



Cynulliad Cenedlaethol Cymru The National Assembly for Wales

Y Pwyllgor Plant a Phobl Ifanc The Children and Young People Committee

**Dydd Mercher, 24 Hydref 2012
Wednesday, 24 October 2012**

Cynnwys Contents

Cyflwyniad, Ymddiheuriadau a Dirprwyon
Introductions, Apologies and Substitutions

Graddau TGAU Saesneg Iaith, Haf 2012
GCSE English Language Grades, Summer 2012

Cynnig o dan Reol Sefydlog Rhif 17.42 i Benderfynu Gwahardd y Cyhoedd o'r Cyfarfod
Motion under Standing Order No. 17.42 to Resolve to Exclude the Public from the Meeting

Cofnodir y trafodion hyn yn yr iaith y llefarwyd hwy ynndi yn y pwyllgor. Yn ogystal,
cynhwysir trawsgrifiad o'r cyfieithu ar y pryd.

These proceedings are reported in the language in which they were spoken in the committee.
In addition, a transcription of the simultaneous interpretation is included.

Aelodau'r pwyllgor yn bresennol Committee members in attendance

Angela Burns

Ceidwadwyr Cymreig
Welsh Conservatives

Christine Chapman

Llafur (Cadeirydd y Pwyllgor)
Labour (Committee Chair)

Jocelyn Davies	Plaid Cymru The Party of Wales
Suzy Davies	Ceidwadwyr Cymreig Welsh Conservatives
Rebecca Evans	Llafur Labour
Julie Morgan	Llafur Labour
Lynne Neagle	Llafur Labour
Jenny Rathbone	Llafur Labour
Aled Roberts	Democratiaid Rhyddfrydol Cymru Welsh Liberal Democrats
Simon Thomas	Plaid Cymru The Party of Wales

Eraill yn bresennol
Others in attendance

Gareth Pierce	Prif Weithredwr, CBAC Chief Executive, WJEC
Jo Richards	Pennaeth Ymchwil, CBAC Head of Research, WJEC
Glenys Stacey	Prif Reolydd a Chadeirydd, Swyddfa Rheoleiddio Cymwysterau ac Arholiadau Chief Regulator and Chair, Office of Qualifications and Examinations Regulation
Cath Jadhav	Cyfarwyddwr Dros Dro Safonau ac Ymchwil, Swyddfa Rheoleiddio Cymwysterau ac Arholiadau Acting Director of Standards and Research, Office of Qualifications and Examinations Regulation

Swyddogion Cynulliad Cenedlaethol Cymru yn bresennol
National Assembly for Wales officials in attendance

Claire Morris	Clerc Clerk
Kayleigh Driscoll	Dirprwy Glerc Deputy Clerk
Chloë Davies	Swyddog Cymorth Pwyllgorau Committee Support Officer
Anne Thomas	Y Gwasanaeth Ymchwil Research Service
Sian Thomas	Y Gwasanaeth Ymchwil Research Service
Liz Wilkinson	Clerc Deddfwriaeth Legislation Clerk
Kath Thomas	Dirprwy Glerc Deddfwriaeth Deputy Legislation Clerk

Dechreuodd y cyfarfod am 10.01 a.m.
The meeting began at 10.01 a.m.

Cyflwyniad, Ymddiheuriadau a Dirprwyon Introductions, Apologies and Substitutions

[1] **Christine Chapman:** Good morning, and welcome to the National Assembly for Wales's Children and Young People Committee. I remind Members who have any mobile phones or BlackBerrys on to switch them off. We have not received any apologies this morning.

10.02 a.m.

Graddau TGAU Saesneg Iaith, Haf 2012 GCSE English Language Grades, Summer 2012

[2] **Christine Chapman:** The purpose of today's session is to discuss the issues that were relevant to the grading of the summer 2012 English language GCSE exams. I welcome representatives from WJEC. First of all, I welcome Gareth Pierce, the chief executive, and Jo Richards, head of research. You are very welcome.

[3] You provided a paper in advance. It arrived yesterday, and we discussed the reasons for that earlier. I will go straight into questions, if you are okay with that, because the issue is a very complex one. I am sure that the questions will start to look at this in more detail.

[4] I want to start off with a general question. Could you give an overview of the process of determining GCSE examination standards, the role of an examining board in this process, and the role of the regulators?

[5] **Mr Pierce:** Yes, it is very much a process that involves a combination of the organisations that you mentioned. The awarding organisations take responsibility for ensuring that all the marking is done, that the data are available and that examples of candidates' work are available—a range of statistical information. An awarding committee then meets, takes account of the whole set of evidence and, essentially, decides, in the case of GCSE, on key grade boundaries, which are at grades A and C. There are then other boundaries that are set arithmetically.

[6] All that happens, then there is the sharing of information with regulators, who set the context for the statistical information that they want us to take into account. These days, as you know from our paper, things called 'predictor models' are a key feature of that landscape. We then meet with the regulators, having completed the full set of GCSE awards—and, typically, that would be very early in August—and we review with the regulators, in the company of other awarding organisations, any issues that people feel have arisen. With regard to maintaining standards, there are sometimes issues in individual subjects, and there are sometimes more generic issues that get discussed at those kinds of meetings. You will also know that regulators have considerable powers, very great powers in fact, to address issues about which they have concerns, including by issuing directions.

[7] **Christine Chapman:** Before we move on to other questions, Jenny Rathbone wanted to ask a specific question on this.

[8] **Jenny Rathbone:** Before we delve into the detail, I want to know what you mean by 'tiered' and 'untiered', in your paper.

[9] **Ms Richards:** GCSE English language is made up of four units. Two of the units are tiered, and that means that those two written papers have a higher tier and a foundation tier—

- [10] **Jenny Rathbone:** What does that mean, a higher tier?
- [11] **Ms Richards:** At the higher tier, you can be awarded grades A* to D, and, in some cases, an E grade. At the foundation tier, the grading system is from C to G. So, candidates who are more able would generally be entered for the higher tier and candidates who are less able would be entered for the foundation tier.
- [12] **Jocelyn Davies:** So, no matter how well they do in the exam, a candidate sitting the foundation tier paper could not get the grades that you would get at the higher tier level. Is that right?
- [13] **Ms Richards:** That is correct. It is tiered at unit level, not at subject level. They can also mix and match the tiers.
- [14] **Jenny Rathbone:** So, an individual candidate would not sit all four exams—written, written, written, and speaking and listening.
- [15] **Ms Richards:** An individual candidate sits four units. There are two written papers, which they can sit at the higher tier or the foundation tier level. Then, there are two controlled assessment units, one of which is a written controlled assessment and one of which is a speaking and listening unit, both of which are un-tiered. They sit all four units and they can choose—or rather their centre or school chooses—whether they sit at the higher tier or foundation tier level for the written papers. Everyone sits the other two un-tiered units.
- [16] **Jenny Rathbone:** Okay, so in the first two exams you can opt whether to go for the A grade, if you think the candidate can do it; otherwise, you go for the lower tier, where the highest grade you can possibly achieve is a C.
- [17] **Ms Richards:** On that unit, yes.
- [18] **Jenny Rathbone:** Yes, on that unit.
- [19] **Suzy Davies:** Can you explain a bit more about your relationship with the regulators—Ofqual, the Welsh Government and the Council for the Curriculum, Examinations and Assessment? Are those relationships different? More specifically, can you explain the different roles of the regulators and the exam boards in the standards and technical issues group meetings?
- [20] **Mr Pierce:** Perhaps I can address the general question and then Jo can comment on the standards and technical issues group, of which she is a member. Our relationship with the regulators spans everything to do with a qualification, from the origins of the qualification onwards. Whenever we, as WJEC, develop a specification, which is the word for ‘syllabus’ these days, it is in the context of regulatory criteria. For GCSEs and A-levels, this is essentially a collective exercise on the part of three regulators working together for England, Wales and Northern Ireland. They will have criteria for, say, GCSE mathematics or English language. Therefore, our work with them starts right from the beginning. We work on specifications that meet the regulators’ criteria. We also have a relationship with regulators on a whole range of things to do with ‘conditions of recognition’, as they are called, under which we operate. So, we deliver assessments, we work with our schools and colleges and centres according to recognition conditions, which are also almost entirely common across the three regulators.
- [21] When we come to awarding a qualification, the standards discussion also involves the three regulators at the table. The one difference that comes in is that, in the case of WJEC, we have some subjects for which our candidature is perhaps entirely or nearly entirely in Wales.

We have some subjects for which the majority of the candidates are in England, and we have very few subjects for which there are a substantial number of candidates in Northern Ireland. That difference has some influence on the extent to which those regulators will want to work in detail with us on the outcomes for particular qualifications. Perhaps Jo can say a little about the standards and technical issues group.

[22] **Suzy Davies:** May I come back on that first? Whatever the subject, if most of the pupils sitting it are from Wales, you will have a more detailed conversation with your Welsh regulator.

[23] **Mr Pierce:** Yes, that is natural. There are some subjects—and obvious examples would be Welsh first language, Welsh second language and Welsh literature—for which it would be unusual if Ofqual wanted to discuss matters of grade data with us to do with those awards, while it would be very natural for the Welsh regulator to wish to do so. There are some other subjects, such as mathematics, for which the vast majority of our candidature tends to be within Wales. We are doing some current work. For example, there is a scheme called the ‘linked pair’ in mathematics, which is a pilot scheme under which youngsters are able to take two GCSEs, essentially, in mathematics. That is running in Wales only, I think, and therefore it is quite natural that the Wales regulator would engage more with us on that, although it is actually a part of a bigger picture as well, on which Ofqual would also have a perspective.

[24] **Ms Richards:** The standards and technical issues group meets probably every six weeks. There is a meeting in the morning between the regulatory authorities, and then the regulatory authorities meet with the awarding bodies in the afternoon, and various issues to do with the regulation, maintaining standards and various rules et cetera are discussed at those meetings.

[25] **Suzy Davies:** So, it is not just the case that the exam boards do all the talking and the regulators just observe.

[26] **Ms Richards:** No.

[27] **Suzy Davies:** There is a conversation.

[28] **Ms Richards:** There is dialogue, yes.

[29] **Lynne Neagle:** Have you seen a change in the way in which regulators, particularly Ofqual, have worked this year compared with previous years, both in their general approach and specifically in the setting of grade boundaries?

[30] **Mr Pierce:** Over a period of time, there has been change. There are fairly well-understood reasons for that. For instance, a major topic of discussion has been the question of whether there has been undue grade inflation. That is possibly linked to the use of qualification outcomes in performance measures, so there is quite a complex interrelationship between what we do in awarding grades and what happens in performance measures. Therefore, understandably, the regulators have jointly introduced a new discipline, perhaps, of looking at ensuring that percentages do not keep creeping upwards, and awarding organisations have bought in to that and we have worked collectively on it. By and large, that works very well. The regulators are able to show that that has had an effect, and there has been a flattening out of some graphs that were previously climbing upwards.

[31] We have some concerns about aspects of the methodology. You will have picked up from our paper that we have some discomfort about prediction models in general, especially as we are an awarding organisation whose cohort of candidates is not a representative sample

of England plus Wales plus Northern Ireland as a whole. We know that, in England, our candidature is not fully representative of England, and we know that we have a large candidature in Wales in almost every subject. We in WJEC are not a representative mix of candidates, and therefore that gives us some nervousness about predictor models in general. Over and above that, we are very much a part of the new landscape of using statistical measures in a way that quite deliberately puts a ceiling on certain percentage outcomes.

[32] **Lynne Neagle:** On the shift that you describe, which you say has taken place over a few years, among the regulators, has there been any particular leader in that, or has it been a completely joint thing?

[33] **Mr Pierce:** In the cases that have been documented, it has been a scenario in which all regulators and all awarding bodies are signed up. That does not mean that Ofqual will not have taken the lead in some of the areas of work, including chairing some of the groups that have met to discuss these things, quite possibly. For example, the technical group that Jo referred to is chaired by Ofqual, and it would naturally have a greater presence in some of those meetings than would the other regulators.

[34] **Aled Roberts:** Hoffwn ddychwelyd at y grŵp safonau a materion technegol. Mae gennym dystiolaeth gan Ofqual sy'n sôn am drafodaethau rheolaidd, ac rydych wedi dweud ei fod yn cyfarfod bob chwe wythnos. Ar ôl hynny, mae'n adrodd yn ôl ar y cyfarfodydd o fis Mawrth ymlaen, ar 14 Mawrth. A oedd trafodaethau ynglŷn â chwrs Saesneg iaith TGAU cyn mis Mawrth o gwbl?

Aled Roberts: I would like to return to the standards and technical issues group. We have evidence from Ofqual that states that discussions take place regularly, and you say that it meets every six weeks. After that, it reports back on the meetings from March, on 14 March. Were there discussions about the English language GCSE course before March at all?

10.15 a.m.

[35] **Mr Pierce:** Rwy'n meddwl bod y trafodaethau ynglŷn â chwrs Saesneg iaith wedi digwydd yn gymharol hwyr yn y flwyddyn. Yn amlwg, roedd manyleb newydd yng Nghymru a Lloegr, felly un cwestiwn diddorol yw a ddylem ni, fel grŵp o bobl—cyrff dyfarnu a rheoleiddwyr—fod wedi sylweddoli ynghynt y byddai materion dyrys yn ein taro o ran dyfarnu graddau Saesneg iaith. Roedd gennym beth profiad o ddyfarnu ar gyfer Saesneg yn ôl y fanyleb newydd yn ôl ym Mehefin 2011 a hefyd yn Ionawr 2012. Mae'n ymddangos bod y dyfarnu hwnnw wedi gweithio'n gymharol hwylus. Roeddem ni yn CBAC wedi cytuno i ddefnyddio'r dull rhagfynegi cyfnod allweddol 2 o ran y Saesneg yn gymharol gynnar yn 2012, a'r ddealltwriaeth oedd y byddai hynny'n cael ei ddefnyddio fel mesur i adrodd yn ei erbyn. Dyna'r unig beth mewn gwirionedd roeddem ni wedi'i drafod a'i gytuno yn wahanol ar gyfer y Saesneg o'i gymharu â phynciau eraill.

Mr Pierce: I believe that the discussions on the English language course happened relatively late in the year. Obviously, there was new guidance in Wales and in England, so one interesting question is whether we, as a group of people—awarding bodies and regulators—should have realised earlier that serious issues would come into play in terms of awarding English language grades. We had some experience of awarding in English according to the new guidelines in June 2011 and also in January 2012. It appears that the awarding in those cases had worked relatively well. We in WJEC had agreed to use the key stage 2 predictive system in terms of English relatively early in 2012, and the understanding was that that would be used as a yardstick against which we would report. That is the only thing in reality that we had discussed and agreed that was different for English compared with other subjects.

- [36] **Aled Roberts:** A oes cofnodion o'r cyfarfodydd grŵp hyn? **Aled Roberts:** Are there any records of those group meetings?
- [37] **Mr Pierce:** Mae cyfarfodydd y grŵp technegol yn cael eu cofnodi, felly bydd cofnodion ar gael ar eu cyfer. Mae grŵp cyfatebol o'r enw STAG—*standards and technical advisory group*—sy'n cynnwys y cyrff dyfarnu yn unig. Mae cofnodion o gyfarfodydd y grŵp hwnnw, ac mae cofnodion o gyfarfodydd y STIG—*standards and technical issues group*. Mae Jo yn mynychu'r ddau grŵp yna. **Mr Pierce:** The technical group meetings are recorded, so there will be records available for them. There is a corresponding group called STAG—the standards and technical advisory group—which includes the awarding bodies only. There are minutes of that group's meetings, and there are minutes of the STIG—the standards and technical issues group. Jo attends both of those groups.
- [38] **Christine Chapman:** I want to move on, because Rebecca's question may touch a little on this.
- [39] **Rebecca Evans:** I want to focus on the factors that might have specifically impacted on the summer 2012 GCSE English language results. What affect did the new combined English GCSE have on the predictive models and subsequently on the results of the other English GCSEs in England and Wales?
- [40] **Mr Pierce:** Could I just check that I understood the question correctly? You are asking how the key stage 2 method related to English. Is that right?
- [41] **Rebecca Evans:** What affect did the new combined GCSE have on the predictive models that you used? That is my first question.
- [42] **Mr Pierce:** The predictive models exist independently of the new GCSE, in a sense, because the predictive model—and this is for England only, of course—takes the candidates' key stage 2 results and then predicts their aggregate outcomes, in whichever GCSE topic. We agreed, for the first time ever, to use that method in one subject only, or in one set of subjects, namely the English specifications, of which there are three in England—English language GCSE, English GCSE and English literature GCSE. So, we agreed to use the key stage 2 methodology for those three.
- [43] We have also agreed to look retrospectively at how that methodology would have affected other GCSEs of ours, where the majority of the candidates taking the exam through WJEC are in England. We are in the process of doing that and we hope to share that with regulators before the end of this week.
- [44] **Rebecca Evans:** Would you be able to share that with the committee as well?
- [45] **Mr Pierce:** That work is still subject to verification, so it is not appropriate for me to provide that to the committee now.
- [46] **Rebecca Evans:** Could you share it in due course?
- [47] **Mr Pierce:** Yes, of course.
- [48] **Rebecca Evans:** Great, thank you. Did a greater proportion of candidates sit the modular examinations in England than in Wales? If so, can the impact of that unitisation on predicting the summer grades in English GCSE be quantified?
- [49] **Mr Pierce:** This is an interesting aspect and perhaps Jo could expand on it. We think

that that is one aspect of how the two countries may have adjusted to the new specification. There is a view that centres in England have adapted in ways that are different to centres in Wales, and some of the evidence that we are just completing this week is to do with that—the extent to which centres in Wales took unitised options or earlier opportunities. Perhaps Jo could comment on that.

[50] **Ms Richards:** The proportion of centres in Wales that took the opportunity to use the modular or unitised scheme was a lot lower than in England, so the centres in England took advantage of the unitised scheme and, therefore, sat some units prior to the summer of 2012. We have done some work on looking at, had the Welsh centres used a similar entry strategy, what potentially would have happened to their results, and it would suggest that there would have been far less of a gap between England and Wales results had those entry strategies been used. We need to remember that, previously in Wales, there was only what we call the linear specification. So, in Wales—and in England, prior to 2012—you could only sit a specification where you sat all of the qualification at the end of the course, and for English, this was the first final session where you could sit it in a modular fashion. It certainly would appear that the England centres used certain entry strategies of the modular set up to their advantage.

[51] **Suzy Davies:** If Welsh schools had been more inclined to use the modular system, would it have been easier to spot problems with the predictive model earlier on?

[52] **Ms Richards:** We do not have the predictive model until the end of the course. The way that it works is that you do not make any predictions until you are sitting the qualification—what we term ‘cashing in’ the qualification. Summer 2012 was the first time the predictive model was used, because that was the first time you could cash in the qualification. Therefore, in any session prior to that where units had been sat, we did not have that model to look at, because none of the candidates were cashing in their qualification.

[53] **Suzy Davies:** You just collect the data and put them in a drawer, basically.

[54] **Ms Richard:** You are banking them, yes.

[55] **Julie Morgan:** Could you explain why Welsh schools did not do the same thing as in England?

[56] **Ms Richard:** I do not know why they did not. You wonder whether it was because, historically, a linear course had been followed. Maybe, in moving to the new specification, they followed it as they had the previous specification.

[57] **Julie Morgan:** So, it is just historical.

[58] **Mr Pierce:** It is curious that it has happened that way, because, of the two countries, it is Wales, in terms of policy, that is favouring the retention of the unitised opportunities, yet—

[59] **Julie Morgan:** It seems out of keeping, really, that Wales did not do it.

[60] **Mr Pierce:** We think that it is a significant factor. There is a Welsh-Government-led group, to which we are contributing, that is doing some field work, having conversations with some English departments in Wales to explore factors such as the one that we are discussing now. It is already emerging that that is an issue, because when an early assessment opportunity is used, it is not just the candidates that benefit, in my experience, but the teaching department as well, in terms of the feedback from candidates’ assessments. It is also a point that our awarding committee noted when it was reconvened in September under direction from the Welsh Government. It discussed, to some extent—this is included in the

note of that meeting—its perception of factors in England that are different to those in Wales. The two things mentioned were: first, that in England there was more evidence of a strategic approach to using those earlier assessment opportunities; and secondly, the amount of time given in many schools in England to this part of the curriculum—namely, things to do with English.

[61] **Aled Roberts:** Os ydych yn sôn am ryw fath o strategaeth yn Lloegr ynglŷn â phobl ifanc yn sefyll eu harholiadau'n gynharach, pwy sy'n gyfrifol am y strategaeth honno? A oes rhyw fath o gyfarwyddyd yn dod oddi wrth y Llywodraeth, awdurdodau lleol neu gonsortia ysgolion?

Aled Roberts: If you are talking about some sort of strategy in England regarding young people sitting their exams at an earlier stage, who is responsible for that strategy? Is there some sort of direction that comes from the Government, local authorities or schools consortia?

[62] **Mr Pierce:** Rwy'n credu mai penderfyniad ysgol neu goleg unigol fyddai'r strategaeth o ran asesiadau. Fodd bynnag, rwy'n meddwl bod y cyd-destun polisi ar gyfer y rhan hon o'r cwricwlwm wedi bod braidd yn wahanol yn Lloegr a Chymru dros nifer o flynyddoedd. Felly, wrth gyflwyno'r fanyleb newydd hon, roedd y cyd-destun braidd yn wahanol yn Lloegr, a oedd efallai yn meddwl am bethau gwahanol a phwyslais gwahanol. Efallai fod cynigion cynnar am asesiadau yn rhan o'r cyd-destun hwnnw.

Mr Pierce: I believe that the decision on a strategy for assessment is for individual schools or colleges. However, I think that the policy context for this part of the curriculum has been a little different in England and Wales over a number of years. Therefore, in introducing the new specification, the context may have been a little different in England, which was perhaps thinking of different issues and with a different emphasis. It is possible that early options for assessment were part of that context.

[63] **Jocelyn Davies:** I want to ask about controlled assessment. I know that you will be aware of the concerns around the consistency of controlled assessment, and I am sure that you would agree that consistency and comparability are crucial when you are talking about examinations. Do you have a view on its introduction, in view of those inconsistencies, and can you quantify the impact?

[64] **Mr Pierce:** We had concerns from the outset about the high weighting of 60% being given to controlled assessment in this particular subject. Obviously, that was a collective decision; WJEC's view back in 2008-09 was a minority view and it went ahead at 60%. Those assessments are done internally by the schools and colleges, and then our moderators look at sample evidence and decide whether the internal assessment is to be accepted or needs to be adjusted. When we do that, we allow a certain amount of tolerance. In retrospect, there is a question about the relationship between tolerance and assessment in this area, and going forward, we are definitely reviewing that, as are all the organisations.

[65] We were in the fortunate position in WJEC of never having assessed these controlled assessments until June 2012, so some of the vigorous debates that are happening elsewhere about whether those boundaries changed between January 2012 and June 2012 do not apply to us. However, that does not ease the problem, because, going forward, we know that we, somehow, need an understanding between us and schools and colleges about where the standards lie in controlled assessments. The only thing that we can do is inspect evidence and adjust if necessary. The whole scheme works so much for the better when there is a good understanding between schools and our examiners of what those standards are. We do our best to communicate those standards, but it is quite a challenge, especially when the weighting is this large.

[66] **Jocelyn Davies:** I want to ask you about the Welsh Government's conclusion in its

report about the characteristics of the summer 2012 cohort of students taking English language GCSE being similar to the 2011 cohort, and that it was considered stable. There was no difference that was significant enough to explain the lower grades; do you agree with that?

[67] **Mr Pierce:** Our view is that, for a number of years, the cohort for English language in Wales has been falling gradually, in line with the population of that age group. So, I do not think that there were any surprises about the cohort. The cohort in England, by the way, has varied rather more; it has not followed the population trend, and there has been some speculation as to why that might be the case. For example, did some candidates take the International GCSE, as it is called? There are various possibilities. However, in Wales, we are in a fairly stable situation in terms of the cohort.

[68] **Jocelyn Davies:** So, there was no significant difference that would explain the difference in results.

[69] **Mr Pierce:** No.

[70] **Jocelyn Davies:** I also want to ask about grade boundaries. Did you have concerns about the difference in grade boundaries for some units between January 2012 and June 2012?

[71] **Mr Pierce:** None that we were uncomfortable with. In any subject, at a unit level, examiners find that they have to debate the possibility of having a slightly different grade boundary to what existed previously.

[72] **Jocelyn Davies:** There was nothing unusual about this year compared with last year or the year before.

[73] **Mr Pierce:** No, nothing unusual.

[74] **Jocelyn Davies:** Are there changes being planned for 2014 onwards to improve the validity and reliability of the process?

10.30 a.m.

[75] **Mr Pierce:** A set of changes are being proposed in Wales, which we have agreed to take forward, in particular, to reduce the weighting of the controlled assessment from 60% to 40%. Alongside that, there will be some changes in the weighting given to elements like spelling, punctuation and grammar. This is being introduced this term. The assessments will be in 2014. We recognise that this is an element of change that was not expected, but on the other hand, as the awarding organisation that will be involved with those assessments in 2014, we are very confident that a sound assessment route will be available for those candidates.

[76] **Christine Chapman:** Before you move on, Jocelyn, there are supplementary questions from Suzy and Simon.

[77] **Suzy Davies:** Could you clarify for me who sits exams in January? Are they part of the unitised or modular system, or are they resits?

[78] **Ms Richards:** Available in January were our unit 1 and unit 2—our two written papers. You would have some year 11 candidates sitting those early to get some of the exams out of the way.

[79] **Suzy Davies:** To bank them, yes?

[80] **Ms Richards:** Yes.

[81] **Suzy Davies:** Okay; thank you. But are some of them resits as well?

[82] **Ms Richards:** Yes, some of them could also be resits.

[83] **Simon Thomas:** Yn eich ateb i Jocelyn Davies, dywedoch nad oedd gennych unrhyw bryderon am y gwahaniaeth rhwng ffiniau'r graddau rhwng mis Ionawr a mis Mehefin. Fodd bynnag, yn eich papur i'r pwyllgor, rydych yn dweud yn glir eich bod yn anesmwyth ynglŷn â'r hyn yr oedd Llywodraeth Cymru yn gofyn i chi ei wneud, sef ailedrych ar y ffiniau rhwng y graddau hynny. Os oeddech yn gysurus â hynny hyd at fis Mehefin, ac yn anesmwyth â'r hyn y gofynnodd y Llywodraeth i chi ei wneud yn yr haf, ym mis Awst, ble ydych chi erbyn hyn? A ydych chi'n gysurus â'r canlyniadau sydd wedi cael eu rhoi i ddisgyblion neu a ydych dal yn anesmwyth gyda'r broses?

Simon Thomas: In your answer to Jocelyn Davies, you said that you had no concerns about the difference between grade boundaries between January and June. However, in your paper to the committee, you say clearly that you were not comfortable with what the Welsh Government was asking you to do, which was to have another look at those grade boundaries. If you were comfortable with that up until June and not comfortable with what the Government was asking you to do in the summer, in August, where are you now? Are you comfortable with the results that have been given to pupils or are you still uncomfortable with the process?

[84] **Mr Pierce:** Mae dau beth a greodd anesmwythder i ni o ddiwedd Gorffennaf ymlaen. Yn ystod Gorffennaf, roedd ein pwyllgor dyfarnu wedi dyfarnu'r graddau ac roedd yn gyfforddus â'r hyn yr oedd wedi'i wneud. Nid oedd wedi bod yn hawdd—er enghraifft, roedd dod yn agos at ofynion rhagfynegydd cyfnod allweddol 2 yn dipyn o her, ac y mae dylanwad hynny wedi bod yn amlwg yn ei waith. Fodd bynnag, roedd yn gyfforddus ei fod wedi gallu dod i gasgliad.

Mr Pierce: There were two issues that caused discomfort to us from the end of July onwards. During July, our awarding committee had awarded the grades and it was comfortable with what it had done. It had not been easy—for example, coming close to the requirements of the key stage 2 predictor was quite a challenge, and the influence of that has been clear in its work. However, it was comfortable that it had been able to come to a conclusion.

[85] Fodd bynnag, yn dilyn hynny, digwyddodd dau beth. Yn gyntaf, gofynnodd y rheoleiddwyr ar y cyd i ni ailystyried, a digwyddodd hynny'n gynnar ym mis Awst. Roeddem yn anesmwyth ynglŷn â hynny oherwydd roeddem yn teimlo, ac yr ydym yn dal i deimlo, bod hynny wedi digwydd trwy roi gorbwyslais ar yr un rhagfynegydd hwn, sef rhagfynegydd cyfnod allweddol 2.

However, following that, two things happened. First, the joint regulators asked us to reconsider, and that happened early in August. We were uncomfortable with that because we felt, and still feel, that that happened by placing too much emphasis on this single predictor, namely the key stage 2 predictor.

[86] **Simon Thomas:** Ac yr oedd hynny'n dod o Loegr.

Simon Thomas: And that was from England.

[87] **Mr Pierce:** Daeth o'r rheoleiddwyr ar y cyd oherwydd ein bod yn gwybod eu bod wedi cwrdd ar y cyd, ond i ba raddau yr oeddent yn cytuno â'i gilydd, ni allaf ddweud wrthydh.

Mr Pierce: It came from the joint regulators because we know that they had met jointly, but to what extent they agreed with each other, I cannot tell you.

[88] **Simon Thomas:** Ond yr oedd y cais

Simon Thomas: But it was a joint request.

ar y cyd.

[89] **Mr Pierce:** Oedd. Gwnaethom ymateb i'r cais hwnnw drwy geisio amddiffyn y safiad yr oedd ein pwyllgor dyfarnu ni wedi ei wneud. Mae hyn yn yr ail atodiad i'r papur. Yn weddol faith, ceisïom roi'n dadleuon dros adael pethau fel ag yr oeddent. Fodd bynnag, ni dderbyniwyd ein dadleuon gan y rheoleiddwyr. Felly, daeth llythyr wedyn yn gofyn i ni addasu, ac rydym yn gwybod faint o bwerau sydd gan y rheoleiddwyr. Felly, yn y sefyllfa honno, gwnaethom addasu ac achosodd hynny elfen o anesmwythder.

[90] Daeth anesmwythder pellach ym mis Medi pan ddaeth cais gan Lywodraeth Cymru i ailraddio ar gyfer ymgeiswyr o Gymru yn unig. Gwnaeth hwnnw greu rhywfaint o anesmwythder oherwydd bod hynny'n golygu torri i ffwrdd o ddull y tri rheoleiddiwr a oedd wedi bod wrth gefn yr holl waith hwn. Dyna pam mai ein hymateb naturiol i Lywodraeth Cymru oedd dweud, 'Iawn, rydych yn codi'r cwestiwn, ond rydym yn credu y dylai hyn gael ei drafod gyda'r rheoleiddwyr ar y cyd oherwydd dyna'r cyd-destun ac felly ein dymuniad ni yw bod trafodaethau'n digwydd ar y cyd.' Fodd bynnag, fel y gwyddoch, nid hynny a ddigwyddodd ac mewn ffordd, cawsom orchymyn i ailraddio.

[91] **Simon Thomas:** Fel corff dyfarnu, a ydych chi'n credu, erbyn hyn, bod y disgyblion hynny wedi derbyn y graddau cywir—y graddau yr oeddech chi'n disgwyl iddynt eu cael—ar ôl yr holl broses?

[92] **Mr Pierce:** Y graddau roeddem ni'n disgwyl iddynt eu cael oedd y rhai a roddwyd gan ein pwyllgor dyfarnu ar ddiwedd mis Gorffennaf.

[93] **Simon Thomas:** Ac roeddent yn wahanol i'r hyn y maent wedi'u cael yn y pen draw.

[94] **Mr Pierce:** Oeddent. Mae anesmwythder ein pwyllgor dyfarnu yn amlwg yn y detholiad o ddyfyniadau rydym wedi eu rhoi i chi, ac y mae'n fwy amlwg fyth yn adroddiad y pwyllgor a gyhoeddwyd ym mis Medi. Y cwestiwn yw: a oedd

Mr Pierce: It was. We responded to that request by trying to defend the stance taken by our awarding committee. That is outlined in the second annex to the paper. We made a fairly lengthy case for trying to keep things as they were. However, our arguments were not accepted by the regulators. Therefore, a letter then came, asking us to adjust, and we know of the scale of the powers available to the regulators. Therefore, in light of that, we made changes and that caused an element of discomfort.

There was further discomfort in September when we were asked by the Welsh Government to regrade for candidates from Wales only. That caused us some discomfort because that meant breaking away from the three-regulator model that had been the foundation for all of this work. That is why our natural response to the Welsh Government was to say, 'Okay, you are raising the question, but we believe that this should be discussed with the joint regulators because that is the context and therefore our wish would be to see joint negotiations.' However, as you know, that did not happen and, in a way, we were ordered to regrade.

Simon Thomas: As an awarding body, do you now believe that those pupils received the correct grades—the grades that you expected them to get—after the whole process?

Mr Pierce: The grades we expected them to get were those awarded to them by our awarding committee at the end of July.

Simon Thomas: And they were different to what they were awarded subsequently.

Mr Pierce: Yes. The discomfort of our awarding committee is obvious in the selection of quotes that we have provided you, and is even more apparent in the committee report that was published in September. The question is: was the Welsh

Llywodraeth Cymru yn codi cwestiwn ynglŷn ag anfantais systemig a effeithiodd ar ymgeiswyr yng Nghymru'n unig? Mae hwnnw'n gwestiwn sy'n werth ei ofyn, ond nid wyf yn credu ei fod wedi ei ofyn yn y ffordd iawn, na'i ateb yn y ffordd iawn. A oedd unrhyw beth ynglŷn â'r gyfundrefn yng Nghymru a wnaeth anfanteisio ymgeiswyr yng Nghymru'n unig? Efallai yr oedd, ond nid yw wedi cael ei drafod. Yn Lloegr, er enghraifft, roedd manyleb o'r enw 'TGAU Saesneg' ar gael; yng Nghymru, dim ond 'TGAU Iaith Saesneg'. Felly, mae'r cyddestun polisi'n wahanol ac rydym wedi trafod rhai agweddau o hynny. Defnyddir yr egwyddor *comparable outcomes* ar gyfer y gwledydd gyda'i gilydd ac nid ar gyfer is-set o wledydd.

Government raising a question about a systemic disadvantage that affected candidates in Wales alone? That is a question that is worth asking, but I do not think that it was asked in the right way, or answered in the right way. Was there anything about the regime in Wales that disadvantaged candidates in Wales alone? Perhaps there was, but that has not been discussed. In England, for example, there was a specification called 'GCSE English' available; in Wales, there is only 'GCSE English Language'. So, the policy context was different and we have discussed aspects of that. The comparable outcomes principle is used for the nations together and not a sub-set of nations.

[95] Felly, mae themâu, sy'n systemig efallai, ond nid oes trafodaeth o rheiny wedi bod. Efallai y dylent gael eu trafod—

So, there are themes, which may be systemic, but there has been no discussion of those. Perhaps they should be discussed—

[96] **Simon Thomas:** Ond, doedd y rheiny ddim yn sail i'r ailraddio.

Simon Thomas: But, they were not the foundation for the regarding.

[97] **Mr Pierce:** Na. Cawsom orchymyn i wneud yr ailraddio ar lwybr ystadegol.

Mr Pierce: No. We were asked to do the regarding on a statistical basis.

[98] **Jocelyn Davies:** So, you had no concerns about the grade boundaries, but were uncomfortable with joint regulators' actions in July and Welsh Government action in September.

[99] I have one last question. You mentioned factors that have been identified that could have impacted on the summer 2012 English language provisional results: a strategic approach and curriculum time; that is identified in the Welsh Government report. You also mentioned adjusting to specification, which might be different in England and Wales and their entry strategies might be different. I am not an educationalist; I have no clue what an 'entry strategy' is—politicians only know about exit strategies. [*Laughter.*]

[100] Are there any factors, other than those you have mentioned or that we already know about, that could have had an impact? You say there is field work being done in terms of conversations to discuss factors; are there any others that we have not yet heard about that could come into play?

[101] **Mr Pierce:** I think those are the main ones.

[102] **Ms Richards:** One of the themes that has come out is the centre variation. Centres have seen—from 2011 to 2012—quite a variation. There has been a decrease of 30% in their outcomes. Interestingly, one of the things we looked at was whether this was happening in other subjects when there is a change of specification, and it does happen. So, there has, previously, been this centre variation, although not in English and English language, due to the fact that it had not changed. That is something else that we have looked at.

[103] **Jocelyn Davies:** So, any particular centre could have a drop in results of up to 30%

when you change the specification of the examination, despite other factors remaining the same: same centre, teacher and cohort, but the results could be 30% different.

[104] **Ms Richards:** Centres have seen that sort of change and some centres will have improved.

[105] **Mr Pierce:** There are some things that we looked at but could pretty much eliminate as factors. For example, we wondered if absence would be a factor, because controlled assessment required youngsters to be there during the interval when the controlled assessment was being done, whereas coursework was a much wider concept. We wondered if absence in schools in Wales had had an impact on controlled assessment, greater than anything previously, but we do not think that is a factor.

[106] We also considered that some pupils in Wales are not entered for the English literature exam. The policy in Wales is that youngsters should follow an English language and an English literature curriculum, therefore we wondered if it would help youngsters if they were excused from the English literature exam to concentrate on English language. Again, there is no evidence of that. So, there are some potential factors to do with the assessment load that is being placed on youngsters or absences, but, so far, we do not see those coming through as factors at all.

[107] **Christine Chapman:** We have less than 20 minutes left and we want to cover as many aspects of this issue as possible; so, I remind Members to be as concise as possible. I call on Angela.

[108] **Angela Burns:** Thank you very much for your paper. Again, I will just go back to the methodology to maintain standards. When you make a prediction based on key stage 2, is that based on an individual?

[109] **Ms Richards:** It is done at candidate level and then aggregated.

[110] **Angela Burns:** So, why is that deemed by one set of regulators to be a better methodology than another? I would have thought that if you had a candidate at key stage 2, you would be able to define what you felt that they were going to achieve. I am kind of confused by this common centre. I note that a common centre is a place where you have a cohort that is based on two years, but you can see, going through any school, that you can have a sort of aberration of cohorts where, suddenly, a school that is consistently at this level will either suddenly jump or go down, and you can follow that particular year group all the way through and they have the same result. I am just trying to understand why go for one method rather than another.

[111] **Mr Pierce:** Our view is that when you bring a range of evidence of these kinds together, each of them contributes parts of the picture. Each of them has its advantages and disadvantages. The common centres aspect, for example, will be based upon putting a lot of common centres together. We would agree with you: if you just take one centre that is common in 2011 and 2012 within our entry, you could get variation for all kinds of reasons. However, if you put 100 common centres together, which were there in 2011 and 2012, you are talking about a sizeable group of candidates. Therefore, that is an interesting indicator of stability or otherwise in the awards. Of course, it is in adjacent years. The issue is to do with any predictor model, especially ones that span many years in the education system. It is all to do with the assumption of added value being uniform for any sub-set of the candidature. That is the bit that I start getting a bit uncomfortable about. Maybe key stage 2 can be shown to correlate with GCSEs in a general kind of way, but is that value added relationship sufficiently robust to use within sub-sets of the candidates that are quite different from each other? We know that even the advocates of predictor models of that kind acknowledge that it

works so poorly for some sub-sets that those sub-sets have to be left out of the model. Our view is that there are probably three or four strands of statistical evidence that are all very important, but are much better when they are looked at together.

[112] Another important piece of work that we do as awarding organisations collectively is what we call statistical screening. We are doing that now for last summer's results. It means that we compare GCSE English with other GCSEs done by the same candidates in the same year. Perhaps Jo could expand a little bit on that because that is a very important piece of evidence.

[113] **Angela Burns:** I wish to ask another question because I am conscious of the time. I think that I follow what you are saying. All things being equal and you get to rule the world, and given the comments that you made to the House of Commons Education Committee on 15 December 2011, where you expressed the view that,

[114] 'standards in any nation's education system need to be based on quality of achievement, the quality of candidates' work, so I would caution against being too comfortable when our standards debate is couched statistically, especially in an international context.'

[115] Where do you think that we should go from here? As awarding people, what would you see as the ideal situation? I also note that you had agreed with Ofqual that you would use both key stage 2 and common centres, which I would have thought would be relatively confusing, but then you might tell me that it was extremely helpful.

[116] **Mr Pierce:** To answer that part first, we would see it as helpful because different types of data bring different perspectives on what is quite a complex exercise of comparing standards from year to year and across awarding organisations. On your first point, I think that several people who work with predictor models would acknowledge that they have a limited short-term purpose.

10.45 a.m.

[117] That is probably to do with stabilising things in the current environment of qualifications being used as performance measures and all kinds of inappropriate incentivisation or whatever. I genuinely feel, going forward, that what is important for any nation is for it to understand what its learning outcomes are for different age groups. Therefore, our assessments should tell us where we are in relation to those learning outcomes. Youngsters need to know where they are in terms of the key learning outcomes from which they were supposed to benefit from a course of study, and employers and higher education also need to know that.

[118] So, the debate is currently happening in England and in Wales, although with different policy backgrounds to that. A qualifications reform is about to happen in both countries, in one way or another, and, in that debate, understanding standards in relation to learning outcomes is one of the things that we should be aiming for. At the same time, the performance measures culture for schools and colleges in both countries should perhaps change as well. That would be a healthy—it would not be a complete divergence, but at least it would be a departure from using the same thing for both purposes. Going back to the point that you made in relation to the quotation of what we said, if Wales needs to make sure that it has young people who are able to contribute economically and socially, in a world that is changing rapidly and that is becoming more competitive, we need to know where we are compared with key learning outcomes.

[119] **Lynne Neagle:** Do you have any comment to make on how the key stage 2 predictor

model worked in other subjects? Did awarding bodies have to make adjustments in other subjects?

[120] **Mr Pierce:** Our experience is limited, because, so far, we have only used it in the English subject. We are doing a retrospective piece of work, which I mentioned earlier, which we will be sharing with regulators—we are happy to share it with you as well. That piece of work will show how much of an adjustment we would have had to make if we had been using key stage 2. The other awarding organisations use key stage 2 more widely as a predictor for GCSEs and, to some extent, we will be aware of the extent to which that causes adjustments from the meetings that we attend.

[121] **Christine Chapman:** We have 10 minutes left before Ofqual comes in. Jenny is next.

[122] **Jenny Rathbone:** You mentioned in your earlier remarks that Ofqual, the English regulator, was concerned about grade inflation. We know from Government statements that there were discussions between Ofqual and the Welsh Government about the way the predictors would be applied to WJEC. However, you, I think, told us that you had no discussions on this until June of this year. Is that right?

[123] **Mr Pierce:** Discussions about grade inflation more generally have been a more continuous process.

[124] **Jenny Rathbone:** So, you were already having discussions with the regulators and the other exam boards about that?

[125] **Mr Pierce:** Yes. I would say that, over a period of two or three years, there has been a discussion with regulators collectively and awarding bodies collectively about the whole question of grade inflation.

[126] **Jenny Rathbone:** Okay, but on the specifics of the summer 2012 GCSE grades, that discussion did not start happening until June, did it?

[127] **Mr Pierce:** The detail for GCSE English certainly did not even materialise until July. We were signed up to using the key stage 2 predictor earlier than that; Jo and her colleagues would have done the preparatory work well before June. However, nothing specific changed in that time period. We have been involved in a more medium-term discussion of grade inflation, and it is pretty clear why there was a need for that, because, for example, in England, between 2007 and 2011, the percentage of A to C grades in GCSEs increased by 7 percentage points. In Wales, where WJEC is the dominant awarding body, it increased by a lot less than that, by about 4 percentage points. So, there clearly was an issue in both countries, but it was a bigger issue in England, where WJEC is a smaller player, and it was a lesser issue in Wales, where WJEC is a major player. You can draw your own conclusions from that.

[128] **Jenny Rathbone:** That would have nothing to do with investment in education and schools and focusing on the quality of teaching. We would hope that that investment and that focus would improve outcomes for students and, therefore, more students would pass the exam.

[129] **Mr Pierce:** Yes, I agree. The point you are making is a fundamental one, is it not?

[130] **Jenny Rathbone:** It is.

[131] **Mr Pierce:** What has been agreed between regulators and awarding bodies and, we think, with Government support in both countries—I guess you can ask about that in one of

your next sessions on this theme—is, we understand, that the priority at the moment was to peg the kind of grade inflation that I have described. However, that goes against allowing space for improved outcomes. That is why, at the end of the day, any nation has to pitch the debate in terms of learning outcomes, because then we know where we are. Otherwise, all you know about, say, our GCSE mathematics is that—wherever it was in Wales—55.5% got a grade C and that is slightly lower than last year. That is all you know, unless I am able to convey to you what that means in learning outcomes. How satisfactory are those learning outcomes for the 55.5% who got grade C in GCSE mathematics? That might be great as far as the regulators are concerned, and as far as the rules that we have agreed with the regulators are concerned, but how great is it in terms of the real learning outcomes? We are not addressing that question at the moment and I think it is your question that points us towards the need to look at those issues.

[132] **Jenny Rathbone:** Indeed. We will, I am sure, come back to that.

[133] **Christine Chapman:** We need to move on now, Jenny, because we have eight minutes left. Simon is first.

[134] **Simon Thomas:** Rwyf eisiau eglurder ar rai cwestiynau sydd dal heb eu hateb. **Simon Thomas:** I want clarity on some questions that have still not been answered.

[135] Rydym wedi cael tystiolaeth gan y Swyddfa Rheoleiddio Cymwysterau ac Arholiadau sy'n sôn am y cyfarfod ar 14 Mawrth eleni, lle y penderfynwyd defnyddio rhagfynegyddion cyfnod allweddol 2. A allwch chi gadarnhau bod rheoleidiwr Cymru, sef Llywodraeth Cymru, yn bresennol yn y cyfarfod hwnnw ac yn cydsynio gyda'r penderfyniad hwnnw? We have had evidence from Ofqual about the meeting on 14 March this year, when the decision was made to use key stage 2 predictors. Could you confirm that the Welsh regulator, namely the Welsh Government, was present at that meeting and agreed with that decision?

[136] **Mr Pierce:** Bydd yn rhaid imi wirio'r cofnodion i gadarnhau hynny. Yn sicr, rydym wedi gweld ar bapur ei fod yn cytuno gyda'r egwyddor y dylem— **Mr Pierce:** I will have to check the minutes to confirm that. Certainly, we have seen on paper that they agreed with the principle that we should—

[137] **Simon Thomas:** Mae Ofqual yn dweud bod e-bost yn dilyn y cyfarfod— **Simon Thomas:** Ofqual said there was an e-mail following the meeting—

[138] **Mr Pierce:** Oedd, siŵr o fod. Rwy'n siŵr fy mod wedi gweld tystiolaeth ar bapur bod cytundeb rhwng y ddau reoleidiwr ein bod yn mynd i wneud hynny. **Mr Pierce:** Yes, I expect so. I am sure I have seen written evidence that there was agreement between the two regulators that we were going to do that.

[139] **Simon Thomas:** A oeddech chi yn y cyfarfod hwnnw o STIG? **Simon Thomas:** Did you attend that STIG meeting?

[140] **Mr Pierce:** Nac oeddwn. Byddai Jo efallai wedi bod yno. **Mr Pierce:** No. Jo might have been there.

[141] **Simon Thomas:** Roedd Jo yn y cyfarfod, oedd hi? Jo, a wnaethoch chi fynegi eich amheuan yn y cyfarfod hwnnw ynglŷn â defnyddio cyfnod allweddol 2, sy'n wahanol **Simon Thomas:** Jo attended, did she? Jo, did you express your doubts at that meeting about using key stage 2, which is different in Wales and England, as the predictor for the

yng Nghymru a Lloegr, fel rhagfynegydd ar GCSE results output in Wales?
gyfer allbwn canlyniadau TGAU Cymru?

[142] **Ms Richards:** I was not at that meeting, as I was still on maternity leave, but my colleague—

[143] **Simon Thomas:** Somebody would have been there.

[144] **Ms Richards:** My colleague, Raymond Tong, was there and he had expressed— orally and in a number of e-mails—some concerns.

[145] **Simon Thomas:** Did the Welsh Government officials express similar concerns?

[146] **Ms Richards:** Again, we would have to look at the minutes.

[147] **Simon Thomas:** Did they express them to you?

[148] **Ms Richards:** They did not express them personally to me, as I was not at the meeting.

[149] **Simon Thomas:** No, but to WJEC? Did they share your concerns?

[150] **Christine Chapman:** Perhaps we can have a note on that.

[151] **Mr Pierce:** The extent to which there were concerns was dwarfed by what happened later. I will make two points, if I may. These predictor models are described as being used for reporting purposes. That means that, if we are outside tolerance, it is clearly flagged up to regulators and they know about it and therefore we would expect a discussion about it and, possibly, we could defend our stance, as we tried to do. The other point is that I think the issue of a disparity between Wales outcomes and England outcomes is a bigger disparity and therefore a bigger problem than anything to do with the use of key stage 2, to be honest. Even if we had not had the use of key stage 2, we would still have had great difficulty balancing our outcomes for England and for Wales, because they are so different.

[152] **Christine Chapman:** Simon, do you have any further questions? If not, Lynne is next.

[153] **Lynne Neagle:** Am I right in thinking that AQA is the only awarding body on the Ofqual standards board? What sort of influence do you think it had on using the key stage 2 predictor model, and do you have any comments on that?

[154] **Mr Pierce:** The Ofqual standards board—I am pretty sure, from memory, that there is a member of AQA staff on it. I think that there is also someone on it from Cambridge Assessment, which is a wider umbrella group and possibly someone from Pearson. However, from memory, in terms of a named awarding body that is operating GCSEs, it may be the case that AQA is the only one. We are certainly not on that group.

[155] **Lynne Neagle:** To what extent do you think that drove the emphasis on the key stage 2 predictor model, and how comfortable are you with that?

[156] **Mr Pierce:** I am sure that the key stage 2 predictors, and others, would be debated by groups of that kind. Our hope would be that there would be a balanced set of expertise around the table, with people expert enough to be able to discuss the advantages and disadvantages of different methods and the different kinds of statistics that we have talked about. For example, I would assume that people around that table would know about the value of such things as

the screening data, which are retrospective, the common centres approach and the predictor model. Our discomfort is the seeming emergence of an emphasis on just one indicator. It is statistically healthier, as well as educationally more valid, to be exploring a range of statistical evidence. It may be that, to go forward with a three-country approach, regulators might find that they have to accommodate that range of indicators, because they are not all available in each country, for one thing.

[157] One thing that we perhaps have not had a chance to say is that we as WJEC, in the interests of candidates in Wales, value the three-country approach. A-levels and GCSEs, which are with us until 2016, are based on a three-country approach. It is a three-country set of standards. So, we are working very hard in WJEC, including through Jo's colleagues, to do all we can to sustain that three-country approach to standards, because that is what will give the currency to the certificate for young people in Wales.

[158] **Christine Chapman:** We are getting really short of time, but I want to allow Aled one final question. Could you try to condense some of your points, Aled?

[159] **Aled Roberts:** Rwyf eisiau sôn am y llythyr i Ofqual sy'n ddyddiedig 9 Awst, lle mae tri opsiwn yn cael eu crybwyll. Nid oedd y rheoleiddwyr yn barod i dderbyn canlyniadau adeg dyfarnu, fel roeddech yn ceisio ei wneud. A oedd y ddau reoleiddiwr o blaid opsiwn 1?

Aled Roberts: I would like to talk about the letter to Ofqual dated 9 August, in which three different options were mentioned. The regulators were not content to accept award outcomes, as you were trying to do. Were both regulators in favour of option 1?

[160] **Mr Pierce:** Dyna yw ein dealltwriaeth. Opsiwn 1, wrth gwrs, oedd y lleiaf niweidiol i'r canlyniadau yn y ddwy wlad. Roedd opsiwn 1 yn gostwng y canran a fyddai wedi cael A* i C yn y ddwy wlad. Yn naturiol, os oes rhaid iddynt fynd am unrhyw opsiwn, byddent yn cytuno ar hwnnw.

Mr Pierce: That is our understanding. Option 1, of course, was the least damaging to the results in the two countries. Option 1 reduced the percentage that would have achieved A* to C in both countries. Naturally, if they were going to have to plump for any option, they would have agreed on that one.

[161] **Aled Roberts:** Yn fyr iawn, roeddech yn esbonio nad oedd yn bosibl ichi anfon y papurau i ni oherwydd bod trafodaethau yn mynd rhagddynt ddydd Gwener diwethaf. Beth yw statws y trafodaethau hynny ar hyn o bryd?

Aled Roberts: Briefly, you explained that it was not possible for you to send the papers to us because discussions were ongoing last Friday. What is the status of those discussions at the moment?

[162] **Mr Pierce:** Mae nifer o bethau yn dal i gael eu trafod. Yr unig beth pendant sydd wedi cael ei benderfynu yw y bydd manyleb newydd yng Nghymru yn 2014, ac mae disgyblion blwyddyn 10 eisoes yn gweithio ar hwnnw. Y pethau sy'n dal i fod dan drafodaeth yw pethau ynglŷn ag Ionawr 2013. Mae cwestiynau ynglŷn ag asesiadau ac asesu yn Ionawr 2013. Mae adroddiad sydd bron wedi ei derfynu gan Ofqual sy'n edrych yn ôl ar haf 2012 ac yn tynnu gwersi allan o'r profiad hwnnw ar gyfer y dyfodol. Felly, bydd hwnnw'n adroddiad pwysig iawn i ni i gyd fel cyrff dyfarnu, yng Nghymru ac

Mr Pierce: A number of things are still being discussed. The only thing that has been decided for certain is that there is to be a new specification in Wales from 2014, and year 10 students are already working to that. The things that are still being discussed are issues relating to January 2013. There are questions about assessments and assessment in January 2013. A report by Ofqual has nearly been finalised. This report reviews the summer of 2012 and draws out lessons for the future from the experiences there. That will be a very important report for all of us as awarding bodies, in Wales and in England.

yn Lloegr.

[163] **Christine Chapman:** Julie has one final question.

[164] **Julie Morgan:** What do you think these events have done to the reputation of GCSE English?

[165] **Mr Pierce:** Overall, because we have had such an intense debate on a range of issues to do with GCSE English, I hope that the next outcome will be a very positive one for this curriculum area. Obviously, it has gone through a very rough time over the last three or four months, but I think that, for the subject area, there are important messages coming out of that turbulence that can only strengthen both the qualifications and, we think, the learning experience for youngsters following GCSEs.

[166] **Christine Chapman:** I thank Gareth Pierce and Jo Richards for attending this morning. A transcript of the meeting will be sent to you to check for factual accuracy. Thank you once again for answering Members' questions.

11.00 a.m.

[167] I now invite our next witnesses to come to the table. I welcome the representatives from Ofqual. Could you please introduce yourselves for the record?

[168] **Ms Stacey:** Yes, of course. My name is Glenys Stacey. I am the chief regulator and chief executive of Ofqual.

[169] **Ms Jadhav:** Hello, I am Cath Jadhav, the acting director of standards and research for Ofqual.

[170] **Christine Chapman:** Thank you very much. We have read your paper and Members have questions for you. I would like to start with a very broad question. Can you give us an overview of the general approach to and process for determining GCSE examination standards, including the roles of the regulators and the examining boards?

[171] **Ms Stacey:** Yes, of course, I will do my best. GCSE qualifications are designed by exam boards. They need to meet qualification criteria and subject criteria that are agreed and published jointly by the regulators. Exam boards design their qualifications and put them to the regulators for approval. The regulators approve them and then they are taught in schools. Exam boards award those qualifications. Now that the qualifications are modular, the awarding happens unit by unit and then comes together for a whole-qualification award at the end. We look at the emerging results from that awarding.

[172] We have common predictor models that are used to evaluate the awards and see how they line up, if you like. The regulators are looking at how they line up against predictions but also how they line up one exam board against another. There are tolerances set, and we are looking to see to what extent these awards are within tolerance. We have a meeting with all of the exam boards to review the whole picture across all GCSEs as soon as we have the preliminary data. That is normally in early August. We discuss with exam boards where anything looks out of line; we try to see whether there is a rationale for that, and sometimes there is. Where there is not, as regulators, we will challenge and ask whether the award can be justified and consider whether it should be reviewed. Through that process, we come out with the final awards for the qualifications.

[173] **Christine Chapman:** Is the level of involvement of WJEC in the process of determining standards different to that of the English examining boards?

[174] **Ms Stacey:** Well, it is certainly invited to the meetings we host as regulators, and its representatives generally come. It is as engaged as others in the process. I would not say that there is a material difference.

[175] **Lynne Neagle:** I want to ask about the standards advisory board and pick up on an issue I raised with WJEC. WJEC was not sure which awarding bodies sit on that board. Can you clarify that for the committee?

[176] **Ms Stacey:** Yes. Perhaps it would help if I clarified its role as well. I think that you are talking about our standards advisory group. It is set up as a committee of the Ofqual board. We set it up last year. You will know that Ofqual is quite a new regulator—we are just a couple of years old. We were looking to really strengthen the understanding of standards, and we wanted an advisory group that pulled together all the best advice and expertise on standards. We have selected 12 or 15 people for that group, including Professor Jo-Anne Baird from Oxford, who is also, I think, an adviser to Pearson or Edexcel. That, however, is not why we have got her; we have got her because she is an expert on assessment. We have academics and people from some examination boards, as it happens, but not many. So, around the table, we have Michelle Meadows, who is head of research at AQA, because of her understanding of research and assessment. We have also recently agreed to and now have Tim Oates as a member. He is a well-known expert on assessment and currently works for Cambridge Assessment Group, which is a mother body of OCR. So, there are a couple of people who are from examination boards or have a relationship with them, but that is not why they are there. It is not a representative body of examination boards; its purpose is to bring together into one place the best possible collection of expertise on standards, and that is what it does. We put to it matters on which we think it can give us good advice and real information about various aspects of standards.

[177] **Lynne Neagle:** So, the person who is on the board from AQA does not represent it; you are saying that it is incidental.

[178] **Ms Stacey:** Absolutely.

[179] **Lynne Neagle:** Can you tell us more generally about your relationship with the Welsh Government and how closely you work with it and with the equivalent body in Northern Ireland?

[180] **Ms Stacey:** We work with the Welsh Government and its equivalent in Northern Ireland as regulators. So, we are not in any way involved in policy or any of those many other things that happen in the Welsh Government. We work together as three countries—we call it three-country regulation. That is a long-established relationship. So, we can look back and see that, for these qualifications, the design rules were agreed among the three regulators five or six years ago. They were translated into qualifications criteria that were agreed by the three regulators at the time, and each qualification has been accredited by each of the three. So, it is a common approach that we have developed.

[181] It is true to say that Ofqual is often the body that chairs meetings and so on. That is because we have by far the biggest resource. We are 160 people, and three of the five exam boards that are in the A-level and GCSE business—the three that are, by far, the biggest, as it happens—are based in England as well. So, there is a recognition in the way that things work that we have more resource, but the big decisions are made jointly and we work hard at that.

[182] **Jenny Rathbone:** To move to what happened with the summer 2012 English language GCSE results, to what extent can you quantify the following factors, which may

have affected predictions methodology: the new combined English GCSE, which was introduced in England only; and the introduction of controlled assessments and variation between centres that introduced these new assessments?

[183] **Ms Stacey:** One of the frustrating things about trying to get to the root of the variations experienced by some schools—variations against what they were expecting and variations, in some cases, against their results last year—is that it is not easily possible to quantify the precise impact of one or other of a number of changes that happened at the same time. You point to controlled assessment, and that is a significant feature. It is different to coursework, which was the way that students' work in schools was done under the old arrangements and it is a significantly bigger proportion in the qualifications this time than it was in the legacy or the old qualifications. So, that is certainly important. We think that the change from two qualifications, two long-established, well-understood, qualifications, namely English language and English literature, to three qualifications, namely English, English language and English literature and the change in that is more significant than, perhaps, has been commonly expressed in the coverage you have seen of the issue so far in the media. So, that is definitely an issue.

[184] You are quite right to point out that these things, for policy reasons, played out differently in England and in Wales. In England, for example, you can take either English or English language and English literature. Those are the two choices. It is commonly assumed that the more able students in England will take English language and English literature, and others will take straight English. In Wales, you can do English literature, you must do English language, and you cannot do English. So, equally, one might assume that the more able students will do English literature as well as English language. Others will just do English language. So, these are assumptions one might make. Certainly, they would be factors that people would have in mind now in reviewing the outcomes. However, it is not possible to quantify that.

[185] **Jenny Rathbone:** The reason that we decided not to go for the combined English GCSE in Wales is because it was thought that we wanted to maintain standards. So, it was about ensuring that people took the English language option, with the more able invited to take English literature as well. However, clearly, maintaining a decent standard in the use of English language is core to the approach that we want to take in Wales. If, in England, they went for what might be an easier option, how would that skew the overall assessment of these modulations of grading in the context of your concern about grade inflation?

[186] **Ms Stacey:** I am not sure that it is right to say that they have gone for an easy option in England, because, in England, one might argue that the more able candidates, as I said, would have gone for English language and English literature. In Wales, candidates must do English language and can do English literature. So, I am not sure that I quite follow. However, one of the fundamental difficulties when we look at achieving a common standard in the two countries is that 65%, I think—I will confirm that figure—of the candidature taking the WJEC qualification were based in England. So, as regulators, when you look at preliminary outcomes from the WJEC qualifications, with that proportion of students based in English schools, and a requirement to make sure that their results compare fairly with the results of other children in other English schools who have taken the qualifications with another exam board—often with a bigger exam board—you have an inherent tension if the results from Welsh students taking WJEC papers are materially different to the results of English students taking those papers. That was what happened this year. So, it is a really tricky position for the three regulators to understand and get to the right standard.

[187] **Jenny Rathbone:** Into this mix we also have the fact that a much greater proportion of candidates in England sat modular exams. There seems to be some concern that by being able to do exams in bitesize chunks, it may enable the student to get the desired outcome,

rather than having to do it all in one hit.

11.15 a.m.

[188] **Ms Stacey:** Again, I am not sure that it is right to say that many more students in England than Wales took one route or another. We are doing a lot of work at the moment to understand the possible influence that different choices about the way through the qualification played out, not just for WJEC, but for the other exam boards as well. There is a view that the ‘route effect’, as it is called, has its part to play in this, but our analysis so far does not suggest that it has as significant an effect as some of the other issues that we can see here. The fact is that you have a modular qualification, whether it is from WJEC or any other board, and schools can choose any which way to go through that. Indeed, as we have identified, there is an increasing tendency for schools to put people through some of these units very early indeed, sometimes even in year 9. That makes it extremely complex not just to award, but to separate out the different route choices and how they impact on the final result. I know that it is not very helpful to you, saying that it is clearly an issue, but it is an issue, just not clearly ‘the’ issue.

[189] **Jenny Rathbone:** We can agree that this is an art rather than a science. My concern is that there has been what some might describe as ‘grade deflation’ going on with WJEC not only in English, but also, I am aware, in maths results. There has been a massive decrease against what was predicted for students with other exam boards, such as Edexcel. Therefore, I am concerned that, overall, there is this notion that we must get the grades down, when the investment in education over the past 15 years might, hopefully, ensure that more students are making the grade. Instead, we seem to be setting the ceiling so that only the same proportion of students will ever get there.

[190] **Ms Stacey:** We need to see the screening data—the after-the-event data—from exam boards to understand whether results actually have gone down or up, and we are expecting those data any day now. What we have so far from WJEC is that, even after the interventions by the three regulators this summer, English language results were still out of tolerance, above predictions, by 3.6% overall, and English was 2.2% above predictions, when the tolerance is 1%. Bear in mind that all those students of GCSE English would have been based in England, so we, as regulators, had legitimate concerns and a proper interest in asking WJEC to explain its provisional results. You will know that, as regulators, we did give WJEC the opportunity to do that, and we had a good exchange with it. It put a number of options to us and, as regulators, we agreed the option that made the least difference to the preliminary results. Indeed, WJEC made it plain to us that it was content to agree it, and did so in writing. So, I would say that what we have done is the right thing in trying to get to the right standard.

[191] What you do see here, however, is that, as policies in the different countries diverge—and having a different approach to the way in which you can combine these subjects is one element of that—and as it gets easier for different routes to happen and so on, it puts a great strain on the arrangements that you can make for getting to a common standard in the qualifications across the two countries.

[192] **Jenny Rathbone:** Would you agree that those who sat the exam in the summer were assessed to a different standard from those who sat the exam in January?

[193] **Ms Stacey:** No, I would not.

[194] **Jenny Rathbone:** Okay. This is what we are here to discuss.

[195] **Christine Chapman:** Some other Members want to ask supplementary questions, but I ask them to be as concise as possible because we can then explore as many aspects of this

issue as possible.

[196] **Julie Morgan:** Briefly, on one of Jenny's questions, I was not certain in your answer about the modular—whether you were agreeing that many more—

[197] **Ms Stacey:** No. Sorry if I was not clear enough about that. We can see that students took different routes through these qualifications, not only in WJEC but in all the other boards as well. We do not have evidence at the moment to say that that difference for WJEC students was materially different in England as compared with Wales. Cath, do you want to provide any more information on that?

[198] **Ms Jadhav:** There were some data in the Welsh Government report showing that there were slightly more students in Wales, so the proportion was slightly greater, doing a linear route with all the examinations at the end. Quantifying the effect of that, as Glenys has said, is very difficult because there are so many different factors.

[199] **Julie Morgan:** So, there were different capacities in each country.

[200] **Ms Stacey:** Yes, but where we perhaps differ is that we do not say that it is a significant difference. There is bound to be a difference. It would be very odd if they were exactly the same, but it is not a big difference, and we do not think that it is significant in the scheme of things.

[201] **Rebecca Evans:** You referred to the various options put forward by WJEC. I was wondering what discussions you had had with the Welsh Government regarding those options. Did it agree with you that option 1 was the best?

[202] **Ms Stacey:** It all happens very quickly, as you will understand, because we need to get results out. The provisional data come in and we are then all working across all the GCSEs to get the results out. My recollection is that it was around 9 August that we received a letter from WJEC setting out these options. We arranged a telephone conference around lunch time with the regulator in Wales to discuss those options. It did present difficulties for us and for the Welsh regulator, because we could see from the data that we had on the results for English students and Welsh students that the difference between them had increased since 2010. In 2010, they were quite well aligned but, by 2012, there was about an 8% difference in the achievement of grades A* to C. We could see that. That was a fact. However, what you could not really see were the reasons why. It could well be policy reasons—and, in fact, it was recognised in the telephone call that there could be a policy rationale. However, the regulators are there trying to set a common standard. We went through the options and agreed that we would go for option 1. We then had a confirmatory letter or e-mail later that day from the Welsh regulator to that effect, so it was an agreed approach—and it was very important to us that we got an agreed approach. We thought that there were certainly arguments to be put for looking more widely at this, but we were very keen to get a common approach, and we agreed option 1.

[203] **Aled Roberts:** So, that telephone call, after 9 August—

[204] **Ms Stacey:** It was on 9 August.

[205] **Aled Roberts:** On 9 August, at lunch time. The Government official here accepted that there was an 8% difference in comparative grades, and accepted that that could be down to policy decisions.

[206] **Ms Stacey:** The Welsh official suggested that it could be to do with policy differences, yes. Thank you very much for speaking in English.

[207] **Aled Roberts:** I do speak both.

[208] **Ms Stacey:** I know, but I do not.

[209] **Christine Chapman:** Obviously, there is a translation available.

[210] **Jocelyn Davies:** You mentioned materially different results between Welsh students and English students, and you said that there has been an increase of 8%. I looked back at the Record of when you appeared before the House of Commons Select Committee, and you said that the results for English students were ‘significantly better’ than the results for students based in Wales. You mentioned a couple of factors, such as that it could be down to a divergence in policy, the route effect and other things. Is it possible that there are any technical factors contributing to the difference in those results?

[211] **Ms Stacey:** I suppose that it depends on what you mean by technical factors. I will ask Cath in a minute, but—

[212] **Jocelyn Davies:** I suppose that what I mean is whether there is anything other than someone drawing the conclusion that children in Wales are not as clever as children in England, or that they are not being taught as well. That is to say, is there some other factor that we could eliminate that could possibly account for this difference?

[213] **Ms Stacey:** I recollect saying in the conversation that I had with the Welsh regulator official at the time that the difference is telling you something, but we cannot define what it is. Policy differences may well have been at the root of it. There are these other aspects of the way in which these qualifications play out that you will be becoming familiar with. So, if Welsh schools do controlled assessment differently to their equivalents in England, that would be a factor, would it not? The way in which the accountability pressures bear on schools might be different in England as compared with Wales, and so that might be a factor. There are the contextual factors, if you like, so it is not just policy, but wider contextual factors that will be relevant. I am not aware of any technical issue that would bear on this, but I will ask Cath.

[214] **Jocelyn Davies:** An 8% difference is very significant, statistically, so I am just wondering whether you can account for that with something that perhaps we will not have taken into consideration yet.

[215] **Ms Jadhav:** This is not necessarily a technical reason, but we have to bear in mind, as Glenys said, the different way in which these qualifications operate in England and in Wales. In England, candidates also have the option of entering just English. Those figures that Glenys was quoting were for English language, so that is all the Welsh candidates in there. It is all the English language candidates for England, but there will be another set of probably generally weaker candidates doing English. If you could somehow put those figures together, I suspect that the difference would be a lot less. The difficulty that we have in this—and we have spent a huge amount of time doing a huge amount of analysis—is that quantifying any of this is virtually impossible.

[216] **Ms Stacey:** That is a good point. What Cath is showing there is that it is quite possible that the 65% of WJEC’s candidates that were from England may have been a more able candidature than the 35% in Wales. That is no reflection on Welsh education; it is a reflection on—

[217] **Jocelyn Davies:** Who took the exams.

[218] **Ms Stacey:** Yes, but it is difficult to quantify. What we had in mind is the difference between those candidates from England taking WJEC and the candidates from England that were taking any of the other exam board offerings.

[219] **Suzy Davies:** I will just turn to the January 2012 WJEC exam. It was the same exam in the two countries, marked by the same people. The Welsh Government said at the time that there was nothing wrong with those grades and that they were fine. You take a different view saying that they have been generously marked. When did you come to that conclusion and how?

11.30 a.m.

[220] **Ms Stacey:** We came to that conclusion in our initial report, which came out about two weeks after concerns were first being expressed about the variations, and we explain in the initial report how we came to that conclusion. We were observing these qualifications, and we had them under what we call a ‘scrutiny programme’, so we were giving them special attention. That involved our observing some of the awarding meetings in January. We could see that it was a testing business, awarding in January. We expected that to be so, because they are new qualifications; they are different. Awarders did not have the past to rely on in the way that they normally do. Awarding is not always easy in a subject like English anyway; it is not like mathematics. For some of the units, there were very few candidates and it was an unusual set of candidates, because there were candidates who were in year 9 and there were candidates who were not representative of your normal June and July cohort. So, we could see that it was difficult. In fact, we saw that at least one of the boards, but possibly more, thought that it was being quite harsh. However, it is only when awarders could see the whole picture that they were able to get to the standard and see with much more certainty how it was.

[221] **Suzy Davies:** That was some time later. Does that explain why you did not insist on the January results being regraded?

[222] **Ms Stacey:** Decisions like that are made by the Ofqual board. The powers do not rest with me personally. The Ofqual board considered that, but as that had only really become apparent when we looked at it after the results, the board took the view that, although there was a technical argument for that, if you like, it would be wrong to do so.

[223] **Suzy Davies:** It would be mean.

[224] **Ms Stacey:** Not so much mean, but regulators have to act on objective evidence and they have to meet their statutory obligation to maintain standards. However, if you can see that candidates had relied on those results and acted in accordance with their results—which they had—the board thought that it would be quite wrong to overturn them. That was the rationale that the board took.

[225] **Suzy Davies:** I have one final question on this. If the exams in January were taken by students who were completing their qualification in June or July and they banked the marks from the generously marked exams, did they not have the benefit of that when it came to the final mark? In fact, the discrepancy of 8% that you talked about might be greater, because a generous mark is hidden in that.

[226] **Ms Stacey:** They will have taken those awards through to the final award and some of them will have benefited from that when compared with other students in the summer. However, the concerns being expressed about GCSE English—and the variations are not about that, which is a relatively small piece of it—are really about the results in the summer and for those students who did not have that benefit. That is the issue of concern that we are investigating.

[227] It is important to note that an inherent part of the design of a modular qualification, and, I suppose, in a way, the rationale for it, is that you can take units when you are ready or judged to be ready, and you can choose routes through and you can choose to bank or not to bank. It is an inherent feature of the design, not just in this qualification, but in all of them. We are learning, together, an awful lot from the experience of GCSE English this year, which we hope to report on, and modularisation is a key part of that.

[228] **Simon Thomas:** I would like to follow up on that. Did the experience of the January assessment affect the way in which Ofqual then approached the summer assessments? In other words, when you saw what was happening with the summer exams, and thought that they were overly generous, did you think 'Right, we know that things were a little bit over-generous in January as well, we have to make sure that we crack down in the summer'?

[229] **Ms Stacey:** No, not at all. We observed some, but not all, of the January awarding meetings, and we could see that it was playing out and there was no reason to be unduly concerned. We then awaited the preliminary outcomes in the summer. When those preliminary outcomes came in they generally looked fine. AQA, which is the other provider that serves the Welsh community, was absolutely aligned with where we expected it to be, as were most of the others. We challenged Edexcel and WJEC, because they were out of tolerance. That was not related to the relationships between January and June; it was to do, simply, with the comparisons with other exam boards. We could see that they were out of line and, in that position, it is the job of the three regulators to ask for the rationale and to challenge it. That is what we did.

[230] **Julie Morgan:** In your written evidence, you say that the view of the experts is that key stage 2 is a better predictor of GCSE achievement at cohort level and common centres. To which experts are you referring, and do you think that this is true in Wales, as it is in England?

[231] **Ms Stacey:** I will ask Cath to speak about this in a minute, but I will just say that key stage 2 data and any other data that you can bring to the table are useful. There is a lot of judgment in awarding qualifications. Anything that you can have to evaluate your judgment, for comparison and reflection, is useful. Key stage 2 data have proved to be very reliable for all qualifications. They are commonly used. Common centre data are also available to bring to the table. So, both are useful and both have been used. This year, we asked WJEC to report to us against common centre and key stage 2 data. The reason that we did that was because of the increased candidature for WJEC from English schools. For the candidature this year, because there has been a gradual change in the market of distribution, we could see that 65% was from English schools. So, as exam boards use key stage 2 data, when they are looking at awarding in English schools, it was right to ask for them. So, in a way, if you bring both to the table, you have as much as you can have.

[232] As for who we discussed it with, we have a standards advisory group, which has looked at our comparable outcomes approach and has recognised and said that it is the best approach known at the moment. That does not mean that it cannot improve, or that it does not need constant review and attention. We have been using it to get a grip on standards; to get to the point where we could get to a steady understanding of standards and not have grade inflation where you cannot see whether it is improvement or benefit of the doubt, or what. However, we certainly agreed to, and still will, review the approach, not only with the standards advisory group, but with other regulators. The difficulty that all regulators have, here, in England and internationally, is that there is no better approach. It is not that we can readily find a better one. I suspect that we are leading the way here in Wales and England in terms of developing the approaches.

[233] **Julie Morgan:** So, the experts are your consultants.

[234] **Ms Stacey:** Yes. Others have expressed views about it; there is plenty of research material around about it.

[235] **Ms Jadhav:** Yes, there is. Several years ago, rather than use key stage 2 in England, we used key stage 3. We had a key stage 3 test, because it is closer to when candidates sit their GCSEs. At the point where the key stage 3 data were no longer going to be available, the awarding bodies collectively did quite a big piece of research work looking at whether key stage 2 would be as reliable a predictor going forward, or whether we should revert back to a common-centres approach, which is what we used to do before there were key stage tests at all. This is probably going back two or three years, but the data then showed that key stage 2 was as reliable as key stage 3, to a couple of decimal points, but that it was much better than common centres. In GCSE English, for all the reasons around early entry strategies and resitting, common centres are not that stable from one year to the next, whereas the key stage 2 data have controls for the ability rate. It is a couple of years' old, but there is evidence to support what we are doing.

[236] **Julie Morgan:** What about the fact that key stage 2 is operated differently in England and Wales?

[237] **Ms Jadhav:** We only have the test data from England. What makes the awarding more challenging for WJEC in particular is that, in England, the awarding bodies have two measures—they can look at common centres and they have key stage 2—but for candidates in Wales, there are only the common centre data. There are no key stage 2 tests in Wales.

[238] **Ms Stacey:** If you look at the particular tensions—the difference between the two countries that the regulators are trying to reconcile when awarding—you will see that there are different policies operating that will affect outcomes in some way, but it is not clear how. Secondly, the historical data that we can rely on is different, because the testing regimes are different between the countries. That also creates a problem for us, which we do our best to work with.

[239] **Lynne Neagle:** I have two quick supplementary questions. You have explained who your experts are, but to what extent has AQA driven the emphasis on the key stage 2 predictive model? Also, you referred to the fact that you felt that this model was working well in other subjects. Have any adjustments been necessary in any other subjects based on this model?

[240] **Ms Stacey:** Was the first question whether AQA drives the use of a predictive model?

[241] **Lynne Neagle:** Yes.

[242] **Ms Stacey:** Not as far as I am aware. We have been around for two years. The predictive model has been around longer than that and it is accepted by all of the exam boards. They all use it. That is the position. I am not aware that AQA is leading on that. I am sorry, what was your second question?

[243] **Lynne Neagle:** You said that you felt that it was a reliable model in other subject areas. I wanted to know whether you had had cause to make any adjustment in any other subject areas.

[244] **Ms Stacey:** Not this year.

[245] **Lynne Neagle:** What about last year?

[246] **Ms Stacey:** No, although we had some odd variations last year in physical education. One might say that, for a subject like PE, there is bound to be more of a stretch for key stage 2 predictions. PE is a very physical subject. Last year, I had only been there a few months, but I recollect that GCSE PE seemed to have a slightly unusual pattern. We should bear in mind that we had been using a comparable outcomes approach to A-levels, as well as GCSEs; A-levels are using GCSE outcomes. So far, we have used it across more than 1,000 qualifications and not had this issue. I absolutely understand your keen interest in the predictive model, but there are other factors at play, which explain the variations that some schools have experienced.

[247] **Christine Chapman:** I will call Aled and Jenny before I come back to you, Julie. I remind Members that we are getting very close to time. We have less than 15 minutes left.

[248] **Aled Roberts:** I will wait for Julie to finish.

[249] **Jenny Rathbone:** Schools use all manner of data to assess individual pupils' attainment, such as the Fischer Family Trust, schools data et cetera, yet experienced teachers have found that the time-honoured way of assessing the likely outcome for individual pupils has produced a completely different result than what your methodology has thrown up. It is therefore perfectly possible, is it not, that your methodology is flawed and that it has produced some really quite strange results for the cohort this summer?

11.45 a.m.

[250] **Ms Stacey:** We have not found that our methodologies are flawed, and I am not aware that anyone else has either when they have looked at it. What we do know is that before these qualifications changed, they were not modular, they had been around for a very long time, they were very well understood by the many who were teaching them and the predictions, in a way, were simpler because you were not dealing with a modular qualification. What we can see now is that many long-established English teachers would have been used to predicting and being able to rely on their predictions. They were in quite an unusual position compared to some other subject teachers who had experienced more changes. So, they would be pretty certain that they would be able to predict well.

[251] However, it turns out that prediction is not as easy as that for these qualifications. Most particularly, if you make assumptions about grade boundaries set very early in the qualification, your students might not have sat that unit, but you will have seen how the grade boundary was set for others, so you might be making assumptions about that. Also, prediction is very difficult to get right if 60% of the qualification is through controlled assessment, and if it is being marked in a school, but graded by the exam board, which is what happened. So, prediction has been much trickier because of the design of the qualification, because of routes through and the assumptions made about grade boundaries.

[252] **Jenny Rathbone:** Obviously, there was a moving of the goalposts because 60% was based on the exam rather than 40%.

[253] **Ms Stacey:** I do not know whether I would describe that as 'a moving of the goalposts'. There was a change in the design, in the balance between written examination and controlled assessment. All three regulators agreed that well before Ofqual's time. The experience this year most certainly brings that balance into question.

[254] **Julie Morgan:** Your evidence states that in the meeting on 14 March 2012,

[255] 'Given that over half of WJEC's GCSE entries were from England, it was proposed that they should report outcomes for those candidates against predictions based on Key Stage 2 (KS2) prior attainment'.

[256] We have already had some discussion about this. Could you confirm whether WJEC was asked to use key stage 2 data for entries from England or for entries from England and Wales?

[257] **Ms Stacey:** We simply asked it to report against both. That is my understanding; I was not at that meeting.

[258] **Ms Jadhav:** It would have been only for the candidates in England who had key stage 2 scores. So, we were asking WJEC to report against common centres, for the common centres that it had, and to report against key stage 2 for those candidates who matched key stage 2, which would have been only candidates in England.

[259] **Ms Stacey:** Thank you. I was not at the meeting, so I was not entirely clear about that.

[260] **Simon Thomas:** I want to stay with this meeting on 14 March, which seems to be a crucial meeting in which these decisions were taken. You have just explained what happened from the point of view of your responsibilities as the English regulator. However, in your evidence, and also in WJEC's evidence earlier, we have confirmation that the Welsh Government sent an e-mail to you after the meeting to confirm support for the proposed approach. Does that mean that the Welsh Government was also agreeing to use key stage 2 data within the 35% of students who were being assessed in Wales?

[261] **Ms Stacey:** The Welsh Government's representative was at the meeting.

[262] **Simon Thomas:** Yes, and that was confirmed via e-mail afterwards, was it not?

[263] **Ms Jadhav:** Yes. There had been a lively debate about the manageability, as well as whether that was the right approach. There is obviously a manageability issue in terms of reporting data. The Welsh Government representative is in a slightly difficult position given that debate, but they confirmed afterwards that they were supportive of the approach that we had taken.

[264] **Simon Thomas:** Can you confirm that WJEC also raised concerns in that meeting?

[265] **Ms Jadhav:** Yes; it did.

[266] **Simon Thomas:** It expressed those concerns, but then subsequently agreed to the process, did it not?

[267] **Ms Jadhav:** It expressed concerns in the meeting. As I say, those concerns were around whether or not that was the right approach, but also whether or not there were risks in having to report additional data and the systems development that that might entail.

[268] **Ms Stacey:** There is always a risk that regulators ask for things like that without recognising the demand and the system risk; for example, as we look at what we would do in relation to future series in November and January, we have to recognise system issues that may lie behind any simple suggestions that we make. That would certainly have been expressed.

[269] **Simon Thomas:** In September, the Welsh Government produced a report on this

whole process, which I am sure that you will have seen. That report states that the Welsh Government is operating as regulator in this context and expresses serious concerns that it was not appropriate for results for Welsh candidates to be determined on the basis of prior achievement by candidates in England. Did the Government express those serious concerns, and how do we square expressing those concerns with the decision, in a subsequent e-mail, to accept this process?

[270] **Ms Stacey:** We are simply the regulator in England. We met with exam boards and our fellow regulators to agree the approach to be adopted for June awarding ahead of June awarding; that approach was adopted.

[271] **Simon Thomas:** As far as you are concerned, all three regulators had a common approach to these exams.

[272] **Ms Stacey:** Yes. We have said that concerns were expressed, but we went into awarding understanding the common approach.

[273] **Simon Thomas:** The Welsh Government Minister, in his statement to the Assembly, said that it was,

[274] ‘clear that the mechanism that had been introduced, at Ofqual’s insistence, to ensure comparable outcomes, had failed in Wales.’

[275] First of all, do you accept that it was done at your insistence? Secondly, do you believe that the mechanism failed in Wales?

[276] **Ms Stacey:** I believe that we have explained how the decision came about.

[277] **Simon Thomas:** So, it is an acceptance by the three regulators; is that so?

[278] **Ms Stacey:** It is not possible for Ofqual to insist and overrule the Welsh regulator in terms of how these qualifications are to be regulated in Wales. Instead, all of the regulators—the three regulators—work very hard to agree an approach. It is what students would expect and it is, after all, what exam boards need. What you have seen being played out in the raw here is how that worked for GCSE English. Exam boards and regulators then applied that approach.

[279] **Simon Thomas:** You know that subsequently, therefore, the Welsh Government, as regulator, issued a directive and insisted on the regrading of this exam—not the AQA exam—in Wales. You did not do the same in England. In your opinion, has this exacerbated the gap between Welsh and English attainment that you mentioned earlier?

[280] **Ms Stacey:** We have already said publicly that we did not necessarily agree that it was the right thing to do. The Welsh regulator has directed WJEC to re-award the qualifications so that the outcomes match those of 2011. Of course, the Welsh regulator is entitled to do that and has done so. It puts three-country regulation in a difficult position, because what we have there is one of the regulators determining after the event, if you like, to set a different standard. We are not, therefore, able to say that we have a common standard for this qualification offered by one exam board across the two countries. That creates very difficult issues for the regulators going forward.

[281] **Aled Roberts:** I just want to be sure that we understand this. You elucidated the difficulty that the three regulators have in trying to ensure common standards. Did the Welsh Government officials explain the difficulties we had with relying on key stage 2 data? You will be aware that, in Wales, Estyn uses teacher assessment at key stage 2. It is a wholly

different scenario here. Given that the Welsh regulator then decided unilaterally to direct regrading, would it have been feasible on 14 March for the three regulators to have reached a situation where agreement was not possible, and where the Welsh regulator could have decided that, as far as students in Wales were concerned, it would stick with a different system to that which they agreed?

[282] **Ms Jadhav:** I think that it was a possibility. As Glenys has said, we work very closely and very hard with the other two regulators to try to ensure that that does not happen, because it puts us in a very difficult position with awarding bodies that are offering the same qualification in England and Wales. However, technically, it was entirely possible that no agreement would be reached at that meeting.

[283] **Aled Roberts:** How soon was it after the meeting that the e-mail came about?

[284] **Ms Stacey:** Which meeting?

[285] **Aled Roberts:** The meeting on 14 March, at which we were told that concerns had been expressed. Then we were told that an e-mail was sent—

[286] **Ms Jadhav:** It was either later that day or the following day. I would have to check, but it was fairly quickly afterwards.

[287] **Aled Roberts:** Okay. Can we then move on to 9 August, when WJEC letter was sent to you outlining three different options to deal with the difficulties you found yourselves in. WJEC indicated that option 1 provided the least different outcomes, but can you explain why you were unable to accept WJEC's at-award option in particular?

[288] **Ms Jadhav:** As Glenys said, what we were trying to do, which was challenging with these new specifications, was to align the standard across the four or five awarding bodies operating in the three countries. Three of those awarding bodies were in line with the predictions and two—Edexcel and WJEC—were apparently generous in relation to the prediction.

[289] **Simon Thomas:** May I just ask a question on that, Chair?

[290] **Christine Chapman:** Yes.

[291] **Simon Thomas:** Can you give us an idea about those bodies in percentage terms? Did those other three awarding bodies represent 80% or 20%, for example, because that makes a difference? It does not just come down to the number of bodies, but the size of the cohort they represent.

[292] **Ms Stacey:** I do not have the figures for Wales, but, overall, the distribution has changed as the new qualifications have come in. Together, WJEC and AQA have 80% of the market for the English GCSE suite. However, that has changed. I will check the figures, but I think that, in the move to the new qualifications, WJEC's market share has increased from 9% to 19%, and almost all of that has come from AQA. I can certainly confirm those distribution figures for you, but that is the broad picture.

[293] **Simon Thomas:** That would be useful.

[294] **Christine Chapman:** If you could do that, that would be useful.

[295] **Ms Stacey:** To exemplify what Cath was saying, in England, for the other boards, the results were within tolerance, if we leave Edexcel to one side for a moment. Certainly, AQA

came in well within tolerance, as did OCR. However, WJEC's provisional results were 2.7% above predictions in English and 4.1% above in English language. Therefore, they were to be challenged.

[296] **Christine Chapman:** Okay. I think Angela has one very brief question.

[297] **Angela Burns:** Yes; I just wanted to make a quick comment. Other changes are being planned for future awarding. Do you think that these changes will improve the validity and reliability of the process? If you believe that the three-regulator model is a way forward, do you think that there ought to be a change to the structures of the three-regulator model?

12.00 p.m.

[298] **Ms Stacey:** That is a big question and do tell me if I do not answer it all. First, we have been looking closely at these qualifications and our preliminary view is that there are weaknesses in the design, particularly in the balance between controlled and written and routes through. However, in a technical sense, when you look at what is covered by the qualifications and the quality of the assessment, the actual examinations and controlled assessment are good, and probably, better, in general terms, than those that they replaced. This is lost in all of this—that you had the potential for a much better assessment of candidates. On modularisation and controlled assessment, there are problems in the design, but in terms of the quality of the qualifications and assessment, that is there and should not be lost in all of this.

[299] Secondly, they have proved to be difficult to award and to protect against some of the pressures in schools. We will act to make improvements and do as much as we can as regulators to protect them as they are running now. So, we are looking closely at the controls around the November resits. We are certainly looking at the controls, fundamentally, with our fellow regulators in relation to January, where we expect quite a large uptake, and in June. There is a longer term issue that we will look at and settle hopefully before Christmas, namely how these qualifications should play out in England in particular from September 2013 onwards. However, we understand that Government here has already made decisions about this year; change is already in hand here.

[300] **Christine Chapman:** Thank you. We will draw this session to a close. I remind Members that we will question the Minister at our next meeting, on 8 November. So, we can have the opportunity to look at some of the evidence and reflect on that at the next meeting.

[301] I thank you both for attending today; we appreciate your coming to the Assembly. There will be a transcript of the meeting, which we will send to you to check for factual accuracy.

[302] **Ms Stacey:** Thank you so very much for giving us the opportunity to do so.

**Cynnig o dan Reol Sefydlog Rhif 17.42 i Benderfynu Gwahardd y Cyhoedd o'r
Cyfarfod**

**Motion under Standing Order No. 17.42 to Resolve to Exclude the Public from
the Meeting**

[303] **Christine Chapman:** I move that

the committee resolves to exclude the public for the remainder of the meeting to discuss the committee's draft report on its inquiry into adoption and the forward work programme in accordance with Standing Order No. 17.42(vi).

[304] Are all Members content? I see that they are.

Derbyniwyd y cynnig.

Motion agreed.

Daeth rhan gyhoeddus y cyfarfod i ben am 12.03 p.m.

The public part of the meeting ended at 12.03 p.m.