Transcript: Michelle Meadows and Jonny Tridgell - A-Level grading

Laura Molway 0:02
Welcome to the Oxford education deanery podcast. In this series we explore the latest research from the Department of Education at the University of Oxford and discuss the real world implications for teachers parents and policymakers.

Hamish Chalmers 0:22
Hello and welcome to the Oxford education deanery podcast. My name is Hamish Chalmers. And I'm joined by Johnny treadle a religious studies teacher with a long standing connection with the department. Hello, Johnny. Hello. And Michelle meadows and associate professor of educational assessment here at the department. Hello, Michelle. Hi. We're releasing this podcast on the eve of a level results day. students and their teachers will be anxiously waiting to find out how they have done keeping their fingers crossed that they will have attained the grades that they need to go on to the next stages of their lives. Be that undergraduate study at university study for a higher technical qualification, a degree apprenticeship or into the workplace. Teachers will be getting ready to congratulate those who get the grades they need to progress and to support those students who don't do as well as they'd hoped. I remember the big day myself, I went into school to collect the results, put the small manila envelope with my results in into my pocket, and then walked in ever decreasing circles around Oxford City Centre for a couple of hours until I built up the nerve to open it to learn how well I'd done. Like most people hoping to head to university, I had to attain specific grades. So while I was building up that nerve, the question in my head was have I done enough and I managed to land in the grade boundaries that would allow me to go on to the place that I had my heart set on. Like me at the time students and teachers today often have questions about how grading works. What's the difference between an A star and an A for example, or our grade boundaries for different subjects set differently is a B in geography worth the same as A B in chemistry for example. Or if I have an identified special educational need is this taken into account when settling on my final grade, and so on. Michelle has a wealth of experience through her earlier work at off coil in AQa where she was responsible for the grading. And as a researcher an educational assessment on how grade setting is achieved. And Johnny also marked for AQa, as a teacher knows what it's like to be supporting pupils through the process. He knows all of the questions teachers and students have. So we thought that this would be an ideal moment to explore some of those questions and demystify for our listeners how it all works. Johnny, what do you think the biggest question for teachers and students is about grade boundaries?

Jonny Tridgell 2:37
I suppose the simplest and biggest one is how they actually set. So I think there's lots of myths. We talked about norm referencing, which I think isn't true. It'd be interesting now how the great boundaries are actually put in place by the examples.

Hamish Chalmers 2:51
Michelle, can you tell us about that? Yeah, sure.

Michelle Meadows 2:54
It's important to appreciate the difference, first of all, between marking and grading. And I say that just because people often do mix up the two terms. So setting the grade boundaries happens once virtually all the marking is done. It happens for each individual subject. And it's done by what's called an awarding committee. So this is a group of senior examiner's, and there'll be a chair, chair of examiner's, and there'll be the principal examiner's who actually wrote each of the question papers. And normally there's a subject manager from the awarding body, the exam board. And there'll be a technical person there who is in charge of statistical information. So they get together. And they're looking at the quality of the student work, but they're doing so also bearing in mind lots of statistical information. In the UK, this technique where you're mixing the examiner judgment of work, and the stats is actually called attainment referencing. So it's not norm referencing that is sometimes mixed up. And depending on how confident you are in the sticks or in the examiner judgment, the balance between the two in deciding the boundaries shifts, for example, if you've got an entirely new qualification, then it's really hard for the examiners to judge what is the quality of work a particular grade that's going to maintain the standard over time. So in that scenario, you're really relying on statistical information to help you judge where to put the boundary. And when you're in sort of steady state and you've had a qualification for many years and and passionate, you've got a really large entry to your qualification, then obviously, the stats are pretty good to them and you can rely on them a bit more. And that's really the strength of the approach that we have in this country. That is kind A mix of examiner judgment and stats that you can play in depending on how confident you are.

Hamish Chalmers 5:05
So does that mean that when we talk about norm referencing, does that mean that it's not a question of you just got to get above 70%? To get an A? Is that Is that how it works? And it's actually a bit more complicated than that.

Michelle Meadows 5:16
Yeah. So Norm referencing is actually a very particular approach, that in my research, I've never found an example of a qualification rather than a test that is norm referenced. So Norm referencing, the best example of this is IQ tests. IQ tests are norm reference, that's where you go, and you get a representative sample of your population, you do a norming study with them, then the test never changes, it's kept steady, and you're constantly referencing the results of the test group, back to the norm that you originally tested. We don't do that do it at all, when it comes to a levels or indeed GCSEs. And so I think what people mean, when they say all this might be norm referenced is, as you say, a particular percentage of the population is always awarded the same grade. Actually, if you look at individual subjects, and the results for each individual a level subjects, you can see that's not the case, because they do bob up and down a little bit over time. Although because we do use statistics in the grade boundary setting process. To a great extent, you can see that once all the subjects are brought together, and you look at the national outcome of all the eight levels, that precious 30, from one year to the next.

Jonny Tridgell 6:44

So if for example, you've got a really tricky paper one year, and then the year before the paper fell a bit more straightforward. How would you make that so that was fair across both years with the gray boundaries,

Michelle Meadows 6:57
this is the strength of the approach that's taken, because you're absolutely right, that no matter how hard a senior examiner tries, it's really difficult to write a paper, that's exactly the same level of demand as the ones that in the previous year, it's really tough. So you've got to set your grade boundaries, such that if you've got hard paper, they need to be slightly lower, if you've got an easier paper needs to be slightly higher, to be fair to the students. And so this is where the examiner is looking at the quality of work, try and bear that in mind. But it's incredibly hard for them. Everything is changing, your questions are changing the papers different, your entry is changing a little bit. So you need to know a bit about your entry. So the students who have entered this year, are they more or less academically able than in previous years? You want to know that, don't you? So this is where the stats come in. So you can find out for example, at a level, what was the mean GCSE score of all of the students that have entered this year? And how does that compare to the previous year? Are they more academically able or less academically able that information, along with things like the mark distributions, and the means and standard deviations of the papers, and of course, the actual student work, all of that information is brought together to enable them to just tweak the boundaries of pull down according to how hard or easy the paper is.

Jonny Tridgell 8:36
So when it finally comes down to it, and we've got this committee that is sitting there, who makes the final call and like fight, this is the mark, this is what the AAA is this is why the years who makes that final decision.

Michelle Meadows 8:51
Yeah, great question. So there are multiple levels of sign off initially. It's the chair of examiner's. So the chair will really try and work hard with the committee with all the information that they're given to come to a consensual decision about where the boundaries are. Once that's happened, then there's a level of sign off within the exam board she can imagine. And there can be different levels actually, depending on which exam board so often, they'll be an initial level where somebody quite technical will really delve into the examiner's reports the boundaries that have been chosen all the information, and they'll be asking themselves, does this look plausible? Does this look right? And hopefully they'll say yes, and sign it off and then it will go usually to the responsible officer of the exam board who is responsible for the maintenance standards. And they will then ask exactly the same kinds of questions you know, do. Do I have confidence confidence in This award, they will hopefully sign it off. And then of course, we have an exams regulator. Whether you're in any jurisdiction in the UK, there'll be an exams regulator, they will then also scrutinize the outcomes from the award. And again, they'll be asking themselves, this is not right, can I kind of have confidence in this? And if the exams regulator things, you know, looks a bit fishy, they will ask some challenging questions, how did you come to this set of grade boundaries and these outcomes, and then there'll be a conversation that goes on. And of course, the exam

board normally has a brilliant rationale? course they do, they have lots of evidence. But that challenge will happen. So there are a number of points at which there's no sense in which the exams regulator signs off boundaries, that's not the case, the boundaries are very much owned by the examples. But there's that challenge there.

Hamish Chalmers 11:01
Sounds like a very sort of multi layered, or at least a lot of checks and balances go through. So this gives us confidence that by the time the results are released, that it's really been sort of well scrutinized, and that everyone is very happy with the results as they've been presented. Yeah,

Michelle Meadows 11:17
absolutely. That's not to say that there haven't been occasions when things have gone wrong. Yeah, just look back to GCSE English in 2012. There are problems then with the grading. But generally speaking, we can have a lot of confidence that the boundaries have been very carefully selected. Great. Now, you

Hamish Chalmers 11:41
mentioned a little bit earlier about the entries, the students taking the the A levels themselves. So I want to move on to talk about though, the potential implications of the people taking the exam. So if a student has, for example, access arrangements, does that affect the grade that they get? Or what about if they have other kinds of special consideration? My particular area of interest in research is children for whom English as an additional language. And so often they get accommodations through the process of taking the exam, like extra time, for example, or in a menu answers for somebody who's not able to write themselves, does this make a difference?

Michelle Meadows 12:16
Well, all of the special arrangements that are put in place occur at the time of the assessment. So the grade boundaries themselves are not adjusted at that later point. So the idea is that by putting in place access arrangements, students are able to show what they can do. Now there is a slight anomaly in the sense that sometimes special consideration is applied. So special consideration is where perhaps there was a temporary kind of issue where perhaps a student, you know, have something really quite awful happened to them the day before the exam. And then there can be an arrangement weighed where a small proportion of additional marks are given. That's actually the stage of marking, but the gray boundaries themselves are not adjusted all the activities to try and ensure fairness put in at the front end of the process.

Hamish Chalmers 13:33
So ever, it was, by the time the exam board has marked the papers, and they're starting to set the great boundaries. And all of those differences in the people taking the exams have essentially been sort of accounted for. That's certainly the aim. Yes.

Jonny Tridgell 13:47
So I think when you're thinking about the work that happens at the end, then with the gray boundaries, the assumptions, and you're looking at, right, we've decided that let's

say six out of 100 is a B, what happens when you're looking at that in terms of student groups? Is there a sense of actual just say, What if we put it at 60? That means X number of children have got to be everybody has 61, that number drops? How much does that drop? Is that something that's considered? And then is it also I suppose, in addition to that, are you aware of things like student ethnicity, or gender, for example, when you're making those calls about it's going to be 60 or 61? And what that would mean for those student populations, or is that information kind of hidden from you?

Michelle Meadows 14:33
And so, while you're setting the boundaries, you're absolutely thinking about, what does this mean in terms of the proportion of students getting a particular grade? That's really at the forefront of everybody's mind. However, it would be very unusual for an example to include information about the protected character stakes of the students at this point. So what should happen is that those considerations occur much, much earlier in the process. So in selecting the questions the way in which they create the assessments, then they should be absolutely thinking, what does this mean for students of different ethnicity? What does it mean for boys versus girls, and so on. So that kind of fairness should be really at the forefront of their minds, then after grading is occurred, there will be lots of analyses of how the question paid, performed, and how different groups performed. And you know, there'll be an opportunity there for doing what's called differential item functioning, which is where you look at, did male students do better than female students on the site, considering how they do on the test overall, that kind of analysis. So the thinking goes on much earlier in the process, and then much later in the process, rather than it being something that's particularly taken into account at the point of setting the grain boundary.

Jonny Tridgell 16:21
So I'd say it's much more about the actual paper itself and the access arrangements, and I suppose the training for the markers and all that kind of end of things, as opposed to what's happening in the raid committees. Yeah,

Michelle Meadows 16:32
Absolutely. Yeah. And that has been discussed over the years and the research literature, would it be right to consider the outcomes for different groups? Should we in the statistics that we use in the awarding meetings actually put in these different characteristics and have them as part of the statistical modeling?

Jonny Tridgell 16:55
And I think it's interesting, looking at contextual offers for university, for example, that universities might give a student with certain characteristics or from a certain type of school a lower offer, for example. And I wonder if that's something that would ever be possible to contextual grade boundaries? Well,

Michelle Meadows 17:13
There's a great research paper by Dr. Neil Stringer, that actually explores exactly that possibility. So you know, what you could do is you could have the standard grades as we have in a normal year. And then you could actually create contextual grades. And you

could do this rather than each individual university doing it for themselves to contextualize their admissions process, you could do this at a national level that perhaps, you know, it could be something that the regulator did, I'm not sure they'd necessarily want that job. Or it could be something that they example did. I think this is, you know, it's interesting, but it's also very, very controversial. Whether or not this kind of information should be included or not, it's very values driven. And, you know, just over recent years, I've been watching with great interest that the changes in attitudes towards standardized assessment in the US, were there really pushing towards much more socio culturally responsive approach to assessment, actually different assessments for different groups. So rather than contextualizing grades, I'm actually changing the assessments so that two groups could set quite different assessments. And here, particularly in England, we're still quite in the standardization framework where our notion of fairness is that everybody gets the same examination. And these are competing versions of what is fair, fairness is one of those very complex concepts.

Hamish Chalmers 19:03
We certainly hear in the EAL English as additional language world, the the idea that what would be the implications of doing first language assessments of core content to allow children to demonstrate if their strongest language is not English, their content knowledge. So you know, if you maybe a brilliant scientist in Ukrainian, but asked to demonstrate that through English might make it look like you're not quite so quite so good at science or whatever. And so I think it opens up a really interesting conversation, but it's also seems to open up a bit of a can of worms, because if you're going to do how do you ensure that each of those different language assessments would be similarly pitched had in the same level of difficulty or, or so on, but it certainly is an interesting

Michelle Meadows 19:48
Yeah, and there's a lot of research on that as well how to create comparable assessments in different languages. Very tricky.

Speaker 2 19:58
I think it's interesting in English, isn't it? Just seeing the impact it's had on students just by a command word in an exam question is one they're not familiar with. And it's kids whose English as a first language, and it's just quite a just a vocabulary word they've not heard before they've not understood and how you adjust for that. But I guess it's when you come back to this thing about setting the great boundaries, using statistics and trying to recognize what's gone wrong, if sudden, you've got a whole group of children who've just misunderstood the question completely.

Hamish Chalmers 20:27
Reminds me. Yeah, a few years ago, I remember there being a controversy over a question that had asked the exam. So just to compare two things. And there was a semantic difference between compare and contrast. And some children had done both they had compared and contrasted other children had taken the question at its word, and just compared it seen what's similar about the two things and so there was this big controversy over whether they had been disadvantaged by the ones who will faithfully answer the question with, with reference to it sort of exact semantic meaning, or those

who have gone the extra mile to, to take into account contrasting as well as comparing. And so those things, I imagined that statistical analysis is something which would be sort of after the fact we'd be helpful in terms of identifying as you say, Johnny, where, where it might have gone wrong, but also looking at those those protected characteristics and so on maybe looking and seeing whether there are things that we need to be doing with it within our teaching. So the does that make these these sort of, essentially summative assessments like a levels and GCSEs to introduce a formative aspect to them. So in as much as we can see that there is an issue here with a particular group over addressing that area of the curriculum. So therefore, we can make curricular changes or pedagogical changes to help address that. Yeah.

Michelle Meadows 21:45
And so examiner reports come out after each exam series, where the Trove examiner's working with their colleagues who set the papers will set out where there have been misunderstandings where perhaps an area of the curriculum hasn't been, perhaps taught or learned as well as they would really hope and where there's been success, too.

Speaker 2 22:12
I guess, following on from that, I suppose, going back to what you were saying earlier about the committee's everyone's involvement there. This thing about how the questions are phrased is really interesting. So rtj level Rs. I've taught both OCR and AQa, the question styles are very different. So the AQa is very kind of regimented, very similar every time OCR much more variable. Choice the questions they create, you don't. Why also, I suppose very interesting for teachers, certainly for me, if I'm trying to pick a spec, there's always that that idea, or when you talk to people about it is one paper, easier to get a in. But that was easy to get in those kinds of ideas. We have an I wonder, first of all, how much of that is just a myth. And we just talked, we're just talking unnecessarily about things that are really true, but also, how you standardize that across the examples of actually in a grade. It OCR philosophy and ethics is the same as an AQa one. So they're worth the same, because I think that's what people worry about a lot when they're making decisions about what exam board to go with. How to make it fair and nice. Is there any sense that one year, it might have been easier to get an A if you've done this paper? But is that board?

Michelle Meadows 23:22
That that's such a good question, I'm fascinated by this, because I've done this work for a long, long time now. And I sort of looked back to you know, over 20 years ago, when I first started working in the field, and actually there were much much greater differences then between the offer of different exam boards. So you know, you would have one exam board offering a modular specification, another offering linear, one might have some coursework the other might be entirely examined. And now, actually, there is a far higher level of similarity between the office and yet still people are concerned that there there might be issues of comparability. The way in which the awarding works is and this is where the, the stats really do come in, because this is the way in which the standards between exam boards are aligned. Given the mean GCSE profile of each exam boards entry, a prediction is given as to what the a level outcomes should look like. So if one

exam board say OCR, attracts an entry with say lower mean GCSE scores and another exam boats, a person attracts an entry in a particular subject that's got higher GCSE scores, then the prediction of how those students should perform in each of the exam boards is adjusted. And this is what the regulator looks at. So the regulator will be looking at has each of the exam boards come relatively close into prediction. And clearly, if one exam board is sticking out, then questions will be asked, there'd have to be a really great rationale, because the exact date the exam regulator will be asking, why is it that all the students in this particular example would have done so much better this year? It's not a national phenomena, we're not seeing it in the other examples. So what's going on here? So the questions will be asked. So it's through the use of these statistical predictions that were able to keep some of the benefits of having different exam boards where you can actually make choices. And frankly, if you're unhappy with the level of service that you've had, you can change exam board, and yet still get comparability of the grades that are actually awarded. So

Jonny Tridgell 26:18
If I'm sort of thinking of switching my GCSE spec, on my a level spec, would you say it's probably not worth worrying too much about what the gray boundaries are? And across spectrum, I'm making that decision, I should focus more on just which content my like, is it? Absolutely,

Michelle Meadows 26:32
that's, I wouldn't be looking at the grade boundaries. Because if you see that one example, it's got lower grade boundaries, that's probably because they've got harder paper, what I'd be thinking about is, what do I enjoy teaching? And what are my students interested in? What's gonna play best for them?

Jonny Tridgell 26:50
I think that's really interesting, because I think when I have conversations with teachers, and heads, the department about what specification they picked, it's interesting how many people do it based on the paper and the gray boundaries? Again, because I suppose we're so focused on wanting to get the best possible outcome for the children. And I wonder if maybe we need to reframe that a little bit as a professional and say, Actually, this believing out that bid isn't the important bit. It's there. It's the other end.

Michelle Meadows 27:16
Yeah, have a look at the question papers have a look at. But I wouldn't worry too much about the grade boundaries themselves. Because give with one hand, take away with the other hard paper, but low boundaries,

Jonny Tridgell 27:29
there's Yeah, it's not necessarily an easy win about sort of the gray boundaries and go, that's interesting.

Hamish Chalmers 27:34
Very good. You spoke there about prediction, and prediction, from GCSE to what one might expect to a level. So I just want to take another different look at prediction. And

that's the use of mock exams. And so are teachers. Should they take into account the grain boundaries? Or what grade boundaries should they use when conducting mocks, where in some cases, you won't be doing a whole paper, you might only be doing a part of a paper?

Michelle Meadows 28:00
This is really tricky, isn't it? Because we just talked about how difficult it is to set a paper that's got the same demand. And that's because even very small changes to questions can really impact on student performance. So just a shift in wording. a different set of numbers in a math question can really, really change how hard students find the paper. Of course, of course, teachers want to use grade boundaries, I think that just have to be very mindful of the fact that they're just an approximation. So, you know, if you're going to use half a paper, the obvious thing to do is just half the gray boundary. But you don't know do you is that the easier half or the harder half? You don't know. So inevitably, you're going to want to do it. But I would just bear in mind the uncertainty, and that they're very, very approximate.

Jonny Tridgell 29:05
So if I was say, then the way that I would do my gray boundaries would be to take the last time I did it took an average of both papers for a level over 18 2018 and 2019. And then gave myself a score out of 100 200 sets of 200 the paper based on all those averages over two papers over two years, and then just divvy that up. So if I mark a 25 mark question, I just do whatever proportion that is of the mark and say that's what the B is, and then usually round up if it's a fraction is that and I think it's those sort of maths that we're all doing as teachers trying to come up with something that feels reasonable. I used to do a really basic 10% Gray boundary and that was an absolute nonsense, I think. Is that a good way to go? Or is it all just shooting in the dark? Well, it's

Michelle Meadows 29:54
Well, it's certainly a better way to go. Look, you know that that doesn't set that sounds reasonably sensible. I think you just have to bear in mind that you're not really sure. And so then it becomes about how you're using those outcomes, and what the students believe about those outcomes. Because students need to feel, need to know that they can nonetheless, do well. And it's about the wash back of those outcomes on that motivation. So I'm sure teachers are very careful about how they then encourage students to, yes,

Jonny Tridgell 30:40
it's not to saying you've got an A Oh, yeah, it's you've got now using these boundaries that I have artificially generated, I suppose, isn't it. But I guess I suppose the real scary thing, particularly at a level is that we use these three that UCaaS predictions. Now with the death of a SS, that becomes scary, doesn't it? If we're saying actually, the gray boundaries start being kind of, oh, this is an interesting thing for us to think about. But really, let's talk about focus on how you can improve to like, actually, the grade I'm going to give you in November of year 13, is a life changing one in terms of your options that

Michelle Meadows 31:14

are scary, because that's why when you look at UCaaS predictions, they tend to be really quite generous. They're about I think it's around a third of UCaaS predictions about our grade too high. A small proportion are too severe. And that's because, understandably and rightly, rarely, given the uncertainty, teachers are given the benefit of the doubt in those that makes sense, doesn't it, I would do exactly the same thing. If I was in that position. I

Jonny Tridgell 31:45
think it's interesting, I think just that bit about how you make that call, but also recognizing, as you say, that the call is always a guess. And based on a tiny number of students, I mean, if you've got particularly at a level, if you've got a class of 10, and you're trying to make some sort of decision went out to the actual datasets of 1000s and 1000s of miles, it's just not to

Michelle Meadows 32:02
me, there's always a disconnect between this. And you know, when grade boundaries are set, and we use predictions in that scenario, it's only when we've got more than 1000 students that we start getting really confident about those predictions, we would never dream of generating predictions on, you know, less than a fewer than 100 students, that would be madness, and certainly not for each individual student. And yet, in another scenario, here, we are expecting teachers to generate predictions. So that

Hamish Chalmers 32:33
will play out very, very sort of, obviously, during the pandemic is underway with the teacher experts, what teachers are predicting against what the now infamous algorithm actually ended up giving? Yeah,

Jonny Tridgell 32:44
speaking of algorithms, do you think AI will come to play a role in this? How

Michelle Meadows 32:48
interesting Yeah, well, I've just come back from a conference where I was discussing for a session that was all about using AI, to mark. So where assessments don't change from one year to the next, unlike here, where they're always changing AI is, has been used for many years. And now there's the possibility of using generative AI, which is much speedier, and it needs less training, even in scenarios like the UK where the questions are changing all the time. Because you don't have to change generative AI in quite the way that you do traditional AI. Having said that, we are years away from being in position. But yeah, I can definitely see that there's a role for AI, not necessarily in first marking, but perhaps in quality assuring the marking that goes on. Is there a role for AI in the generation of gray boundaries? Wow, that's a really interesting idea. I mean, the bit that we really struggle with is this issue of comparing performances over time. Because obviously, the statistics that were used tell us nothing about whether the actual student performance has got better or worse. It just tells us about who the cohort are. So whether there's a role for AI there in actually analyzing the qualities of the scripts. That's a very exciting possibility. But we're a long way away from it.

Jonny Tridgell 33:42
And with lots of risks, I suppose as well of the way AI can reproduce bias as well. I think sometimes that assumption that AI will do it neutrally is unhelpful as well, isn't it actually. Yeah.

Michelle Meadows 34:37
So people tend to think AI is going to make everything more efficient. Yeah, maybe on the the cost side, there's all the validation work. Really, really careful validation work that has to go On to, as you say, check the validity of the outcomes and look for bias. And so although you're making a saving in one place, there's going to have to be a massive investments elsewhere to take real care over any potential use of AI, Brave New World.

Hamish Chalmers 35:16
Johnny, when the A Level results come in, I'm just now reflecting back on my moment when I picked up that manila envelope, and then was too scared to open it, what's your experience of that with with the young people that you work with and what might happen if, in essence, they get the wrong grade, or they don't get the grade that you predicted? Or they they don't make the the the grades that they need to go to where they want to go next?

Jonny Tridgell 35:37
I've come at this from a few different angles. So I was heterosexual and for a long time, so we used to get the results the day before. So that sense, you're sort of caught, you have to fly, right? I know that this might not have gone well. There's discussion says plans in place, it will depend what the school have got in place, perhaps for that young person, if they know the results haven't got quite there if they put some plans in place for what the conversation would be like. And then are sort of just done as a classroom teacher who doesn't have that, that extra knowledge or just being in the room on results day and seeing them open and basically just watching their faces and see what happens next. I think the wrong grade is is brutal, I think good. That kind of what that means for that person. And it's always interesting to see because things sometimes it doesn't matter, because I've already heard from UCaaS that are being told they've got in and they will be completely unbothered, they'll be upset about it later. Because they'll be like, I'm annoyed, but I got in. So I don't care. Like that's fine. I think the bit when it's I've got the wrong grade. And that's changed my future. It's really difficult. And I think for teachers is obviously difficult for us in a different way. It's not our lives that have been changed. But also there's a sense of feeling quite helpless to do anything about it. And I think also lots of worry about is this my fault? If I done something today teach this badly? Have I let this child down and then have I ruin their life is there's what's happened there. So I think there's a lot going on on results day for teachers. It's quite a roller coaster, actually very empathetic with the children. You've got a kid celebrating and then you feel brilliant. If you've got a child crying in front of you, then that's obviously just heartbreaking. And then also yes, your own kind of emotional reaction if my class did really well, oh, God, it all went horribly wrong. What happened? So I think it's there's a lot and I think a key people, people results day when I was head of Sixth Form or anything else was desperately trying to stay calm, and be kind of rights, we're

going to be calm, we're going to help solve problems in the top people off ledges, we're going to make sure everything's okay. But also recognizing there's not that much power to do that.

Hamish Chalmers 37:41
Yeah. And that brings the question for Michelle. So given all of the work that goes into trying to be as fair as possible in setting great boundaries and marking the work in the first place, is there any sort of avenue for children who have young people who want to challenge the grade they've been given? Or is it so sort of meticulously done, there is no space for challenge? Oh,

Michelle Meadows 38:00
But look, assessment is never perfect. And, you know, we're talking about getting the wrong grade. And you can get the wrong grade for all sorts of different reasons, of course, because unfortunately, some kids will just have a bad day. And in way, in a way, they've got the wrong grade. You know, you might feel that as a teacher, and, and they may feel it, because you, you know, they could do better under different conditions. And that's really sad. And then there's, you know, the questions that come up on the day, played some people's strengths, and not others, you know, assessment is an imperfect thing. But often, what we're talking about really is where we think there's a problem with the marking, so something has gone wrong with marking, and then absolutely, then there is an opportunity to appeal and have the marking checked. And you know, where we're talking about a levels there, you know, you can get that turned around really quite quickly. So that, you know, the student can still go off to the university that they want to go to. We have a system where for a levels in particular, we have long essays, etc. There is room there for differences of opinion, that actually are entirely legitimate about quality of the work, but there's also room for error. And if there's a mistake that's happened, then, you know, getting the example to fix that mistake is really important. And that's what the appeal process is for.

Hamish Chalmers 39:40
Presumably that takes resources of some sort. If not time, then money that people have to pay to do this. Is there a difference between for example, better resource schools being able to or more willing to challenge results than less well Resource schools, is there an issue around equity in that part of the process?

Michelle Meadows 39:58
Yeah, I think there really is a, it's one of these really wicked problems. The way it works is that if indeed, the exam board finds that there is an error in the marking, even if it's just one mark, if there is an error, then you don't have to pay. Nonetheless, well resourced schools are in a better position to be able to appeal more of their grades. And that feels naturally unfair is unfair. There have been conversations about whether it's right that anybody has to pay for their marketing to be checked. The issue is, imagine the volume of appeals there would be if nobody had paid for you get everybody's challenging, right. And then essentially, what would be doing is remarking everything, reviewing everything, rather, in a very short period of time, what's the quality of that process going to be like? If anybody's got any smart answers to this, that would be fantastic. So I think it is hugely

problematic. It feels dreadfully unfair. But without any resource implication for the school or the family, then we would have a system where everybody, well, perhaps 90% of people would appeal, understandably. And then we just have second marking, and I'm not sure it'd be any better than the first marking.

Hamish Chalmers 41:32
I want to finish up now on research that's been done on this whole process. And I understand that you've been doing research with a qualification Wales, on setting great boundaries. So perhaps as a takeaway for our listeners, we can you can tell us a little bit about the research has been done on gray boundary setting and point people in the direction of any materials that they might want to follow up this podcast with.

Michelle Meadows 41:53
Yeah, well, there's a huge history of research in this area into the quality of human judgment, you know, How good can an examiner be in terms of trying to identify the exact mark that a boundary should lie on and research into the quality of the statistics that she used as well, you know, what's the best kind of statistical information? And how should the two be blended? And actually, if you look around the world as well, there are different approaches to setting standards. So the approach that we use here is attainment referencing, it's got that blend of judgment and stats around the world, we see criterion reference systems that vary in terms of how strictly criterion reference they are, we see criterion referencing in vocational assessments. And we've already talked a bit about norm referencing. There are other approaches such as cohort referencing, which is where a specific proportion of students are given a grade, regardless of the particular entry for a subject that's used around the world. So there are all these different approaches. And the particular piece of research that we did, funded by qualification, Wales, was to look at all of these different approaches, summarize their strengths and weaknesses. And actually, this was worth focusing on GCSE rather than a level. And I was essentially asking, if we were to take a different approach to standards, if we were to say, this criterion reference thing for GCSEs. What would that mean, for GCSEs? How would they be designed? How would each each stage of the qualification lifecycle look different? If they're a criterion reference? What if we norm referenced? How would it look then? And that's been incredibly helpful, I think, because it just brings together what is a huge literature across many countries and asked that question. So you might be interested to know that GCSEs were originally intended to be criterion referenced. And however, it failed, because of the sheer volume of criteria that had to be created for any particular subject, and the level of specificity associated with each of these criteria.

Jonny Tridgell 44:24
So it's criterion referencing a bit like when you do your driving test, like sort of tick off, you can do X, Y, and Zed therefore, you get this. Yeah, did that.

Michelle Meadows 44:30
Well, that's the idea behind it, that if we can just specify all of the criteria that a student has to meet to get a particular grade, and then it's normally based on teacher assessment normally rather than exams, and the teachers will judge Yes, this student

can do this particular piece of maths. The problem is that actually just because of the vagaries of language, specifying the criteria in such a way that teachers can apply them consistently is harder than you might imagine. So you get differences between schools, etc, which is why you end up with this massive volume, a very, very specific criteria. And then that begins to have a wash back on teaching and learning. So there's a little phrase I like, which is that the assessment becomes learning assessment as learning and that's what can happen. I think

Hamish Chalmers 45:33
anyone who's used a scoring rubric for anything will find that familiar. But we will link to that research in the show notes to this podcast. Before we sign off, anything that we've missed today,

Michelle Meadows 45:45
no, I could talk about this for ever.

Hamish Chalmers 45:50
Well, I reckon we'll have you on again for another conversation. I think that's a good place to leave it. So I just want to thank you, Michelle. And thank you, Johnny, very much for joining us today and sharing those valuable insights into the process of a level grading. I want to finish by saying good luck to everyone getting their results tomorrow and to their teachers who will be supporting you through that. Or if you're listening to this podcast after the fact. I hope that you and your students got what you were hoping for. future starts here.

Michelle Meadows 46:16
Indeed.

Laura Molway 46:19
You've been listening to the Oxford Education Deanery Podcast. To find out more about the research featured in today's episode and everything else that deanery has to offer. Visit our website and sign up to our newsletter. You'll find the links in the show notes.