explanations for these differences. In a standards-referenced assessment system, there is no reason why these percentages should be the same or even similar in different courses. However, given the frequency with which this matter was raised, the Board may wish to investigate some of the possibilities summarised above, particularly in relation to courses at and near the top of Table 5.1.

Common scales within subjects

Another general issue raised by submissions related to the reporting of courses within the same subject on a common scale. In 2001, only two courses were reported on the same scale, Standard and Advanced English, but these courses were the subject of considerable discussion and raised some general questions about how best to report results in different courses within the same subject.

The decision to develop a single set of standards against which performances in both Standard and Advanced English could be assessed and reported was an attempt to recognise that students taking these two courses are studying the same subject at different levels. It was always anticipated that most Standard English students would perform at levels below most Advanced English students, but that there would be some Standard students performing at high levels and some Advanced students performing at low levels.

Although the described standards in English were conceptualised as levels of increasing achievement in the subject English, some submissions clearly regarded these standards as course grades. They saw Band 6 as the highest available grade in Standard English and so argued that it should be awarded to a significant percentage of students taking that course. From this perspective, the low percentage of Standard students in Bands 5 and 6 was an indication that the higher grades in the course had been 'under-awarded', not that very few Standard students performed at the highest levels of achievement in the subject English.

On the basis of the 2001 results, some submissions concluded that no student could ever achieve Band 6 in Standard English and expressed concern that this had not been made clear to teachers and students. In fact, there is no reason why a Standard English student could not achieve Band 6, although the content of the English syllabuses makes a Band 6 performance much more likely for students in the Advanced course.

Submissions to the review raised several matters that warrant consideration by the Board:

- ◆ It was reported that some schools were now encouraging Year 11 students to take Advanced rather than Standard English. This outcome may be desirable in some cases, but it would be unfortunate if students for whom Standard was the more appropriate course undertook the less appropriate Advanced course because it was believed to increase their chances of higher marks.
- ◆ If this is the advice being provided by some teachers, then there could be a risk that Standard English will become unavailable to students in some schools.
- ◆ In 2001, no student enrolled in Standard English was awarded a First in Course because no student scored above 90. This raises the question of whether students who enrol in Standard English and perform well should have those achievements recognised, even if they have not performed as well in the subject as some students in the Advanced course.

It seems desirable for the Board to continue its efforts to explain that the achievement standards developed for Standard and Advanced English are to be understood as standards for the subject English, rather than as grades for each course separately. This distinction appears to have been an

important distinction in the McGaw Review's recommendations for the reporting of courses within subjects against common scales.

However, attention also should be paid to ways of addressing unintended consequences of the current arrangements for Standard and Advanced English with a view to maintaining the assessment and reporting of achievements in these two courses against a common set of standards.

Constructing a common scale

One way of placing results in Standard and Advanced English on the same scale is to conduct standards-setting exercises to determine the mark on each examination corresponding to the boundary between Bands 5 and 6; the boundary between Bands 4 and 5; and so on.

A second way of establishing mark equivalences between Standard and Advanced English is to use students' performances on the common English paper. This process establishes statistically the mark on one paper representing the same level of English achievement as a particular mark on the other.

Both these methods were implemented by the Board in 2001. The first was used to convert raw marks in Standard and Advanced English to the same HSC marks scale. The second was used to establish relativities between the two papers for UAI purposes.

Figure 5.4 shows scaled marks in the five English courses. In this picture, the relative locations of Standard and Advanced English are based on students' performances on the common paper. The relative locations of all other courses are based on between-course scaling. This figure shows that, when students' performances on the common English paper are used to place marks from Standard and Advanced English on the same scale, the cut-off marks between bands on this common scale are lower for Advanced English than for Standard English. In other words, the two methods of placing these two courses on a common scale produced somewhat different results and an apparent anomaly in 2001.

It is suggested that the Board undertake further exploration of these two methods. Ideally, if courses within a subject are to be placed on the same scale, this should be done by the Board using only one method, thereby avoiding potential anomalies (see Technical Note, page 56).

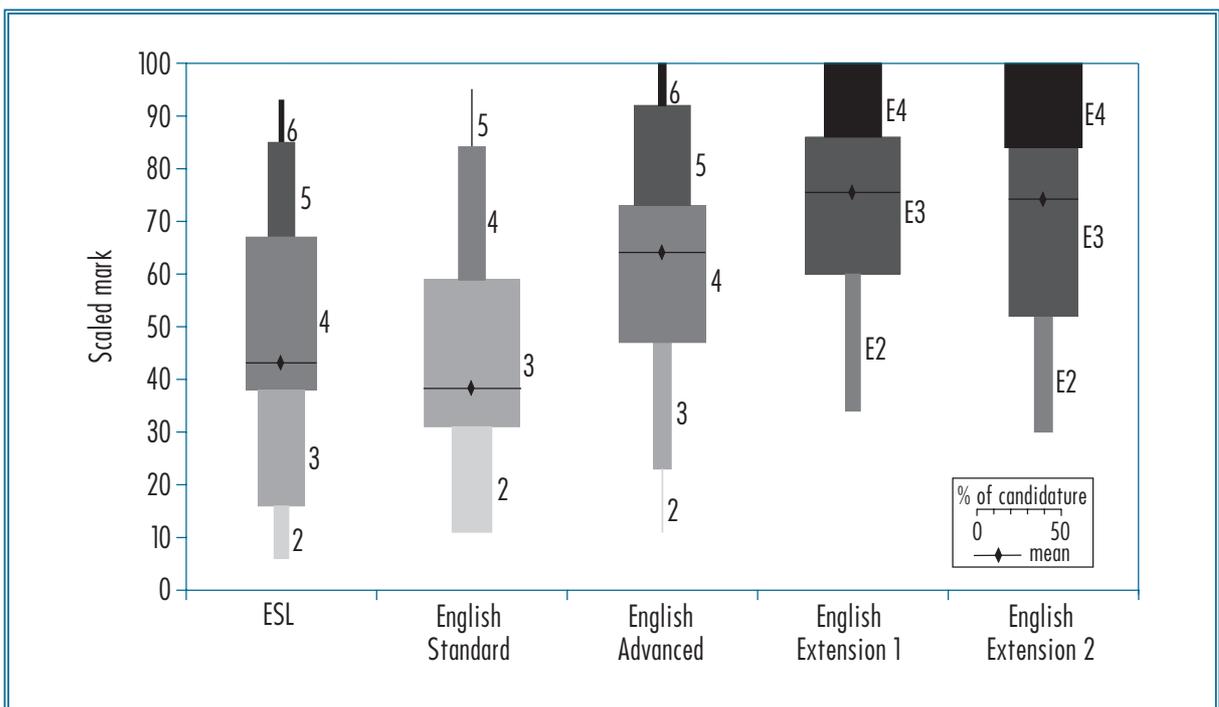


Figure 5.4 Performance bands for English courses (after between-course scaling).

Technical note: common scale

One method that the Board may wish to explore for placing Standard and Advanced English on the same scale is outlined below. This method appears to have the advantages of minimising potential anomalies, providing reliable equivalences between the Standard and Advanced papers, and setting standards once for the subject English rather than for each course separately.

Steps:

- ◆ jointly calibrate questions on the Standard and Advanced papers using the common paper as the link (if the Rasch measurement model is used, the difficulty estimates for all questions will be expressed in logits on the same scale);
- ◆ undertake judging exercises as in 2001;
- ◆ use the results of the judging process to calculate the cut-off mark between Bands 5 and 6 expected of somebody who attempted all examination questions (ie, sum expected marks over all Standard, Advanced and common questions), and repeat this for all boundaries between bands;
- ◆ establish the logit values equivalent to these cut-off marks;
- ◆ read off the raw scores on the Standard and Advanced examinations corresponding to these cut-points in logits.

In this way the boundaries between bands would be at the same level in Standard and Advanced English (not at different levels as in Figure 5.4).

Extension courses

Figure 5.4 also shows the performance bands for English Extension 1 and English Extension 2. It can be seen that the boundary between Bands E3 and E4 is lower than the boundary between Bands 5 and 6 in Advanced English. This is consistent with reports of students often achieving E4 in Extension English but not achieving Band 6 in Advanced: the standard required to achieve E4 appears to be lower.

The inclusion of Extension Courses in Figure 5.4 also re-introduces a general question raised in several submissions: are some Extension courses usefully conceptualised as building on to and going beyond 2-unit courses, and if so, should consideration be given to reporting these Extension courses on the same scale as the corresponding 2-unit course?

Some Extension courses are not designed simply as 'more difficult' versions of 2-unit courses, but are qualitatively different in the content they address. It seems unlikely that these courses should be reported on the same scale as the 2-unit course. In other cases, an Extension course may be intended simply to extend student learning beyond the level of the 2-unit course. In those cases, it may be useful to consider reporting both courses on the same scale. Whether this is possible or desirable would depend on the results of a detailed analysis of the content and intended outcomes of the 2-unit and Extension courses.

Reducing the emphasis on bands

In the introduction of a standards-referenced approach to the Higher School Certificate, considerable emphasis has been placed on the notion and role of 'bands'. This is perhaps not surprising given the use of 'bands' in other reporting contexts in New South Wales, including the School Certificate and Basic Skills Tests.

Despite the high profile that has been given to bands, they are far from essential to a 'standards-referenced' system of assessment. The essence of a standards-referenced system is the ability to interpret students' performances in terms of a hierarchy of knowledge, skills and understandings

typically demonstrated by students at different levels of attainment. This can be done without using the term 'bands', without attaching labels to bands, and without calculating and reporting the percentage of students in each band.

The term 'bands' was used in the McGaw Review, but that review cautioned against labelling bands and treating them as 'grades':

The bands on the scale could be given grade labels [such as A, B, C etc or High Distinction, Distinction, etc], but there are two problems with that. One is that the location of a student's performance on a standards-referenced achievement scale carries more information than any performance grade. To label the band with a grade nomenclature would invite too much focus on the band and attach too much significance to the location of the boundaries between bands. It would, for example, encourage the interpretation that students just separated by being on opposite sides of a boundary are more different than they are.

The other problem with use of grades is that it invites the strong assumption that a grade has a common meaning across subjects. There is no way of knowing that an 'A' or a 'High Distinction' in one subject has the same meaning as in another¹².

In practice, the bands introduced as part of the new Higher School Certificate have been given labels ('Band 1', 'Band 2', 'Band 3', etc.) and are being treated as grades with all the consequences anticipated by the McGaw Review. In particular,

- ◆ there is now a significant focus on bands;
- ◆ considerable significance is attached to the boundaries between bands (in the sense that students separated by one mark are being assigned to different bands and being interpreted as more different than they are); and
- ◆ there is an assumption by many that a band should have a common meaning across courses and subjects.

A further difficulty with labelling bands is that band labels often end up being attached not only to levels of achievement, but also to students.

The current labelled bands provide no information not already provided by students' HSC marks. Students with marks in the nineties are in Band 6; students with marks in the eighties are in Band 5; in the seventies, in Band 4; etc. In fact, the band labels provide less information than HSC marks, because marks indicate where students are located within bands. The numbering of bands thus introduces an unnecessary and redundant metric (1, 2, 3, 4, 5, 6) that sits alongside HSC marks (0 to 100) and the UAI (0 to 100).

It is not obvious that much would be lost, and there may be much to be gained, by not using the term 'band' at all in relation to HSC results, but referring instead to HSC mark ranges. The described achievement levels (standards) in each course would then simply become descriptions of mark ranges (eg, marks of 70 to 79) in each course, consistent with the new Higher School Certificate's intention to give 'meaning to marks'. The implications of this suggestion for the reporting of HSC results are considered in Section 6.

Maintaining the meaning of marks

The construction of a standards-referenced HSC marks scale anchored to standards (with the values 50, 60, 70, 80 and 90 on the scale defined by the boundaries between standards) introduces the possibility of making direct comparisons of HSC results from year to year. All that is required is that marks on particular examinations be mapped on to this unchanging standards-referenced scale.

¹² McGaw, (1997) *op.cit.*, pp 99-100.

This approach to constructing an HSC marks scale is very different from past practice. In the past, HSC marks scales were anchored to the group of students taking the course in each year and so moved from year to year with the group. For example, the value of 60 on the scale was always anchored to the group median. This practice of referencing scores to the cohort of students taking the course guaranteed that small but steady improvements or deteriorations in course achievement levels over time were hidden. The new system is designed to reveal trends of this kind.

In each year the Board proposes to undertake a judging exercise similar to the 2001 exercise to establish the raw examination mark corresponding to each boundary between the course standards. This process will allow examination marks in each year to be aligned with the standards-referenced marks of 50, 60, 70, 80 and 90.

The year-to-year comparability of HSC marks resulting from these judgements will depend on the consistency with which judges interpret the course standards over time. In this process, judges are asked to develop a mental 'image' of a student at each of the boundaries and to make their judgements against this image. As experience develops and examples of student work at each boundary are assembled, increased consistency should be achieved.

As a further check on the comparability of HSC marks from year to year, the Board might explore the possibility of making direct comparisons of the examinations used in different years. As an illustration, the judging process might lead to the decision to align a raw mark of 87 on one year's examination to the standards-referenced mark of 90, but to align a mark of 92 to the standards-referenced mark of 90 in the following year. These two judging exercises in effect establish that a mark of 87 on the first year's examination is equivalent to (ie, represents the same level of achievement in the course as) a mark of 92 on the second year's examination. There are more direct approaches available for establishing mark equivalences between two examinations (see Technical Note below). The Board may wish to explore the use of one of these more direct methods as part of its processes to ensure comparability of HSC marks from year to year.

Technical note: establishing equivalences

The process of establishing the mark on one examination that represents the same level of achievement as a particular mark on another examination in the same course is known as 'equating'.

There are various approaches to establishing mark equivalences between two tests, but the most common involve administering the two tests to the same group of students (eg, in another state or country) or incorporating a common set of items into both tests. These two procedures (known as 'common person' and 'common item' equating respectively) use students' test performances to establish mark equivalences statistically. Neither seems particularly feasible in the HSC context.

An alternative method that is sometimes used is to ask a group of judges to compare the difficulties of the questions on two tests. These comparisons are made by presenting judges with two questions at a time, one from each test, and asking them to judge which is more difficult. When a number of judges make a number of pairwise comparisons of this kind, it is possible to use their judgements to establish statistically the mark equivalences between two tests.

Recommendations

It is recommended that the Board:

- ◆ cease using the term 'bands' and refer instead to HSC mark ranges;
- ◆ for courses with unusually low or unusually high percentages of students in Bands 5 and 6 in 2001, investigate possible explanations (such as examination, marking and judging processes and the standards themselves) and make changes to future processes if appropriate;
- ◆ investigate methods currently used to place Standard and Advanced English results on the same scale with a view to using only one method for HSC reporting and UAI purposes from 2002;
- ◆ investigate the feasibility of using the judging process to establish one set of cut-off marks on the common scale for Standard and Advanced English;
- ◆ investigate whether there are subjects in which it might be desirable to report results in an Extension course on the same scale as the 2-unit course; and
- ◆ explore the possibility of using direct comparisons of examination questions in different years to enhance year-to-year comparability of HSC marks.

6 Reporting HSC results

In 2001 students received reports showing how they had performed in each course against a set of described standards of achievement. These new reports represent a significant change from previous HSC reporting practices. In the past, students received reports that showed only how they had performed in relation to the group of students taking each course. For example, an HSC mark of 60 indicated only that a student had performed at the median of that year's group; a score of 90, only that the student had performed in the top 1% to 2% of the group.

The intention in reporting HSC results in terms of achievement standards is that the reports will provide users of HSC results, including post-secondary education and training institutions and employers, with information about the kinds of knowledge, skills and understandings that individual students have demonstrated in the Higher School Certificate.

Each student's HSC mark is obtained by averaging their examination mark and their school assessment mark. Prior to obtaining this average, school assessments – which are intended to provide evidence about a broader range of learning than can be assessed through examinations – are statistically moderated against students' examinations marks. This process places school assessment marks on the same scale as the examination marks, allowing school assessments also to be interpreted in terms of the described course standards.

An important change under the new Higher School Certificate is that students receive, through a set of Course Reports, a description of the hierarchy of achievement standards in each course they have taken, and an indication of where their examination result, school assessment mark and overall HSC mark in the course are located within this hierarchy of standards.

The process

Once examination marks are available, the Board moderates school assessments against the examination marks, averages each student's examination and school assessment mark to provide their HSC mark in the course, and then prints and distributes a set of reports for each student.

Combining examination marks and school assessments

Schools submit students' school assessments in each course to the Board of Studies. These school assessment marks are intended to be based on assessments made throughout the course and to provide evidence about a broader range of syllabus outcomes than can be assessed through written examination. School assessments are based on an assessment program that each school is required to develop and must include mandatory components specified in Board syllabuses.

School assessment marks, when submitted by schools, are not intended to be comparable across schools, but should indicate the rank order and relative differences between students in their course achievements.

School assessment marks are statistically moderated against students' performances on the course examination. The process used in 2001 was identical to the process used in previous years, with school assessment marks being moderated against examination marks prior to the alignment of examination marks with the course achievement standards. The moderation process sets the mean of a school's assessment marks in a course equal to the mean examination mark of those students, the top assessment to their top examination mark and, where possible, the lowest assessment to the lowest examination mark for that group.

With statistical moderation complete, the school assessment marks are then aligned with the course standards in the same way as the examination marks. The result is that examination marks and school assessment marks both are expressed on a scale that has 50, 60, 70, 80 and 90 anchored to the boundaries between the standards.

Each student's examination mark and school assessment mark are then averaged to provide that student's HSC mark in the course.

Student reports

Each student receives a portfolio of credentials consisting of:

- ◆ *The HSC Testamur* – the official certificate confirming achievement of all requirements for the award.
- ◆ *The Record of Achievement* – a document that lists the courses studied and reports the marks and bands achieved.
- ◆ *Course Reports* – for every examinable 240-hour Board Developed course, students receive a Course Report showing:
 - ◆ their examination mark, school assessment mark and HSC mark;
 - ◆ the band descriptions for the course; and
 - ◆ the statewide distribution of marks in the course and the student's position in that distribution.

Students' Course Reports provide descriptions for Bands 2 to 6, but not for Band 1. The minimum standard expected in each course is represented by a score of 50. Students who score below 50 (Band 1) do not receive a description of their achievement in a course but receive a mark and recognition that they completed the course.

Issues

A number of submissions raised matters relating to the reporting of HSC results. The most frequently raised issues related to the numerical marks awarded in 2001 and to the adjustments made to school assessments to bring them into line with examination marks.

Reported HSC marks

Concerns about the marks awarded in 2001 were of two kinds: concerns about the reduced mark range in comparison with previous years, and concerns about the percentage of students scoring 50 or better.

Some submissions noted that the range of marks awarded in 2001 was significantly smaller than the range of marks awarded in the same courses in previous years. It was suggested that in many courses, the range of marks had in effect been reduced to marks between 50 and 90. Some believed the smaller range of marks in 2001 reduced the ability of the HSC to discriminate among students and made students who had significantly different levels of achievement look more similar than they were.

There were concerns that, because so few students had marks below 50 in 2001, the marks under the new HSC were 'inflated'. It was argued that students' achieved marks often were higher than the marks teachers gave and that students were used to receiving, as well as being above expectations based on previous years' results. Some noted that even students who were 'struggling' achieved reasonable HSC marks in 2001, but that these inflated marks had been illusory and had resulted in disappointment when UAI results were released.

A particular concern was the use of 50 as the minimum standard expected in the new Higher School Certificate. Some submissions argued that students who, in previous years, might have scored in the 30s achieved a mark of 50. Given the widespread interpretation of 50 as a 'pass', there was a risk under the new HSC that students would feel they could achieve a respectable result with minimal effort.

Reporting school assessments

Not surprisingly, there were related concerns about the way in which school assessment marks had been rescaled to bring them into line with examination marks. Some were concerned that their school assessments, which had been distributed over a scale of 0 to 100, had been 'compressed' by statistical moderation; others were concerned that they had been 'inflated'.

Several submissions referred to the 'hybrid system' now operating in the HSC, with the marking and reporting of examinations being based on a standards-referenced approach, but school assessments still using norm-referencing and being adjusted by statistical moderation.

Other general issues raised about school assessments were:

- ◆ concerns that, although school assessments are intended to include outcomes that cannot easily be measured by written examination, the moderation of school assessments against examinations has increasingly led schools to use assessment tasks resembling examinations;
- ◆ a questioning of the need for statistical moderation at all, particularly in a standards-referenced system where performances can be judged and reported in terms of standards;
- ◆ claims that the type of school or the nature of their candidature could bias the moderation process; and
- ◆ assertions that disparities between students' assessment and examination marks were greater in 2001 than in previous years.

Information reported

Although there was general support for reporting the standards students had achieved, the Course Reports introduced as part of the new Higher School Certificate attracted little comment. When this matter was pursued in consultations, the general view was that, while the reports were potentially useful, many employers may find the descriptions too verbose and have difficulties comparing results across courses. Some speculated that the difficulties of using descriptive interpretations of HSC marks could result in increased use of the UAI as a simple index for comparing students and schools.

Some submissions noted that changes to the results provided under the new Higher School Certificate had required schools to change the summary information they use to report HSC results to their communities. The absence of percentiles and other normative information had been an issue for some schools, and some submissions included requests for additional information to be reported. The most frequent requests were for the inclusion of percentile bands, separate marks for different components of examinations (such as practical, performance and project work) and course merit lists in rank order.

Comment

As noted earlier, the HSC marks scale is defined by assigning values of 50, 60, 70, 80 and 90 to the boundaries between the achievement standards in each course. A mark of 50 is the 'minimum standard expected' and separates minimally adequate performance in a course from less than adequate performance.

In the past, the HSC marks scale in each course was established by setting the median mark in the course to 60 and then adjusting marks so that about 25 per cent of students scored below 50 and about 1% to 2% of students scored above 90.

These two methods of defining the HSC marks scale are illustrated in Figure 6.1. The upper part of this figure shows the distribution of marks awarded in 2-unit Biology in 2001. The lower part of the figure shows the distribution of marks these 2-unit Biology students would have received if the old rules for assigning marks had been applied to their 2001 examination results¹³.

It can be seen from Figure 6.1 that, under the new procedure, most Biology students met the minimum standard expected in 2-unit Biology and so had marks of 50 or above. While the median score under the old method was set at 60, the median under the new method appears to be between 70 and 80. Figure 6.1 illustrates comments made in submissions about marks being 'elevated' and 'compressed' under the new HSC.

Importantly, these two methods of assigning HSC marks do not change the order of students. The order is unchanged from the order provided by students' raw examination marks. (Each column in the two distributions corresponds to a raw examination mark.)

One way of thinking about the difference between old and new HSC marks is that they are reported on different scales. When measuring temperature, a difference of 90 units on the Fahrenheit scale is a difference of only 50 units when measured on the Celsius scale; when measuring length, a difference of 90 units on the centimetres scale is a difference of only 35 units on the inches scale. The two distributions in Figure 6.1 are very similar in appearance to a set of temperature or length measurements made in different units.

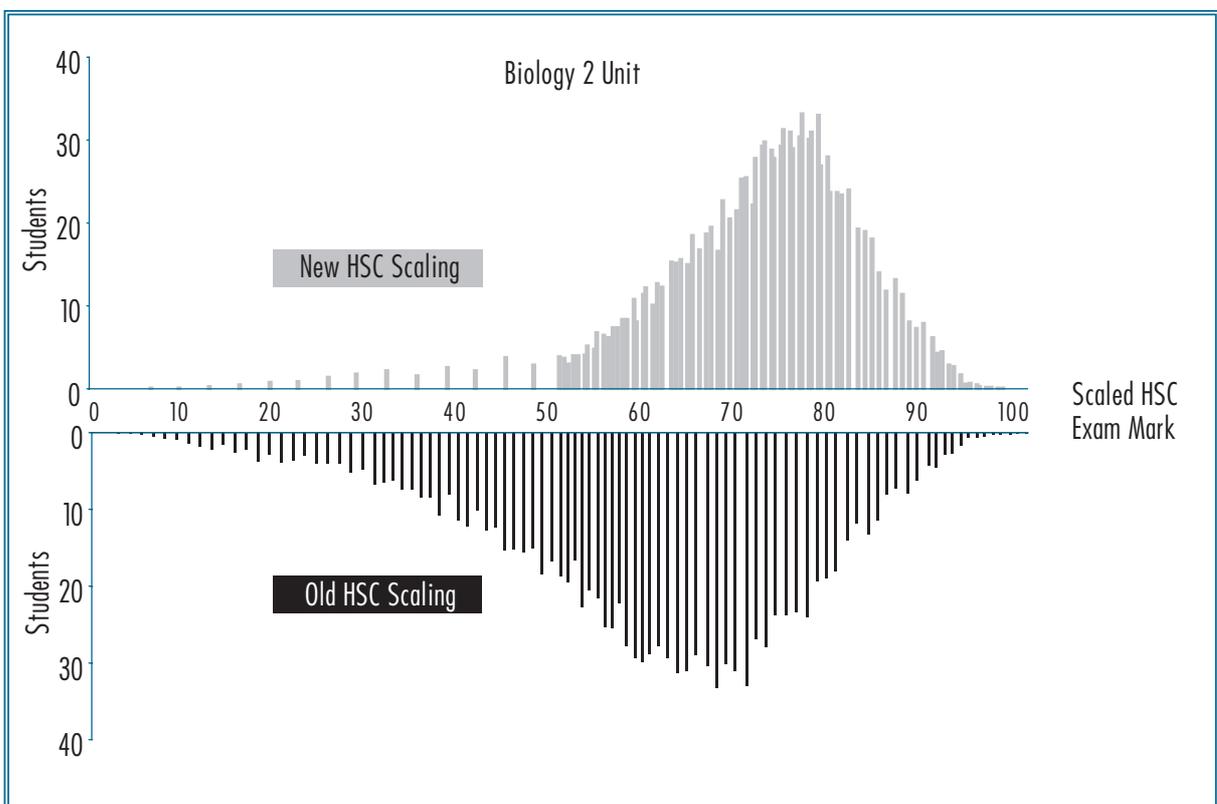


Figure 6.1 Marks for students taking 2-unit Biology in 2001 scaled by new and old methods

There is thus a sense in which new HSC marks are simply expressed on a different scale from earlier marks. The conversion of 10 miles to 16 kilometres is not usually described as an 'inflation' of distance measures; the fact that a temperature difference of 9 degrees on the Fahrenheit scale is only 5 degrees on the Celsius scale is not usually described as a 'compression' of temperature measures. To the extent that new HSC marks are merely expressed on a new and unfamiliar scale, some of the concerns raised about the 2001 marks may simple reflect transitional issues as schools and users of HSC marks become familiar with the new HSC marks scale.

¹³ It is not possible to say exactly what the lower distribution would have been. The distribution shown here has been constructed by applying the pre-2001 rules to the 2001 Biology marks.

However, given that HSC marks are reported to the nearest integer, it is relevant to ask whether the new HSC scale retains sufficient discrimination among candidates. While a number of submissions noted that students' marks were less varied than in the past, this in itself is less important than the possibility that useful distinctions between students' examination performances were lost by assigning the vast majority of students to one of about 40 marks in the range 50 to 90. No evidence was presented to the review showing that important distinctions between students were lost as a result of the new HSC marks scale. In any case, concerns to maintain fine distinctions between students' examination performances are more likely to arise in the context of the UAI than in the reporting of HSC marks. And because students' UAIs were constructed directly from raw examination marks (and school assessment marks moderated against them), the rescaling of HSC marks had no influence on the UAI.

The other concern about the new marks scale was the high percentage of students achieving the 'minimum standard expected' and the possibility that this would lead to reduced student effort.

It is common at the present time in Australia to identify minimally acceptable standards of student achievement. For example, through a national collaborative exercise involving the States, Territories and Commonwealth, minimally acceptable standards of achievement in literacy and numeracy have been established for all Year 3 and Year 5 students. Although these minimum standards ('benchmarks') are not defined in a norm-referenced manner, it is common for about ten percent of students not to achieve them.

Sitting alongside this effort to establish the minimum standards expected of students in particular areas of learning are traditional and widely held notions of 'passing' and 'failing'. In the public's mind, passing usually is associated with a score of 50, presumably originating in the notion that mastering 50 per cent of the material in a course should be good enough.

In the new HSC, these two related ideas were brought together by establishing the minimum standard expected in each course and then assigning this standard a mark of 50. The McGaw Review recommended against reporting new HSC marks on a scale of 0 to 100 that could be confused with percentages or with the UAI. That review also recommended against setting a pass/fail mark in each course. Some submissions noted the wisdom of that advice in light of the confusion resulting from attempts to compare new HSC marks with old HSC marks, and new HSC marks with the UAI in 2001.

The concern that students might set their sights on achieving only the minimum standard expected (50) seems unjustified for the vast majority of HSC students. Nevertheless, experience in other contexts suggests that, for some teachers and students, 'minimum competency' standards can focus teaching and learning on satisfying minimal requirements only. This is a matter that the Board will need to watch. And trends towards improved overall levels of achievement in the longer term may mean that it will become necessary to review and perhaps raise the minimum standard expected in some courses.

School assessments

Concerns about the 'inflation' and 'compression' of students' school assessment marks to bring them into line with examination marks in 2001 also are understandable given the transition to the new reporting scale. Teachers no doubt awarded school assessments as in the past, providing results distributed over much of the 0 to 100 scale (perhaps not very different from the lower distribution in Figure 6.1). The Board's statistical moderation process would have adjusted these school assessments, giving them a distribution more like the upper distribution in Figure 6.1.

Because 2001 was the first year in which HSC examination results were reported on the new standards-referenced scale, it was inevitable that it would be a 'hybrid' year in which school assessment marks were reported as in the past, and examination results were reported against the new scale. No student should have been disadvantaged by this difference, but some teachers clearly were surprised at the rescaling of the marks they submitted.

Some submissions claimed that, because of the different systems operating for HSC examinations and school assessments in 2001, disparities between students' school assessments and examination marks were greater than in the past. An analysis of correlations between examination marks and school assessment marks (summarised in Table 6.1) appears not to support these claims.

Year	Number of courses	Median correlation	Minimum correlation	Maximum correlation
1995	91	0.88	0.48	0.99
1996	92	0.88	0.55	0.98
1997	93	0.88	0.56	0.97
1998	91	0.90	0.66	0.98
1999	91	0.90	0.63	0.99
2000	87	0.90	0.47	0.97
2001	61	0.89	0.48	0.97

Table 6.1 Correlations between reported examination marks and moderated school assessments 1995–2001 (all courses with 100 or more candidates)

The Board of Studies has made clear its intention that schools should begin making more use of new standards-referenced reporting scales for their own assessments. The release of the Board's standards packages to schools should assist teachers in their understanding and use of these standards.

Another issue raised in submissions concerned the adequacy of current achievement standards as a frame of reference for both examinations and school assessments. There was a belief that band descriptions sometimes included skills and understandings that were difficult to assess in an examination and that were best assessed over a period of time. Others questioned whether current band descriptions would be helpful as frames of reference for assessing and reporting the broader range of outcomes addressed in school assessments. This is a matter that the Board probably should keep under review. Over time, it may be desirable to expand the descriptions of standards to include learning outcomes that can be assessed by examination and also outcomes that can be assessed only in school settings. In this way, standards would become descriptions of achievement *in the course*, and judgements would be made against whichever learning outcomes were addressed by the examination or by school assessments.

There was some interest in the possibility of using described achievement standards as a basis for comparable school assessments and as an alternative to statistical moderation. Some believed that assessments against standards could provide a more valid basis for establishing between-school comparability because examinations and school assessments are intended to address somewhat different learning outcomes.

As course achievement standards become more widely disseminated, more richly illustrated and better understood in schools, it can be expected that they will play an increasing role as frames of reference for schools' assessments. To the extent that teachers develop common understandings of standards and report students' achievements against those standards, school assessments are likely to become increasingly comparable across schools. A long-term objective could be to support teachers to make comparable assessments against standards with a view to being able to use school assessments without statistical moderation. However, any move in this direction would need to be taken cautiously and, even if consensus moderation eventually were used as an alternative to statistical moderation, some monitoring of school marks against examination marks would be desirable, with statistical moderation being used in cases of unacceptable school-examination discrepancies.

Removing references to 'bands'

It was recommended on page 59 that the Board of Studies cease using the term 'bands'. As an alternative to this terminology it is proposed that the current Bands 2 to 5 be referred to as described 'standards' in each course. These standards describe typical student achievement within particular HSC marks ranges: 50–59, 60–69, 70–79, 80–89 and 90+.

A first implication of this proposal is that bands would not be reported on students' Statements of Results. In 2001, this piece of information – on the far right of the report – was largely redundant because the Band was completely determined by the student's HSC mark.

A second implication is that the band labels ('Band 1', 'Band 2', etc) would be removed from the far left of each Course Report. The statement above the descriptions of students' levels of achievement on the Course Report could be changed to read: 'The typical performance in this mark range'. The Course Reports would otherwise be unchanged (see Figure 6.2).

Thirdly, and importantly, the Board would not report the percentages of students in bands. For the purpose of monitoring standards of achievement in a course from year to year, the percentage of students in a band (whether identified as 'Band 5' or 'marks of 80 to 89') is likely to be a less reliable indicator than the mean HSC mark in the course. For the purposes of monitoring trends over time, it is suggested that the Board use statistics describing the distribution of student marks (eg, the mean and key percentile points) rather than percentages in bands.

If the Board adopts this recommendation not to report percentages in bands, then it probably will be important to provide schools with an alternative basis for comparing and reporting the achievements of their students in each course with statewide performances. A simple way to do this would be to provide the marks in each course corresponding to key percentile points. Table 6.2 shows one way in which this might be done. The marks shown here correspond to some selected percentile points for a number of HSC courses in 2001¹⁴.

¹⁴ 2-unit courses with candidatures greater than 7000.

HIGHER SCHOOL CERTIFICATE

HIGHER SCHOOL CERTIFICATE

2002 Course Report



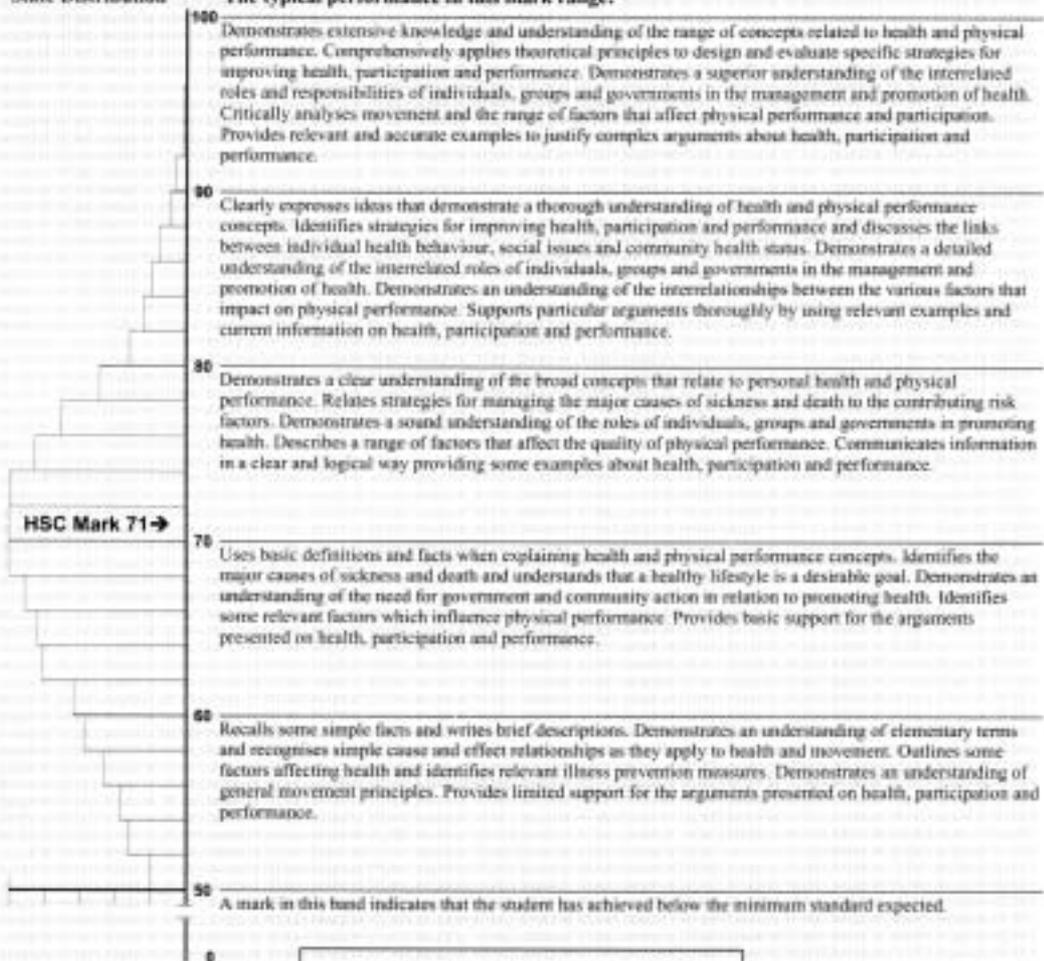
Personal Development, Health and Physical Education SAMPLE STUDENT

Examination Mark: 72

Assessment Mark: 70

State Distribution

The typical performance in this mark range:



Student Number:

Dated at Sydney on 18th January 2003
Issued by the Board of Studies without alteration or erasure.

Justin Stanley
President

Figure 6.2 HSC course report incorporating proposed changes

Course	Percentile			
	50th	75th	90th	95th
Ancient History	75	83	89	92
Biology	73	78	83	87
Business Studies	74	82	88	91
Chemistry	69	77	85	88
English Standard	63	67	71	73
English Advanced	77	82	86	89
Information Process & T	64	72	78	82
Legal Studies	67	74	80	84
General Mathematics	68	75	81	85
Mathematics	76	85	90	93
Modern History	76	83	89	91
PDHPE	69	75	80	84
Physics	70	77	84	88
Visual Arts	73	80	86	89

Table 6.2 HSC marks at key percentile points

Descriptive reporting

A central intention of standards-referencing under the new Higher School Certificate is to provide more informative reports of students' performances against the learning outcomes identified in course syllabuses. Through the provided Course Reports, HSC students in 2001 were given much better descriptions of what they had achieved than students had been given in the past. However, very few submissions made reference to this new information.

The relatively few comments made on standards-referenced Course Reports may reflect the limited opportunity there had been at the time of the review for students and others to make use of this information. Over the next few years, it may be useful for the Board to seek systematic feedback from employers and tertiary institutions on the value of the new reports and on the ways in which they are being used. It also may be useful to consider what advice the Board might develop to assist users in the interpretation and use of HSC reports.

Ultimately, the development, description and illustration of clear achievement standards in HSC courses may be of greatest value to classroom teaching and learning. Standards have the potential to provide a basis for conversations between students and teachers about the nature of increasing achievement in a course and about the progress that individual students are making.

Recommendations

It is recommended that the Board:

- ◆ remove references to 'bands' from reports (eg, Statement of Results and Course Reports) and not report percentages in bands;
- ◆ continue efforts to explain the nature of the scale on which new HSC results are reported and to show how differences between past and current HSC marks can be understood in terms of the changed scale;
- ◆ monitor the possibility that the number of students achieving low standards in a course will be increased by the decision to assign a mark of 50 to students achieving the minimum standard expected;
- ◆ continue efforts to ensure that described standards in all courses can be used with both examinations and school assessments;
- ◆ continue to consider how standards packages, professional development activities and consensus moderation exercises can be used to enhance understandings of achievement standards and the comparability of school assessment marks; and
- ◆ from 2002, provide a report of the marks in each course corresponding to some selected percentile points.

7 Using HSC results

Students' HSC results are used for a variety of purposes, including entry into employment and selection into education and training programs. Although issues concerning the use of HSC results are outside the terms of reference for this review, many of the submissions to the review raised questions about the relationship between HSC marks and the Universities Admission Index (UAI) in 2001.

Under the new Higher School Certificate, an attempt has been made to separate clearly the reporting of students' HSC achievements from the use of HSC results by universities in the ranking and selection of candidates for entry to university courses. The reporting of achievement in the Higher School Certificate is the responsibility of the Board of Studies; the construction of the UAI is part of universities' selection processes and is under university control.

The UAI, which is constructed entirely from HSC results, ranks all students applying for university entry on the basis of their overall HSC performance. It is a number between 0.00 and 100.00 in increments of 0.05. The reasons universities commonly give for constructing an overall rank order of applicants are: the need to manage a competition for university places that is seen to be fair and open; the belief that an aggregate of student achievement is less likely to impact on students' choices of HSC subjects than some other alternatives; and evidence that the best single predictor of success in tertiary study is achievement in the last year of schooling. Although the UAI is used as the basis of selection into most university courses, universities also use interviews, portfolios, principals' recommendations and special tests where appropriate.

The process

HSC results are provided by the Board of Studies to the Chair of the universities' Technical Committee on Scaling¹⁵. Students' HSC results are provided for the purpose of calculating UAIs, and arrangements are in place to protect the confidentiality of student and school results. The Chair of the Technical Committee on Scaling is able to use HSC data for research associated with the UAI.

It has been agreed that the reporting of HSC results and UAIs will be separated, with the UAI being released on a day after the release of HSC results.

There are several steps in the calculation of the UAI. The main steps are:

Scaling HSC marks

The universities scale HSC marks to adjust for the fact that HSC courses are taken by different groups of students (candidatures). The basic intention underpinning the scaling process is that students should be neither advantaged nor disadvantaged by the courses they take.

In the scaling process, the quality of the candidature for any given subject is defined in terms of their performance in the other subjects they have taken. This process in effect estimates what students' marks would have been in a subject if that subject had been taken by all HSC students rather than by the subgroup of students who chose it.

The scaling process modifies the mean, the standard deviation and the maximum mark in a course, changing the scale on which students' marks are expressed. The resulting 'scaled scores', which are not reported to students, do not change the order of merit in a course.

¹⁵ The Technical Committee on Scaling is a committee of the NSW Vice-Chancellors' Conference.

A difference introduced into the scaling process in 2001 was to relate the maximum mark in a course to the mean of the scaled marks in that course. This change was made in response to criticism that previous procedures encouraged students to take courses that were perceived as 'easy' in order to obtain high marks. In 2001 many courses had maximum scores less than 100.

Aggregating scaled marks

Each student's scaled HSC marks are then summed to produce an aggregate score for that student. This aggregate is an intermediate step in the calculation of a student's UAI and is not reported. An aggregate and UAI are calculated for a student only if the student meets certain university requirements.

To be eligible for a UAI a student must have completed at least ten units from Board Developed courses, including at least two units of English. The Board Developed courses must include at least three courses of two units or greater, and at least four subjects.

The universities have classified Board Developed courses as either Category A or Category B courses. Category A courses are considered to have academic rigour, to develop a depth of knowledge, and to contribute to knowledge assumed by tertiary courses. The cognitive and performance demands of Category B courses are considered to be less satisfactory for university selection purposes.

The aggregate is based on the student's best two units of English and their next best eight units, provided that these units include no more than two units of Category B courses.

Determining the UAI (rank)

Once aggregate scores are available for all students requesting a UAI, these scores provide a rank order of students. Each student's ranking is then expressed as a position in the entire age cohort (in other words, an estimate of the ranking as it would have been if all Year 10 students had completed Year 12 and been eligible for a UAI). This ranking is the UAI.

Students with the same UAI are located together in the rank order of students, but do not necessarily have exactly the same aggregate score. Students with a UAI of 100 in 2001, for example, had aggregate scores spread across a range of 16 marks.

Once calculated, students' UAIs are confidential and are released only to the Universities Admissions Centre (UAC) which then provides them to students who have requested a UAI and to the universities to which students have applied¹⁶.

Issues

Issues associated with the use of HSC results were almost exclusively concerned with the Universities Admission Index (UAI). Most commonly, submissions described confusion about the meaning of the UAI, the way in which it is generated and the relationship between HSC marks and UAIs in 2001.

In particular, there was concern over widespread disappointment following the release of UAIs with many students receiving UAIs significantly below those anticipated on the basis of their HSC results.

A number of submissions called for greater openness and transparency in the processes used to generate UAIs, and for better public explanations of these processes. Some requested that HSC and UAI results be released on the same day to minimise the kind of disappointment that occurred in 2001. A number called for the release of students' UAI results to schools to give schools a better

¹⁶ More information about the calculation of the Universities Admission Index (UAI) is available from the website of the Universities Admissions Centre: www.uac.edu.au

understanding of how their students had performed and to put an end to the current practice of telephoning all students in a school to obtain their UAI results.

There was concern that the UAI was gaining prominence in other settings, such as in selection for employment. Some submissions called on tertiary institutions to use a wider range of evidence for student selection, thereby reducing the critical nature of the UAI.

A number of submissions raised technical issues associated with the computation of the UAI in 2001, particularly apparent discrepancies in the scaling of marks for Advanced and Standard English and for Extension Mathematics and the procedures for determining maximum scaled marks in each course.

Comment

It seems likely that, for many students and their parents, the most significant concern surrounding the 2001 Higher School Certificate results was the disappointment and frustration that followed the release of the UAI. On the basis of their HSC marks, many students anticipated a significantly better UAI. When the UAI failed to match their expectations, they reached one of several conclusions: that there was a problem with the calculation of UAIs; that HSC marks in 2001 had been inaccurate and inflated; or that there was something amiss with the whole system.

With the benefit of hindsight, much of the disappointment over UAIs might have been avoided if the relationship between HSC marks and UAI results in 2001 had been anticipated and explained prior to the release of results. In the past, with HSC marks distributed over a wider range of the 0 to 100 marks scale, there was a closer relationship between students' HSC marks and their percentile rankings in the cohort. In 2001, HSC marks tended to be concentrated in the range 50 to 90, meaning that there was less ability to 'predict' UAIs from HSC marks.

This difference can be illustrated using the Biology marks in Figure 6.1. A student's Biology result can be reported either as their HSC mark or as their standing (percentile rank) within the Biology cohort. For example, a student with a Biology mark of 60 under the old system would have achieved a better result than 50 per cent of the cohort and so received a percentile rank of 50. A student with a Biology mark of 60 under the new system would have achieved a better result than about 15 per cent of the cohort and so received the much lower percentile rank of 15. UAIs are percentile ranks based on aggregated HSC results. The different scale used for the reporting of HSC marks in 2001 meant that HSC marks often corresponded to lower percentile ranks than in the past.

This point has also been made by the Chair of the universities' Technical Committee on Scaling:

In previous years there was some correspondence between average HSC marks and the UAI, since students who received HSC marks in the 60s (around the course average) were also in the middle of the HSC cohort (a UAI around 63). In 2001 this did not apply. To be in the middle of their HSC cohort (and obtain a UAI in the 60s) students still needed to have HSC marks around the average of their courses, but in 2001 these marks were higher, in the 70s rather than the 60s¹⁷.

It should be possible to address concerns over the unanticipated relationship between HSC marks and UAI ranks by providing better explanations of this relationship in the future. The major problem in 2001 appears to have been that teachers, students and parents attempted to attach the same meaning to new HSC marks that they had attached to old HSC marks, and so expected them to have the same relationship with UAI results. This clearly was inappropriate and was the source of the disappointment. It is desirable that the Board of Studies works with tertiary authorities to ensure that the relationship between new HSC marks and UAI results is better understood.

¹⁷ Cooney, G (2002). *The Universities Admission Index 2001*, p3.

Perhaps arising out of the disappointment experienced by some students, teachers and parents in 2001, a number of submissions expressed frustration with the UAI and the dominating role it continues to play in the reporting of HSC results. Certainly, in consultations conducted for this review, conversations often turned quickly to discussions of the UAI. For many, the UAI appears to be the measure that matters.

Although the public prominence given to a single summary index that indicates only how students have performed in relation to each other may be somewhat disappointing from the point of view of efforts to build a standards-referenced system that provides rich and detailed information about what students know, understand and can do, it is important to bear in mind that:

- ◆ a major use of HSC results is for tertiary entry, with the final years of secondary school playing an important role in preparing a large proportion of students for post-secondary study;
- ◆ there was a clear preference arising from the McGaw Review for the continued use of the HSC as the basis for university selection (in preference to university-designed alternatives that may have an even greater influence on senior secondary schools);
- ◆ the UAI is based entirely on, and is a particular way of summarising, HSC results; and
- ◆ the method used to construct the UAI provides a level of flexibility in HSC subject choice, encouraging students to take the courses best matched to their interests and abilities.

Other concerns relating specifically to the construction of the UAI – including issues relating to maximum scaled marks and the scaling of the Mathematics Extension papers – are addressed in reports of the Technical Committee on Scaling available from the Universities Admissions Centre, and are not addressed here.

Recommendation

It is recommended that the Board:

- ◆ continue to work with tertiary institutions and authorities to ensure that students, parents and teachers understand the relationship between new HSC marks and the Universities Admission Index.

8 Conclusion

Fundamental changes were made to Higher School Certificate examinations in 2001. These changes, which form part of the New South Wales Government's broader reforms of the Higher School Certificate, have implications for all stages of examination development and marking and for the reporting of student results.

Two aspects of these broader reforms have particular implications for the assessment of student achievement in the Higher School Certificate. The first is the decision to develop greater clarity and explicitness about the intended learning outcomes in each HSC course. These intended outcomes are made explicit in each course syllabus. The second is the decision to develop for each course a description of increasing achievement in relation to the syllabus outcomes. This description provides a framework of course 'standards' against which students' achievements can be assessed and reported.

This move to interpret students' Higher School Certificate achievements in terms of a hierarchy of described standards (a process known as 'standards-referencing') is in line with international best practice in educational measurement. Many large-scale testing programs now provide descriptive interpretations of students' marks. The reporting of results in terms of what individuals know, understand and can do – rather than reporting only where students stand in relation to each other – also is consistent with approaches adopted in vocational education and so has the potential to provide greater consistency between general and vocational learning.

Giving 'meaning to marks'

At the time of this review, many students, parents and teachers in New South Wales were still coming to terms with the extent and significance of the changes made in 2001. In particular, new HSC marks were only partially understood, with many trying to compare them directly with old HSC marks, and to apply old relationships between HSC marks and UAI results to new arrangements.

Under the new Higher School Certificate, a marks scale extending from 0 to 100 has been introduced to mark out and indicate positions along a continuum of described achievements in each course. Achievements in the ranges 50 to 59, 60 to 69, 70 to 79, 80 to 89 and 90 to 100 on this continuum are described in terms of the kinds of knowledge, skills and understandings typical of students at those levels of achievement.

The use of marks to indicate students' locations on a described continuum of achievement was referred to in the McGaw Review as giving 'meaning to marks'. This use of marks is very different from the previous practice of defining the HSC marks scale in terms of the group of students who happened to take a course in a particular year. Previous marks were not referenced to standards of achievement but were anchored to, and moved with, each year's group of students.

The intention to assess and report students' achievements in terms of described standards has implications for how examinations are developed. Under the new Higher School Certificate, the primary purpose of examinations is not to sort students into an order of merit, but to provide evidence about each student's level of achievement in relation to intended learning outcomes. It follows that the main requirement of examination questions and tasks, and of their marking schemes, is that they be useful for this purpose.

This intention also has implications for the reporting of results. Under the new Higher School Certificate, the primary purpose of HSC marks is not to indicate the relative achievements of students, but to indicate each student's standing on a described continuum of achievement in each course.

The procedures introduced in 2001 made significant and impressive progress towards a fundamentally different, standards-referenced approach to the assessment and reporting of student achievement in the Higher School Certificate. These procedures also raised a number of issues. Some of these issues are probably best understood as transitional issues; others are more fundamental questions about the best ways to implement a standards-referenced approach in a large public examination.

General observations

Submissions to this review raised a great deal of detail about the 2001 HSC examinations. Accompanying this detail were some general messages that should not be lost.

- ◆ There is widespread support for the reforms underlying the new Higher School Certificate.

Despite concerns over particular aspects of the 2001 HSC examinations, there is widespread support for the changes that have occurred as part of the new Higher School Certificate. The review received numerous comments to this effect. For example, one peak professional body commented that its members were 'pleased that the new HSC had brought about improvements in teaching and learning for both students and teachers' and had 'found that the new HSC had challenged teachers to be more creative and to rethink their subjects'.

There was also considerable support for the Board of Studies and its work in implementing the new Higher School Certificate. A number of positive comments were made about the effectiveness of the Board's efforts to communicate its syllabus development processes through information bulletins and explanations by Board staff at meetings and conferences. There was also appreciation of the Board's efforts to make the new examination system transparent through communications relating to marking guidelines, the Glossary of Terms and draft achievement standards for courses.

- ◆ There is widespread support for the HSC examination system.

It is also worth noting that there was strong support for the HSC examination system. Although one or two submissions questioned whether examinations were an effective way to assess student achievement in the Higher School Certificate, the vast majority of submissions were supportive of the current examination system and were focused on suggesting improvements to a system that they believed already worked well.

- ◆ There was a general expectation that further refinement would be required.

Many submissions referred to their expectation that there would need to be ongoing changes to the new HSC examination procedures. There was an appreciation that the introduction of any new system involves transitional issues as participants develop an understanding of new procedures and as those procedures are fine-tuned. There was an expectation that the new HSC examinations would require review and refinement.

There was also a view – held particularly by those who had argued for the delayed introduction of the new HSC – that if more time had been taken to introduce new procedures, then fewer issues would have arisen and there would have been less need for fine-tuning. Whether a delay would have helped is questionable; some issues are likely to have become clear only upon implementation.

- ◆ The 2001 examinations ran relatively smoothly.

Although a number of concerns about the 2001 examinations were aired in the media, these concerns focused on only a small number of papers. The evidence suggests that, overall, the 2001 examinations ran relatively smoothly.

A number of submissions to the review described the 2001 examination papers as clear and consistent and providing students with opportunities to demonstrate their knowledge, skills and understandings. An analysis of comments by teachers and students in the media immediately

following the examinations shows that they were similar to comments made in previous years, with most (but not all) examinations being seen as fair and meeting prior expectations.

At the time of releasing HSC results each year, the Board of Studies conducts a toll-free telephone inquiry service enabling students to discuss their results. Approximately 5000 calls were received following the release of the 2001 HSC results, about 2000 fewer than in recent years.

The Office of the Board also receives written inquiries from students, parents, teachers and principals relating to HSC results. Principals are able to request an explanation of student results in any course in which the school performance appears to be inexplicably below expectation. Following the release of the 2001 results, the Board received 62 appeals of this kind, 50 per cent fewer than in recent years.

In the course of the review it became clear that Board officers were already aware of many of the issues being raised in submissions to the review. In some cases, as part of their day-to-day work, they were developing solutions and strategies for addressing these issues.

- ◆ There is strong support for the move to standards-referencing.

Finally, submissions to the review expressed strong support for the decision to introduce standards-referenced assessments into the Higher School Certificate. Despite concerns over particular aspects of the new system, nobody called for a return to an earlier system. Some submissions noted that this was the system that had been recommended to the McGaw Review and reaffirmed their commitment to seeing it work. Several argued that the new examination system 'needs to be given a chance'.

One school commented in its submission: 'The new Higher School Certificate promotes good teaching practice. Quite specific outcomes for each of the bands assist teachers in planning their lessons and the system of descriptive reporting is a definite improvement. Everything about the school-based assessment procedures is transparent. The student knows exactly what they have to do in order to achieve a specific band. Standards-based assessment is a much fairer system of assessing what a student knows, understands and can do.'

Developing examinations

The task in a standards-referenced examination is to develop questions and other assessment activities capable of providing information about students' levels of achievement of the knowledge, skills and understandings identified in the relevant course syllabus. Marking guidelines and marking schemes also must be designed to provide evidence about the learning outcomes that questions/tasks are constructed to address.

This close relationship between syllabus outcomes, examination questions/tasks and marking schemes required by a standards-referenced assessment system is very different from the loose relationship found in traditional public examinations. In many traditional examinations, questions are developed with little explicit attention to the learning outcomes they address. Instead, they often are written in an open-ended way to allow students to demonstrate whatever they know about a topic (referred to in the course of this review as a 'brain dump'). How students' responses to a question are to be marked often is not considered as part of the examination development process, but is the later responsibility of a separate group of markers.

The challenge for the Board of Studies in developing standards-referenced examinations will be to maintain an appropriate balance between being clear and explicit on the one hand, and being overly specific and prescriptive on the other. There are many examples of 'outcomes-based' assessment systems that have degenerated into checklists of outcomes that are either 'achieved' or 'not achieved'. These approaches commonly fragment curricula into increasingly superficial pieces of knowledge and skill. Too frequently, the result is poor pedagogy and even poorer assessment. The Board of Studies has not gone down this path, and there is no reason to believe it will.

Nevertheless, there were concerns expressed in some submissions that the introduction of standards-referenced examinations in some HSC courses had been accompanied by an increased emphasis on the assessment of peripheral and unimportant detail. Some of these comments were linked to concerns, particular in the sciences, that there is currently too much content in syllabuses. Others expressed concern that, because the 2001 examinations addressed only a sample of syllabus content, students sometimes were left feeling that much of what they had learned had not been covered.

In developing examinations to more explicitly address intended learning outcomes, it is inevitable that only a sample of these outcomes will be addressed in any one examination. It will be important to ensure that syllabus content is sampled in a balanced and representative way. It probably also is inevitable that there will be concerns from time to time that examinations have sampled unimportant content. The general challenge will be to ensure that HSC examinations focus on underpinning and central knowledge rather than peripheral and superficial detail, and that they assess students' deep understandings of content and their ability to apply what they have learned in meaningful and worthwhile contexts.

Recommendations relating to the examination development process appear on page 30.

Marking examinations

The task in marking examinations in a standards-referenced context is to judge and record the extent to which students have demonstrated the learning outcomes that questions are designed to address. This task requires marking guidelines and marking schemes that are tightly linked to the intentions of examination questions.

This close relationship between the development of examination papers and the marking of those papers is in contrast to much traditional practice. It has not been uncommon in examinations for papers to be developed by one group, and for students' responses to be marked by another. In many traditional examinations, particularly those based largely on extended written responses, marking is driven by the attempt to produce an adequate spread of student marks (typically a normal distribution or 'bell curve') based on impressionistic judgements of the 'quality' of student responses.

The general challenge for the Board of Studies will be to continue efforts to integrate the development and marking of papers to ensure that the marking process provides the best possible information about students' performances in relation to syllabus outcomes. The inclusion of the development of marking guidelines in the examination development process is an important step in this direction.

It is clear that, in some instances, there was uncertainty in 2001 about the extent to which marking guidelines could be changed and about the flexibility and professional judgement markers were able to use in interpreting guidelines and schemes. In at least one case, marking guidelines were interpreted and applied in an overly-rigid fashion. But there is no evidence that marking guidelines were 'thrown out' as claimed in the media. Many of the concerns relating to marking in 2001 appear to have had their origins in changed procedures and responsibilities and some uncertainty about what alterations to guidelines and flexibility of interpretation were possible.

Specific challenges will be to ensure that, in all courses, experienced markers are involved in the development of guidelines, that these guidelines are adequately tested against students' responses, and that changes to guidelines, the development of more detailed marking schemes and the identification of benchmark scripts remain faithful to the learning outcomes that questions/tasks are designed to address.

Recommendations relating to the marking of HSC papers appear on page 41.

Standards setting

The task in standards setting is to interpret performances on a particular set of assessment activities in terms of a hierarchy of described standards. In other words, the task is to map the specific (performances on a particular examination or other assessment activity) on to the general (a described continuum of achievement in a course). In the HSC context, standards setting is a process of deciding how raw marks on a particular examination relate to marks on a standards-referenced HSC marks scale.

Following the release of the 2001 HSC results, there was a considerable focus on 'bands'. The percentages of students achieving Band 6 in different courses were compared and debated at length. No other issue was raised as often with this review as the variability in the percentage of students achieving Band 6.

The problem with this discussion is that it was based on the assumption that Band 6 can be compared across courses. In general, it cannot. In each course, Band 6 represents a particular level of achievement defined in terms of the learning outcomes for that course, and set by experts in the subject. Because there is no way of knowing that Band 6 in one course represents the same level of achievement as Band 6 in any other course, it is difficult to compare meaningfully the percentage of students achieving Band 6 across courses¹⁸. The single biggest issue raised with the review was based on an attempt to compare percentages that are fundamentally unable to be compared.

Interestingly, the McGaw Review anticipated this problem and warned against labelling described standards of achievement on a continuum, arguing that this would invite the strong assumption that labelled levels had a common meaning across courses and would result in too much focus on 'bands'.

The challenge confronting the Board is to take steps to discourage invalid comparisons of this kind. The most effective way to do this probably is to reduce the public focus on bands and to focus instead on the McGaw intention of giving 'meaning to marks'. The important relationship in this standards-referenced system is between marks and their substantive meaning. In this relationship, labelled 'bands' are an unhelpful distraction.

Bands also are redundant. Reporting the bands students have achieved provides no information not already contained in their HSC marks. In fact, bands provide a less precise indication than marks of individuals' levels of attainment on the achievement continuum that underlies both bands and marks.

Another challenge for the Board will be to continue to support schools in their understanding and use of standards-referenced HSC marks scales. The dissemination of standards packages will assist schools in this regard. As schools become more familiar with course standards, it is not unreasonable to expect that school assessment marks also might be provided directly on the standards-referenced HSC marks scale in each course.

Recommendations relating to the standards-referenced interpretation of HSC marks appear on page 59.

Reporting HSC results

The reporting task in a standards-referenced system is to report individual and group results in terms of a described scale of increasing achievements.

The standards-referenced reporting system developed by the Board provides a basis for interpreting students' HSC achievements against described standards as well as allowing students' achievements in a course to be interpreted in terms of the performances of other students. In other words, it allows *both* standards-referenced and cohort-referenced interpretations of achievement. Although this point is well illustrated in the Course Reports that the Board provides to each student

¹⁸ except in Standard and Advanced English which are reported on the same scale.

(see page 67), it was not always understood in consultations conducted for this review. For example, some seemed to believe that any referencing of students' HSC marks to the distribution of student results in a course (eg, a comparison with key percentile points) would be a return to the norm-referenced system of the past. This clearly is not the case.

Although the main focus of the Board's attention in relation to HSC examinations is on assessing and reporting individuals' achievements of the intended learning outcomes in the courses they have taken, schools and education systems also are users of HSC results for their own monitoring and reporting purposes.

The review was shown examples of ways in which schools had used 2001 HSC results in reporting to their communities. Some schools included in their newsletters a report of the percentage of students achieving Band 6 in each course. In some cases, these percentages were reported without accompanying information showing the percentage of students in the State achieving Band 6. In other cases, more complex reports were provided showing both school and State percentages in Band 6 in each course.

If the Board is to discourage invalid comparisons of percentages in bands across courses, then it will be important that this method of reporting also is replaced by a more appropriate method in analyses of school results prepared by the Board. A more useful method of analysing and reporting results at school level may be to compare results in a school with key percentile points in the statewide distribution: for example, to provide schools with the ability to calculate the percentage of their students achieving above the State median in a course, in the top 10% of students, or in the top 5% of students in the State.

For the purposes of system monitoring, the percentage of students achieving a particular band provides a less reliable basis for monitoring trends over time than the average HSC mark in a course. The course mean could be calculated and monitored for particular subgroups of the student population and, if more detail is sought, a system could monitor the achievements of particular groups of students (eg, top 10%; bottom 15%) measured on the standards-referenced HSC marks scale.

Recommendations relating to the reporting of HSC marks appear on page 69.

Using HSC results

A final general challenge for the Board is to continue its efforts to support the appropriate use of HSC marks.

A major reason for introducing standards-referenced assessment and reporting of HSC achievements was to provide users of reports with better information about what students know, understand and can do. The introduction of a standards-referenced marks scale in each course is a significant initiative to provide more informative reports of students' achievements.

However, in the evidence presented to this review, there was very little indication that the new Course Reports were being used by employers or tertiary institutions. This may be a reflection of the limited time there had been for these reports to be used. It also may indicate that Course Reports are providing a level of detail that employers and selection officers currently are finding difficult to use. The Board probably should monitor the use of this new level of information and explore ways of supporting users in its interpretation and use.

A further challenge will be to continue to clarify the different purposes of HSC results (ie, reporting what students have achieved) and UAI results (ie, ranking applicants for tertiary entry). In addressing this challenge, it will be important for the Board to continue to work with the Universities Admissions Centre to ensure that these different purposes and the different ways in which HSC and UAI results are reported are understood.

A recommendation relating to the reporting of HSC marks appears on page 73.

Appendix

The review met with representatives of the following organisations:

Association of Catholic School Principals (NSW & ACT)
Association of Heads of Independent Schools of Australia (NSW)
Association of Independent Schools (NSW)
Catholic Education Commission, New South Wales
Committee of Chairs of Academic Boards/Senates of Universities in NSW and the ACT
Council of Catholic School Parents
Federation of Parents and Citizens' Associations of New South Wales
NSW Department of Education and Training
NSW Parents Council
NSW Secondary Principals Council
NSW Teachers Federation
NSW/ACT Independent Education Union
Professional Teachers Council NSW
Technical Committee on Scaling

Written submissions to the review were received from:

Individuals

Ms C Abigail, Port Hacking High School
Mr C & Mrs D Agnew
Dr I Ali, Malek Fahd Islamic School Greenacre
Mrs L Allen, Armidale
Ms F Altinok
Mr J Alvaro, Croydon
A Amr, Auburn
Dr A Anderson, St Euphemia College Bankstown
Mr B Anderson, Tweed River High School
Mr U Badar, Merrylands
Ms S Ban, Vaucluse High School
Ms W Barel, Masada College St Ives
Dr T Bavaro, The Scots College Bellevue Hill
Ms R Bews, Central Coast Adventist School
Mr W Biddle, Epping Boys High School
Ms J Brassel, Cabarita
Mr M Brown, Wagga Wagga High School
Mr R Burton, Coffs Harbour Senior College
Dr M Butler, Gosford High School
Ms R Carlson, Coffs Harbour Senior College
Mr R Clarke, SHORE Sydney Church of England Grammar School
Ms M Clemson
Mr A Clucas, Covenant Christian School Frenchs Forest
Mr G Cooper, Sydney Technical High School
Ms L Craig, Lugarno
Ms S Crawford, St Paul's College Manly
Ms R Deeley
Ms C Del Gallo, North Sydney Girls High School

Mr I Denman, Canberra Grammar School
Mr J Dowle, Goulburn
Mrs N Drayton, Pacific Hills Christian School
Ms C Dunlop, Turrumurra
Mrs A Ellis, Ambarvale High School
Mr A Faulks, St Joseph's College Hunters Hill
Ms A Fell
Associate Professor T Gagen, University of Sydney
Ms J Gebels, Mosman
Mr R George, Western Institute of TAFE
A L Godden, SHORE Sydney Church of England Grammar School
Mr P Graham, Sydney Technical High School
Ms P Grocholsky, Elderslie High School
Mr P Grover
Mr G Haley, North Nowra
Mr S Henry, Cherrybrook Technology High School
Ms M Hunter, St Luke's Grammar School Dee Why
Mr R Ireland, Toronto High School
Ms H Kam, TAFE NSW Access Division
Mr T Kelly
Mr P F Kidd, The Forest High School
Ms J King, Riverside Girls High School
Mr R Kirk, Christian Community School Regents Park
Mr M Larkin, Tamworth
Mr P Lentern, Our Lady of Mercy College Parramatta
Ms K Lew, Mater Maria Catholic College Warriewood
Mr R Lidbetter, Coffs Harbour Christian Community School
Mr B Lindbeck, Kempsey
Mr D Lumley
Mr J McGrath, Marrickville
Ms J McIlwain, Killara High School
Ms R McKie, Newcastle TAFE
Ms S Melville, Erskine Park High School
Ms L Michalk
Ms D Middlemiss, Blue Mountains Christian School
Ms C Moses, Loreto Normanhurst
Mr R Murray, Carinya Christian School Tamworth
M Naito, SCECGS Redlands
Ms H Neeson, Bega High School
Emeritus Professor G Nettheim, University of NSW
Ms A O'Donnell, Georges River College
Mr G O'Donnell, Lane Cove
Ms N Parson, Central Coast Grammar School
Ms S Paton, St Francis Xavier's College Hamilton
Mr K Phillips, St Francis Xavier's College Hamilton
Mr W Powell, Whitebridge High School
Mr K Pradhan
Ms N Pradhan
Ms D Reading, Ravenswood School for Girls
Mr W Richards, The Armidale School
Mr M Riley, Sydney Technical High School
Mr J Robertson, Port Macquarie High School

Ms S Russo, Ryde
Mr A Scott, Wenona School
Mr D Sharpe, Keira Technology High School
Mr D Shephard, Kiama High School
Mr G Shrubbs, Bradfield College North Sydney
Ms H Simmelhaig, Carss Park
Mr J Simmonds, Sydney Technical High School
Mrs P Smith, Lindisfarne Anglican School Tweed Heads
Mr A Stevens, St Francis Xavier's College Hamilton
Mr A Stone, St Francis Xavier's College Hamilton
Mr N Strugnell, Heathcote High School
Ms T Tagg, Concord High School
Mr G Thickett, Peakhurst
Ms L Thickett, Burwood Girls High School
Ms G Thomas, Newtown High School of the Performing Arts
Ms R Thomson, Coffs Harbour Senior College
Mr G Thwaites, St Johns Park High School
Mr B & Mrs P Willis, Putney
Mr D Wilson, Cherrybrook Technology High School
Mr H Yu, Seaforth
Mr L Zietsch, Sydney Technical High School
Anonymous (2)

Schools

Barker College
English Faculty, Birrong Girls High School
Blue Mountains Grammar School
Business Services Learning Area, Bradfield College
Callaghan College - Jesmond Campus
Canabalos Rural Technology High School
Caroline Chisholm College Glenmore Park
Casino High School
Curriculum Committee, Central Coast Grammar School
Cranbrook School
East Hills Girls High School
Science staff, Elderslie High School
Executive staff, Erina High School
English and Mathematics senior teachers, Glenaeon Rudolf Steiner School
Green Point Christian College
Greystanes High School
English staff, Hawkesbury High School
Science teachers, James Sheehan Catholic High School Orange
Kincoppal-Rose Bay School
Japanese teachers, Kingscliff High School
Business Studies teachers, Loreto Normanhurst
History teachers, Loreto Normanhurst
Biology teachers, Mackellar Girls High School
Newtown High School of the Performing Arts
Science Faculty, North Sydney Boys High School
Modern Greek teachers, Open High School
Oxley College Bowral
Pymble Ladies College

Redfield College Dural
Rooty Hill High School
Santa Sabina College, Strathfield
Languages staff, St Catherine's School Waverley
St Clare's College, Waverley
English Faculty, St Francis Xavier's College Hamilton
HSC Review Committee, St George Girls High School
St Gregory's College Campbelltown
St Luke's Grammar School Dee Why
St Vincent's College Potts Point
Sydney Grammar School
Executive staff, Sydney Technical High School
Tempe High Languages School
Tenterfield High School
Trinity Catholic College Lismore
Westfields Sports High School

Organisations

Association of Heads of Independent Schools of Australia (NSW)
Australian Society of Music Educators
Catholic Education Commission, New South Wales
Catholic Education Office, Parramatta
Catholic Education Office, Sydney
Computing Studies Teachers' Association
English Teachers' Association (NSW)
Federation of Parents and Citizens' Associations of New South Wales
History Teachers Association of NSW
Legal Studies Association of NSW
Mathematical Association of NSW
Metropolitan East Social Sciences Teachers Association (MESSTA)
NSW Modern Greek Teachers Association
NSW Parents Council
NSW Secondary Principals Council
NSW Teachers Federation
NSW/ACT Independent Education Union
Parramatta Diocese Secondary Principals Association
NSW Personal Development, Health and Physical Education Teachers Association
People with Disabilities (NSW)
Professional Teachers Council NSW
Science Teachers' Association of NSW
VET in Schools Directorate, NSW Department of Education and Training