

# Regulating A Level standards

## *Mathematics in Education and Industry (MEI)*

### **This paper**

At a recent conference Ofqual raised the following question.

*What does it mean to maintain standards in qualifications when the qualification outcomes are predetermined by unit outcomes ?*

This paper addresses this question in the context of A Levels. Many of the points made are also relevant to other qualifications.

### **The authors**

MEI can claim considerable expertise in this area. MEI Structured Mathematics was the first modular A Level in any subject. It received accreditation in February 1990 and teaching began in September of that year. The first A Level awards in Mathematics and in Further Mathematics using this scheme took place in 1992. This was a year ahead of any other modular A Level.

The MEI scheme was designed from first principles, starting with a blank sheet of paper. Other credit accumulation frameworks, such as that used for awarding Open University degrees, were considered and judged to be inappropriate for A Levels. The scheme which MEI developed is in essence the one that is used today; the term “uniform mark” was coined by MEI. By contrast, other schemes that were devised for other syllabuses and implemented in 1991, or shortly afterwards, have long since disappeared.

### **The purpose of A Level qualifications**

Before considering how best to regulate standards, it is important to step back and consider the purpose of A Level qualifications. Two different views can be heard.

- A Level should be a competitive examination for university entrance.
- A Level should measure and report candidates’ attainment over a range of reasonably well understood standards.

These two purposes would require quite different accreditation regimes. The first is supported by norm referencing, the second by criterion referencing.

Our present system is designed to encourage young people to expand their horizons by learning as much as they possibly can. This aim is supported by our modular system which provides students with realistic goals and thereby promotes achievement. Philosophically it is in tune with criterion referencing and a very long way from norm referencing.

## Norm and Criterion referencing

Norm referencing was used for A Level until 1987 when, following a conscious decision, it was replaced by criterion referencing. No decision has been made to revert to norm referencing but in practice it has been used since the introduction of Curriculum 2000.

Awarding bodies are instructed to maintain the same pass rate in any subject year on year. Only small variations are allowed from one year to the next. Appendix A, giving AS pass rates from 2001 to 2008, shows just how effective this regime is as a way of maintaining the status quo. The data show that that essentially the same group of subjects always have the highest pass rates and another group of subjects always have the lowest pass rate; the positions of subjects in the middle are similarly stable. There is a well-defined rank order.

## The present regime

The present regime is imposed at the unit level. Thus, although in theory an awarding committee sets the grade thresholds by looking at the quality of borderline candidates' work, in practice they are more usually determined by the mark distribution. In many cases an awarding committee is asked to decide between two adjacent marks that have been predetermined statistically, or even just to confirm a particular mark. In the unusual event of an awarding committee recommending a different threshold, this can be overturned by the awarding body's accountable officer.

The fact that unit by unit the grade distribution is the same from one year to the next means that the same is true when the unit scores are aggregated to give qualification grades.

The rank order of subjects shown in Appendix A is that which happened to occur in 2001, the first year of new specifications, and has nothing to do with the relative ability of the candidature for the different subjects. Indeed the order that is being imposed is incorrect. This is particularly serious for STEM subjects, all of which have high failure rates. It reinforces the message that these are hard subjects and so discourages many students from taking them; this is strongly against the national interest.

The present system does not allow for changes in the ability or level of achievement of the candidature. Thus, while various initiatives encourage better teaching, the assessment system effectively denies that these can have any effect.

The only answer that can be give to the question "*Have standards changed ?*" is that the same percentage of candidates are getting a particular grade as before. The information is not available to say whether candidates' essays are of comparable quality, whether they can do their algebra as accurately or whether they can still apply scientific principles as well.

## Problems with norm referencing

Norm referencing encourages lazy regulation. It allows a cursory glance at a mark distribution to become a substitute for monitoring how well an examination is assessing the content and skills that candidates are supposed to have mastered.

Consequently, the return to norm referencing in the last 8 years can be seen as a retrograde step on behalf of the regulators. They no longer routinely receive the information they need for a proper understanding of what is going on, both at unit and qualification level. This makes it impossible for them to meet the requirements of Clause 125(2) of the Apprenticeship, Skills, Children and Learning Bill.

*The qualifications standards objective is to secure that regulated qualifications –*

- (a) give a reliable indication of knowledge, skills and understanding, and*
- (b) indicate a consistent level of attainment (including over time) between comparable regulated qualifications.*

Since it is manifestly not possible to regulate standards while norm referencing remains in place, the answer to the question at the start of this paper must involve a return to criterion referencing. In the absence of any decision to revert to norm referencing, this is what should have been happening anyway.

## Criterion referencing

The essential first step is to recruit people with appropriate subject knowledge and examining experience to allow them, collectively, to be arbiters of standards. Such people must be of high calibre; they will be required to exercise deep professional judgement when comparing examinations from different sittings and different specifications.

It should be recognised that criterion referencing can degenerate into an exercise in ticking boxes. While this may be appropriate for the driving test, it is not what education, and its assessment through school examinations, is about.

In recent years there has been a tendency to rely less on people and more on procedures and systems. This should be reversed; it needs to be recognised that every aspect of education, including its assessment, is a human activity. Ultimately, standards can only be determined and maintained by the collective wisdom and experience of teams of competent people.

Since standards are about candidates' subject knowledge and their ability to apply it, criterion referencing must operate at unit level. The regulator should ensure that suitable examination papers are set in the first place and that the thresholds determined at the awards are consistent with the demands of those papers. This has happened in the past but not for quite a number of years. These days, scrutineers do sometimes attend revising meetings or awards but their sole concern is whether procedures are being followed; the question of whether the outcomes are of high quality is beyond their remit.

## **Other considerations**

While a move back to criterion referencing is essential, it is not the only change that is needed to allow good regulation. The two points below are particularly important.

### ***Uniformity across subjects***

The data in Appendix A highlight discrepancies in standards between subjects. It is self-evident that this situation should not be allowed to continue, particularly when performance tables and the UCAS tariff both depend on the assumption that all subjects are at the same standard. It would be inconsistent with their role if those responsible for regulation were to do nothing, and worse still if this were to be justified on the grounds that the situation is too difficult and so nothing can be done about it.

The problems arise because the same procedures, including those involving such diverse matters as syllabus design and the interpretation of uniform marks, are applied across subjects with very different assessment requirements and conventions.

It is essential that regulators understand these differences and so how applying the same regulations to all subjects makes it impossible to ensure comparability between them.

### ***External influences***

External pressures, not least from performance tables, have probably increased the level of re-sitting AS units. If so, the impact of this on their mark distributions, together with their knock-on effects on qualification grades, needs to be considered seriously. In the first place research is needed to establish if this really is a problem and, if so, in what subjects. This is not an argument for restricting re-sitting but for looking at what other changes could be made to accommodate the changed situation.

## **In conclusion**

This paper would not be complete without a final note of caution. In recent years a number of really valuable syllabuses have had to be discontinued because they did not meet the particular regulatory requirements of the day; this should never happen again.

Regulation is not the purpose of education. Those responsible for regulation have a responsibility to ensure that their work supports the wider aim of giving young people a rich and diverse experience that will prepare them for the rest of their lives, and this applies both across the curriculum and within subjects.

*Roger Porkess,  
MEI Chief Executive.  
23/10/2009*

## Appendix A

## Rankings of AS subjects by Pass Rates, 2001 – 2008

Source: JCQ

Subject	2001		2002		2003		2004		2005		2006		2007		2008	
	Rank	Fail	Rank	Fail	Rank	Fail	Rank	Fail	Rank	Fail	Rank	Fail	Rank	Fail	Rank	Fail
Welsh	1	2.8	1	2.2	1	0.9	1	1.7	1	1.6	1	1.7	1	2.0	1	1.1
Classics	2	4.0	2	4.2	3=	4.7	4	4.6	4=	4.4	6	4.9	5	4.8	5	4.7
Express Arts	3	4.3	3	4.4	3=	4.7	2	4.0	2	3.5	2	3.4	2	3.2	2	2.8
Music	4	5.2	9=	8.5	8=	8.0	12	8.2	15	8.8	14	8.0	14	8.1	14	7.3
English	5	5.3	5	6.2	5	5.3	6	5.2	4=	4.4	3	4.0	3=	3.9	4	3.6
History	6=	6.4	8	7.9	7	7.2	8=	7.0	9	6.1	11	7.0	8	6.2	8	6.2
Media St	6=	6.4	4	5.9	6	5.1	5	4.8	6	4.6	5	4.8	6	5.0	6	4.8
French	11=	9.0	16=	10.4	14	8.8	15	9.1	14	8.6	15	8.3	15	8.2	15	7.6
Spanish	11=	9.0	14=	9.9	11	8.4	11	8.0	10	7.1	9=	6.8	10	6.6	12	6.8
Geography	13	9.2	14=	9.9	16	9.2	16	9.7	16	9.7	16	9.2	16	9.2	16	9.2
Business St	20=	13.0	21	13.1	21	13.2	22	13.7	21	13.4	20	13.1	20	12.7	20	12.5
Chemistry	20=	13.0	22	13.3	22	14.4	23	15.1	23	14.9	23	14.4	23	14.6	23	14.4
Sociology	22	13.5	23	14.8	23	14.8	21	13.5	22	13.6	21	13.2	22	13.2	22	13.3
Physics	23	13.9	24	15.4	25	16.0	25	16.5	24	16.1	17	15.8	24	15.4	25	15.1
Biology	26	15.6	26	17.1	27	18.3	27	18.0	27	17.3	27	18.1	27	17.0	27	16.6
Psychology	27	17.2	27	17.3	26	18.0	26	17.9	28	17.9	29	18.5	28	18.4	29	18.1
General St	28	18.1	28	19.9	29	19.3	30	19.3	30	19.0	30	18.9	29	19.1	30	18.5
Computing	29	19.5	30	21.7	31	21.9	31	20.0	31	20.3	31	19.4	31	19.5	28	17.1
Law	30	20.5	29	20.6	28	18.7	28	18.6	29	18.1	28	18.3	30	19.7	31	19.3
Mathematics	31	28.6	31	22.1	29	19.9	29	19.1	26	17.4	26	16.7	26	16.7	26	15.7

Note This table is based on the 2001 subject definitions. In a few cases, e.g. Further Mathematics, these differ from those currently reported.