Guide to Ofqual Technical Report
on A Level Assessment and related matters

"Exam boards will be required to ensure that their question papers differentiate in a more reliable way between the more able students in addition to addressing concerns regarding the design and underlying principles behind their mark schemes. These recommendations are being made to achieve greater fairness in the grades students receive.

Changes to exam papers are required for summer 2015 with changes to the assessment of speaking being required in later years.

These changes will not require any change to the way the subjects are taught or the way students are prepared for their exams"   Ofqual Press Release (26th Sept 2014)

   http://ofqual.gov.uk/news/improvements-made-level-foreign-languages/

Introduction and Documentation

Some of the text and illustrations in this document are taken from the Ofqual Report; others come from analysis published by ASCL / ALL London - references are given to make it clear.

ALL and other organisations have been raising concerns for many years about severe and unreliable grading at both GCSE and A-level for many years. Since 2010, concern has also been raised about the disproportionately low number of A* grades awarded in ML compared with other facilitating subjects. *http://www.all-london.org.uk/severe_grading.htm* and Appendix A

JCQ investigated the concerns at AL and published a report in July which was very thorough and wide-ranging, "A review of Modern Foreign Languages at A level: A* grade and low take up" *http://www.jcq.org.uk/media-centre/news-releases*

ASCL / ALL London produced a paper in August "ML AL grade boundaries for June 2014 for AQA + others - v4b.docx" both in response to the JCQ Report and the change in June 2014 in the AQA AL grade boundaries and its impact on A*.

The Ofqual Report   *http://ofqual.gov.uk/news/improvements-made-level-foreign-languages/* looks primarily at the issue of unreliable grading at AL especially for able candidates, and its recommendations should also bring a greater fairness to the number of A* grades awarded.

It considers in detail the questions and mark schemes for AL French, German and Spanish for the 4 main boards (AQA, OCR, Pearson and WJEC).

The findings and recommendations are presented in Chapter 10 (pp 89 - 91), and, for convenience, are in Appendix B to this document

Such a level of detail is fascinating as it shows how a specific action or requirement can have unintended consequences, and also valuable for the teachers of each Board to see the differences between the different papers/skills and questions within each Board, as well as between the Boards. Some of the recommendations are Board / skill specific.

There are some key over-arching messages though which need to be explored and understood, as there is significant risk of misunderstanding both the message and the relevant recommendation. For simplicity and ease of following through, the examples are taken from AQA which is the has the highest number of candidates, but the situation is similar in all of the Boards.

Other key points

Distribution of raw marks

The spreading out of the raw marks (which links with increased discrimination in the assessment sense of the word) will NOT affect the distribution of other grades, as the grade boundaries will shift to match the wider distribution of marks. Note that in the Technical Report, words such as "demand" appear, but these are being used in a strictly technical sense.

AS / A2 mark split

This has been one of the fundamental reasons of the relatively low number of * grades in ML compared with other subjects. Of course, the very issue of an AS/A2 split in marks will automatically disappear with the new exams as all the AL grades will be determined solely by performance at the end of course. The new definition of A* is still under consideration.

Overall severe grading

Ofqual recognise (as did the JCQ report) that this is a fundamental and critical issue for ML at both GCSE and AL, which must be addressed to ensure a "level playing field" as students and parents compare their grades in ML with those in other subjects. The challenge now is to determine a mechanism for change, and so Ofqual are undertaking an ongoing programme of work.

Change in cohort profile over time

One of these over-arching messages is the current mismatch between the "ability" profile of the candidates and the "difficulty" profile of the assessment.

It is important to look back in time to better understand this.

Since Curriculum 2000 was introduced and the first A2 exams in 2002, there has been a steady decline in the overall numbers taking AL ML, in contrast to
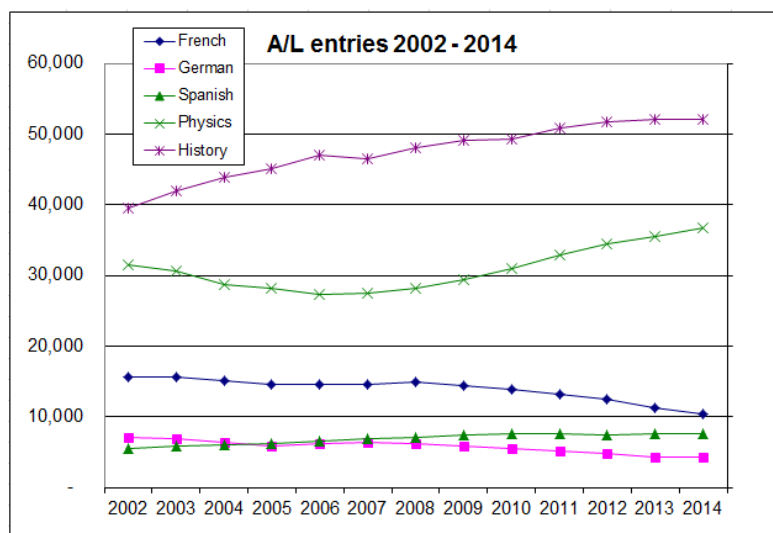
other "facilitating" subjects such as History, Maths and Physics, as illustrated on the right *(source for all figures in this section: ASCL/ ALL London from JCQ UK figures)*

So French has declined from 15,130 in 2002 to 10,433 in 2014, and German from 6,810 to 4,158, with Spanish increasing from 5,416 to 7,555

But within this overall decline from 2002 to 2014, there is a particularly sharp drop in grade D and E candidates, whilst the number obtaining A & A* and B grades remained reasonably constant (with a dip from 2012 once the problems with the introduction of A* became clear). This is clear in the graph on the right below:

no. students each grade AL French 2002-14
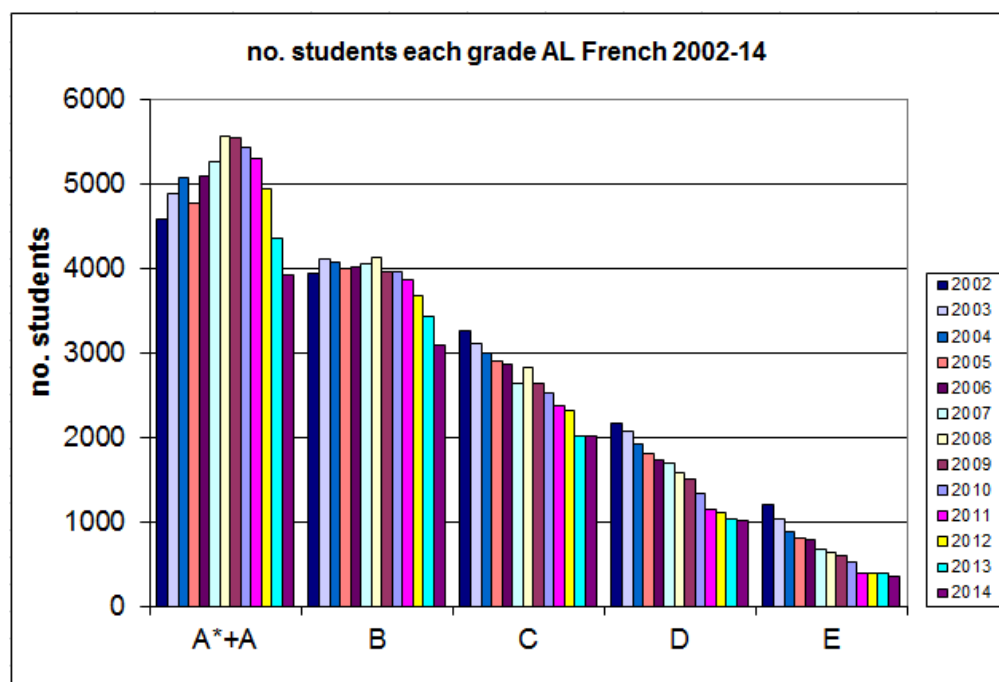
	A*	A	A*+A	B	C	D	E	
2002			4575	3935	3263	2155	1202	first year of A2 and AS
2010	1066	4349	5415	3947	2521	1330	512	first year of A*
2014	689	3234	3923	3078	2003	1012	344	latest results
2014 as % of 2002			*86%*	*78%*	*61%*	*47%*	*29%*	

source: ASCL / ALL London from JCQ UK figures

The percentage of D and E grades has shrunk from 21.5% to 13% in this period. Therefore it is not surprising that there is a mismatch between the appropriateness of the questions and the ability profile of the candidates if there is to be good differentiation of candidates across the ability range ("discrimination"). This means that the style of paper would need to change, and that is something to which there is probably a natural reluctance, as normally teachers and students want the papers to be similar from year to year to help with preparation.

Please note that it is vital to keep completely separate
1.   the range and difficulty of the questions (which leads to RAW marks)
2.   grades which are awarded - the examiners make a judgement as to where the grade boundaries are set in terms of raw marks ("standards")

The Ofqual Report is looking ONLY at the assessment process and NOT at overall grading

Mark distribution

The examples are from AQA for convenience. The illustration on the right below shows the proportion of candidates getting how many marks on the French 3 paper (Listening + Reading + Writing). Note that the questions have different numbers of marks, so both the horizontal and vertical scales differ from question to question. Importantly though, for many of the questions, many of the candidates are scoring highly, as displayed in the table below, leading to a bunching of the overall RAW marks to the high end. Technically, this is referred to as a "negative skew".

There is nothing wrong per se in having a negative (or positive) skew in an exam, but it does make it more difficult to reliably and fairly discriminate amongst candidates, as small fluctuations in marks can lead to disproportionate change in grades.

Table 3.5: Item facility indices for FREN3, GERM3 and SPAN3 for certificating candidates in June 2013

	FREN3	GERM3	SPAN3
Sect A			
1	0.93	0.90	0.98
2	0.71	0.75	0.84
3	0.72	0.86	0.98
4	0.62	0.91	0.86
5	0.82	0.80	0.66
6	0.72	0.87	0.84
7	0.73	0.55	0.69
8	0.67	0.82	0.55
9	0.54	0.62	0.57
10	-	0.55	-
Sect B			
1	0.59	0.68	0.64
2	0.54	0.70	0.64
3	0.62	0.73	0.68
4	0.63	0.75	0.70

Figure 3.3: Item-level mark distributions for FREN3 for candidates sitting the unit and certificating candidates at A level in the June 2013 exam series

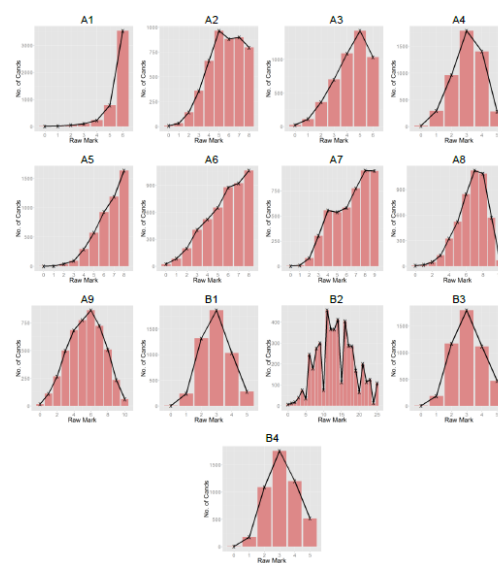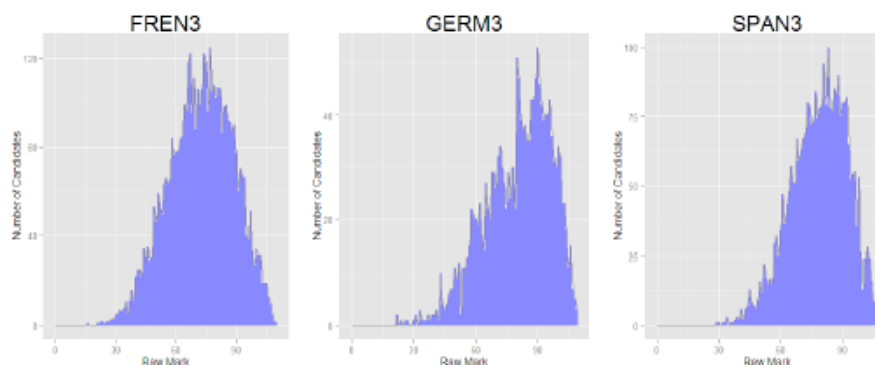**Figure 3.6:** Unit-level mark distributions for the AQA A2 written units for A level certificating candidates in June 2013

	Mean	Standard Deviation	Skew
FREN3	72.5 (65.9%)	16.0 (14.5%)	-0.22
GERM3	80.5 (73.2%)	16.0 (14.5%)	-0.63
SPAN3	78.5 (71.4%)	13.7 (12.5%)	-0.42

## Imbalance between the questions and the ability profile

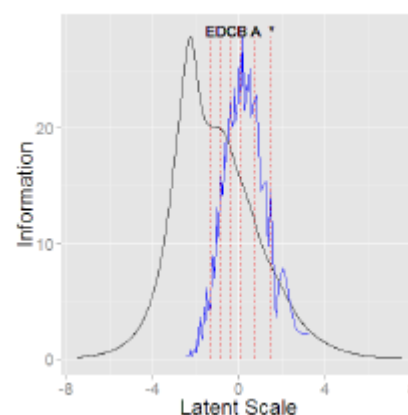This imbalance is shown in **Figure 3.7:** *Test information functions (black) for a) FREN3, b) GERM3 and c) SPAN3 from June 2013. Superimposed are the unit-level grade boundaries (dotted) and the distribution of candidate person parameters relative to the information functions (blue).* (Ofqual Rpt p.23)

a) FREN3   b)GERM3 c) SPAN3
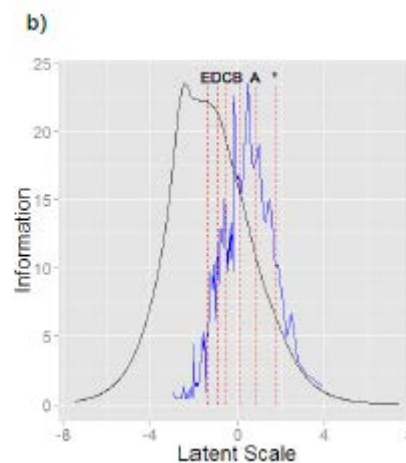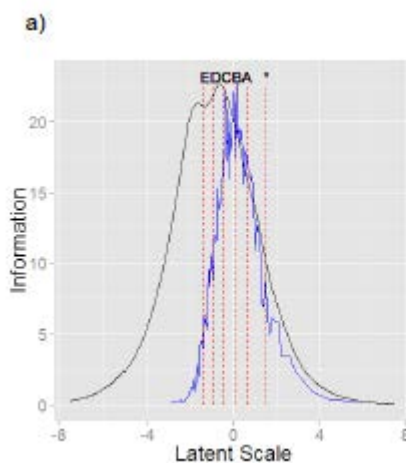
These are complex diagrams, but in simple terms, the range of "difficulty" of the questions in the paper (black) tends to be skewed to the left, whereas the ability profile of the candidates (as we have seen above) is skewed to the right

How the changes might increase the number of A* grades whilst leaving overall A+A* same

Ofqual Report Chapter 7 Impact of assessment functioning on A outcomes (p.79)*

This is an extremely technical section, but in summary, because of the particular way that A* is defined, a bunching of RAW marks mismatched with the ability profile of the candidates has a mathematical link with the proportion of A* within the overall A+A*.
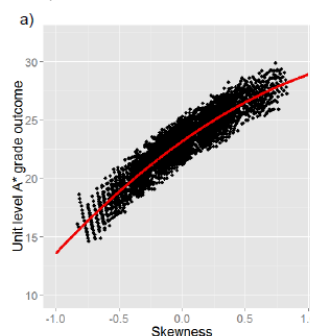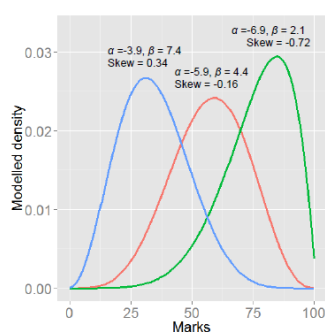
In the graph on the left, the AL ML distribution is typically that of the green curve The graph on the right shows that by making the distribution of marks less bunched (e.g. the red curve with zero skew), then the RELATIVE proportion of A* within the unchanged total A & A* will increase.

Appendix A - The A* problem for ML

2014 figures

|	A*	A	A+A*	A* as % of A+A*
French	6.6	31	37.6	**18%**
Spanish	7.7	27.9	35.6	**22%**
German	8.9	31.7	40.6	**22%**
History	6.3	19	25.3	**25%**
Economics	8.5	23.1	31.6	**27%**
Physics	10	20.6	30.6	**33%**
Biology	9.4	18.1	27.5	**34%**
Mathematics	17.3	24.8	42.1	**41%**

	A*	A	A+A*	A* as % of A+A*
History	6.3	19	**25.3**	25%
Biology	9.4	18.1	**27.5**	34%
Physics	10	20.6	**30.6**	33%
Economics	8.5	23.1	**31.6**	27%
Spanish	7.7	27.9	**35.6**	22%
French	6.6	31	**37.6**	18%
German	8.9	31.7	**40.6**	22%
Mathematics	17.3	24.8	**42.1**	41%

June 14 - A* as % of A+A*
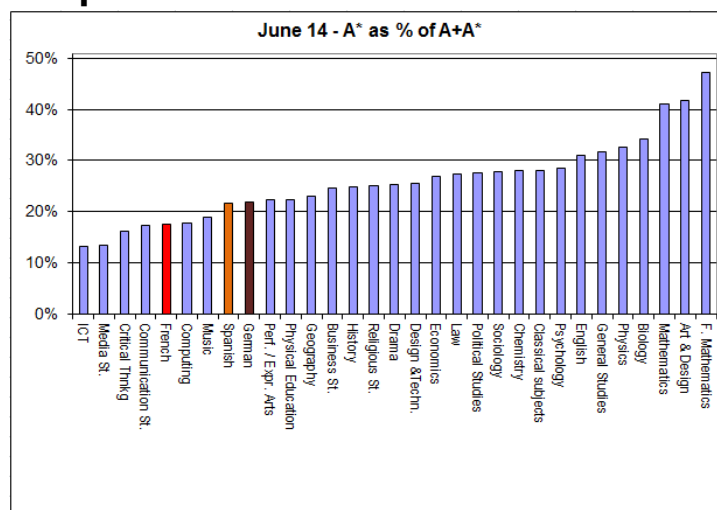
Statement in 2012

Figures are for England and have been taken from JCQ statistics. There is an accompanying Excel spreadsheet with all the figures and graphs. These documents and many others relating to "severe grading" are at the ALL London website: *http://www.all-london.org.uk/severe_grading.htm*

- The serious issue regarding the number of A* awarded in Modern Language subjects especially **French has continued and become even more pronounced.**

- The intake profile of French has more higher-attaining students (like other subjects such as Maths and Physics), and so they have had a higher percentage of students gaining grade A at A-level (e.g. 38.6% of entries in 2009 [the last pre-A* year] – German 40%, Maths 45%, Physics 32% etc compared with 26.7% for all subjects)

2012 sorted by A* as % A+A*

	A*	A	A+A*	A* as % of A+A*
Media / Film	1.6	9.4	11	15%
French	6.8	32.6	39.4	17%
Critical Thin	2.3	10.1	12.4	19%
Spanish	6.8	29.7	36.5	19%
ICT	2.4	10.3	12.7	19%
German	7.9	33.5	41.4	19%
Religious S	5	20.5	25.5	20%
Business S	3.3	12.2	15.5	21%
Geography	6.3	22.8	29.1	22%
Computing	3.4	12.2	15.6	22%
Performing	3.9	13.9	17.8	22%
Music	4.1	14.6	18.7	22%
Drama	3.7	13	16.7	22%
Economics	7.8	26.1	33.9	23%
Sociology	4.2	13.9	18.1	23%
Political St	7	23.1	30.1	23%
Psychology	4.2	13.3	17.5	24%
Chemistry	8.8	25.6	34.4	26%
Physical Ed	3.8	11	14.8	26%
History	6.9	19.6	26.5	26%
Biology	7.9	20.7	28.6	28%
Design and	4.9	12.5	17.4	28%
Law	5.1	12.8	17.9	28%
Classical su	10.7	26.4	37.1	29%
Communic	5.4	12.1	17.5	31%
Physics	9.9	22	31.9	31%
English	6.8	14.4	21.2	32%
General Stu	4.2	8.2	12.4	34%
Mathematic	17.4	26.5	43.9	40%
Art and Des	13.1	17.1	30.2	43%
F. Mathem.	28.6	28.8	57.4	50%
All Subje	**7.9**	**18.7**	**26.6**	**30%**

2011 sorted by A* as % A+A*

	A*	A	A+A*	A* as % of
Media /	1.6	10.2	11.8	14%
Critical	2.7	11.5	14.2	19%
French	7.7	32.4	40.1	19%
Drama	3.6	14.5	18.1	20%
Religiou	5.5	21.8	27.3	20%
ICT	2.5	9.6	12.1	21%
German	9.1	32.9	42	22%
Spanish	8.3	29.5	37.8	22%
Music	4.2	14.8	19	22%
Compu	3.7	12.8	16.5	22%
Busine	3.8	12.6	16.4	23%
Geogra	7.1	23	30.1	24%
Psycho	4.3	13.9	18.2	24%
Econon	8.6	26.6	35.2	24%
Sociolo	4.8	14.6	19.4	25%
Law	4.7	13.7	18.4	26%
History	7.2	20.2	27.4	26%
Chemis	9.4	24.9	34.3	27%
Politica	8.8	23.2	32	28%
Classic	10.8	28.2	39	28%
Perform	4.7	12.1	16.8	28%
Commu	5.5	13.8	19.3	28%
Physica	4.4	10.8	15.2	29%
Design	5.3	12.5	17.8	30%
Biology	8.8	19.5	28.3	31%
Genera	4	8.8	12.8	31%
Physics	10.4	22.6	33	32%
English	7.1	15.2	22.3	32%
Mathen	17.8	26.9	44.7	40%
Art and	14.2	17	31.2	46%
Mathen	27.5	30.2	57.7	48%
All Sul	**8.2**	**18.8**	**27.0**	**30%**

2010 sorted by A* as % A+A*

	A*	A	A+A*	A* as % of A+A*
Media / Film	1.8	10.7	12.5	14%
ICT	2	9.3	11.3	18%
French	7.7	31.4	39.1	20%
Music	4	15.1	19.1	21%
Computing	3.5	12.8	16.3	21%
Religious Stu	6	21.5	27.5	22%
Business Stud	3.7	13.1	16.8	22%
Spanish	8.5	29.9	38.4	22%
All Other Sub	4.2	14.5	18.7	22%
Geography	6.8	23.4	30.2	23%
Economics	8.8	27.9	36.7	24%
German	9.9	30.9	40.8	24%
Classical Sub	9.6	28.6	38.2	25%
History	7	20.7	27.7	25%
Political Stu	8.7	25.1	33.8	26%
Chemistry	9.2	25.4	34.6	27%
Sociology	5.4	14.9	20.3	27%
Sport / PE St	4.2	11.5	15.7	27%
Law	5.5	14.8	20.3	27%
Critical Thin	3.6	9.6	13.2	27%
Psychology	5.2	13.5	18.7	28%
Biology	8	20.5	28.5	28%
Drama	5.6	13.8	19.4	29%
Technology St	5.3	12.5	17.8	30%
Physics	10.3	22.6	32.9	31%
Science Subje	9.5	20.8	30.3	31%
English (1)	7.4	15.7	23.1	32%
Communication	6.5	13.4	19.9	33%
Performing /	5.9	12	17.9	33%
General Studi	4.4	8.6	13.0	34%
Mathematics	17.2	27.6	44.8	38%
Art and Desig	13.3	18.4	31.7	42%
Maths (Further	29.9	28.9	58.8	51%
All Subjects	**8.1**	**18.9**	**27.0**	**30%**

- One would therefore expect to also have a higher than average percentage of students gaining A* in such subjects. However, given the definition of A*, it would seem plausible that the ratio of A* to A*+A students would be similar across a range of subjects

- But this is not the case for French which has 39.4% of students with A or A*, but only 6.8 (7.7% in '11 and '10) with A* - this gives a very low ratio of 17% of A* to A+A* (19% in '11, 20% in '10)

- The front page of TES on 13th August 2010 had the headline "Exam boards massage A* A-level marks, Ofqual admits" http://www.tes.co.uk/article.aspx?storycode=6054157; Geoff Lucas (Secretary to HMC) wrote a measured critique of the process of introducing the new A* grade in TES on 20th August 2010

- The situation in 2012 has deteriorated for ML. The proportion of A* to A+A* is given below:

	2010	2011	2012		2010	2011	2012
Biology	28%	28%	**28%**	French	20%	19%	**17%**
Chemistry	27%	27%	**26%**	German	24%	22%	**19%**
Mathematics	38%	40%	**40%**	Spanish	22%	22%	**19%**
Mathematics (Further)	51%	48%	**50%**				
Physics	31%	32%	**26%**	All Subjects	30%	30%	**30%**

- Several universities now require an A* so this issue is increasing in importance.

- There are some very technical issues regarding the process of assigning grade boundaries at A-level using the A/B boundary and the E/U boundary with subsequent interpolation (and now extrapolation for A*) which were raised at the Ofqual inter-subject comparability seminar in Oct' 08, together with the "2% rule" referred to by Geoff Lucas.

Appendix B - Ofqual Report Findings and Recommendations

Findings

10.1 AQA

☐ The approach to awarding candidates marks for quality of language in the extended writing tasks does not appear to have a sound basis. The current approach is highly likely to be having a negative impact on the rank order of candidates and therefore the validity of the assessment.

☐ The prevalence of items that are of relatively low demand for those candidates sitting the written assessment is having a negative impact on the valid discrimination between candidates, especially for the most able candidates.

☐ The targeting of the written assessments relative to the required standard is suboptimal. This means that there is a greater amount of information collected to differentiate between candidates at the lower-ability range (where there are fewer candidates) than those of higher ability. This is more pronounced for German and Spanish than for French.

☐ The tendency for the lower-demand items to be concentrated in the assessment of listening for all languages has impacted on the extent to which that skill exerts influence over candidates' final outcomes. This has led to systematic differences between the intended and achieved weighting of skills, with listening being consistently underweighted and writing overweighted.

☐ The raw mark distributions for the speaking assessments are highly negatively skewed, with a large number of candidates achieving very high marks. This is likely impacting on the discrimination between the more able candidates on these assessments.

10.2 OCR

☐ Despite many items spreading candidates well across the mark distributions, the written exams contain a high proportion of items with a relatively high facility index, with some offering little discrimination between candidates.

☐ The targeting of sections A and B of the written exams relative to the required standard is suboptimal. This means that there is a greater amount of information collected to differentiate between candidates at the lower-ability range (where there are fewer candidates) than those of higher ability. This is more pronounced for German than for French and Spanish.

☐ There appears to be a lack of clarity and principle regarding the definition of acceptable responses for the translation task. Inconsistent principles may impact on the validity of the rank order of candidates.

☐ The raw mark distributions for the speaking assessments are negatively skewed, with the mark distribution for German containing a large number of candidates achieving maximum marks. This results in a lack of discrimination between the most able candidates.

10.3 Pearson

☐ The correlations between candidates' reading and writing marks at AS and A2 level are low. This suggests a potentially low level of marking reliability that is impacting on the rank order of candidates and, therefore, validity of the mark distribution.

☐ A scaling factor of less than one is applied to marks resulting from the translation tasks. This leads to an unnecessary reduction in the discrimination between candidates on this element of the written assessments.

☐ The raw mark distributions for the speaking assessments are negatively skewed, with truncated distributions at the top of the mark scale. This suggests that the discrimination between candidates at the top end of the ability is reduced.

10.4 WJEC

☐ The targeting of the combined listening, reading and (compulsory) writing sections of the written exams relative to the required standard is suboptimal, with a greater amount of information collected to differentiate between candidates at the lower-ability range (where there

are fewer candidates) than those of higher ability. However, this is extremely marginal for Spanish where the targeting of the exam appears to be broadly appropriate.

☐ The written assessments in French and Spanish are well targeted to the ability of candidates sitting, with candidates being spread across the mark distribution. This is less so the case for the German exam where there are a number of items with high facility indices meaning they offer little to the discrimination between candidates.

☐ Even when accounting for the relatively short mark scale, the relationship between AS and A2 marks for listening is weak for all languages. Given that the marking of these items is largely objective, this may suggest issues with item design in this area that require further investigation.

☐ Marks for the quality of response element of the writing task at AS level are doubled, as the mark scheme has a maximum of 10 marks yet the design is for this element to carry 20 marks. This approach does not, therefore, discriminate between candidates with the resolution that is likely possible when marking this task.

☐ The raw mark distributions for the speaking assessments are highly negatively skewed, with a large number of candidates achieving very high marks and large regions of the mark distribution being unused. This is likely impacting on the discrimination between the more able candidates on these assessments across all languages.

10.5 Wider findings

☐ All exam boards, to varying degrees, assess the content of the responses provided in the writing tasks in addition to the quality of the written response. This aspect is not reflected in the assessment objectives for the current specification. Given its inclusion in the Aims and Objectives of the current subject criteria and its proposed inclusion as an assessment objective in the reformed specifications, this is not viewed as compromising the validity of the assessments.

☐ In an attempt to prevent candidates from being rewarded for pre-prepared responses, a number of mark schemes articulate the manner in which these responses should be credited. These strategies represent a significant risk if the rationale to identifying a pre-prepared response is not clear and justified by evidence. Misidentification or misapplication of an approach would have a negative impact on the rank order of candidates and therefore the validity of the assessment.

Recommendations

10.6 Recommendations

Given these findings, summarised below are the recommendations from this report and the organisation to which those recommendations are relevant. Those marked with * should be considered by exam boards to be required actions that will be followed up by Ofqual

	Recommendation	Organis-ation	Justification
*1	The demand of the written assessments must be reviewed in line with the evidence presented in this report. It is strongly recommended that the demand be increased to facilitate more effective measurement of the abilities of the more able candidates. Exam boards must report to Ofqual their approach to addressing this for the assessments to be delivered from summer 2015, along with an action plan and rationale for their approach.	AQA OCR WJEC	Suboptimal targeting of assessment demand relative to the required standard. A high prevalence of items that are relatively low demand for the cohort. Systematic differences between intended and achieved weight of skills.
*2	Consideration must be given to how the assessments (and supporting processes such	AQA OCR	Raw mark distributions with high mean marks and negative skew in

		Pearson WJEC	addition to unused parts of the mark scale and truncation of the distribution for high-ability candidates.
	as standardisation and moderation) of spoken language can be better designed to address the issue of poor discrimination between candidates. It is not expected that spoken language assessments/arrangements are modified from summer 2015, however, opportunities must be sought to improve these assessments in the lifetime of the current specifications in addition to considering alternative approaches in the reformed specifications. Exam boards' reviews and action plans in relation to the current specifications will be followed up.		
3	Consideration must be given to how the assessments (and supporting processes) of spoken language can be better designed in the reformed specifications to improve, monitor and intervene in the quality of marking/consistency of marking standard.	AQA OCR Pearson WJEC	Low correlations suggesting low quality of marking and/or poor discrimination between candidates.
*4	The rationale for capping candidates' quality of language marks in the writing task based on marks achieved for content must be revisited and appropriate modifications to the approach made for the summer 2015 assessments.	AQA	Distorted item-level mark distributions and misapplication of marking rules affecting the rank order on invalid grounds.
*5	Further exploration of additional operational data and assessment/mark scheme design must be performed to understand the low correlation between writing marks, which suggest unsatisfactory item design or quality of marking.	AQA Pearson	Low writing intra-skill correlation.
*6	Further exploration of additional operational data and assessment/mark scheme design must be performed to understand the low correlation between listening marks, which suggests unsatisfactory item design or quality of marking.	WJEC	Low listening intra-skill correlation.
*7	The application of a scaling factor less than 1 to marks from the translation task should be revisited and alternative approaches sought in time for the 2015 assessments.	Pearson	Loss of discrimination through scaling factor.
*8	The approach to up-scaling quality of response marks (10 marks x 2) rather than applying a mark scheme with a sufficient length (20 marks) must be reviewed and addressed in time for the 2015 assessments.	WJEC	Potential loss of resolution in the mark scale.
9	The absence of cultural aspects of knowledge and understanding from the assessment objectives should be considered in the criteria	Ofqual/ ALCAB	Evidence that these elements are valued as relevant areas of understanding.

	for the reformed specifications as part of the on-going consultation process.		
*10	The principles underlying the design of the mark scheme and determination of what constitutes an acceptable response must be reviewed for the 2015 assessments and the principles clearly articulated. This will support transparency and future item development.	AQA OCR Pearson WJEC	AQA: expert review finding 2a. OCR: expert review findings 2c, 5a. Pearson: expert review finding 1a. WJEC: expert review findings 2b, 2d.
*11	The principles for defining and crediting pre-prepared responses and targeted lifts from resources must be clarified and articulated for the 2015 assessments reflecting on the findings of the expert reviewers.	OCR Pearson WJEC	OCR: expert review finding 2a. Pearson: expert review finding 2a. WJEC: expert review finding 2c.
*12	Exam boards must monitor the impact of making modifications to the assessments considered here using appropriate metrics as a basis for reporting to Ofqual. Processes should also be put in place for the on-going monitoring of assessment functioning/quality.	AQA OCR Pearson WJEC	Impact of any modifications is necessary for monitoring purposes. On-going good practice in assessment quality monitoring.
13	The principles and practice of handling word limits must be reviewed, clearly articulated and evidence based.	Pearson WJEC	Pearson: expert review findings 1b, 4a. WJEC: expert review finding 2a.
*14	The design of levels-of-response mark schemes must be reviewed including consideration of the comments of the expert reviewers to achieve consistent application of best practice across all languages/mark schemes/optional questions. This must be considered for the written assessments in time for the 2015 assessments.	AQA OCR Pearson WJEC	AQA: expert review findings 1a, 1b, 1c, 1d, 3c, 4c, 4e, 6a, 6b. OCR: expert review findings 1a, 1b, 2b, 3a, 4a, 6a. Pearson: expert review findings 3a, 3c, 3d, 6a. WJEC: expert review findings 1a, 3a.
*15	The comparability of the different optional routes through the assessment must be reviewed in light of the qualitative findings. This must be performed ready for the assessments to be delivered in summer 2015.	OCR	OCR: expert review findings 9a.

10.7 Implications of findings and recommendations for teaching and learning
Increasing the demand of the assessments in line with the recommendations outlined above will improve the validity of the rank order of candidates. There will likely be implications for teaching and learning and the perceptions of users, however, as no change to content or approach is being proposed, modification of what candidates are taught or how they are prepared for exams is not required. It is unlikely that the changes in demand required to effect an improvement in the validity of the assessments will be substantial. However, consideration should be given to how to provide support in these circumstances. While some of these recommendations may appear to have the potential to impact on the grades of candidates, awarding will account for any increase in demand, therefore protecting outcomes.