

Education Committee

Oral evidence: [The use of Artificial Intelligence and EdTech in Education](#), HC 151

Tuesday 8 September 2026

Ordered by the House of Commons to be published on 8 September 2026.

[Watch the meeting](#)

Members present: Helen Hayes (Chair); Sureena Brackenridge; Jodie Gosling; Darren Paffey; Manuela Perteghella; Mark Swards; Peter Swallow.

Questions 34-80

Witnesses

[I](#): Professor Alison Clark-Wilson, Professorial Research Associate at UCL Knowledge Lab, UCL Institute of Education, Professor Rebecca Eynon, Principal Investigator at the Oxford Internet Institute and the Department of Education at the University of Oxford, Darren Northcott, National Official for Education, NASUWT and Pip Sanderson, Director of Teacher Development, National Institute of Teaching.

[II](#): Andrew Freeman, Executive Partner, Pearson Qualifications, Gillian Hillier OBE, Chief Executive, Standards and Testing Agency and Colin Hughes, Chief Executive, AQA.

[III](#): Sir Ian Bauckham, CBE, Chief Regulator, Ofqual.

Examination of witnesses

Witnesses: Professor Alison Clark-Wilson, Professor Rebecca Eynon, Darren Northcott and Pip Sanderson.

Q34 **Chair:** Welcome to this morning's evidence session, the second in our inquiry on AI, EdTech and their role in our education system. I should say at the outset that we have three panels this morning, meaning an unusually congested agenda for the Committee, so I will be reasonably strict on time and would appreciate it if witnesses and Members could keep their contributions as succinct as possible, to help us get through all the topics we need to cover. I will invite our witnesses to introduce themselves.

Professor Clark-Wilson: Good morning, everybody. I am Alison Clark-Wilson, a professorial research associate at University College London's Institute of Education, with a background in mathematics education as a former school maths teacher in the UK.

Professor Eynon: My name is Rebecca Eynon. I am professor of education, the internet and society at the University of Oxford, where I am jointly appointed by the Department of Education and the Oxford Internet Institute.

Darren Northcott: Good morning. My name is Darren Northcott. I am the national official for education at the NASUWT.

Chair: Welcome back for the second week in a row.

Pip Sanderson: Good morning. I am Pip Sanderson, director of teacher development at the National Institute of Teaching. I also head up our new Centre for Digital Information Literacy in Schools. I started my career in education as an English teacher.

Q35 **Chair:** Thank you very much. To start our questioning, what impact is AI having on teachers?

Darren Northcott: I think it is having an increasingly significant impact, in terms of both the AI that children are accessing at home and the increasing prevalence of AI in the education system. We might explore this in a little more detail later on, but although there are definite advantages and positives to that, there are also risks that need to be managed. Clearly, AI is becoming a much more important part of the lives of children and young people, and an important part of their education.

Pip Sanderson: The Commission into Countering Online Conspiracies in Schools first published in 2024 to report on the growing problems of misinformation and conspiracy theories in schools. They recently did a resurvey and published new findings that showed that there has been a significant growth in the problems that conspiracy theories are presenting in schools, from pupils to teachers. That growth has largely been identified as being driven by AI, so I think that that is a particular aspect that is often overlooked when we think about the impact of AI in schools and what we need to consider.

A second point to add is that, from the teachers we speak to daily, we know that many fantastic teachers are using the technology in a very judicious and thoughtful way to try to enhance the quality of their planning and resources, but we also know that there are a huge number of teachers who are ethically deciding to opt out of the use of AI because of their environmental, ethical and cognitive concerns. We are at a very mixed point for the sector at the moment as regards how AI is impacting in schools.

Professor Clark-Wilson: I totally agree with Darren that it is impacting, and the rate and scale to which it is impacting depends very much on the approach to innovation in the different institutions, schools and colleges. Those that are always at the front and are more innovation-focused, with processes to think about how it is impacting, are in one place, whereas we have many institutions that are holding back to see what others are doing. It is quite a mixed bag.

Professor Eynon: At the moment, a lot of the technology we are seeing in schools is very much designed to automate and then theoretically augment teachers' work, but at the heart of a lot of these claims is a very strong efficiency logic that does not really think about teacher autonomy or teacher expertise.

In some extensive ethnographic work we have done in secondary schools, we have really seen how these automation logics have compromised teacher expertise. This can happen, for example, in the ways they remove information for teachers about how to teach their subjects. For example, when they are using AI for lesson planning it can remove information about how to teach their subject through the black box nature of the AI that they are using.

In contrast to augmentation, we are also seeing teachers use an awful lot of time making this technology work for them—dealing with tech failures, passwords, log-ins and all these kinds of things. That means that it is not really augmenting in any meaningful sense and it is reducing the potential for teachers to be autonomous and expert in their teaching.

The last point I will make, very briefly, is that there is this claim that technology will supplement teachers and enhance what they do, but what we are seeing, particularly in schools that serve more disadvantaged communities and have fewer resources, is that they are starting to rely on these technologies to some degree and there is some form of automation happening in some schools.

Q36 **Chair:** Do you think there is a need to think about the fundamental purpose and role of teaching in the light of the prevalence of this technology? Rather than equipping teachers with traditional teaching skills and hope for the best in the context of AI, should we think in a more profound way about what teaching means in this context, and if that is the case, what would that look like?

Professor Eynon: Very much so. We need to start to change the way we think about what teacher skills might mean and what teaching looks like in these contexts. I would like to see a much stronger basis for these discussions based on teacher expertise. We are seeing a lot of the agenda being set by the commercial companies that build these products rather than the teachers or the education community. What we need to encourage in our schools and our education system is a critical literacy for teachers that allows them to have the language, space and time to critique and push back, as well as to refuse sometimes and to ask for alternatives or different kinds of technology. There is something about a shift from an agenda set by technological and commercial agendas to something much more educative.

Professor Clark-Wilson: I would go beyond the teaching practices in terms of the changes that are needed, because unless we fundamentally look at the curriculum and the purpose of education more broadly and really think about what we value and how we assess it, we are tinkering rather than taking the bigger piece. It is undoubtedly changing everything,

so we have to really think about what we teach and what we assess, because we know that that drives much of what goes on in schools and colleges.

Darren Northcott: One of the things that our members report to us is the risk that they encounter that technology, rather than pedagogy and thinking about the aims and objectives that we want children to secure, leads what happens in classrooms. That is quite a long-standing problem. I think back to when I was teaching and interactive whiteboards came in. They were installed in every classroom, and then afterwards we thought, “Well, how are we going to make use of this?” Rather than thinking about how technology can support learning and achievement, the technology arrives and then we try to retrofit educational objectives to it. As I say, that is quite a long-standing pattern in the education system in England, but we see that very strongly in respect of AI.

Pip Sanderson: The idea that the use of AI in education is inevitable is one that we need to question. It is a choice that we as a sector are going to have to make. While many schools are choosing to use it, there are many that are choosing not to use it. We need to look at the reasons why, and not assume that we need to adapt what teaching is to accommodate a technology that we have a choice about whether to use. I am not denying that AI is prevalent in changing society at large, but we do have a choice about whether we want to bring it into our school system.

Q37 **Manuela Perteghella:** Is there a risk that increased reliance on AI and EdTech could reduce teachers’ professional autonomy or expertise over time?

Professor Clark-Wilson: The answer is undoubtedly yes. Without really understanding the professional working practices of educators that we might offload—essentially we are offloading to a technology—the risk comes from a lack of trust in the school, classroom and parental environments around who is making judgments, who is providing feedback and who is creating the lessons. If we do not have that very explicit, and discussed and agreed within schools and institutions, I think that is where we can lose the trust, and that will not help the profession.

Professor Eynon: I would echo what I said in my opening response. We absolutely need to think about these questions of expertise and autonomy. In the work that I have done, where we have supported teachers, recognised their expertise and, crucially, supported their autonomy in deciding how and when to use technologies and how to teach, that is when we have seen the most educative uses of technology.

Darren Northcott: I cannot but agree with those two contributions. The idea of that technology, and particularly AI-driven technology, replacing that professional judgment and discretion over children’s learning, pedagogy, assessment, planning—all those key professional functions—is something that teachers are very concerned about. Many teachers are nervous that what could happen, and perhaps in some instances is starting to happen, is that those professional functions are displaced by AI in a way

that not only undermines teacher professionalism, but will undermine the learning experiences of children and young people.

Pip Sanderson: We urgently need more investment for test and learn in this space, and an expansion of the remit of the funding that has been committed for research in this area to include teacher development, which is currently not defined within those parameters. We know there are real problems in the sector that we want to fix. One is how to ensure that all trainee and early-career teachers have access to high-quality, consistent mentoring. It is really difficult for schools to provide that, so is there an opportunity for AI-assisted observations to allow expert teachers to see what is happening in novice teachers' classrooms and provide support? The simple answer is that we do not know what the impact of that will be, but that is a problem that needs to be solved somehow, so why don't we test the technology and evaluate the impact? There could be great gains, but there could be unintended consequences that cancel out those great gains.

Q38 **Manuela Perteghella:** That brings me to my next question. What safeguards would you like to see to ensure that AI and EdTech have a supportive role in the classroom, rather than undermining teachers' professional judgment and expertise, and reducing their role in the classroom of inspiring students' critical thinking skills?

Professor Clark-Wilson: There are a number of safeguards across that piece, but there is one that I think is a quick win. China introduced bold legislation on the role of agentic AI inside all EdTech products for minors. That means that they have banned the use of human-like personas and avatars. That is a subtle but important distinction that means that young people are not going to start forming relationships with agents that are not humans in ways that we know are problematic in later years. There are some interesting ways of thinking about different parts of the piece, but across all of it we have data governance, data sources, data security and interoperability—lots of things that affect how data flows in and out of these different AI tools.

Professor Eynon: We also need to encourage a lot of precision about what we are talking about when we talk about AI and technology, and what we are rejecting or accepting. It is really important to think about the specific affordances that we find acceptable or not within education. It might well be that these are certain kinds of AI that use data-extractive practices that cause a lot of privacy concerns or that it is impossible to understand the level of bias or inaccuracies in these technologies. We want to be precise about the kinds of things we are unhappy about with technology and what we want to protect and design for.

Alongside other things that we have talked about already, we need to expand the kinds of evidence that we see as important and to think not just about the kinds of outcomes of particular educational interventions using technology, but also about the processes, unintended consequences and a range of other things. When we use technology, what does that mean to the experience of young people engaging in education? When we

have a stronger way of accounting for teacher knowledge and ways of thinking about evidence, we can start to shift the procurement requirements of schools and so on. That starts to change the demand structure around technologies and starts to shift the market in ways that we think are important, thereby safeguarding the things that we think are educationally important.

Darren Northcott: That point about schools being informed consumers of the technology they buy is important. There is a real risk that expensive packages are being pushed hard on schools—we have finite resources in schools—and that those investments will not be as effective or safe as we would want them to be. I agree with what my colleagues have said, but I would add two more things.

One is around the interaction between AI and data protection and privacy, which concerns us a lot. Again, we are not confident that trusts and schools across the system really understand those data protection and privacy risks that can be attached to the use of AI: the uploading of information and data that may contravene GDPR. Part of the training and support that we are thinking about here has to involve how you use AI—if you are going to use it—in ways that respect the data protection and privacy rights of staff and learners. I think that the fundamental principle in terms of safety that we would want to see applied is pedagogical primacy: the idea that technology, when it is used, supports teaching and learning, as well as the professional decisions that a teacher makes, and does not supplant them. To me, that is a really important safeguard for, among other things, the quality of children's educational experiences.

Pip Sanderson: The prominent cognitive scientist Dan Willingham wrote about this topic recently, and he had a simple principle that I think is useful to say out loud. He said that we need to ensure that AI tools are not used for “tasks wherein students would benefit from doing the mental work themselves”. If we have a guiding principle as simple as that, it might help shape the work we do.

Picking up on Darren's point, the simple answer here for safeguarding is to ensure the quality of training. A massive burden has been put on schools to try to reconcile and understand what on earth to do with this technology. They have found money from unbelievably finite, small pots to create AI champion roles; they have colleagues who are working—unpaid—above and beyond to try to upskill and understand this technology and to think about what that means for their pupils. We need to support the sector and ensure that every teacher in this country has access to high-quality, evidence-informed training, not just in how to use the tools but in how the tools work, the ethics of use and the considerations around cognitive offloading. We want to upskill those in our profession to make these choices and informed, intentional, evidence-led decisions themselves.

There is an opportunity to do that: we are lucky in this country, because we have national statutory frameworks. But we have nothing at all in those frameworks that talks about AI. An urgent update to those

frameworks would ensure that the many thousands of teachers already accessing professional development in this country would immediately start to get high-quality AI training, rather than leaving it largely to tech companies that are obviously providing that training from a particular angle.

- Q39 **Manuela Perteghella:** As a lecturer, it was fundamental for me to plan my lectures, and particularly my seminars. That was because I knew my students. I knew the class dynamics and I knew their needs. It was not just an admin task; it was really the foundations to then foster critical thinking and share knowledge in the classroom. Are there particular areas of teaching where AI should play only a limited role, regardless of how capable the technology becomes?

Pip Sanderson: Absolutely. The particular challenge this makes me think about is the novice new teacher versus the experienced teacher. Headteachers have an absolute challenge at the moment to try and think of a policy that works for both those colleagues.

We know that helping an expert to work more effectively is not the same thing as helping a novice become an expert. How on earth are we meant to have an all-encompassing policy in a school for somebody who needs to do the hard yards—the cognitive grapple—of learning how to plan, which we know is foundational to being a great teacher, but also ensure that our colleague with 20 or 15 years of experience is able to experiment with how AI might enhance the quality of their lesson planning? How on earth we wrestle with that is a really tricky situation for the sector to try and unpick.

Darren Northcott: That is right. You are absolutely right about planning as well. A lesson plan is a product. Planning a lesson is a professional process that a teacher goes through, and it is very important. That point about how teachers coming through initial teacher training now might not be having those experiences of thinking about planning and doing that planning as a kind of cognitive professional exercise, particularly when they are in schools, is a concern.

However, that is not just AI. We hear from student teachers who go on school experience—what used to be called teaching practice when I was training to teach—that when they are deployed in schools the plans are simply given to them to follow.

We know that some initial teacher training providers are very concerned because, among other things, to meet the standards to be awarded QTS, teachers have to be able to demonstrate that they can plan lessons. If the school you are in is handing you lessons and basically telling you to read that off a script, that is concerning.

AI just adds to that concern because, clearly, if you are just presented with something that AI has churned out, you are not going through the process that you described that is so important to high-quality provision.



Professor Eynon: I would echo all of those comments. It is definitely something that we have seen in our research. People are very concerned about AI for planning lessons, particularly for less experienced teachers and what that might mean. Thinking about this intersection between expertise and AI is going to be crucial. As you say, we also really have to think about what is worthwhile to people—what is meaningful to people in doing their job well—when we think about encouraging people into the profession and encouraging people to stay there. We therefore have to think more carefully about how we use AI as part of that process.

Professor Clark-Wilson: I am involved in two teacher choices studies that the EEF is currently conducting. One is looking at the use of co-pilot to support lesson planning and the other is looking at a particular EdTech product that is there to support assessing in specific marking of student work.

One of the red lines in providing guidance for teachers is around grading. Currently that sits in different institutions' policies. Some are very hard line in that you cannot use an AI to actually grade students' work, particularly in anything that has a summative piece of coursework, for example.

It is about thinking about where AI sits in the process of assessing. Is it a red line that you should always have reviewed the work yourself first before putting it into an AI engine to take that task on? This is where we really have to understand where the benefits are.

The allure is there, and I can see many teachers looking to cut time by fully automating those processes, but all of the same points around lesson planning have to go alongside the process of assessing. It is critical to a teacher or lecturer deciding next steps for that group of students. We have to look really judiciously at how we embed some of these tools if we see there are opportunities.

- Q40 **Sureena Brackenridge:** Following on from those points, the Government have said that AI can reduce teacher workload by automating routine tasks, but will it impact on wellbeing? Will it be a sustained reduction of workload? We have also heard evidence that it may just be a reconfiguration of workload, with additional responsibilities put on to teachers. How strong is the evidence that AI is reducing teacher workload in practice?

Darren Northcott: I do not think that the evidence is very strong at all that it is reducing teacher workload. There is a concept, if you like, of task efficiency—can a task be done more quickly? We surveyed our members earlier this year. Around three quarters had used AI in one form or another in their work, but only 7% had said it had reduced workload. We need to explore why so few teachers are reporting that, overall, the use of AI is reducing their workload.

There could be a number of reasons for that. One is that just using AI in and of itself creates additional workload burdens. For example, if you are



HOUSE OF COMMONS

using AI to support planning, that can be very helpful, but if you are doing that appropriately, you are then going to check all the citations and all the evidence that the AI platform has produced in developing that lesson plan. So, as you said, it is kind of transforming the workload so it takes another form, but the overall amount of workload is not reduced.

Another point, which goes to the point that we were just discussing, is that many teachers—if they have the choice, and we know that some don't—are very wary of just using an AI platform to generate a lesson plan that they then deliver. That seems to be sold as the biggest workload reduction benefit, but for very good professional reasons, teachers are nervous about doing that, for all the reasons that we have discussed.

The jury is very much still out on the extent to which AI currently reduces workload and also the extent to which it will, in the long term, lead to sustainable workload reductions for teachers.

Pip Sanderson: I very much agree that there is limited evidence in this space and an urgency to do more evaluation.

One research trial done by the EEF looked at 259 secondary school teachers and their year 7 and 8 marking. They found that those colleagues using AI to help plan their lessons saved 25 minutes a week. What they did with that 25 minutes was not part of the trial. There really is a great need to delve deeper into this area to understand.

One of the reasons that is worth exploring is the lack of training, meaning that when people are using AI, they are maybe not using it in a judicious way. If you don't understand how it works, you are maybe going to ask it to do the wrong kind of tasks that don't actually help you.

To your earlier point at the start of your question, in terms of back office function, we have spoken to many schools who are doing brilliant, innovative things with back office, non-people-facing functions, which are speeding up administration tasks and efficiencies there. Again, we do not know what the time saved turns into, in terms of pupil impact. Our north star has to always be: what is the impact on pupils?

I very much agree with what Darren said about this urgency; you will get bored of me saying it, but we need to desperately do more evaluation in this space urgently.

Professor Eynon: I would also very much support a reconfiguring of teachers work. I am not sure it is ever going to save anybody any time—does email save us time? I don't feel like it does.

Technologies often promise to be time-saving, but in our research we very much saw teachers saving time to some degree and using it for very repetitive or non-teaching tasks—maybe a bit of lesson planning, particularly for lessons that are not their core responsibilities, so when they are covering lessons, for example. But it is also taking time. It is also taking time to think about how we teach to ensure that we do not have cognitive shortcuts. How do we manage all of the tech and the distraction



in the classroom? How do we manage all of the laptops? It is giving and taking time in the same moment.

Reconfiguring allows us to see the complexity of when we put technology and work together. One of the challenges with the research in this space is not just that we have not got enough of it; it is also that it is very hard for people to measure time.

It is also about looking at it conceptually. For example, in that EEF study, which is a good quality study, we are looking at the quality of the lesson plan, but we are not looking at the implications on actual lesson practice. Even if it saves time, we don't know the downstream effects of that. It is a complicated space.

Professor Clark-Wilson: I don't have much more to add to that.

Q41 **Darren Paffey:** Looking ahead, learning from what we are seeing now, where do you see the biggest potential to make sure that AI is reducing workload and improving efficiency? Where should we be looking for that in the school sector?

Darren Northcott: I will pick up on what Pip said on the potential around what you might call back office functions and some of teachers' administrative and clerical burdens. That might be an area to explore because, compared with some of the higher stakes things we have been discussing, the stakes there are lower. Those might be things to consider.

It still has to be done extremely carefully. I would go back to the point about schools, trusts and employers not really understanding things around data protection and privacy. A lot of work has to be done to upskill and enhance the knowledge of schools and employers on that, but it might be an area where we can focus our attention.

We need a lot more evidence, as colleagues have said, around how some of the higher stakes stuff, particularly around pupil-teacher interactions, can be done safely and effectively. There are some opportunities, but we still need to be quite cautious.

Pip Sanderson: Marking is one opportunity that is worth examining. If it is done well and judiciously and is a complement to the human-led marking, potentially there is a role there. In certain subjects, the marking burden is huge. As an English teacher, I can speak to how long it takes to mark a class of 30 mock exams.

If there is an opportunity to use some of the really good technology out there that is using comparative judgment along with AI—actually, some of the earlier studies are finding that there is a reduction in some biases and more consistency in the distribution of marks—we need to test that and see what the consequences are of introducing a judicious deployment of AI marking, in areas where it can be an aid to the teacher and free them up.

It comes back to the point about what a teacher can do with that free time. What do we imagine is going to suddenly happen with that free

time? If nothing is put in place in terms of what we want—whether a chance for increased professional development or an opportunity to rehearse and practise lessons more—we will not see the gains land on our pupils.

- Q42 **Chair:** We had evidence in our first evidence session from an organisation that had surveyed teachers, which concluded that what teachers did with the extra time was mostly more assessment. That points to the need for exactly that guidance and further information.

Professor Clark-Wilson: I would bring in my experience of working with the sector at this point, because one of the things I found in working alongside EdTech companies—some of them at the early stages of developing their tools and services and others quite the big guys—is that they are all unclear of what the impact goals for the technologies they are building actually are. They are there to sell a hype, really, and people are buying into it.

The opportunities that the new DfE EdTech impact testbed project is giving us are worth recognising because it will enable us to work more closely with the tech partners, as well as the schools and practitioners, in part to generate the much-needed evidence, but also to understand those different perspectives. There are different goals involved. We have to be the guardians of the pupils and the system, and that is where guardrails and legislation come in, because we are much closer to seeing what the problems are when we work alongside the developers, rather than always seeing their tools as something to be externally evaluated.

That test-and-learn model, which will allow us to take promising practices from schools and colleges, work with researchers and their companies to really home in on what the impact we are looking for is. If it is workload, what aspect of workload are we talking about? We will have a much finer body of evidence, hopefully at more scale, to draw on. I would love to see that project be sustained long after the funding is going.

Can we change the culture in our schools and colleges to enable them to have an evaluation mindset on any innovation? It is AI now, but it was the whiteboards 20 years ago. We have seen these cycles regularly, and we have never quite found the right school-level and college-level processes that are manageable to help us do that evidence generation at home base.

Professor Eynon: I would still campaign for a way of thinking about this as a way of reconfiguring work and not saving time. That then encourages us to think about the meaningful work that teachers do and trying to increase that through technology, rather than efficiency and time saving. I do not think work is like a cake that we can cut up into bits and save a quarter; I think it happens in a much more complicated way. Thinking more about meaningful work would be much more useful for the sector.

- Q43 **Darren Paffey:** Do you see the benefits of AI varying between primary and secondary phases at all? Is there one part of the educational journey where you see benefits land more or they are able to take more advantage



of that reconfiguring? Are there any thoughts on that from anyone?

Professor Eynon: Most of my work is in secondary school, so it is a little bit difficult for me to answer that question, but the nature of teachers' work in secondary is obviously quite different and therefore I think the possibility for reconfiguring is maybe greater.

Pip Sanderson: If we think about the teacher's use of AI, it is irrelevant whether it is primary or secondary if we keep focused on the principle of whether they are using it to support their development and their expertise or whether we are using it as a shortcut to replace that. I think we could do some good work if we keep that guiding principle front of mind.

- Q44 **Darren Paffey:** If you make the distinction between the benefits to the teacher process and the ultimate benefit to the learner—the child—do you see any difference between primary and secondary?

Professor Clark-Wilson: It is about where we place the education about AI. Even if we decide to steer clear from it, we still need to educate our young people about it. How we do that in the earlier stages looks very different to the later stages. Ultimately, we want young people to come out able to programme it. We have a huge need there to upskill, and to think about how that education reform will allow us to upskill our population to live in the world that lies ahead of us.

Pip Sanderson: I think it is worth thinking about AI in the context of the age that we live in. We are living in a time of information system change. When our information systems change, we change as a society. It happened before with the invention of writing and the printing press and it is happening again now; we are living through that moment. That teacher-pupil relationship is becoming ever more important in this age, where we see that trust is decreasing; we have an abundance of information but an increasing scarcity of trust. Whether you are primary or secondary, a core focus for our teachers is giving them the support to focus on that interpersonal relationship development and how we can build those relationships with our pupils in the information climate that we live in. The good news is that pupils generally trust their teachers, particularly at primary, so there is something interesting to think about at primary in terms of being really careful about how AI possibly could be an interference in that really beautiful teacher-pupil relationship, which sets young people up to go on to do great things.

- Q45 **Jodie Gosling:** The Government have highlighted how AI can help out in retaining the 6,500 new teachers. How significant a contribution can AI actually make in the crisis that we have?

Darren Northcott: It is about exploring on what basis the Government hold that belief. There are two strands to it; I am much more sceptical about one than the other. The first is workload. As we have just discussed, the Government's narrative seems to be that AI can exert downward pressure on workload. We know that workload is the biggest driver of exit from the profession and I think it deters lots of graduates from thinking about teaching as a career. I do not think the case is proven that the

increasing use of AI will reduce workload in a way that supports addressing some of the retention challenges that we face.

The other area may be more interesting to explore. Perhaps the Government think that people coming into the workplace and graduates coming into the labour market now would expect to have AI tools available and would want those to be used as part of their work; that becomes part of the expectation. That is reasonable, but I do not think any of it addresses the fundamental drivers of the recruitment and retention crisis, which are really well established. We know that workload, teacher wellbeing and pupil behaviour are cited, frankly, year after year by teachers and leaders as reasons why there is such a challenge in recruitment and retention of teachers in the profession. AI may have a role to play but at the moment I think the safest thing we can say is that it is going to be, at best, at the margins. There are root causes of that crisis that we, and the Government, need to address.

Professor Clark-Wilson: I tend to think more about the incentives than about the challenges and the issues. I look back to some of the incentives that were around that successfully helped teachers to stay in the profession. We heard mention of AI champions and digital champions. It is about recognising in the system the value of these new people and the education workforce that we need in our schools now, be they education data analysts or education AI specialists. These new leadership roles, which do not necessarily fit in the existing structures, are going to help us get on top and stay ahead.

To think about what that looks like in the professional routes for teachers, in some ways, it opens up opportunities to stay because the workload is often same old, same old—you do not necessarily have interesting progression routes within the system as it is now. There is most certainly an opportunity to see how this aligns with teachers being able to gain masters credits or do something professionally rewarding. Often, or inevitably, it takes them out of the classroom, but they stay in the system.

Q46 **Jodie Gosling:** Do you think that the increased use of AI or EdTech affects the desirability of teaching for new entrants or existing teachers?

Professor Clark-Wilson: Not yet, but I think that, with some really good marketing, PR and case studies showcasing how it is transforming what is going into schools, it could.

Pip Sanderson: In an odd way, we might see an increase in the number of people wanting to become teachers, because it becomes one of the professions where there is still a lot of human interaction. Where that work appeals and it is largely AI free, that may become a massive plus point for teaching. In that way, we may see a side benefit from AI in terms of teacher recruitment and retention.

From talking to teachers, one thing that I think is important to say is that we are often focused—rightly so—on evaluating efficacy and impact on young people, but there is also something important about how AI makes



HOUSE OF COMMONS

teachers feel about being a teacher. That sounds a bit soft and woolly, but it is really important. How do I feel as a teacher? If you take away my lesson planning, which is my chance for creativity and my ability to think hard about my pupils, do I still want to be a teacher, just delivering something that somebody else has created? That might appeal to some people who would not have thought about teaching, but I think we would see a lot of people who currently teach well suddenly finding that career path much less desirable.

Professor Eynon: I agree with all the things that have been said. That is why I am arguing for the idea of meaningful work and a reconfiguring of work, rather than efficiency gains, because it is about creating meaningful work that will attract people to, and retain them in, the profession.

Another aspect of this involves thinking about public perceptions of teaching, which is really important. There are some really interesting tensions that our research has shown. Teachers have talked to us about their concern that the public might have conflicting ideas about when they use AI. On the one hand, they are forward thinking, modern and so on, but on the other, are they being lazy if they use AI to create reports? Do they not have expertise if they are using AI to generate their lesson planning? There is this tension in terms of how the public are also thinking about AI and then thinking about the profession, which may well have implications for people being attracted to it.

Q47 **Jodie Gosling:** Pip, the National Institute of Teaching has said that AI may be used to identify and support trainee teachers who are at risk of leaving the profession. What evidence is there showing that those systems improve that?

Pip Sanderson: Very little. The evidence is theoretical at the moment. Some studies were done recently, in 2024, but they looked at how AI was used to help identify students on degree courses who were underperforming and likely to drop out, so targeted support could then be put in place. As far as we understand it, no testing has been done with teachers specifically to see how that could be done, but theoretically it could be done. However, that opens up a whole host of issues in terms of that level of monitoring of a student's or a teacher's performance.

Q48 **Jodie Gosling:** We have heard that many teachers lack confidence in using AI and EdTech effectively. What training do teachers currently receive, and do you think it is adequate?

Professor Clark-Wilson: I can talk from the perspective of initial teacher education. It is inadequate in respect of digital technologies. There just isn't enough time in the programmes, or expertise in the schools where students are placed, to provide that. That is a problem and has been for 25 years; it is not a new one.

Q49 **Jodie Gosling:** What do you think teachers need to use it effectively?

Professor Clark-Wilson: There is a set of competencies around AI that are a good starting point. These have been developed by UNESCO as an

international set of competencies. They outline things quite generally. Making sense of them is in the practice, so it is about mentorship and support in the school setting, at all stages of a teacher's professional career, for them to continue to grow and learn.

Pip Sanderson: We have a real problem in that, if we do not train the mentors and the experienced colleagues, and train only the trainees, we will get a divergence and not a cumulative coming together.

In terms of what AI training should contain, the one thing missing in most provision from tech companies is anything on the social and psychological effects of AI use. It is not talked about in the broader landscape of the changing digital information age in which we live. There is very little attention to the cognitive science of how learning happens in relation to what we know about how LLMs work with probabilistic inference. We have found from our own courses that those two things together help the lightbulb moment come from trainees in thinking, "This is why AI might limit my own learning, and actually, maybe it could be useful in this little bit of use over here."

Unless we give people a true understanding of how it works, how the science of learning works, and the social and psychological impacts of AI, we are never going to get a well-informed generation of teachers who can make intentional choices. You cannot do that in a 30-minute session. All the HEI providers and SCITT providers across the country are working their socks off to upskill themselves and find space in the margins to deliver 30 minutes, or maybe an hour or two, here and there. We need to change the frameworks to make it an absolute statutory entitlement that every trainee teacher, every ECT and every NPQ participant in this country has access to high-quality, evidence-informed AI training. That will be a really good step towards safeguarding our pupils and teachers.

Q50 Jodie Gosling: How well prepared do you think schools and teachers are for choosing safe and trustworthy AI or EdTech platform products?

Darren Northcott: I do not think they are. This is where the DfE's role is really important. We are asking schools—sometimes schools on their own, not as part of trusts or with local authority support—to make complex decisions about how AI is deployed by their staff in teaching and learning, and in respect of their pupils. There is a massive issue with confidence. The DfE has a role to play in that.

Much of what the DfE has produced—some of it is quite interesting and helpful—seems quite laissez-faire. It says to schools, "You should do this" or "You could do this". Schools are looking for much clearer rules of the road about AI. If it is to be integrated into teaching and learning and in other functions within schools, how is that best done? What are the DfE's expectations of that?

Briefly picking up the earlier point about training, across the piece, our members are profoundly dissatisfied with the CPD they experience in all areas and in all respects. Part of that is because of excessive workload,

and part of it is because CPD is marginalised into small spaces. That is what Pip's organisation exists to try to address.

Another point about the training that teachers actually experience—this goes back to a point we discussed earlier—is that a lot of it comes from the suppliers of the AI platforms themselves. That is from either the ones being used—maybe there is a case for that—or the ones being pitched at schools. There is perhaps an issue there with the objectivity of some of the advice that schools are receiving. I cannot speak for all teachers, but if I were sitting in a staff room listening to this advice coming from an organisation attempting to sell products into schools, I would have a little bit of scepticism about what I was being told and why.

Professor Clark-Wilson: I would just add, on that last point about working with the industry, that there is a lever. We can set in frameworks how we want EdTech tools to look for our learners. We can also set in processes. Even if you are selling, do you sell with a pilot evaluation toolkit that means you provide some guidance on the specifics of the impacts, allowing the schools and colleges to refine it, but giving them a little instrument or tool to help them decide whether to go and procure? These sorts of approaches mean that you do not have to make such big decisions with no evidence. Even if something has been evidenced elsewhere, you can essentially repilot it in your setting and then make a decision.

Pip Sanderson: When we say, “working with the tech companies,” we need to make sure that teachers are compensated to do that. Currently, it is often, “Here is an opportunity to come and work with a tech company and have some input on the tool,” and they are doing it above and beyond in their own free time, or their school is covering them by paying a supply teacher to take that lesson. Ultimately, who is going to benefit from the product? Hopefully, it will be the sector, but we also know that other people are going to benefit. Unless we start compensating our teachers properly to be true partners, it will always be a one-sided relationship.

Professor Eynon: Very briefly, it is important for us to start to capture expertise from teachers, to build and grow it, and to encourage participatory forms of design that are not tokenistic or too quick. While we think about training, we need to think about culture change and how we learn from each other, share knowledge and build that knowledge to shape the industry.

Q51 **Peter Swallow:** Darren, without putting words in your mouth, you have been clear that the Government need to step in and support schools to establish clear policies governing the use of AI by staff and pupils. You also said that schools do not understand data privacy concerns. Does anyone want to take the counterfactual on that? If too much decision making is taken centrally by Government on the policy governing this, is there a risk that schools and colleges will not be in a position to innovate, adapt and apply innovative uses of AI to their specific settings?

Pip Sanderson: I agree with Darren. It is difficult. The Council of Europe just published a set of 28 AI recommendations—

Q52 **Peter Swallow:** The framework?

Pip Sanderson: Yes. They are excellent; they allow enough room for space and innovation, but they give the guidance. It is not just for education; it lets businesses and families think about AI.

Schools are desperate. I was running an NPQ AI session just before the summer holidays, and the feedback from that was, "This is all great and really interesting, but what do I do?" They are desperate for really clear guidance and advice around what to do. The sooner we can provide parameters and core principles, and exemplify them with case studies of how schools interpret them for their unique contexts—because every school has different needs—the better. The sector is scrambling to try to find this guidance. If you happen to know a great school that is doing it, or you happen to learn about it, that is not equity for the system. We need a way to share the best practice effectively across the whole school estate.

Q53 **Peter Swallow:** That Council of Europe framework is very helpful, but, again, it is about principles, not implementation.

Pip Sanderson: You need both. You absolutely need the case studies to show it.

Darren Northcott: Just to confirm, I do not think that you put words in my mouth; that was a fair reflection of our position. The point is about balance. The implied point you are making about suffocating innovation and constraining schools and teachers too much is fair and well made, but my assessment would be that we do not have the right balance at the moment, particularly in circumstances where there is, as we have discussed, so much uncertainty and so many good questions to which there are not particularly good answers.

Peter Swallow: You are slightly putting words in my mouth; I was playing devil's advocate rather than advocating for that position—just for clarity, in case anyone critiques me later.

Professor Eynon: If I could add something briefly, it is about getting the balance, but we have to remember that there are significant inequities in the schooling system and in how schools can respond to policy and technology. What the DfE is doing for digital infrastructure is so important; we are definitely not there. Finding ways to provide space for teachers to work out what is important in their context remains important, but I agree that there is a huge amount of duplication and churn in the system where schools are individually trying to write DPIAs, evaluate evidence and all the rest of it.

Q54 **Peter Swallow:** I wanted to come to disadvantage next, so that was a helpful segue. The Government have announced plans to expand AI tutoring tools to deliver personalised one-to-one support, particularly targeted at students who need additional support. How do you all see that working in practice? What are the benefits and opportunities, and what are the risks?



Professor Clark-Wilson: I have worked with a number of companies that are already involved in some of those funded projects, and I have seen under the engines of the technologies they are using. We have to be so cautious about any AI-driven systems that are directly facing students and learners.

Right now, we have human tutors who are between the AI tools on tutoring platforms and the students, so the AI is advising the tutor on what move to make next with the learner and is not directly advising the learner or directing them in any way. We know that the difficulties that underachieving children have are usually quite deep and highly interconnected with other aspects. Developing technologies tend to focus on the subject content. In other words, it is a version of the textbook with some fancy ins and outs and ups and downs, but many of the issues go broader than that. It is conceptual problems and challenges that underpin the difficulties rather than quick fixes. It is cautious on where the technology is at, but also cautious on whether that is the best use of time with those students and whether those students are the ones who most need human tutor interaction, rather than agentic AI tutor interaction.

Q55 Peter Swallow: I want to come back to that human element, because it is really important, but very briefly, as a former teacher I know that one of the most challenging things to do in a classroom setting with 30 different students in front of you is differentiation. I also have seen some AI tools that are very good at differentiation. It is one of the things that it seems best suited to.

Professor Clark-Wilson: There is a diagnostic aspect, which can be much quicker than a human teacher in some ways, because they can connect up a bigger picture. But many of them assume that the learner has been on the platforms for hours and hours, to have sufficient data to do the fine tuning, and rarely does that happen—there is an assumption. One of those fundamental assumptions in how these AIs will work well is not satisfied, because you have to start somewhere.

Q56 Peter Swallow: Pip Sanderson, earlier you were very strong on the human aspect of teaching. Again, as a former teacher, I would agree. If we are overly reliant on AI tutor tools to support the most disadvantaged students, is there a risk that they will miss out on fundamental and valuable face-to-face time with teachers?

Pip Sanderson: I think so. Given everything we know about how learning happens and how relationships develop, we know there is a massive risk of that. I wish there was a magic bullet where we could suddenly turn on great teaching automatically for disadvantaged pupils, but I do not think it exists, and I think it is direct. It is sending our attention in the wrong way when we should be thinking about how we really solve the issue of disadvantage. How do we get more high-quality teachers into the classroom, rather than trying to use AI chatbots to solve that problem?

What often gets conflated is the idea of efficiency. AI chatbots are more efficient, and there are schools in America that are on that model of



efficient learning, but do we want children to have efficient learning? Is there an efficient poem or an efficient song to sing? Teachers are not just about transmitting information. There is the relationship and showing people how to be kind and courteous and how to wait. If we have a hyper-personalised learning bubble that the child sits in with a screen, when do they learn to tolerate others? When do they learn that it is okay to be a bit bored and okay not to have your every single need met every moment of the day? What kind of expectation are we then creating for them for the rest of their lives? We will be doing them a great disservice if we switch them over to a screen AI tutor relationship, rather than having a teacher relationship.

Darren Northcott: I agree wholeheartedly with what has just been said.

Professor Eynon: I would just say that in our ethnographic work in secondary schools, we have definitely seen how these adaptive personalised systems do not work that well. In particular, they do not work well for learners who struggle academically or students with SEND. They find these very demotivating and soulless experiences. What is particularly important is that schools that serve more economically disadvantaged communities rely on these tools more often. You are getting this problematic effect where they tend to rely on those options because they do not have enough teaching assistants or they have teacher shortages in some areas, compounding the problem. We have many ethnographic stories of students saying, "I just want someone to sit with me while I go through this tool," and teachers talking about how they are now sitting in a room with students doing their literacy intervention on an adaptive tech, and they feel like there is a kind of third actor in the room where they are not engaging properly with their students. There are some really important dynamics here that we have to think about.

Q57 **Peter Swallow:** So the current evidence is that these tools are worst placed to support the most disadvantaged.

Professor Eynon: I think so, yes. From the qualitative work that we have—it is very strong, very rich ethnographic data—they are not doing the job that we hope they will do.

Pip Sanderson: If you look at the corporate side and the big models that are out there, like the Khan Academy, and the whole promise of what Khanmigo would produce and what Sal Khan said himself about the promise of this tech, he has stepped back and said, "It cannot deliver. This personalised learning is not going to be what we thought it could be." He is the most invested in saying the opposite. If he is to say that, we need to pay attention to that point.

Professor Eynon: We are conflating good evidence about human tutors with assumptions that AI does the same thing. Of course, it is a completely different thing; technology does not care about students or know about them in the same way that teachers do.

Chair: Thank you. It has been a really fascinating panel discussion this

morning. I am grateful to all of you for taking the time to be with us. If there is anything that we did not have the time to get into in as much detail as you would have liked or if there are additional points you want to make, please do write to us after the session. We would welcome hearing from you again, but for now, thank you.

Examination of witnesses

Witnesses: Gillian Hillier, Colin Hughes and Andrew Freeman.

Q58 **Chair:** We are now turning our attention to the question of assessment. Can our three witnesses on that topic please introduce yourselves, starting with Gillian Hillier?

Gillian Hillier: My name is Gillian Hillier, and I am chief executive of the Standards and Testing Agency, which is an executive agency at the Department for Education. We are responsible for assessment up to key stage 3.

Colin Hughes: Good morning. I am the chief executive of AQA, which is the largest exam board in the UK and the only one that is an independent education charity.

Andrew Freeman: I am Andrew Freeman and I am an executive partner at Pearson qualifications. My teams develop GCSEs, A-levels, vocational qualifications and teaching and learning resources. We are also one of the teams developing one of the tutor pilots with DfE and DSIT.

Q59 **Chair:** How would you describe the overall impact that AI is having on the assessment system so far?

Colin Hughes: There is an enormous range, actually. In many ways, a key thing for the Committee's inquiry is that it ranges from deeply negative at one end—where we have got to be really wary and cautious and put in place mitigations or even bans and preventions—through to massive opportunity at the other end of the scale, where we have got to be open to exploring the possibility for significantly improving quality.

In your previous body of evidence just taken, you had quite a lot of talk about saving time and so on. I am much more exercised about generally lifting the quality and performance of the education system. That is where the greatest gains are going to lie, not in cutting costs or diminishing the human teacher-student interaction but actually enhancing that. For me, the litmus test is if it enhances the human teacher-student relationship, we should absolutely be looking at that. If it seeks to displace, be very wary.

Andrew Freeman: I would agree it is mixed and I would categorise how it is mixed in slightly different ways. In terms of the impact, our "Pearson School and College" report showed that 64% of pupils are using AI for their homework and 80% of them are using it every week. If you look at the overall assessment system, exams are fairly similar to how they were 10 years ago, so there is a mix of impacts.

To Colin's point, there is also a mix of good and bad. The key point is that AI is not inherently good or bad. If it is pedagogically led, educationally led and based on learning science, it has the potential to be good. To make that happen, it is important that there is a strategy in place across teaching, learning and assessment. One of the findings from the "Pearson School and College" report was only 19% of teachers say that their institution has an AI strategy in place.

Gillian Hillier: It is a bit different for the work we do, because obviously the impact in primary is lesser than in secondary. I look at it in the context of two things: how we deliver and develop our tests. On the former, we are making very limited use of AI to enhance the development of tests—only in a very small way, for things such as auto-marking. All the things facing secondary schools around the potential use of AI in coursework and homework are not having an impact in primary, because we do not have coursework involving a large, high-stakes essay done at home. We are just not seeing that in the same way.

Q60 Chair: How should the assessment system evolve in response to the presence of AI in society? I would like to think about both the skills and capabilities it should prioritise to enable young people to thrive in workplaces, many of which will be dominated and enabled by AI in future, and those needed to enable our young people to build a good society in a future where AI is such a dominant and influential feature of life. I might start with Andrew Freeman this time, if that is okay.

Andrew Freeman: The first point is that the challenge is not preventing AI in education and assessment—it is already happening. About 64% of pupils are using AI every week for their homework, and probably, since that survey was done, that number has gone up. The challenge is how we ensure that we still have valid, reliable, meaningful and fair assessment in that AI-enabled world. Related to that, those students will be in that AI-enabled world in future and, to thrive, they need to be able to work with those AI tools.

The core thing within assessment is to move from assessing output to looking at the process of learning. One of the key skills—learning to learn—will be hugely important for their future success: those higher-order thinking skills such as problem-solving, reasoning, application and within that, communication. The potential for AI within the assessment system is to make assessments more authentic and enable such things to be assessed.

As an example, we are doing a lot of work with AI personas internally. We take a lot of data and market research about what a science teacher might think and then allow our teams to talk to them to get good insight based on the collective knowledge we have. You can see that it is quite a small step to saying, "Could you do that in an assessment world as well?" For example, you could have an AI persona with dementia that you could use to interact with a student doing health and social care.



HOUSE OF COMMONS

Colin Hughes: Picking up on what Andrew is saying, there is an interesting paradox, isn't there? It is arguable that the skills people are really going to need to use AI effectively in their subsequent post-education lives—whether there is a post-education life is a different question—are actually the kind of higher-order thinking and critical skills that Andrew is talking about.

I do think—let's call it out—we have a problem with coursework. There is an issue around the degree to which it is possible for students essentially to replicate the appearance of having thought about something when, in fact, they have used AI to do that. We need to head in the other direction and use AI to enable teachers to track students developing those thinking and critical skills and so forth.

That is partly because, as students move on, the people who are now deploying AI effectively—the people in my organisation who are really good at it—are the people with those higher-tier skills. That leads me on to my second point, which is that digital literacy is absolutely fundamental here. We have really been quite backward in developing digital literacy in our schools and colleges. The problem there is partly that there is—in my view—a rather slack view that all young people are digital natives, which is utter nonsense.

Our own research illustrates that three out of four students are extremely comfortable with TikTok and Snapchat, but actually only about half will say that they are comfortable even using Word, which is probably the most basic thing they will use out there in the workplace—they are not comfortable with PowerPoint or Excel. I think there is a big job to be done, and we as AQA are frankly sceptical about simply hoping it will just happen in the process of teaching other subjects.

Further down the line, some of the things that your previous witnesses were talking about offer really serious opportunities to ascertain a much more rounded picture of where a student is at any given point in time. For example, using artificial intelligence tools for diagnostics could provide a body of data that enables us to properly respond to each individual student, so there is real opportunity there. Those are my three points.

Gillian Hillier: Again, the context is very different for primary. Obviously, STA does not test computing or IT; we do the basic skills. However, the research has highlighted the importance of, alongside developing digital literacy, having those basic skills and basic knowledge. We support that by ensuring that pupils have basic maths and literacy.

The other thing relevant to the work of STA is ensuring that we do not do anything that extends the digital divide. In all the work we have done in relation to more digital tests, we make sure that we do not assume that any pupil has the level of knowledge of using a screen or digital devices. There are options to do things off screen, as well as familiarisation work, so that all pupils come at it from the same level.



The third point I would make is the potential risk of having data that is somehow exposed to AI, or having pupil data somehow put into AI systems. That is something that all of us who have databases with pupils' results take really seriously. I think one of the risks is having an AI model that has access to our pupil data, which is something that we need to be very careful of.

Q61 Chair: Finally from me, the Government made a big announcement over the summer about technical education in schools and wanting to see an expansion of that. Do you have any brief thoughts on what implications the growing use of AI might have for that desire to expand technical education in schools?

Colin Hughes: First, this is very welcome news. We have long believed as an organisation that we need to significantly improve in this area, and schools are frankly begging for it. The other interesting thing here is that the use of technology in general—or digital tools, if you like—is that it actually has a longer, better and deeper customisation in vocational teaching and learning than it has so far in more academic teaching and learning.

I think this is one of the areas where there is actually real opportunity—for example, the opportunities for simulation and practical activities are very high. These things are getting steeply cheaper to do, and they are fundamentally not about replacing teachers with machines but using machines to enable the teacher to stand to one side and to educate and assess. It really meets my test, if you like, of whether it enhances the quality of the human student interaction. I think there is lots of opportunity there. Having said that, we also need to be very clear about the risk that we end up measuring a student's ability to use artificial intelligence and not their ability to understand the material that they are grappling with.

Andrew Freeman: I agree with a lot of what Colin has said—there is tremendous opportunity in that space. With the assessment of vocational education, the key point is that it is properly practical. Vocational education, as with AI, is so varied and means different things in different places. Sometimes AI may be a useful additional tool within the assessment process and sometimes it may not be, but it needs to be practical, relevant and authentic.

It also needs checkpoints. Again, it is not just about the output, and we have to move away from that. Someone may have produced something very shiny at the end, but that is not what is being assessed. Assessment needs to be done right the way through. That is partly a safeguard and triangulation that the work has been done and is of that quality right the way through, but it also goes to the point about learning to learn being fundamentally important for their futures. There is also probably an element of supervision for some elements of it, as part of that safeguard.

At the moment, another thing we are looking at is an AI Viva tool. That is an opportunity to push on those oracy skills and to have another

checkpoint for how a student is able to talk about their work, as well as the final product that they produce.

Gillian Hillier: I will not add anything, because, as I said, the limits of my responsibility are up to key stage 3.

Chair: That is great; thank you very much.

Q62 **Jodie Gosling:** We have heard that on-screen assessments could better reflect the digital nature of modern learning and employment while improving accessibility for some learners. What opportunities and risks are associated with the greater use of digital examinations?

Andrew Freeman: AI is not all one thing, nor is on-screen. We have been doing the Pearson test of English for 15 years now, using not generative AI but machine learning as an offshoot of AI. We assess 1 million or so learners a year using that on-screen methodology. We have a lot of evidence that that works and is robust, and it is accepted by Government and institutions around the world. We also have a lot of evidence from our international GCSEs, which we do thousands of on screen.

For me, the biggest risk in assessment is that we lock down and are left behind on the assessment side compared with what is happening on the teaching and learning side, where it is very digital. Two of the big opportunities are around authenticity—having authenticity to what a student does every day, not only in teaching and learning but outside of that, and authenticity to what they end up doing after they leave that education sector and go to the next stage, whether that is university or work. Again, digital and AI are increasingly implicit in so many roles and in further study.

On the accessibility point, we have seen real benefits on the accessibility side, compared with, for example, the typed responses that are already done. That experience is fundamentally not ideal because it is a paper-based thing that is typed into, as opposed to something that has been created to have lots of accessibility features that are much more beneficial for a learner.

Colin Hughes: You touched on one of the benefits: it is potentially fairer. For example, not all but most special needs issues can be dealt with better through on-screen assessments. It is also more resilient. I am really worried that people have already forgotten that I was asked in November 2020 whether I could deliver maths and English GCSEs online in 2021, and the answer was, “No, because nobody has ever asked us to do that.” We would have been in a much more resilient place during the pandemic if we had already had experience of delivering digital exams.

Security is also probably going to be better. I support everything that Andrew said and I am trying to add to it. The really big appeal is potentially a more rounded assessment. We will be able to assess different things in different ways.



HOUSE OF COMMONS

We absolutely must not replace all our examinations with digital or on-screen examination; that is a really terrible place to go. First, as you can see, I am very fond of pen and paper. We know it has lots of advantages. It has cognitive advantages and it is one of the media that we should continue to use, but there are certain things that we cannot easily examine with pen and paper. Think about the business of interacting with maps in geography or replicating scenarios in physics and so on and so forth. There are all sorts of things that we could potentially do online in due course. The key thing is to match the media to the thing that we actually want to assess, and not get obsessed with the idea that it must all be pen and paper or it must all go digital. That is not the right way to think about it.

There are very significant gains to be had. We are strongly in favour of taking moderate steps in this direction. We are not in favour of jumping hell for leather to a total on-screen examination approach, but I am concerned that we are at risk of moving too slowly and at too small a scale. We need to do things on a slightly bigger scale than what is currently being discussed, and we need to do a slightly wider range of subjects for the reasons that Andrew has just given.

It is going to be different in each subject domain area, and we need the opportunity to start doing this, not with hundreds of thousands of students—AQA examines 800,000 GCSE English language students a year; I am not proposing to go there—but I think we should look at tens of thousands, not 2,000 or 3,000, and we should look at four or five subjects, not just two. But having made that one relatively small critique, I think the step-by-step, in-moderation approach is definitely the right approach.

Gillian Hillier: We already have two on-screen assessments that are used with around 600,000 pupils a year, and we have seen quite a lot of benefits. One reason why we introduced them as fully digital assessments was accessibility. In contrast, with our key stage 2 assessments, or SATs, we send out hundreds of millions of pieces of paper around the country and they all have to be collected in. There is a vast workload for the teachers and admin staff in schools in relation to doing that sort of support for tests. For the on-screen tests, there is none of that. There is also an environmental impact. Obviously digital has an environmental impact—I am not saying that it does not—but there is an environmental benefit through not sending out a lot of paper.

The most important one, which you touched on, is accessibility for pupils with SEND. Again, with our paper-based key stage 2 assessments, we print some in large print and we do some in braille, which schools can request, but some pupils have visual impairments that involve needing different colour paper, slightly larger paper, a slightly different format or a more simplified image. If we have those digitally, that can all be done through the system itself. Teachers are not having to photocopy on to different colour paper and that kind of thing, which happens a lot with the key stage 2 SATs.



Through the tests that we are already doing, we are seeing benefits. On top of that, there is the ability to collect data pretty much instantaneously. Both those tests have an element of auto-marking only where there is a single, unequivocal answer. Where that happens and with the reception baseline—the rest is inputted by a teacher on-screen—the data collection is instantaneous. There is no waiting for weeks while marking takes place.

We can also get data back on how the system is being interacted with, so that we can improve it year on year. Those are quite significant benefits that we have seen through the two statutory tests that we have already done using on-screen testing.

Q63 Jodie Gosling: Are you planning to expand your use of on-screen in future? If so, what consideration will be given to the extra infrastructure, equipment and capacity in schools?

Gillian Hillier: We have no specific plans to do that at the moment. We have done various bits of research over time about extending how we use digital—for example, with more auto-marking. But in terms of something like the key stage 2 test, which is very high stakes, we would take it very cautiously. We would have to do a lot of comparability tests to make sure that like for like was being measured. We have a standard for the key stage 2 test that is set back to the 2016 standard. We would have to test all of that. We would not want to be putting the test behind glass, which is one option; we would want to develop a test that was purposefully for on-screen, and we do not have plans to do any of that yet. It is a possibility for the future, but I think we would have to have a lot of assurance that it was as good a test.

You have also mentioned infrastructure. With our existing on-screen tests, they do not require a class of 30 pupils, or two classes of 30 pupils, sitting down at the same time, which for a primary school implies quite a lot of technical infrastructure that they probably do not have and connectivity that might be beyond what they have. I don't think we would, at the moment, feel confident. Our expansion of digital is, I would say, cautious.

Colin Hughes: Subject to regulatory approval—we are in constant, and I have to say friendly, discussion with Ofqual around this topic. Anything I say in this area is not an argument with Ofqual; we are in a good place in our discussion about this. As AQA, we are absolutely ready to start doing this in a range of areas and want to get on with it. But one of the reasons we want to do it at moderate scale to begin with is precisely the point you talk about.

Actually schools, by and large, do have the machinery and so on. They have the connectivity, because they will be downloading assessments and uploading them afterwards; they won't be online during the exam. The issue really, overwhelmingly, where schools do have problems is space. Having 250 students in an exam hall all doing one big examination is definitely a problem tomorrow morning. But, for example, doing a minority language GCSE or a relatively small A-level like philosophy is not particularly problematic. We can take those incremental steps, if you like,



and demonstrate. Because the key thing we have got to do is carry public confidence with us around this, so going gung-ho is not going to be a good idea.

The interesting thing when we talk to teachers, school leaders and MAT leaders is that they have on the one hand a wariness—a little bit of reluctance: “Have we got all the kit? Are we ready for this?” Then you say to them, “But do you think we should go there?”, and they say overwhelmingly, “We have to.”

Andrew Freeman: Yes, we are carrying on with our ongoing expansion of digital on-screen assessment. A lot of that is international, so in different jurisdictions. Our experience on the international GCSE side has been very much as Colin has said—it is kind of incremental. Schools take time to think through whether they want to be part of a programme to go on-screen. They move in a slow but considered manner. It is not from zero to an enormous number in a very short period of time. Our international programme will continue.

Within the UK system for GCSEs, that is again subject to Ofqual’s programme. We would call for that to go a bit faster in terms of the number and size of subjects, because one of the key things that we have found is that by doing it you learn a huge amount, and you can incrementally improve that on-screen assessment experience. We need to do that in those different subjects, and at potentially a big enough scale, to mean that you can learn more quickly.

Q64 Jodie Gosling: What impact might on-screen exams have on different groups of learners, especially those with SEND or disabilities that were touched on before?

Chair: Sorry to interject, but we have about 15 minutes left. I need to end this panel promptly at 11.45 am so that we can get in our final witness. I need to ask you to be extremely brief, if that is okay, because we have quite a few more topics to cover that we want to ask you about.

Colin Hughes: The big one is this question about the so-called digital divide. For me, one of the arguments in favour of moving exams to digital is that it will then mean that schools need to act on that. In terms of special needs and so on, I think there is largely gain—it will not be total, but it will mostly be gain.

Andrew Freeman: We look closely at the comparability between on-screen and paper-based assessments—that is a key focus. The evidence we have had so far is that the SEN experience is much better and stronger within that, but we are always alive to comparability of outcomes, and that is an important part of the research that is done when we are moving to an on-screen delivery mode.

Q65 Darren Paffey: Ofqual have reported that, supposedly, they are detecting very few incidents of AI misuse, but they are concerned about the scale of the problem. From your perspective, how significant an issue is AI misuse for things like assessments outside the classroom or the exam hall? What



mechanisms do you think we need?

Colin Hughes: There are two things. One is that Ofqual are right; it is not massive. This year, we had 96 malpractice cases reported on AI, and we have acted on all of them already. That is out of 3.5 million qualifications that AQA delivers and about 10.5 million scripts—that gives you some notion of what a small proportion it is. It is only 2.4% of the total malpractice issues raised with us. It is relatively minor at this stage.

The concern is much more the degree to which teachers looking at coursework—for example, in A-level English literature or A-level history—now have an extra burden on them to try to work out, “Has my student used AI here?” It is a genuine struggle for them. What is tending to happen is that moderators are slightly over-reporting their concerns about that, which I kind of understand—that seems logical and natural.

During this process of curriculum and assessment reform, all of this is being reviewed at the moment, as you very well know, and my personal view is that we will probably need look at that question, not because we think there is a lot of cheating going on—we do treble somersaults backwards to identify cheating, and we are very good at it. We kind of know what we are doing, and we have a long history in doing that.

The problem is more about the burden on teachers and the difficulties they have and the worry they experience in trying to identify AI. As I said earlier, there is also the concern that we might end up examining a student’s ability to play with AI and not their ability to show that they have really thought about and understood a Shakespeare play or whatever it may be.

There is a simple solution to that, which is just to switch it all into a closed exam room. There is a much more exciting answer to that, which I think Andrew was touching on: the potential for having digital platforms watching a student develop their work as they are doing it and being able to report a more rounded picture of their understanding and ability. That is very exciting. This is the area where I see big risk over here and significant opportunity upstream.

Gillian Hillier: We have a slightly different context in primary, because we have some teacher-assessed work on writing, which is part of the school accountability system and part of the national headline statistics. We have not previously seen a problem with this, and I would imagine that we would not, because teachers, through that process, assess a portfolio of writing, so they will see a whole range of writings, not a single piece of high-stakes coursework. Also, the teacher is the one doing the assessing, so they will understand what their pupil can and cannot do and hopefully be able to identify where something does not look quite right.

This year, though, for the first time, we have had reports from the local authority moderators, who moderate the teacher judgments, saying that they have started to detect the use of AI in the writing or the editing of that work. We have not had any maladministration reports yet and we

have not seen any evidence of it, but we have built clarity into our guidance, where it was probably not clear before, that independent writing cannot include anything that is using AI.

Of course, the other thing that we have to allow for is that every year we get maladministration reports about teachers doing things they should not do, such as correcting work, for example. If it is the teacher who may be doing something using AI to correct that work, that is potentially more of an issue, and one that we are aware of. As I say, there have been no reports of maladministration yet, but I would not be surprised if we saw some.

Andrew Freeman: Similarly, our numbers are very, very low—of a similar order—in terms of reported AI abuse to date. We have checks and balances in place. We work with JCQ and Ofqual to ensure that the centre is checking both at a human level right the way through the process and then using automated detection tools as well. We then do similar things as an awarding organisation.

Taking a step back, looking at assessment holistically and going back to that stat of 64% of children having used an LLM for their homework, clearly, AI is in the mix in assessment in totality. One of the things we asked about in the Pearson school report was whether teachers felt they had prepared their students for an AI-enabled world. Only 8% of teachers said that they felt confident that they are doing that. There are a number of things associated with that. We talked about infrastructure a little bit, but professional development is really critical here. Professional development in relation to assessment is really important.

Q66 **Darren Paffey:** On biases in AI systems, particularly AI detection tools, there are concerns about reflecting or even amplifying existing societal biases. How do we maintain confidence in the fairness and the impartiality of AI-assisted assessment?

Colin Hughes: Our solution to that is relatively simple: we are not talking about using AI to mark high-stakes examinations. What we are doing—we have piloted it for two years now, and we are very nearly ready to go with it—is having AI mark the markers; it is sitting on the human's shoulder and watching the human. The evidence is crystal clear that we can massively jack up the quality of marking, because the machine is watching the marker in real time, and then alerting a sort of super-examiner human if there is a question of consistency, for example. That is saving a huge amount of time and lifting the quality of the human marking, and I think that is where the game is at the moment as far as high-stakes examinations are concerned.

It must be clear to you, of course, though, that while we are doing that, the machine is getting to understand a huge amount about how marking is done. There will be, of course, a second phase to all of this. Interestingly, when I ask multi-academy trust CEOs, which I did recently, how they feel about the introduction of AI to marking, their response is that it is hugely welcome. We have talked to them about what we are doing, and they can

see the immense attraction because part of the issue at the moment is that we have 38,000 teachers marking exam papers every summer. What we are hoping AI will do—in fact, the evidence is that it will do it—is enable us to be far more consistent across that very large number of human markers.

Andrew Freeman: We have been doing the Pearson test of English for such a long time, and a million assessments happen every year, so we have a lot of research on it, which we would be very happy to share. Broadly, it is a combination of an expert panel, sensitivity reviews, piloting and statistical analysis. We can share more information on that if that is helpful.

Gillian Hillier: Human interaction with the AI is the crucial thing. Where we have used AI to train the moderators of the markers, that has needed a lot of human interaction because AI makes things up sometimes. Not only does it amplify biases, but it invents stuff. The crucial thing in all of it is using humans to cross-check the AI.

Q67 **Mark Swards:** It is pretty clear that the benefits of AI in EdTech depend on not only access to the technology but the actual skills that you need to use it effectively. How can AI literacy be better integrated into the curriculum and assessment system so that pupils can learn to use AI safely and, of course, critically?

Andrew Freeman: Our Pearson school report found that 53% of teachers would welcome CPT on AI. Related to that, secondary school teachers said that the top barrier to embracing technology was access to devices. Those two are completely interrelated, as is a whole strategy that covers infrastructure, leadership and teaching practice, and curriculum and assessment.

We would not see AI as something that you have as a discrete subject; it has to be embedded across the curriculum in all subjects because it is applicable in different ways. We are very proud to be part of the Government AI skills boost programme, so I have experience on that front. I am also looking at a programme with Dr Fiona Aubrey-Smith on AI literacy and how best to embed that. Again, we would be very happy to share further information on that.

Mark Swards: Yes, please. Given that time is short, that would be helpful.

Colin Hughes: I would be wary of embedding AI in teaching and learning, but I would be strongly in favour of enhancing the teaching of digital literacy, and not just hoping, because it is over-optimistic to expect it to simply arise out of subject teaching; we need to do something specific and distinct in terms of enhancing digital literacy. We are also in favour of looking at the potential for AI or future technology qualifications, which could assist with all of this.

The risks in teaching and learning are quite high. The gains in assessment are already demonstrated to be pretty positive. I have not mentioned, for



example, that a subsidiary of AQA delivers all the adaptive testing at key stage 3 in Wales. There is huge teacher wariness about it until it is actually done, and then they realise that it is just providing them with the most gorgeous body of data that they have ever had and it is completely wonderful. Then they embrace it and are happy to go forward with it.

Part of the issue is that we need to not be too shy about actually getting out there and making some stuff happen. Some of it will work; some of it will not. We do not take high risks with high-stakes exams, but we do things like adaptive diagnostics. We also have a platform that looks at key stage 4 mathematics diagnostics; it is called GCSE Stride and is, incidentally, free to schools. That uses both adaptive and machine learning systems. The machine can do things that the teacher finds really difficult as a maths teacher, which is really to understand, when students arrive at the beginning of year 10, where the gaps are in their conceptual understanding of mathematics. That is the biggest single problem.

All of our diagnostic testing in other parts—we are also currently doing it in Scotland—shows us that something like one in four students are no further forward in mathematics in year 11 than they were in year 7. Identifying that through just human interaction is very difficult; using these kinds of diagnostic technologies is telling us stuff that should be having a major policy impact—but, in brackets, is not having enough of a policy impact right now.

Chair: I am going to move us on to a final question from Darren—you have two minutes, if that is okay.

Q68 Darren Paffey: We all know how quickly AI is advancing. If you could make one observation about our current regulatory framework around the use of AI in qualifications and assessment—either where we are or where we need to get to—what would that be in terms of trying to keep up with the speed of AI?

Gillian Hillier: Again, I think the question is not as relevant to the work we do in primary. Where we have used AI, the key thing that the regulatory system will need to be sure of is that there is fair assessment of pupils—fair, not because one pupil is better at using the internet or AI and can write a better answer to a test, but because it is their own knowledge. That is the bit that is really challenging for everyone at the moment. I imagine that Sir Ian will talk about how challenging that is for Ofqual to oversee.

Colin Hughes: As Gillian says, you are about to hear from the chief regulator. I am pretty confident that he and I would not disagree about this. As I said earlier, we think there are issues with regard to coursework and how we manage that going forward. We need to pay close attention to that as we develop the new assessment regime over the next two, three and four years. That is going to be a primary area. I am afraid the answer is actually that we just need to enhance digital literacy among teachers and students more broadly, including the ethical background to all of that, which is critical, as your earlier witnesses said.

Andrew Freeman: On the high-stakes sides, we are working very closely with Ofqual, and lots has already been said on that. On the more general assessment within schools, there is quite a lot of different requirements and frameworks. All of that being brought together, which is the plan, will be very helpful. There are things around data processing, child-appropriate design, cognitive safety, gen AI safety, keeping children safe in education, GDPR, DfE's AI product safety guidance and the ISA's guidelines. There is an awful lot of different things. The plan to bring that together, which is in place, is really welcome.

Chair: Thank you very much indeed. Apologies that we were a bit rushed. If there is anything that you would like to tell us about in more detail, please write to the Committee after the session. We would very much welcome that input.

Examination of witness

Witness: Sir Ian Bauckham.

Q69 **Chair:** Welcome to the final segment of this evidence session on AI and EdTech. We will now hear from the regulator of qualifications, the head of Ofqual, Sir Ian Bauckham. Sir Ian, would you like to introduce yourself?

Sir Ian Bauckham: Yes, absolutely. Good morning. It is very good to be here. Thank you for inviting me to give evidence. I am the chief regulator of qualifications and assessments in England and the chief executive of Ofqual.

Q70 **Chair:** Thank you for being with us this morning. I will lead on our questioning. How would you describe the impact of AI and EdTech overall on the assessment system that we have?

Sir Ian Bauckham: Before giving a detailed answer, it is interesting to step back for a moment and just reflect on what the purpose of education itself is, and therefore what the purpose of educational assessment is. In education, which is an enormously expensive and important endeavour that we undertake collectively as a society, we not only prepare young people for work, we also transmit our civilisation to them, introduce them to our civilisation and facilitate their cognitive development.

Given the importance of all of that for us as a society, it is essential that educational assessment accurately measures at a granular level how successful we are being in that undertaking. If we do not do that, we are not able to make informed decisions about which young people should go to which destination, and neither are we able to understand which parts of our system are working well and which are not.

To achieve all of that, the next step in the logic chain is to say that educational assessment must tell the truth—it must have authenticity. Across our suite of qualifications and assessments, we have many that are undertaken in supervised conditions. They have a high level of authenticity. We have, for example, written tests. We have the fabrication



of physical artefacts that can be evaluated by an expert, and spoken tests that can be recorded. There are high levels of authenticity.

In a smaller number of cases, we have assessments that are essentially extended writing assessments, which are undertaken with only very light-touch supervision or very little supervision at all. In those cases, my view is that we are exposed to risk in the shape of AI being marshalled to support students to undertake those assessments. That compromises the degree to which those assessments really tell us the truth. We can track this all the way back to the importance of education as an endeavour, and therefore the risk that we are exposing ourselves to if we cannot guarantee authenticity in assessment.

- Q71 Chair:** Given what you have said, and the widespread use of AI in education, which we have been hearing about in this inquiry, do existing qualifications and assessment models remain fit for purpose, or should we be looking at a more fundamental change?

Sir Ian Bauckham: I do not think that fundamental change is quite the right way to conceptualise this. We have some assumptions that we apply when we think about education reform. We look, for example, at the undoubted benefits of coursework or non-examined assessment for motivating students over time, developing their research skills and so on. When we are thinking about that, we must now understand that that comes with a very significant downside, which I would say is not just greater in volume but greater in kind than was the case until the advent of AI. We have to look critically at some parts of our assessment landscape in the light of the pervasive use of artificial intelligence.

- Q72 Mark Sowards:** Ofqual has consulted on the increased use of on-screen assessments in schools. What are the main opportunities and risks associated with moving away from traditional paper-based exams?

Sir Ian Bauckham: When I talk with people about on-screen assessment, I often hear a confusion between on-screen and online assessments. I regard those as two very different things. Good on-screen assessment utilises a secure portal that means students cannot access the wider internet and cannot communicate with other students or other individuals, whether inside or outside the exam room. That is fundamentally not the same as online assessment.

There are both risks and opportunities. You are quite right that we have recently consulted on an approach to the judicious introduction of some on-screen assessment to GCSEs and A-levels. The potential benefits are that there are some kinds of assessment that can be more accurately modelled using an on-screen format than is possible on paper. It is also the case that some special educational needs—I say “some” advisedly, because it is a very broad spectrum—can be more easily catered for on-screen than is possible on paper. Those are a couple of benefits.

In terms of risks, the first is the scale and complexity of implementation. We have a large number of schools and colleges where qualifications that we regulate are taken up and down the country and we have uneven



access to high-quality digital infrastructure and devices. There are also possible social and disadvantage-related risks. It is the case that there is still a digital divide in the country, and some students have had greater and more fluent access to sophisticated online systems and applications over a protracted period than others. We have to bear that in mind as well.

There is also something to note about the signalling that the widespread use of on-screen assessment in terminal assessment gives to schools and the system—the signal being that we expect learning and assessment to be on-screen. That has a washback effect into schools and increases the expectation that teaching and learning should be on-screen, but that is a contested question. Fortunately, it is outside my remit, but we have to be mindful of it. To conclude that answer on special educational needs and reasonable adjustments, it is already possible under equalities law for on-screen assessment to be used where there is a demonstrable need for it related to a particular need that a student has, so that already exists.

- Q73 Mark Swards:** Very briefly on SEND, you talked about some of the advantages and I can imagine some specific scenarios where somebody has a need that can be met by an on-screen assessment. Has any of the evidence of your consultation said that there are disadvantages for certain SEND pupils?

Sir Ian Bauckham: Yes, absolutely. SEND is a very broad spectrum. Not all special educational needs are the same, by any stretch of the imagination. Some simple and obvious things that an on-screen format can achieve that paper cannot include different font colours, different text sizes and reduction of the amount of text on the screen. That is very useful for some students with some needs, but other SEND students actually find the digital interaction and the navigation of a keyboard more difficult than a pen-and-paper format, so it is not an even pattern and decisions must be taken on a case-by-case basis.

Mark Swards: That is very helpful; thank you.

- Q74 Darren Paffey:** Ofqual has expressed concern about the scale of pupils' misuse of AI tools, although there are still relatively few detected cases. How confident are you that the qualification system is currently able to identify and address AI-enabled malpractice?

Sir Ian Bauckham: Thank you; that is a very important question. AI is developing faster than the detection tools can keep up with. You are quite right to say that our reported cases of detected and proven AI misuse remain relatively few—they are reported in our annual report every autumn—but I visit schools and colleges around the country on multiple occasions every term and every time I talk to teachers, and every time I speak at a conference and am approached by teachers afterwards, the line is always the same: "We know our students and their work because as human teachers we interact with them, so we know that they are using AI but we cannot prove it. We put it through the detection software and the AI is now too clever for it."



I think we must caveat the small number of detected and reported cases very heavily. I have a very strong and I think reasonably well founded view that AI use by students is much greater. Indeed, when I visit schools and colleges I always take the opportunity to talk with a group of students—or pupils, in the case of schools—and try to do it without the deputy headteacher being in the room with them. I ask them to tell me honestly how often they are using AI not just to support them with their homework and their classwork but for actual assessed work. I usually get a grudging admission that that is happening on quite a large scale.

Q75 Darren Paffey: What does all that consultation with both the professionals and the students tell you that you need to do, and what are you doing, to protect the integrity of the qualifications?

Sir Ian Bauckham: I think we can take a threefold approach here. We are doing several things already. I wrote to the chief executives of all the main exam boards earlier this year to ask them to take this issue even more seriously and to step up even further their approaches to checking for malpractice. However, given what I just said, I am not naive about that. Within the current qualification system there will be limits to what they can do. Internally, we just created a new post and stood up a new team for qualifications integrity that will take a lead on this work—that is an important thing to do.

In terms of what we do with assessments now, the threefold approach to which I referred is basically: first, avoid extended writing non-examined assessment if we can, because that is where the risk is greatest. Where it is a question of making a physical artefact that can be supervised and whose provenance can be more securely guaranteed, the risk is lower. If it is something that can be recorded either by a teacher or by an electronic device—for example, a student doing a dance, undertaking PE or a drama performance—that is all much more secure.

However, extended-writing coursework should be avoided wherever we can. Where we cannot avoid it, we should adapt it to make it as secure as it can possibly be, bearing in mind what we are trying to achieve. At the same time as all of that, we need to deter. We need to send very strong messaging to students about the consequences and sanctions of malpractice involving AI, which is effectively cheating. We should send strong messages to reduce the occurrence of that. It is a threefold approach: avoid, adapt and deter.

Finally, as part of my answer to that question, in the current round of qualification reform, which is being led by the Department for Education, we are providing very clear and unequivocal evidence about the risks that some forms of extended-writing, non-examined assessment are now open to. There are always trade-offs in assessment. For a bit more of one thing you have to do a bit less of something else. It is important that those making decisions about qualification reform are very clear about what the risks are, because they track right back to the fundamental purpose of educational assessment that I set out at the beginning.

Q76 Jodie Gosling: You have highlighted some of the risks of bias in AI systems and in AI detection tools. What assurances can we give pupils and parents that the use of AI in assessment systems remains fair and impartial?

Sir Ian Bauckham: I am very concerned about biases built into AI systems. It is always the case that, when there is new technology, there is a kind of amplification of the views and representations of those who own and design the technology. That is certainly the case with AI. The two examples I would give are the use of AI by awarding organisations in setting exam papers—writing exam questions—and in marking exam questions.

On the setting of exam questions, there probably are some very constructive potential uses of AI to help generate good-quality questions. Question papers, by definition, once the exam has been sat, are effectively open to public scrutiny. It would be possible for the public to identify if there were any biases creeping into the framing of questions, so I am a little bit less concerned about that side of it.

Marking is an altogether more difficult question. Marking, by definition, must happen behind closed doors. There is the question of permitting the unaided use of AI for the marking of assessments. By AI use, I am not talking about machine marking, where there is just a very limited set of possible right or wrong answers; I think that is okay, and that has been happening for a while quite securely without concern.

By AI in marking, I am talking about AI mimicking expert human judgment over sophisticated assessment outputs, such as pieces of extended writing or artistic artefacts—that more subtle, nuanced mimicking of human judgment. At the moment, my position is that I would not want to allow AI to do that unaided, because there is a very real risk that bias could enter the system if that were to be allowed.

The fact is we do not allow that. We have been very clear about that. AI can be used as an ancillary in marking to support quality assurance and so on—that is okay. It can identify rogue markers whose marking is out of kilter with other markers. All of that is okay. However, unassisted replication of sophisticated expert human judgment over artefacts that determine a young person's future? No.

Q77 Jodie Gosling: How are you evaluating the potential impacts of these technologies on different groups?

Sir Ian Bauckham: I think the safeguards we are putting in place are very good for ensuring that we do not have differential impacts on different groups. We are essentially not allowing AI to do marking—that has to be a primarily human activity. We are factoring AI out of those primary judgments. We have a whole other range of requirements that we place on exam boards to ensure that bias does not enter the human marking process. For example, when markers mark exam questions, they do not know who the pupil is, their sex, where they're from or what school

they went to. It is possible to factor out a lot of potential bias through that mechanism.

Q78 Chair: Thank you. Some of the evidence we have seen argues that there is currently no coherent framework governing the use of AI in assessments. Is there a risk that AI is developing faster than the Government's ability to make policy about it and faster than Ofqual's ability to regulate it?

Sir Ian Bauckham: No, I do not think that is the case. We are taking a very cautious and precautionary approach with the extent to which we are permitting AI to be used for high-stakes qualifications, so I do not think we are lagging behind what is happening. There may, of course, be awarding organisations that operate effectively in a competitive market that would like to do more—and that may or may not be a reasonable aspiration—but my job is not to do what the awarding organisations would like me to do. My job is to do what Parliament has tasked me to do, which is to ensure that our qualifications give a reliable and accurate representation of students' achievement consistently over time, and command public confidence. To achieve those statutory aims that Parliament gave me, it is incumbent on me to take a highly precautionary approach, so we will not at any time soon—and maybe at any time ever—be allowing unassisted use of AI for marking.

Q79 Chair: What steps are you taking, as you think about how the landscape is evolving, to strengthen the regulation of AI in assessment, and to make sure that we have a comprehensive framework that is fit for purpose for the future?

Sir Ian Bauckham: As I say, we have recently stood up a new team that will be looking at qualification integrity—really making sure that the integrity of qualifications can be safeguarded in the AI world. We are an outcomes-based regulator, so our duty and way of working is to ensure that the results that students get when they take a regulated assessment can be trusted and accurately represent the quality of work that the student has evidenced at the point where they took the assessment.

We take a permissive approach to much of what happens upstream within the qualifications industry, which is what it is. If awarding organisations want to use AI for data analysis or to help them generate question papers or assessment plans, providing what we see meets the very demanding quality bar that we set both when an awarding organisation applies for recognition and when it submits individual qualifications for accreditation, we do not mind how it is done.

What is critical is that, at the point a student is assessed—where a sophisticated judgment is needed—their work is marked by an expert human marker. That is also important for contestability. It is a natural right that if someone is making an important, consequential judgment about me on the basis of my work, and I do not agree with their judgment, I can contest it. That simply does not work if you allow AI to do that instead.

Q80 Chair: Finally, as the Committee considers this big and challenging topic of AI and edtech in education, is there anything else you would like to draw our attention to that we should think about seeking evidence on and consider as we make our recommendations to the Government?

Sir Ian Bauckham: This might be partially a recap of some of the points that I made before but, if you will permit me, I will restate them. This is not a question of being for or against AI; it is about making conscious decisions about where we as human beings want AI to be deployed in our education system for the benefit of young people, for the benefit of education and for the benefit of humanity, and where we absolutely don't. We have to do that consciously and discriminatingly.

We have to be aware of the risks of bias and unfairness through both on-screen assessment and AI. We have to be conscious of the signalling or washback that the extensive use of AI in high-stakes assessment and the extensive use of on-screen assessment send into the system.

I would contend that there is increasing consensus that collectively we made a grave mistake when from 2012 onwards we allowed our young people unfettered and unsupervised use of smartphones and social media, and we now have to live with the consequences of that catastrophic decision. I do not want us to make the same decision and the same mistake when it comes to AI and computer use in schools, so we have to be very careful and proceed on a precautionary basis.

We also have to be very conscious that, if you were ever to allow AI unsupervised sole marking rights of students' work, it would make edge decisions about what was and was not allowed to get a mark. We would effectively be surrendering judgment of human learning to a machine, and I do not think we should do that. I think that is a fundamental moral question.

Finally, there is a question of accountability. You can call me for an accountability hearing at this panel. If my work was outsourced to a bot, you would not be able to do that, and I do not think that is right in a democratic society, so this is a high-order, very significant question.

Chair: Thank you very much indeed. That has been very interesting and has given the Committee a lot of food for thought. If there is anything else that you would like to draw to our attention, please write to us as we progress with this inquiry. We are very grateful to you for being with us at short notice.