

Is it grade inflation or real change in student achievement?

Dr John Bennett
Office of the Board of Studies NSW

Introduction

Many education systems use standards-setting procedures based on professional judgement to assess and report student achievement. What features can be built into such standards-setting procedures to ensure that if over time more students are being awarded higher levels of achievement, this is happening because of genuine improvements in student learning? Without suitable checks and safeguards it is difficult to counter claims that an increase in the proportion of students achieving the higher standards is simply “grade inflation”.

The NSW Higher School Certificate is awarded at the end of secondary school. It is examination-based and also incorporates a significant school assessment component. Following a review of the Higher School Certificate in the late 1990s (McGaw 1997), the decision was taken to change from the reporting of student achievement using a norm-referenced approach to one based on explicit achievement standards. Since 2001 the achievements of students in each course have been reported in terms of a mark and one of six performance bands. An Angoff-based standards-setting procedure is used each year to determine the examination marks that correspond to the borderlines between the performance bands. A number of key strategies have been built into the procedure and the associated processes to ensure that significant changes in the distribution of performance bands over time are the result of real changes in student achievement.

This paper first provides a brief overview of the NSW Higher School Certificate program and the standards-setting procedure employed. It then highlights some of the strategies and checks employed to ensure the integrity of the application of the procedure. Finally, data on the implementation of the procedure over time are provided.

The NSW Higher School Certificate

The Higher School Certificate (HSC) is the credential awarded to New South Wales students at the end of their secondary education. During Year 12, the final year of school, students typically study courses in five or six subjects. Each course has a unit value, where a unit refers to an indicative 60 hours of class time in the year. Most courses are of two units in scope. In order to receive a Higher School Certificate a student must satisfactorily complete courses totalling a minimum of 10 units in Year 12. At least two units of English must be studied.

Student achievement in a course studied in Year 12 is assessed through two components – an external examination and a school-based assessment program.

The HSC examinations employ a variety of different item types, as appropriate for each course. Most examinations contain short and extended written response-type items that are scored polytomously. Many examinations also include objective items. In some courses the examinations include other substantial manifestations of student work. For example, in Visual Arts students submit artworks they have created. In a number of technology courses students submit projects they have designed and built. In the examinations for foreign languages, items that assess listening and speaking skills are employed. In Music and Drama, students are assessed on the quality of their performances of pieces of music or short plays they have prepared. Nearly all examinations involve some choice in the questions students will answer.

The school assessment marks are based on a program of assessment activities developed and administered by each school in accordance with requirements and guidelines provided by the Board of Studies. The assessment marks submitted by each school are statistically moderated by changing the mean, top mark and bottom mark to match those of the examination marks obtained by the students from that school in that course.

The marks that report a student's achievement in each course are determined through the standards-setting procedure described below. This procedure has the effect of aligning the initial examination marks and the moderated assessment marks to the reporting scale based on the achievement standards established for that course. Hence, students' results in a course consist of one of six performance bands and a mark that shows whether they were near the top, middle or bottom of that band.

The results a student receives, before they are aligned to the standards, are used in the calculation of the student's Universities Admission Index (UAI). The marks in each course are re-scaled by the universities taking account of the general academic ability of the candidature of each course. These scaled marks are then used to calculate an aggregate mark for each student, and subsequently, a rank based on that aggregate. It is on the basis of this rank that those students interested in proceeding to university are offered places in particular courses.

One of the advantages of separating the reporting of achievement in the HSC from the university selection process is that there is less focus on the proportions of students achieving each band than in a system where students need to achieve particular bands in order to be selected for a particular university program.

A further feature that also takes some of the focus off the band distributions is the reporting of marks as well as the performance band. It is clear for all to see the small difference in the standard of achievement of a student who receives a mark of 72 which places them in Band 4 and the achievement of a student who receives a mark of 69 which places them in Band 3. If only the band was reported this would tend to imply a more significant gap in achievement. The practice of reporting marks as well as performance bands also assists in the interpretation of a student's achievement.

An overview of the standards-setting procedure

Since 2001 the initial mark from the examination and the statistically moderated assessment mark from the school have been 'aligned' to a standards-based

performance scale in order to obtain the mark reported to the student. The alignment process, described below, consists of the application of a structured, multi-stage Angoff-based standards-setting procedure involving teams of highly experienced teachers, referred to as ‘judges’ (Angoff, 1971). Each year the judges determine the examination marks that correspond to the borderlines between the different levels of achievement (which are referred to as ‘performance bands’). A multilinear mapping process, which adjusts these cut-off marks to the borderline marks of 50, 60, 70, 80 and 90 used as part of the reporting scale is then applied to all the examination marks and moderated school assessment marks for a course. In this way students’ HSC results are related to the knowledge, skills and understanding they have achieved in each course. The standards-setting procedure was especially developed to suit the nature and form of the HSC examinations (Bennett, 1998). The procedure and its implementation were reviewed as part of a wider review conducted in 2001 (Masters, 2002).

Following the 2001 and 2002 examinations, the HSC Standards Packages were produced. These packages consist of a CD-ROM containing the three elements needed to properly clarify the standards. They contain:

- descriptions of the levels of achievement that are part of the performance scale
- the examination paper and its marking guidelines
- samples of the responses of students who received marks for a question the judges believed would be scored by students at the borderline of each pair of adjacent levels of achievement.

Since 2006, these packages have been also published on the Board of Studies’ website as part of the Assessment Resource Centre.

These standards packages encapsulate the standards of performance that have been created for each course. They are used by the teams of judges each year to develop a thorough understanding of the standards associated with the borderlines between each pair of adjacent performance bands. In this way, we can be confident that those charged with the responsibility for establishing the cut-off marks each year will be basing their decisions on the *same* standards of performance even though the examination paper may vary in difficulty from year to year, the marking schemes will vary from year to year and student rates of achievement may vary from year to year;.

The standards-setting procedure consists of three distinct stages that typically occur at least several days apart.

Stage 1

After a close study of the materials in the relevant standards package, each judge independently forms “mental images” of the knowledge and skills of students whose achievements would place them on the borderline between each pair of adjacent bands.

Then, for each question on this year’s examination each judge records the mark a student on the borderline between a Band 6 and Band 5 would receive. Adding up the marks for individual questions gives the total examination mark that the judge believes corresponds to the borderline or cut-off mark between Band 6 and Band 5. Averaging the cut-off marks between Band 6 and Band 5 proposed by all the judges

produces the first estimate of the examination mark that will represent the borderline between those bands. The judges follow the identical procedure for the Band 5/Band 4, Band 4/Band 3, Band 3/Band 2 and Band 2/Band 1 borderlines.

Stage 2

The judges come together to discuss the decisions they made individually. At the same time they are given specially designed statistical reports that are very effective in showing how students of different abilities have performed on each question in the examination. The judges then use this information to assist their discussions about the decisions they made during Stage 1. During this process a judge has the opportunity to modify any of the decisions he or she recorded during the first stage. It is made clear to the judges that there is no requirement or expectation that they will reach consensus on their cut-off marks.

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

Stage 3

Clerical staff collect a sample of students' examination responses which have achieved the marks for a question equal to the band cut-off marks determined by the average of the judges' decisions for that question.

The judges then meet again to review and discuss these examination responses. They are asked to confirm that the responses produced by these students are typical of what they would expect of students placed at the borderline between a pair of bands. The judges also review student works slightly above and below their proposed cut-off marks. During this process the judges have a final opportunity to further refine their individual cut-off marks for any question.

When they have completed this third stage the average of the decisions concerning each borderline made by the individual judges becomes the recommended cut-off mark for each performance band.

These marks, after review by the HSC Consultative Committee as outlined below, are then used to finalise the marks that are to be reported to students. This is done by adjusting the mark that is judged to be the borderline between Band 6 and Band 5 to 90, the mark adjudged to be the borderline between Band 5 and Band 4 to 80, and so on. Marks between these borderlines are simply adjusted using linear interpolation.

Key features of the procedure that protect its integrity

The HSC Standards Packages

The HSC Standards Packages are a critical component of the standards-setting procedure. They provide a clear and public illustration of the standards of achievement associated with each HSC course. They form a concrete basis for the procedure, from the training of the judges through to the end of the procedure. They do this by enabling the judges to reference their decisions and inform their discussions. As the standards packages are publicly available, teachers and students can also use them to gain an understanding of the quality of student work associated with each standard.

Supervising and supporting the standards-setting procedure

Before the standards-setting operation begins judges are given training in the procedure. Each team usually consists of a mix of experienced judges and others involved in the process for the first time. Each judge is trained in the application of the procedure, including a session where they focus on the standards packages and sharpen their understanding of the standards associated with each borderline between the performance bands.

Throughout the application of the procedure the judges are supported by a person referred to as a Field Officer. Field Officers are trained in the application of the procedure, but are not subject specialists and are not involved in the decisions made by the judges. The roles of the Field Officer are to:

- observe the application of the judging procedure and ensure it is applied correctly
- ensure the judges receive the materials and reports necessary to do their job
- identify any issues that arise during the operation that may impact on its outcomes
- provide a written report at the end of the operation.

Restricting access to certain statistical data and delaying the publication of band distributions

As part of the standards-setting procedure the judges are given special statistical reports that indicate how students of particular ability levels have performed in each question in the examination. These reports assist in their discussions and consideration of their initial decisions.

The judges, however, are not given any reports that provide frequency distributions or summary statistics on individual questions or on the examination as a whole. This prevents the judges from simply using this information to establish their cut-off marks in such a way as to place a certain percentage of students in a particular performance band. If it becomes apparent to a Field Officer that a team of judges has had access to such information this would be noted in their report for later consideration by the Consultative Committee. This is an important element in maintaining the integrity of the exercise.

Even when the judges have finalised their decisions they are not given any indication as to what the distribution of performance bands will be when their decisions are applied. The publication of information showing the proportions of students who receive each Band in each course is delayed until after the HSC results are released to students.

Final cut-off marks are not published

Even after the HSC results and associated reports, including the proportions of students placed in each performance band, are released, the final band cut-off marks for each course are not released.

The Board has consistently maintained the position since 2001 that the band cut-off marks each year should remain confidential so that the standards-setting procedure can operate the way it was designed. That is, each year the judges need to focus on the

standards themselves using the standards packages, rather than simply start with the cut-off marks from the previous year and decide whether they should increase or decrease each cut-off mark.

Although those judges who were involved in the standards-setting operation in the previous year may recall the decisions they made, they can not be absolutely sure that the values they established were the final values that were used. (Following the completion of the standards-setting operation the HSC Consultative Committee may have taken the decision, through its rigorous review of the operation and other information, to alter some or all of the cut-off marks recommended by the team of judges.)

Maintaining the confidentiality of the cut-off marks is an important element in maintaining the integrity of the HSC program.

The HSC Consultative Committee

The HSC Consultative Committee is an expert committee appointed by the Board of Studies to review the outcomes of the standards-setting procedure in each course. The committee consists of 10 to 12 highly experienced academics and others with expertise in educational measurement. All members have a thorough understanding of the Board's assessment policies and programs.

It is the role of the HSC Consultative Committee, acting on behalf of the Board of Studies, to either accept the cut-off marks recommended by the judges or to make minor adjustments where the committee believes it is necessary.

Following the completion of the standards-setting operations at least two members of the Consultative Committee meet with the representatives from each course. They also review the reports produced by the judges, supervisors of marking, Chief Examiner and Field Officers. Through these meetings the Consultative Committee members will confirm that the standards-setting procedure followed by each team was in accordance with the Board's requirements, or identify any departure from the specified procedure. The committee will also identify any issues that may have impacted on the effectiveness of the standards-setting procedure and on the quality of the judges' decisions.

Having met with the judges and other course representatives the Consultative Committee then uses a purpose-built software package to review the data associated with the course. This information includes the proportions of students who will receive each band if the judges' recommended cut-off marks are applied; the judges' individual decisions following each stage of the procedure; the summary statistics of the examination marks; and a graphical representation of the distribution of examination marks. The Consultative Committee also can access the same information from previous years.

The Consultative Committee as a group discusses each course in turn, taking into account the information provided in the reports and from the meetings with the course representatives, and the statistical information provided. Following its deliberations the committee either confirms the judges' recommendations or modifies one or more

of the cut-off marks. In making any amendments the committee is governed by the following considerations:

- Do the recommended cut-off marks place the current percentage of the course candidature within the variability observed in the bands over time? If so, changes to the recommended cut-off marks would only be made where other available information identifies a cause for a shift in percentages outside the observed variability.
- If a change is deemed necessary then it is usually to be within the variability of the judges' decisions recorded after Stage 1 of the standards-setting procedure.

If the Consultative Committee makes a change to any of the judges' recommended cut-off marks it is required to record the reason it has made the particular change.

Some statistics on the distribution of bands over time

This section contains some information that provides evidence to support the claim that differences in the distribution of performance bands in courses from year to year are within the natural variability that can be expected from a standards-setting procedure such as this. Furthermore, there is data to support the claims from teachers that the availability of the HSC standards packages and the opportunities they provide for teachers and students to compare a student's work to the samples of work in the packages have had a positive impact on learning. Teachers are able to demonstrate to students how they can improve the standard of work they are currently producing in order to achieve higher performance bands. This aspect is explored in Bennett and Taylor (2003) and Bennett (2004).

In interpreting the information provided in the tables below it is important to remember three important points.

- In establishing the standards for each course in the NSW Higher School Certificate it was not a requirement that courses would have the same proportions of students in each performance band, although the distribution of bands in like courses is usually fairly similar.
- In any standards-setting procedure based primarily on informed professional judgement it must be expected there will be changes in the proportions of students awarded each band from one year to the next, and so care must be taken in making claims about changes in patterns of student performance over such a short timeframe.
- The fact that the underlying initial mark distribution is aligned to a performance scale where the great majority of students receive a mark between 50 and 100 means that for many courses a difference of one mark, or even a fraction of a mark, in a cut-off may lead to a difference of several percent of the candidature receiving a particular performance band. That is, some variation between the percentages of students receiving each percentage band from year to year is to be expected..

1. *The distribution of performance bands in a selection of courses 2003–2007*

Each year there are around 110 courses for which students can enrol. The candidature of courses varies from less than 10 to more than 30 000 students.

Tables 1a to 1g show the distribution of bands in a number of courses over time. The courses chosen are taken from a range of learning areas. Percentages of students who entered but did not receive a result in the course are not reported.

Table 1a English (Advanced) (2007 candidature 28 113)

Band	2003	2004	2005	2006	2007
6	6.8	7.6	8.0	6.0	9.2
5	34.8	42.5	37.8	32.8	37.7
4	46.5	38.9	44.1	43.6	42.6
3	11.1	9.0	8.9	15.8	9.6
2	0.5	0.7	1.0	1.6	0.9
1	0.1	0.1	0.1	0.2	0.1

Table 1b Mathematics (2007 candidature 17 825)

Band	2003	2004	2005	2006	2007
6	14.5	15.5	15.1	14.6	15.4
5	25.2	26.5	23.7	24.1	24.1
4	28.3	26.5	24.9	26.2	30.3
3	17.9	17.6	18.9	18.2	18.2
2	8.5	9.0	10.6	8.9	8.0
1	5.3	4.5	6.1	7.5	3.6

Table 1c Chemistry (2007 candidature 10 333)

Band	2003	2004	2005	2006	2007
6	6.7	8.3	8.3	8.9	10.8
5	19.2	26.8	23.6	27.2	28.4
4	26.4	30.5	30.6	27.4	29.7
3	26.6	19.7	23.5	25.9	22.2
2	13.9	8.6	10.2	8.0	6.1
1	6.9	5.6	3.3	2.3	2.3

Table 1d Economics (2007 candidature 5716)

Band	2003	2004	2005	2006	2007
6	12.6	13.5	14.1	13.9	14.6
5	31.2	34.7	34.9	32.8	32.0
4	29.2	31.3	25.7	28.1	25.9
3	15.9	12.3	13.9	15.6	15.6
2	7.1	5.6	5.8	6.2	7.6
1	3.7	2.2	5.1	2.7	3.7

Table 1e French Continuers (2007 candidature 842)

Band	2003	2004	2005	2006	2007
6	22.6	22.7	20.8	27.8	28.7
5	26.5	23.4	27.7	31.0	28.7
4	26.2	27.6	29.4	28.0	22.8
3	14.0	16.8	17.3	7.7	13.4
2	6.6	5.9	4.1	3.9	4.9
1	4.0	3.5	0.7	1.5	1.4

Table 1f Design & Technology (2007 candidature 3916)

Band	2003	2004	2005	2006	2007
6	3.1	2.8	4.0	6.1	5.5
5	18.5	16.9	16.6	18.9	18.3
4	39.7	39.4	34.0	37.9	39.3
3	28.5	30.4	33.3	29.8	29.5
2	8.4	8.5	10.0	6.6	6.6
1	1.5	1.6	1.8	0.7	0.5

Table 1g Visual Arts (2007 candidature 9369)

Band	2003	2004	2005	2006	2007
6	11.3	11.0	11.3	13.3	11.4
5	38.3	36.9	40.0	42.4	40.5
4	33.4	37.8	35.9	35.4	38.4
3	14.1	12.8	11.2	8.2	8.9
2	2.1	1.1	1.2	0.5	0.5
1	0.6	0.1	0.1	0.1	0.1

2. General Indicators of Performance 2001 to 2007

The values reported in Table 2 were calculated by dividing the total number in each performance band in all courses by the total number of performance band results awarded in that year. While the points made above about variations in the distribution of bands that result when the standards are created are still relevant, these indicators nevertheless give a general indication of student performance across the whole HSC program.

Table 2 General Performance Indicators over time

% of Course Entries	2001	2002	2003	2004	2005	2006	2007
Band 6	6.1	7.7	7.9	8.6	8.9	9.0	9.6
Bands 5 & 6	24.5	30.5	31.5	33.8	35.3	35.5	36.3
Bands 4, 5 & 6	55.2	61.1	62.8	64.6	67.4	66.2	68.1
Above MSE ¹	95.4	95.0	95.4	95.8	96.1	95.5	96.1

These indicators show a steady increase in the performance of students in the HSC since the introduction of the standards-based approach to reporting student

¹ "Above MSE" means in Band 2, 3, 4, 5 or 6.

achievement in 2001. Furthermore, after the first few years when the approach and the standards were being “bedded down”, the figures have become quite stable.

Given the measures that are put in place to protect the integrity of the standards-setting procedure, it is reasonable to claim that the small increase in the proportion of students achieving the higher performance bands over the past four or five years can be attributed to the clear, publicly available standards and teachers’ and students’ engagement with them.

Further work

Berk (1996) states that the validity of a judgemental standards-setting procedure is dependent on the expertise of the judges, the procedure being thorough and it being rigorously applied.

He claims that when the procedure is thorough and rigorous the final standards will be “whatever the judges say [they are]” (p.230). In a situation where the standards-setting procedure is based on professional judgement, the outcomes should be accepted if it is a good quality procedure that is properly followed by expert judges.

The strategies described in the paper protect the integrity of the NSW Higher School Certificate standards-setting procedure and have delivered the necessary level of confidence in the outcomes produced. As with all aspects of the HSC, we regularly review, assess and improve our processes. To this end, some studies will now be conducted to confirm the effectiveness of the judging procedure. These studies will involve additional judges undertaking some activities in a sample of courses to validate the work of the judging team.

Conclusion

The standards-based approach used in the NSW Higher School Certificate to report student achievement in relation to the publicly available standards is widely accepted in the educational and wider communities. The use of trained professionals to implement the multistage standards-setting procedure each year is well regarded. While representatives of some subject areas from time to time comment on the fact that their course does not have as high a proportion of students who receive, say, Band 6 as some other course, they nevertheless accept that the system is based on clear standards appropriate for their course.

In any system that uses a standards-setting procedure based on professional judgement it must be expected that there will be some variation in band distributions from year to year. This is especially the case when such a procedure is applied to examination papers with the level of complexity of those used as part of the NSW Higher School Certificate. If the variation over several years always showed an increase in the top bands, one might be excused for supposing, in the absence of reasons to think otherwise, that this was due either to a faulty procedure or some manipulation of the procedure. In these circumstances, a claim of “grade inflation” might have some substance.

In the case of the NSW Higher School Certificate, the range of features built into the standards-setting procedure itself and the various oversight measures that are used all play a part in ensuring the outcomes of the process are fair and reasonable. These measures protect the system from being able to be easily manipulated to produce a particular result.

Although care needs to be exercised in reading too much into changes in band distributions over short periods of time, the data provided above support the anecdotal evidence provided by teachers that the public availability of the achievement standards, and the opportunity that teachers and students have to engage with them, is delivering genuine overall improvements in student learning.

References

Angoff, W. (1971). "Scales, Norms and Equivalent Scores" in R.L. Thorndike (ed.), *Educational Measurement* (2nd ed., pp 508-600), American Council on Education, Washington, DC.

Bennett, J. (1998). *A Procedure for Equating Curriculum-based Public Examinations Using Professional Judgment Informed by the Psychometric Analysis of Response Data and Student Scripts*. Unpublished doctoral thesis, University of New South Wales.

Bennett, J. and Taylor, C. (2003). *Is Assessment for Learning in a High-Stakes Environment a Reasonable Expectation?* Paper presented at ACACA Conference Adelaide 2003.

Bennett, J. (2004). *How can assessment be used to improve student learning in a high-stakes environment?* Paper presented at IAEA Conference Philadelphia 2004.

Berk, R. (1996). "Standards Setting: The Next Generation" in *Applied Measurement in Education*, 9, 215-235.

Masters, G. (2002). *Fair And Meaningful Measures? A review of examination procedures in the NSW Higher School Certificate*. Australian Council for Educational Research Ltd.

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