

Education Committee

Oral evidence: [Primary Assessment](#), HC 682

Wednesday 22 February 2017

Ordered by the House of Commons to be published on 22 February 2017.

[Watch the meeting](#)

Members present: Neil Carmichael (Chair); Marion Fellows; Lilian Greenwood; William Wragg.

Questions 157 - 242

Witnesses

I: Claire Burton, Chief Executive Officer, Standards and Testing Agency, **Sally Collier**, Chief Regulator, Ofqual, and **Dr Michelle Meadows**, Executive Director for Strategy, Risk and Research, Ofqual.

II: Nick Gibb MP, Minister for School Standards, Department for Education.

Written evidence from witnesses:

- Ofqual [[PRI0399](#)]
- Department for Education [[PRI0403](#)]

Examination of witnesses

Witnesses: Claire Burton, Sally Collier and Dr Michelle Meadows.

Q157 Chair: Good morning and welcome to what will be our final day of hearing evidence for our inquiry into primary assessment. Today obviously we are looking at the Standards and Testing Agency and Ofqual to see what they think about the way forward. One of the distinctions that we have to make is accountability of schools and performance of children. That is the important distinction in this process. Without further ado, would you like to, for the viewers who are tuning in—many have been having an early breakfast to make sure that they are ready for this—say who you are and from which organisation you come, starting off with Claire?

Claire Burton: Hello. My name is Claire Burton. I am Chief Executive of the Standards and Testing Agency.

Sally Collier: Good morning. I am Sally Collier. I am the Chief Regulator at Ofqual.

Dr Meadows: I am Michelle Meadows and I am Deputy Chief Regulator at Ofqual.

Q158 Chair: Okay. The first key question is: at a previous session we heard that the division of responsibilities between the STA and Ofqual is unclear. Is that correct or not, Claire?

Claire Burton: I do not think it is unclear. The Standards and Testing Agency's objectives are set out and published on the gov.uk website. We are responsible for the development and the delivery of testing in primary up to the end of key stage 3, and I think that that remit is quite clear. Sally obviously can talk about the Ofqual remit. They are very distinct.

Q159 Chair: Sally, do you think they are clear?

Sally Collier: Yes. They are perhaps not clear to the outside world, but I am happy to elaborate. We have a very different set of responsibilities for the national assessments than we do for other types of qualification. In our primary role in general and vocational qualifications we have a whole suite of powers, from requiring exam boards to do things, being able to force, if you like, or direct that awarding organisations do certain things, and we, of course, have fining powers. We do not have that range of powers in relation to national assessments. We have a very narrow range of powers and duties, including a duty to inform Ministers if there is likely to be or has been a serious failure. We have power to request information from the Standards and Testing Agency. It is quite limited, what we do in this respect.

Q160 Chair: Michelle, you can adjudicate on this question, because you are not



under the spotlight. What do you think?

Dr Meadows: I agree with Sally that it is not obvious to the outside world what Ofqual's role is relative to the STA's, but I think that the primary players in the system do understand the distinction. I would just add to what Sally has said that our particular focus is on the validity of the assessments and the setting and maintenance of standards. That is our core concern in a regulatory sense.

Q161 **Chair:** In short, none of you really think we need to re-clarify or restate the roles of the STA or Ofqual?

Claire Burton: I don't think so, although I do agree with Sally: it is not always obvious to the outside world where the difference is between Ofqual's regulatory role for general qualifications and the role on primary assessment. It is possible that that might be an area that we could be clear about publicly, although I think it is quite clear in private.

Sally Collier: If I may add, it is not clear, if our roles were changed—although the evidence from this Committee may help—what extra value we would add. When we started in this role of regulating national assessments, we concentrated both on the delivery of those assessments and on the validity and standard setting. I think my predecessor wrote to you in August 2014 and said we were going to change our role to concentrate more on the validity and standard setting and not on the delivery, because at the time we were simply marking, duplicating, and adding bureaucracy to a system that has been set up with a Government agency reporting to Ministers. It is not entirely clear to me—but I would not rule it out—what value we would add by changing the remit.

Q162 **Chair:** Okay. Thank you. Claire, do you think you should have more independence from the Department?

Claire Burton: I think the balance we have at the moment is a good one. As you probably know, the history of the Standards and Testing Agency has gone through various permutations. We have been part of the Qualifications and Curriculum Development Authority, and before that the National Assessment Agency. The balance that we have at the moment is that I have access to discussions around curriculum in particular and accountability, which are really critical in developing our assessments. I value being part of those discussions and being able to access them through our relationship as part of the Department for Education.

I think, though, we are very clear that I have an entirely independent role in the development of tests and in the setting of the standards. It is very clear, and we have been clear publicly, that that is an area where there is no role for Ministers. Ministers have no engagement. That feels like a good balance. I realise that there are some tensions there—but I think those are tensions for me to manage—around our contribution and



our involvement in policy. Overall, I think they lead to a better system and a better outcome.

Q163 **Chair:** The situation that effectively produced the root-and-branch review that the Government announced last year was obviously created by circumstances that we all know about. Do you think that Ofqual could have been more involved in solving some of the problems at an earlier point?

Claire Burton: I think the issues that led to the STA review were primarily around failures in our process and our internal processes, and I do not think, even if we had been very closely scrutinised by Ofqual, they could have predicted or warned us about those. The support that I have had from the Department in terms of its internal audit functions and support and challenge from the Permanent Secretary has been a thing that has really helped us to put better processes in place and to be clear about where our clearance sits. Probably Ofqual would have just muddied the waters a bit there. Sorry, Sally. I do see that there is a question, because Ofqual clearly do have an interest in the delivery of the tests as well.

Q164 **Chair:** Michelle, how would you tackle strategy and risk?

Dr Meadows: We asked ourselves this very question: if we had been monitoring delivery more closely would we have spotted what were essentially human errors? We came to the conclusion that we probably would not have. We do attend some of the key operational meetings, we scrutinise process documents and reassure ourselves that they look appropriate. Would we have spotted occasional human error? If you take it in the context of the annual cycle of assessment that is done year-on-year for the whole cohort, we suspect not. We think that where we add the most value is not getting in and monitoring delivery, and maybe being able to pick up potential for human error, but rather scrutinising things like the standard-setting process, the maintenance for standards and so on, where our independence from Government adds value.

Q165 **Chair:** Okay. Thank you. You obviously value your independence.

Dr Meadows: Indeed. Of course.

Q166 **Chair:** You think you are sufficiently independent?

Dr Meadows: Yes.

Q167 **Lilian Greenwood:** Can I ask a quick supplementary question? I know that some of the difficulties with the test related to human errors and mistakes, but some things, like the delays in the production of guidance to teachers, are not down to human error. That is down to poor planning and process. I am just interested. You say that you do not think you would have picked up those things —



HOUSE OF COMMONS

Dr Meadows: Sorry, I was referring specifically to the two security breaches that occurred in the summer rather than other issues that have been raised with the Committee.

Claire Burton: I agree. In terms of the role that Ofqual could have played, there were some delays to guidance in the last year. We had set out and we had plans in place, but we had to respond to what ended up being some important feedback in the year, which did delay some of our guidance. Again, I do not think that that is something that Ofqual could particularly have played a part in either challenging or, indeed, supporting us to sort it out. It is something the STA had to take responsibility for and sort out ourselves.

Chair: Okay. Any further questions on this issue? William on the implementation of a new assessment system.

Q168 **William Wragg:** Indeed. I think from that supplementary question from my colleague it is fair to say that the implementation of the new assessment system last year did not quite go as smoothly as hoped, and it was widely criticised by many teachers. How confident are you all that primary assessment will be delivered successfully this year?

Claire Burton: I would start by saying that last year was a particularly challenging year, and it was always going to be a very challenging year because we were delivering change over a number of fronts. It did not go as smoothly, and I take responsibility for the failings that the agency played in that, but some of that was also to some extent inevitable when you consider the degree of change that teachers had to incorporate.

Q169 **William Wragg:** Had you communicated that degree of change to the Department and expressed that it was not deliverable?

Claire Burton: I do not think we had expressed that it was not deliverable. I think that we had communicated, both to the Department and more widely to the teaching profession and the sector. Since the introduction of the new curriculum in 2014 we have had regular communications updating on where we were with the delivery of the test, with the standards that could be expected to be seen in the new tests, and then throughout the year in 2015 and 2016 we also put out regular communications, including to schools, parents and more widely. I think that we communicated that the change was coming. That does not necessarily mean that people were still prepared for what that would actually mean, and I think to some extent that is very difficult to achieve.

This year I think we have a far greater degree of stability, that teachers know what to expect. There is very little change. We have improved some of our guidance working with the teaching unions. But broadly I expect it to go far more smoothly this year, and I certainly do not have any indications at the moment in our monitoring that there are similar issues emerging. But this is always a very volatile area, and I think that that is almost inevitable.



Q170 **William Wragg:** I wonder if I could ask Sally and Michelle what advice Ofqual offered the STA during the implementation of the test last year and if that differs this year.

Sally Collier: Michelle has outlined what our role was last year, which was to focus on where we thought the highest risks were, which were the development of the test, the test-development process, and the standard-setting process. That was where most of our activity was, other than, obviously, being alerted when particular issues happened. I will leave Michelle to say exactly what we did and when. I don't think we are planning to do anything different from that in respect of the validity of the test and the standard-setting process this year. We are—and we may come on to the specifics of individual tests—doing some particular pieces of work in respect of some particular tests. Michelle, do you want to add to that exactly what we did and when?

Dr Meadows: Indeed. Whenever there are changes to assessment arrangements, we offer advice to the STA about those changes. On the introduction, for instance, of the interim teacher frameworks, we offered advice to make them robust for teachers, to be able to apply as consistently as possible, especially given the fact that they are operating in a system of accountability. We also looked, as Sally has explained, very hard at the test construction process, and assured ourselves that teachers, and representatives from disability groups and so on, were involved appropriately at all the stages. We attended the standard-setting meetings. We looked at the process by which standards were set, ensured that we were happy that a fairly standard approach, a book-marking approach, was being taken, that it was technically robust, and that teachers were involved at each stage. Then we scrutinised the outcomes of the tests, so the data coming out of the tests, to check that they had performed, at least on the basis of this evidence, as assessment instruments.

Q171 **William Wragg:** Okay. I just wonder about the writing guidance in particular. My colleague asked you about it earlier, but what was the cause of the delay in the publication of that guidance?

Claire Burton: The guidance specifically on the writing framework?

William Wragg: Yes.

Claire Burton: On the interim teacher assessment frameworks, we first started consulting on performance descriptors—which were intended to be used as teacher assessment frameworks—in December 2014 when the consultation closed. What we discovered was that the frameworks that we consulted on simply did not work in the way that we intended them to. There was quite a significant risk that we were going to be introducing something that would effectively bring back the levels that we had just spent several years trying to remove from the system. We found ourselves in the spring and summer of 2015 having to significantly revise our approach.



What we did was we took it right back to the national curriculum. We developed interim teacher assessment frameworks that are firmly and wholly rooted in the national curriculum itself and the approach behind the national curriculum. That process did delay us. Our intention had been that we would consult on the performance descriptors and that we would come out with them in a near-final version, trial them in schools in the summer term, and then we would publish them in September. It was the trialling in schools that was missed out. The intention was always to publish the frameworks in September, but we did miss a step, and it was because of that change in approach that we took.

Q172 **William Wragg:** Did you recommend to the Department, therefore, that there should be a year's delay if it has not gone through the thing that you initially set out to test these things?

Claire Burton: I think that we felt that since the approach that we had taken was familiar—there is nothing in the interim frameworks that is not in the national curriculum, so we were using a familiar material—that was a relatively limited risk. I can understand that many people disagree with that, but we felt that the risk was one that we could take. The alternative was to have nothing because, of course, we have removed national curriculum levels; we have no way of using teacher assessment otherwise, unless we had some sort of standardised approach that we could take in 2015 and 2016. I do accept that it was not an ideal position. It was where we found ourselves, having listened to feedback and considered our approach, but it did lead to those delays.

Chair: Okay. Thank you very much. Lilian on the STA review that Nick Gibb announced.

Q173 **Lilian Greenwood:** I just wanted to come back again on that question, because what you said was essentially you had to go back and start again. Has that changed how you would approach it in the future? Do you think there should be a longer lead-in time in case that happened again, so that you would not be left with that situation of thinking, "Do we delay or not?"

Claire Burton: I do think that, particularly in areas where we are talking about teacher assessment, which is difficult to get right. We had the Bew Review in 2011 for that very reason. Ofqual, over the last year or two, have also pointed out many of the issues with trying to ensure they are robust. I do think that if I was doing the teacher assessment again I would want to try to build in a time for more testing and trialling with teachers.

Q174 **Lilian Greenwood:** Thank you. We have already talked about some of the criticisms of the tests. When the NAHT gave us written evidence they talked about them being poorly designed and poorly administered, and obviously picked up on some of the mistakes. That is what led to the root-and-branch review that found that the STA has considerable areas for improvement. Are you confident that you are going to be able to



make sufficient improvements to the organisation?

Claire Burton: I am, yes. I participated, and all my staff participated, in the STA review. It was a very inclusive review. I think the review team, over the course of the summer, spoke to pretty much every single one of my staff. It covers a broad range of areas, but they are also areas that we recognise in many cases and had already begun to respond to. I do think that we have the capacity within the agency to make the changes that are needed. Some of those I think we can make quite quickly. Others, culture, leadership, vision, they take longer. We have plans in place to really change the way in which we behave, particularly with regard to risk and value for money, over the course of the next year. I am confident, and I have quite a degree of support from the Department, in making sure that those are happening and that I am held to account for them.

Q175 **Lilian Greenwood:** One of the key recommendations of the review was around leadership and management. Maybe you could say a bit more about those areas and how you are going about making changes to meet these recommendations, including the sorts of changes in culture and style that you have just alluded to.

Claire Burton: What I think is one of the most significant findings was around the fact that there was a silo culture, and the understanding of what was happening across the whole agency really only happened with me. That was a failing, because it meant that my senior leadership team did not necessarily feel that they were involved in decisions across the agency. We have taken the decision to restructure slightly, so I now have a strategy team who work within the agency, sit within the assessment policy, but who work with me on trying to ensure that we have a common vision across the senior leadership team, which is my three deputy directors, and then across the senior management team, which is across the wider range of managers within the agency.

Part of that has been about trying to ensure that we have a common idea of what we are doing and a shared appreciation of each other's work, because one of the issues that came out was the sense of a division between the delivery part of the agency and the test-development part. Again, that has been about building an appreciation of each other's skills and work. I suppose the way in which I have tried to lead the response to the review is to engage as many of my staff as possible in coming together to discuss some of the issues. I have made some top-down changes to my leadership team and to my senior management team, but for the bulk of the changes we are putting together groups from across the agency, so representation from across our sites, across the various divisions, to really come together to discuss and then put in place the products and the activities that need to happen to embed the change.

I think that approach will probably take longer, but I think it is more sustainable, and more holistic, and that it engages. At the moment we



HOUSE OF COMMONS

have about half my staff in the agency involved in various different strands of the review.

Q176 **Lilian Greenwood:** Do you think staff feel confident to raise concerns if they think things are not right or going wrong?

Claire Burton: I do. We have quite a strong culture building now of looking at risk, but also looking at near misses, and rather than just looking at things that might happen, looking at what has happened and how we can learn from the experiences. We are beginning to build that culture. It is not there yet. It takes time, but I am confident that we will get there.

Q177 **Lilian Greenwood:** Are you being given extra resource or support from the Department to ensure that mistakes are not repeated?

Claire Burton: We have managed so far within our existing resource. It has been largely through recruiting to fill vacancies that we had. The Department have been clear that they will support us. If there are specific areas where I need more resource, I know that I have the Department's backing and I can go to the Permanent Secretary, who has been very clear about that, but so far I have not needed to. It may be that I will at some point, but I want to be able to manage this within my staff and within the agency, if I possibly can.

Q178 **Lilian Greenwood:** We noted as a Committee that there had been a £6.8 million reduction, I think, in the STA programme costs for external marking. We had a concern about whether cost savings were made at the expense of quality. What assurances can you give us in relation to that change?

Claire Burton: The cost savings made in our programme budgets are largely due to contract negotiations that we have, and our commercial relationships with our partners. They do not relate to staffing costs or resource costs. I am very clear—and this was one of the findings of the STA review, which I think is quite difficult to balance—that we constantly have to balance the fact that we are using taxpayers' money against the fact that we have an accountability system that relies on the data that we produce. Therefore, it is right that we are robust in making sure that data are as rigorous as possible, which in some cases does mean that it is not the cheapest option that we could possibly go for. Those are balances that I think I have to make.

Q179 **Lilian Greenwood:** But you are confident that you have not gone for cheap at the expense of not being of sufficient quality?

Claire Burton: I am confident that there is no compromising on quality.

Q180 **Chair:** Michelle told us that Ofqual, with its strategy and risk management, would not have been able to spot what was happening last year. Do you have adequate risk analysis in your structure?



HOUSE OF COMMONS

Claire Burton: At that time we had very close scrutiny and very adequate risk assessment of the risks in our supply chain and our commercial areas. We had processes in place within the agency, but what we had not done was ensure that they were being refreshed and used, and that we had really embedded them, particularly with new staff coming in. It would be fair to say that we did have a gap in our risk assessment, and it was a gap in looking at our internal risks rather than our external ones. That is certainly something that we have filled, and we have filled very rapidly over the period of last year.

I am now confident in our risk processes. I joined in the January of 2015 and I inherited a very strong culture of continuous improvement and, because our work is cyclical, finishing every cycle with lessons learned, which were then embedded in the next cycle. What I tried to do is bring that into our day-to-day business, as well as the end of year. But we were, and have always been, an agency that has learned from its mistakes.

Chair: Okay. Thank you. Marion, you are going to talk about support for schools.

Q181 **Marion Fellows:** Yes, I am. Good morning. Ofsted told the Committee that the quality of assessment systems now used in schools is mixed. What more should have been done to avoid this situation, and what do you think can be done now to improve it?

Claire Burton: I think one of the most significant changes that I referred to earlier was the removal of national curriculum levels, which had undoubtedly become embedded in a large majority of schools' internal assessment methodologies, and that was having a damaging effect on pedagogy. It encouraged teaching to the levels, rather than teaching to the curriculum. But removing that has removed a significant aid that teachers were using. The risk at the time was that if we had put in place a prescribed alternative we would have had an utterly chilling effect on the ambition, which was that schools should have the freedom, flexibility, and autonomy to deliver and develop in-school assessment that meets their own approach to the curriculum, their own community and their own context.

What we did was we commissioned the Commission on Assessment Without Levels, chaired by John McIntosh. I think that has been well received, and it came up with recommendations for schools looking at things that they could think about when they are developing their own methodologies, the sorts of areas that they might want to look at and some advice and guidance on buying commercial products if they really felt that was necessary. I think that was well received. We have also looked at, and are in the process of looking at, the recommendations and delivering them. So I think for us it was a balance against really encouraging schools and trying to create that peer-to-peer approach where they are learning from the best—I think we will see that going



HOUSE OF COMMONS

forward; last year was the first year—and also providing the support through the commission to give schools an idea of where to look.

Q182 **Marion Fellows:** Sally, do you have anything you would like to add to that?

Sally Collier: Not with respect to a formative assessment in the classroom, no.

Q183 **Marion Fellows:** Okay. Claire, you have supported schools to implement the new systems following the removal. How are you going to continue doing that? You have already touched on it, but how are you actually going to continue this change this year?

Claire Burton: Just setting aside the statutory testing, which I am responsible for, on the more formative in-school assessment I think what we are really doing is focusing on those recommendations that came out of the Commission on Assessment Without Levels, which were around developing more units for initial teacher training so that teachers could be more confident in the use of assessment in the classroom. That is certainly something that has already been done. We will also be looking at areas like the item bank that they suggested. The idea would be for a bank of test papers that teachers could draw off to help them, by giving them a standardised approach to know where their pupils are in relation to the rest of the country. I think those are areas that we can certainly look at. I have seen some really impressive teacher-assessment models beginning to develop now and, as they become more widely known, and as specialist leaders in education who work confidently move to teach in other schools and spread them, we will be seeing more of that.

Q184 **Marion Fellows:** Okay. Do all three of you think the Government should consider implementing a national assessment system to replace levels?

Claire Burton: I don't. My personal view is that we should try to keep statutory assessment, which covers the whole country, as far as possible from normal classroom practice. There is a real value in schools having freedom and flexibility to fit their own school assessment systems to the curriculum that they have developed and that they are working to. I do think that the moment you start putting a national system in place, it freezes that development. Supports, such as an item bank, are useful then. That is something that is being developed in Scotland, and I think those are really interesting developments. But a one-size-fits-all approach would be something that I think would recreate all the things we tried to remove in the past.

Q185 **Marion Fellows:** Do I hear any disagreement from Sally or Michelle?

Dr Meadows: No, I agree that the richness of formative assessment and what it can do for teaching and learning could easily be quashed by a national system. The purpose would shift. I think I agree.

Q186 **Chair:** Thank you. Before I start talking about the design of SATs, this



HOUSE OF COMMONS

Committee has heard about the idea that teachers might really benefit from more training in assessment. Dr Becky Allen notably told us that. Sally, do you have any thoughts on that?

Sally Collier: Certainly, as you know, for someone new into the assessment area myself, it is obviously far more complex than one at first thinks. I certainly would advocate more training and assessment. It is probably not, though, a one-off thing. I have read the evidence to the Committee and I think there are different ways that you could deliver that training, whether it is early on in teacher training or continuously. I certainly think it would be helpful; the question is how might you deliver that and how frequently, because the world changes constantly and the system and framework needs to adapt.

We are now seeing potentially more online testing and potentially more use of computers. How do we keep up that training and make sure that the whole system has the right balance? Perhaps I could come back and make a comment about statutory assessment versus all the other types. No testing system anywhere is perfect. There are imperfections and the question is: have we got the balance right—that is what this Committee is looking at—on what is statutory? Because it does place pressure on teachers and teacher assessment, depending on how those tests are going to be used. Sorry, that was a bit of a long-winded response. Yes, I think it is important.

Q187 **Chair:** Thank you. Claire, what are your thoughts on training?

Claire Burton: This is not my specific area of responsibility, but certainly when I go out and talk to teachers about the formative assessment they are using in classrooms, it is not an area that is necessarily intuitive. It is not something that people find is always easy to engage with. It is something where it is really important that school senior leadership teams have a strong view about the culture assessment that they want to create in their schools, and that they support teachers to really understand what assessment can do for them, which is tremendous amounts. There is a huge body of evidence that says that effective in-school assessment can have a really significant impact of children's progress.

Q188 **Chair:** Thank you. On the design of SATs, how are SATs tests piloted with children and teachers?

Claire Burton: I am really pleased you asked this, because I know that you have had a lot of evidence around the test development process. I wanted to talk about this because it is an area where I think, certainly for the STA, we are really proud of our development process. We develop tests over a two-year cycle. If we go back to the 2016 tests, we published the test frameworks in 2014 and questions were written on the basis of those test frameworks. The test frameworks set out what we would be testing, what the content of the tests would be, what the cognitive domains were, and they also set out a description of what the



HOUSE OF COMMONS

standard of the test would be. Over the summer of 2014 we trialled in schools. At the same time as pupils were taking their live key stage 2 national curriculum tests, they also trialled in between 300 and 400 schools, so about 4,000 to 6,000 pupils will have sat questions that were intended for the 2016 cycle.

We take the data from that trial and we run it past various panels. We have a panel of teachers who are based in classrooms, a panel of inclusion experts and a panel of curriculum experts, and they all comment. They look at the data and they look at the questions. Sometimes questions do not perform as you intend them to, so they might be changed or they might be discarded. The questions that are left are compiled into tests and set again in a wider trial, so that would have been in 2015, with usually between 10,000 and 15,000 pupils chosen from a representative sample of schools. It is chosen randomly so that we can get a representative sample.

Again, the data from that trial goes in and is then scrutinised by the same three groups, and finally compiled into tests that are sat in the live test cycle. There is a degree of rigour that goes into ensuring that those questions are as good as we can possibly make them, and that when they are compiled the tests are doing what we intended them to do. So they are testing what we intended to be tested and they are at a level that will allow pupils who are performing at the lower end of the achievement scales to get enough marks, but also they will really stretch the ones at the upper end. I think we have a very good idea even before we go into the live tests what those tests are doing, and where their strengths and weaknesses are.

Q189 **Chair:** How did it go for the reading test for 2016?

Claire Burton: The reading test is the one where I know we have had the most feedback. I suppose, and Ofqual can comment on this, it went through exactly the same process. It was scrutinised by teachers, inclusion experts, it had been sat and trialled in schools beforehand, and broadly the test did perform as we expected it to. It had sufficient marks at the lower end of the scale that we were able to differentiate pupils there. It also included that higher-level content, so we were able to look at the pupils who had previously perhaps been performing at that level 6 test that had been removed. It did all of those things.

However, this is where you come to the: how does the test feel? I do completely accept the feedback that some children really struggled, particularly with the first and second questions where it got quite difficult quite quickly. There is more that we can do around the child's experience of the test, but in terms of how the 2016 test performed, it performed well. I think that we can reassure people that the data is reliable and valid. Perhaps Michelle, who is the technical expert, might be able to say more.

Q190 **Chair:** Okay, Michelle, let's hear about the reading test.



Dr Meadows: We did scrutinise the process by which the test was developed, and what I would say is that is a very robust process. It is far more robust than what is used for general qualifications, GCSEs and A-levels. We looked at the distribution of marks that pupils scored, and the test seemed to be well targeted at the population of pupils taking it. In other words, it was nicely and normally distributed. What I would say is that that is a change on previous years. In the past the distribution of marks, if you like, has been skewed to one side, inasmuch as it has been easier, which of course we knew. It was a change in the challenge and the demand of the test. I do not want to disregard the feedback from pupils, parents and teachers, because while we can look and see that as an assessment instrument it seems to have performed well, there may be access issues that need to be looked at. It will be really important to look at how this test now beds down in future years as schools get used to this new, more challenging standard.

Sally Collier: We are looking in particular at this.

Dr Meadows: Yes, we are. We are conducting, at the moment, research to look at the validity of both this reading test and maths test at key stage 2. We are looking at the way in which the curriculum content has been sampled by both the reading and maths test, and also the cognitive challenges of the tests, and we will be pulling in SEN expertise to look at the test. It will be interesting to see how our findings differ from what was intended or not.

Q191 **Chair:** Thank you. We heard from pupils about a strong focus on the technical aspects of writing in key stage 2. Do you think that is the best approach to judge writing at that point, Sally?

Sally Collier: This is a very difficult one. First, would you use teacher assessment in statutory tests? Then, if you are using teacher assessment, would you use this type of framework? I come back to: no system is perfect. In the first one, should you use teacher assessment? In these types of high-stakes accountability tests there are risks to using these sorts of teacher assessments, because teachers will know unconsciously what these tests are being used for, and that is how they will make their judgments and mark. Everybody you have had in these seats has said that, including the teaching unions. I think there is a question about whether they should be used at all, and if they are used, are the frameworks the right thing? These are questions to be looked at.

There has been much talk of alternative use of different systems and of using comparative judgment to assess writing. It has its strengths; it has also introduced other risks. Certainly this is an area that needs to be looked at. We have not assessed the interim frameworks; they were interim frameworks. We will wait to see what this Committee says, what any Government consultation says, what happens and the STA's plans for this year, and then we will look at what further work we need to do.

Q192 **Chair:** Claire, what are the plans?



Claire Burton: Going back to your original question, which is around the double assessment of technical aspects. When we constructed the test frameworks we believed that it was important that children are able to use the technical aspects of writing that they demonstrate in the grammar, punctuation and spelling test, and that they can also bring those across into their everyday writing. That is the rationale, and I think that is the right thing to do. What we did see last year—and this goes back to the earlier points around guidance being late or not being complete—is some extremely contrived behaviours in classrooms where children were writing pieces specifically to meet the frameworks, and that was never the intention.

What we have done this year is make it much clearer, and we published guidance that we had worked through in quite close collaboration with teachers and with the teachers' unions, making it very clear that we are not expecting to see one piece of writing that meets everything in the framework. That would lead to children trying to shoehorn semicolons and commas into inappropriate use. What we expect to see is children demonstrating over the course of the year, and this is the benefit of teacher assessment: it allows children to grow over the course of the year, and it shows what they can do over that year. It means teachers are looking at multiple pieces. They might be looking at evidence from a whole range that demonstrates those areas.

I do accept that there were issues with the way in which the interim frameworks performed last year, because I think it did in some cases drive some quite poor experiences for children, who found themselves effectively having to write for a framework. We have tackled that for this year and I think we have made it much clearer. The other thing that we have done for this year is really looked at moderation. Sally described some of the issues around when you use teacher assessment you have to really make sure that the moderators have a very common understanding of what the expected standard is. Something we took feedback on last year was that the experience of moderation had been, in some cases, quite patchy.

This year we have held moderation events where we have trained moderators from every single local authority and they are now in the process of going through a standardisation exercise. If they pass, they will then be able to pass on the training; if they do not, we will come up with other alternatives. But we are really trying to ensure that that expected standard has been understood and internalised by the people who are really looking at teachers to say, "Have you made that judgment? Is that quite the right one there?"

Q193 **Chair:** You have touched on the training issue again obviously in connection with moderators. That is an important point, too, is it not?

Claire Burton: Yes.

Q194 **Chair:** What alternative methods of assessing writing are you thinking



HOUSE OF COMMONS

about?

Claire Burton: The method that we have at the moment is a result of the Bew Review in 2011, which recommended the grammar, punctuation and spelling test and the teacher assessment. That is the model that we are using. The Secretary of State in October last year set out her intention to consult on a broad range of areas, including the primary assessment and the teacher assessment, so we are expecting to consult on it shortly.

Q195 **Chair:** Okay. Generally speaking, as Sally noted, there are changing circumstances ahead. What kind of attitudes will you have to approach in terms of designing and overseeing testing in the future?

Claire Burton: In terms of how we run the tests this year?

Chair: Yes.

Claire Burton: The test development model that we have at the moment is well embedded, but the world of testing is always changing. We continue to incorporate some of the latest thinking in terms of the methodologies that sit behind some of our tests. We have quite an interesting new area with the multiplication tables check coming in in 2018-19, which gives us our first whole cohort for the on-screen test. That is certainly an area where we are seeing internationally there are developments in on-screen testing, and the real potential for that I think we have yet to unlock.

What I would say is that certainly for whole cohort primary assessment we have to be really cautious in how we approach wholesale change. I think last year showed us that and I would certainly think that moving to any new formats of tests we would want to be really careful about how we did that. My staff are for the most part test developers and they are aware of those sorts of innovations and interested in bringing them in.

Q196 **Chair:** One last question on this subject. We have just launched a report on teacher recruitment and retention, and one of the things, of course, is workload and constant changes from the Department and, indeed, by implication from STA and Ofqual. So far, we have been hearing about potential changes and more changes. Sally, how do you think we can allay the fears of the profession that there is going to be more change, given the need that you have both basically set out for change?

Sally Collier: I think we have to be extremely cautious about wholesale more change. It has been very clear from the feedback from the witnesses you have had and generally that we ought to be very cautious about too much change too quickly. I think we have changes in primary assessment. We have many changes in general qualifications, in vocational qualifications. Our job as the regulator is to ensure that there is a robust framework of regulation in place and proper safeguards and proper assessment of risks. We in Ofqual are extremely alert to the potential burden that we may be putting on schools. We seek feedback



and we are always assessing any new regulations in respect of the burden that we add. In fact, we have a statutory duty to look at what burden we are adding. So, I think extremely cautiously, and we would certainly recommend trialling any new approaches.

Q197 Lilian Greenwood: Chair, can I just bring it back with a quick question about the reading test? You said earlier—I think it was you, Sally—that it worked as an assessment tool and it produced a nice bell curve. Do you think that doing that, even though it works as an assessment tool, has unintended consequences either for individual pupils or for schools that have a large proportion of pupils from perhaps a low socio-economic background or children with an additional language? In evidence it was described to us by a deputy head as virtually inaccessible for a good chunk of children. I understand it works as an assessment tool, but do you think there are problems if that is how it is perceived?

Dr Meadows: That was our concern. You can look at a distribution of marks, see that they are normal and draw one conclusion, but if perhaps some students have found it particularly difficult to access, then although the distribution of marks looks right, the rank ordering of pupils could be suboptimal. That is why we are doing the piece of research that we are doing to see whether from the point of view of SEN experts, teacher experts and subject experts that test could be improved. We are very much interested in this.

Claire Burton: Some of the criticism I have heard is around the test not being trialled in schools that had high proportions of disadvantage, and I think that is not the case. We trial in a representative sample of schools, so those tests have been seen and they have been trialled with pupils. I welcome the work that Ofqual is doing because I think it is important and, as I recognised before, I do think that the feedback on the child's experience is something that we can take on board for future tests. As you say, you can develop something that works extremely well as an assessment, but we want children to also experience having as positive an experience as possible of that test. I do think that is something we can look at.

In terms of the reading test, it has been scrutinised by panels of inclusion experts, by teachers, so I am confident that the process it went through was rigorous, but I am interested to see what Ofqual can find because we can always learn from that.

Q198 Lilian Greenwood: Is there any research and evidence on what impact it has on young children who come to do a test and find they cannot do it and what that does for their future performance?

Claire Burton: There is some research. I cannot quote it to you at the moment. There is a limited amount. It has not been done by my agency. The thing I would say also about the reading test, though, is that the percentages of children attempting the last question in the test, which is broadly what we tend to use as our sign of whether children have



engaged with it, is pretty much exactly the same as it was in 2015 and 2014. You can draw whatever conclusions you like from that, but what I would say is that children were engaged and were working through the test, answering the questions that they could, exactly as their teachers will have supported them and helped them to do. It is not necessarily the case that we saw widespread dropping off of children as a result of difficult first questions.

Dr Meadows: I have just recalled one of the things that Ofqual scrutinised was the way in which STA looked at how individual questions had functioned for different groups of students—boys versus girls, and EAL versus non-EAL—and there wasn't evidence from the available analyses that had been done to suggest that the individual questions were biased and inaccessible for those groups. I do want to see what our research finds, but the evidence that we interrogated so far does not suggest a particular problem.

The other thing to note, of course, was that the reading test was the first test to be sat, I believe, so it was the first experience for schools of this new challenging standard.

Chair: Okay, Marion on accountability.

Q199 **Marion Fellows:** We are looking at baseline measures, I believe, sorry. Do you think the introduction of a baseline measure at age four would be a positive step and, if so, what do you think should be assessed in this measure?

Claire Burton: This is one of the areas that the Secretary of State has said that she is going to consult on as to what the baseline should look like and when it should be. You will know that we introduced an optimal reception baseline in 2015, but we did not proceed to make it statutory. That was largely because of the model that we had chosen to introduce it for. Technically, it is possible to create an assessment at age four. We have seen there are various examples of that available already commercially. It is also possible to create something that does not look like a test and that is still valid and reliable. I don't think that there are any barriers in those terms to creating something that will work as a reception entrance baseline, but it might be that Ofqual have a bit more experience around what that would look like.

Q200 **Marion Fellows:** Do you think it would be a positive step as well, though, Claire?

Claire Burton: I think the benefit of having a baseline at reception is, of course, that it allows schools to be credited with the progress that they make across the whole of the seven years of primary education, whereas at the moment with key stage 1 we are really only looking at the progress between key stages 1 and 2. I do think that it has the potential to be very positive in looking at those early years.



HOUSE OF COMMONS

Dr Meadows: I think this could be positive. Ofqual's view would be that it would be important to trial this carefully. I think it is true that one can assess these young children reliably, but what you can assess reliably might be relatively narrow. There is a question about the extent to which that would be teacher assessment or some form of test perhaps with some teacher support. If it is teacher assessment, we have to take into account the pressures of the accountability system.

Then there is the whole question about how having measured something relatively narrow at this young age, how that correlates later with key stage 2 performance. There are some unknowns I think that we would want to see it trialled because it is potentially something that would be a really positive step, but we would want to see how it actually operates once it becomes used as part of an accountability system.

Q201 **Marion Fellows:** We had some evidence that there could be difficulties with that because if you set the baseline there, the baseline might be set lower so that progress looked higher. That would be a difficulty. Do you think overall the accountability system should put more emphasis on progress rather than attainment?

Claire Burton: I think the accountability system as it is set out now does put a strong emphasis on progress, but it maintains that balance. For me, the important thing is that we need to know that children are finishing primary and that they are ready to go on and achieve in secondary. That attainment is really important if they are going to go on and be successful in later life. Equally, I think the current system really does recognise that schools are making progress and that progress is incredibly valuable to those children.

I personally think that the balance at the moment feels broadly right. I would recognise Michelle's point that the accountability has an impact on assessment when you use it for statutory purposes, so having a balance of progress and attainment feels like a good compromise.

Q202 **Marion Fellows:** Do you have anything else to add to that, Michelle or Sally?

Dr Meadows: Of course, accountability is a matter for Government rather than Ofqual, but I think Ofqual's perspective is always that where high stakes judgments are being made on the basis of testing, it is always better to use not just one test but multiple indicators so as to ameliorate the pressure on any one individual test.

Sally Collier: Clearly, if we are going to measure progress we need a reliable baseline. I think you have also heard evidence about whether we could improve the use of the key stage 1 test. We would certainly advocate doing what we can to get to a more reliable baseline, whether that is a new reception test or whether that is making improvements to the current system.

Q203 **Chair:** That is a key issue, isn't it, because what does a good baseline



HOUSE OF COMMONS

measure look like? Do you have any thoughts you can share with the Committee, Claire?

Claire Burton: To function effectively as a baseline, it needs to be reliable, in that it needs to apply to all schools and be able to measure a standard approach. It also needs to correlate to key stage 2 or wherever you are finishing your baseline as well. Those for me would be the two functions, the extent to which you can demonstrate that it holds up as a national standard and the extent to which you can correlate the findings into future academic career.

Q204 **Chair:** Do you concur with that, Sally? Presumably you do.

Sally Collier: Yes.

Chair: Michelle?

Dr Meadows: Yes, and I think the big question is that correlation and how wide or narrow that baseline test can be, in terms of what you can reliably assess with four-year-old children.

Chair: Well, that is going to be a very interesting topic for us to dwell on. Thank you very much indeed for coming before us today and answering our questions. It is much appreciated.

Examination of witnesses

Witness: Nick Gibb MP

Q205 **Chair:** Good morning, Nick.

Nick Gibb: Good morning, Neil, Mr Chair.

Chair: It is always good to see you here. You are a regular performer, so you know how we operate.

Nick Gibb: I certainly do and am looking forward to it very much.

Q206 **Chair:** Thank you. The overall process that we have just been discussing was featured in the media and not in a complimentary way. It was I think described as chaotic generally. What do you think really went wrong?

Nick Gibb: It was always going to be a challenging year, as we heard from Claire, because this was the culmination of six years of work. We launched the national curriculum review in early 2011. We worked solidly on that for two years. We started the informal consultations on that curriculum in 2012. We published the final version in July 2013 and then again in September 2013, and that gave schools a year in which to prepare for the statutory implementation of the national curriculum in September 2014. Then two years on from that were going to be the first SATs.



This was always going to be a challenging year. This is a significant step up, in terms of the level of expectations on schools and pupils. We were concerned that the old national curriculum did not have sufficiently high expectations because, for example, if you were not to achieve a level 4, the expected standard in the SATs up until 2015, only 7% of those pupils went on to get five or more good GCSEs. We had to raise the standard of attainment, the standard of the curriculum, raise the expectations in that curriculum so that it was more on a par with other high-performing countries around the world.

It was always going to be a challenging year, and then on top of that the curriculum review recommended the removal of levels. We heard in the previous session the importance of removing those levels because levels had been introduced as a form of summative assessment at the end of a key stage, but they had turned into an assessment system within the classroom per pupil and it was beginning to damage the teaching of those children. They had to go. It was recommended by the expert panel that they went. That then meant a whole different approach to grading and assessing those tests, which led then to standardised scores in terms of the marking as well.

Q207 **Chair:** So, there was quite a lot going on and you basically think that that was the fundamental cause of the difficulties?

Nick Gibb: Yes, because we had to publish all the different frameworks, but there were significant lead times despite that. We published the test materials in June 2015. The frameworks were available at the beginning of that academic year, September 2015, so schools did have the materials. There was criticism about the quantum of materials that came through during 2016, but those were intended to be exemplification and helpful materials for schools.

There was an issue about the writing assessment exemplification materials—Leigh versus Morgan, if you remember that issue. It was about the real work of children that was meant to demonstrate the standard that schools should be aiming for, and the child who was named Leigh—it was not the child's real name, but it was his or her real work—was at a standard that was seen to be very high, beyond the 4B that people had been expecting. That is because Leigh was actually meant to be someone towards the top end of the expected standard and Morgan was meant to be the standard of work that had just gone into the expected standard. But because L came before M in the alphabet people looked at Leigh and did not look at Morgan and that did cause some concerns.

Q208 **Chair:** How would you describe the implementation of the new national curriculum assessments for last year?

Nick Gibb: It did not go as smoothly as we would have hoped—I absolutely have to say that—and we would have liked things to have been smoother. We would have liked there to have been things published



earlier than they were published. Nonetheless, the curriculum was published in July and September 2013. Schools had a long time. They knew what the curriculum was and there were test materials published in June 2015. In retrospect, perhaps it would have been better had we highlighted those publications more—the communications might have been better—so that when they saw the later materials there was not the impression that that was the first information being supplied to schools, because it was not.

Q209 Chair: What will be different this year?

Nick Gibb: We have now implemented this new assessment, so schools are used to it. As we have heard from the Standards and Testing Agency, they will be more timely in terms of when the different materials are published.

Q210 Chair: Can I ask you about baseline measure? What is your general view of a baseline measure?

Nick Gibb: You have to have a baseline if you want to measure progress for a school, and it is felt that it is fairer to assess a school in terms of the progress they make for all their pupils rather than on simply raw attainment. In primary, of course, raw attainment is very important because every child needs to leave primary school as a fluent reader. Every child needs to leave primary school fluent in mathematics and have a good grasp of grammar, punctuation and spelling. That is why we do have a combination of attainment and progress in the floor standards and the coasting standards. Notwithstanding that, it is important that schools receive credit for the progress that they are able to help children make during their time at that school, and that is why you need a baseline from which to measure that progress.

Q211 Chair: How would you describe a baseline that would be appropriate for the task that you have just set out?

Nick Gibb: It does need to be reliable, and when you are engaged in what will become high-stakes accountability measures it is important that it is reliable and that Ofqual, for example, is able to verify that they can trust it.

Q212 Lilian Greenwood: Can I follow up on that? Is it your view that that means it should be a testing model rather than an observational model, given that we are talking about children who are only four or five years of age?

Nick Gibb: That is something we are going to be consulting on. We want to consult as widely as possible about these very issues, and not just about that, but about the timeline. When should you introduce the baseline in terms of getting the best progress measure possible? That is something that we will be consulting on very shortly.

Q213 Lilian Greenwood: Are you open-minded about whether it is a testing



model or an observational model, or do you have a view?

Nick Gibb: Open-minded. The key factor is whether it is reliable and that will be the most important thing.

Q214 **Chair:** Just to probe that point further, do you have an instinct that observation or testing would be the more dominant for reliability purposes?

Nick Gibb: These are the issues that we are going to be consulting on. From my perspective, this is a baseline that has to be accurate and there should not be incentives for it to be too low or too high. It needs to reflect accurately the level of attainment of the children at the start of the process from which you are measuring progress. One of the issues is the timing, but it is also about accuracy and that is the issue that we will be genuinely consulting on and taking evidence to ensure that we have a system that is robust and that is sustainable.

Chair: Okay, thanks very much, Nick. I think the next person to comment is William on the removal of levels.

Q215 **William Wragg:** That's correct. Good morning, Nick. With the benefit of hindsight, did schools have the support they needed to design their own assessment systems following the removal of levels?

Nick Gibb: That was the reason why we asked John McIntosh to chair the Commission on Assessment Without Levels, to ensure that schools were being advised about how to deal with assessment without levels.

Q216 **William Wragg:** A lot of the evidence that we have, and anecdotally as well particularly from my personal background and within the constituency, is that there was not that support and there was a great deal of confusion. In a sense, as somebody put it to me, in compelling them to be free—and you will spot the contradiction in that statement—when people have been so dependent on the national curriculum levels and sub-levels, they did not feel that they had the support and, indeed, with high-stakes accountability there was still a need for a measure. Do you not accept that, given the massive change that it was, more support should have been given or, indeed, greater time?

Nick Gibb: We did not want to return to the Michael Barber cascading national strategies approach. We are, in terms of school reform, generally trying to move towards a school-led system and a system where professional autonomy is paramount. Throughout our reforms we have not gone for the cascading big conferences where these issues are conveyed from Whitehall to 23,000 schools or 16,000 primary schools. That has not been the philosophy behind our school reform measures. Therefore, there is always a reluctance to deliver that kind of approach when making reforms. However, we did take on board the concerns that were being expressed and that is why we established the commission that was chaired by John McIntosh.

Q217 **William Wragg:** Still, when assessment is so closely linked to



accountability, which of course is a standard national model of accountability, is it right to leave individual schools to design their own assessment systems?

Nick Gibb: Yes, because we want schools to design their own curriculum. There is a national curriculum, but that is a minimum level and what we are trying to establish throughout the school system is a school-led system with professional autonomy where schools can design their own curriculum beyond the national curriculum. There is a distinction between a national curriculum and a school curriculum, and the assessment process needs to be linked closely to that curriculum.

Q218 **William Wragg:** We have also had evidence in terms of the commercial products available for schools that are of varying quality and use. Does that concern you? Have you thought of, or maybe already done, sharing best examples and best practice?

Nick Gibb: Yes, it does concern me. It was a concern highlighted by the Commission on Assessment Without Levels and we are working with the STA on having an item bank—again, that was one of the recommendations that came out of John McIntosh’s commission—of questions and so on that can be used by schools. That is something that the STA is working on at the moment.

Q219 **Lilian Greenwood:** The assessment of writing has been heavily criticised this year. There have been concerns about the role and nature of teacher assessment, about secure fit and about the necessity and technicality of the spelling, punctuation and grammar tests. What alternative methods of assessing writing are you considering?

Nick Gibb: The Bew Review in 2011 made it very clear that assessment of writing should be by teacher assessment, and I think that is something I certainly agreed with at the time. I have seen nothing that would lead me to change that view. Teacher assessment is by its nature a complicated issue. It involves moderation. It involves having to establish a framework, and in establishing that framework we were very keen to avoid the danger of re-establishing levels because the national curriculum review was very emphatic in its opinion that levels were having a very damaging effect on pedagogy within the school.

We did not want to have an interim teacher assessment framework that somehow reintroduced levels, so what it did introduce was these can-do statements. For example, the pupil can write using paragraphs to organise ideas, can describe settings and characters; that is working towards or working at the expected standard. The pupil can write for a range of purposes and audiences, creating an atmosphere, selecting vocabulary and grammar and construction. We have a series of “can” statements that are directly linked to the national curriculum.

One of the dangers with the level system was that it was a best-fit system, so you could get a level 4 even if there were some elements of that level that you were not able to do. That meant that there was a



HOUSE OF COMMONS

danger of children being pushed through the curriculum when they had not mastered some of the key elements of the curriculum. That was one of the reasons why the expert panel recommended removing levels.

That is why we went for the “can” statements and why secure fit was the approach taken in 2016, because we wanted to be sure that pupils were able to do all these things. We heard from Claire that there was a danger then that in providing evidence for moderation, schools were trying to shoehorn a piece of work that reflected all the “can” statements, which was absurd and was never intended. This is an issue that we are going to be consulting on very shortly.

Q220 Lilian Greenwood: Can I press you a little further? Are you considering alternative methods of assessing writing and, if so, what are those? What considerations are you making?

Nick Gibb: Certainly not moving away from teacher assessment. Teacher assessment of writing is, in my view, the correct approach for a whole raft of reasons. What we are going to be consulting on is how that teacher assessment should be conducted. We want to make sure the moderation is right. We want to make sure that how teachers prepare the evidence for that moderation is right. These are the kinds of issues that we will be consulting on very shortly.

Q221 Lilian Greenwood: You touched on some of the issues around secure fit. Are you concerned about and are you planning in any way to address the perceived impact on children with perhaps dyslexia that affects their spelling or dyspraxia that affects their handwriting?

Nick Gibb: This is an issue that has been raised by a large number of people. In terms of the spelling element, it is only an element of the marks within the writing assessment. It is important that those children are supported and part of the purpose of assessment is to identify children that do have needs in that area. As Claire pointed out, we have panels of teachers that assess the frameworks and the tests, and that includes experts in the field of special educational needs and disabilities.

Q222 Lilian Greenwood: Can you tell us what evidence there is that a strong focus on grammar, spelling and handwriting rather than on composition and creativity improves pupils’ writing?

Nick Gibb: During the curriculum review, and also the Bew Review, the introduction of the grammar, punctuation and spelling test as a separate test came out of the Bew Review when he took evidence. The expert panel also produced a report from their survey of international systems throughout the world and their recommendations came out of that and led to the curriculum that we now have in place. From that curriculum we then produced the tests.

Q223 Lilian Greenwood: Finally, can you explain to me or set out what your role as the Minister is in the design of tests or what is tested? Is that something you think you have a role in?



HOUSE OF COMMONS

Nick Gibb: We do not have a role. We have a role in the curriculum. We can see and we comment on some of the documents that are published, the framework documents. In terms of the preparation of the three-year cycle for the preparation of these tests, it is very guarded. I do not get to see a test until after it has been sat and even then, even after it has been sat, it is some time before I am able to see a copy of the actual test paper. I have now seen, of course, the 2016 tests and have taken on board the comments that people have made and they have been fed back to the STA. They will ensure, and I will ensure, that they reflect those in future years' testing.

Q224 **Lilian Greenwood:** I have heard in the House of Commons people call for there to be tests of times tables. Is that something you think Ministers should comment on, or should it be down to expert evidence about whether that would be a helpful thing in the assessment of children's mathematical learning?

Nick Gibb: I think it is an issue of policy and it is my view that there should be a multiplication check. It was in our manifesto in 2015 that there would be. We think times tables are a very important part of mathematical knowledge. E.D. Hirsch talks about automaticity, that the working memory is small, and Daniel Willingham also talks about what is an educated person. An educated person is somebody who has knowledge in his or her long-term memory. Automaticity is very important in terms of mathematics. If you are trying to perform long multiplication or long division and you are dividing seven into a four or five-digit number, you need to know your seven times table to be able to do that calculation. That is why it was in our manifesto. It is why we are introducing a multiplication check in 2018-19.

Q225 **Chair:** The answer you have just given to Lilian, Nick, raises the question of whether or not the STA is properly independent from your Department and why the national curriculum assessment is not done by something more independent, as other qualifications normally are.

Nick Gibb: It is independent. They are very guarded about the tests. They are very guarded about the development of those tests. I have no influence, nor does the Secretary of State, on the choice of questions or the choice of texts that are used in the reading tests and so on. Where we do have a role is in the national curriculum and the policy leading up to the national curriculum, albeit we consult very widely on developing that national curriculum.

Chair: Thank you. Marion, you are going to be talking about negative impact, aren't you?

Q226 **Marion Fellows:** Yes, I am indeed. Good morning. We have heard that subjects such as science, humanities and art are squeezed out of the curriculum because of the accountability and assessment system. How are you ensuring that schools still continue to teach a broad curriculum when they are under so much pressure to improve results in English and



maths?

Nick Gibb: The national curriculum is very clear about those other subjects. They are all compulsory in the curriculum leading up to end of key stage 3, and beyond in some of those subjects. I make the case that those schools that do well in their English are the schools that have a broad and balanced curriculum. What we wanted was a SATs system that you could not teach to, and the way to do well in the reading SAT is to have a good vocabulary. The way to establish a good vocabulary is to have read a lot. My view is that the more knowledge-rich your curriculum is at school, the more your vocabulary will develop. The stronger your vocabulary is as a child, the more accessible more challenging books become, the easier they are to read, and the more challenging books and the more books that you have read, the more effective will be your result in the reading SAT, reading comprehension, that is taken in May. I think it is a shortcut and a wrong approach to just focus the curriculum on those subjects that are tested in the SATs. I think that is not the way to do well in those tests.

Q227 **Marion Fellows:** I don't think anyone around this table would disagree with the words you have just said, but how are you ensuring that this is what is happening in schools?

Nick Gibb: The law is clear, so maintained schools are required to abide by the national curriculum. We also sample test science. It is not tested in the same way as maths and English and writing and grammar; it is tested using similar tests but on a sample basis.

Q228 **Marion Fellows:** You mentioned local authority maintained schools, so how are you doing it in academies?

Nick Gibb: Under the funding agreement they are required to deliver a broad and balanced curriculum, and Ofsted do assess whether a school is delivering a broad and balanced curriculum.

Q229 **Marion Fellows:** Okay. Pupil wellbeing is a real concern for parents and teachers. Do you think it is right to set the standards and the SATs so that nearly half of pupils are labelled as failures when they leave primary school?

Nick Gibb: Well, they are not. These SATs are about accountability for the schools and, although there is a requirement for the school to report to parents the results of the reading and the results of the maths and so on, that needs to be delivered to the parents in the context of the overall achievements of those pupils at the school, not simply on the SATs results. If you segregate out the different subjects, 66% of pupils achieve the expected standard in reading, 70% achieve the expected standard in maths, and 74% achieve the expected standard in writing. These are high percentages at a time when this was a significantly more challenging curriculum and set of SATs than in previous years.

Q230 **Marion Fellows:** Would you agree that some parents only see the fact



HOUSE OF COMMONS

that they did not do well in the SATs? They are told their children failed this. That can have an impact on the child.

Nick Gibb: Well, they should not be told that they have failed. If they do not achieve at the expectation, that should be reported to parents in the context of their broader achievement at school, not simply the results of that particular SAT test.

Q231 **Marion Fellows:** Do you believe that is what is happening?

Nick Gibb: It is what ought to be happening at schools. We do not want to be labelling any child as a failure, and the same applied when we had level 3s. One of the concerns I had about the level system is that there was a labelling happening. I went to schools and a child would tell me, "I am a level 3 child". To move away from that was one of the reasons for removing levels from our system. We want children to do their best, to work hard, but we do not want them put under undue pressure in these SATs. We have to keep reminding ourselves that the purpose of SATs, the reason why they were introduced originally in the late 1980s, was to hold schools to account and that is the purpose of them.

Marion Fellows: I press the point because we did have evidence from one school who said that many teachers felt that they were concentrating on English and maths because of the SATs and that pupils were feeling that they had failed because they did not do well in the SATs. I think you still maybe have an issue there.

Q232 **William Wragg:** One of the consistent themes throughout our inquiry has been the link between assessment and accountability. Is it necessary to have such a high-stakes accountability system at primary school?

Nick Gibb: I think schools need to be held to account, and combining progress and attainment is a fairer system than the previous system that was simply based on attainment alone, albeit that attainment is important because we want all children to be leaving primary school as comfortable readers and to be fluent in mathematics and literacy.

Q233 **William Wragg:** Given that the basis of much of Ofsted's judgment will be upon the attainment at the end of key stage 2, and bearing in mind the recommendations of the Bew Review, has the Department considered using key stage 2 data differently or considered a wider range of measures to avoid any negative impacts of high-stakes accountability?

Nick Gibb: I am not sure what you are hinting at. The way that the floor standard works is that you have to either achieve 65% achieving the expected level or at least above a minus five in English reading and a minus five in mathematics and a minus seven in writing in the progress measure. I think that is a fair system of determining whether we regard a school as underperforming.

Q234 **William Wragg:** Okay. Perhaps going back to the question that cropped up earlier in terms of a baseline assessment, and particularly for accountability purposes, what would make it reliable and when would any



HOUSE OF COMMONS

consultation begin on determining the nature of that baseline assessment?

Nick Gibb: We have been working on the assessment approach ready for consultation and we will be consulting on that shortly. The issue is about timing. As we heard from Claire earlier, if you want to measure progress across the whole of the primary school phase, you would need to have it in reception. Then the question is: when in reception? At the moment we already assess pupils in reception. We assess them at the end through the early years foundation stage profile. There are other points at which you can determine a baseline. For example, at the moment it is determined from key stage 1, the key stage 1 teacher assessment. The key thing is that it should be reliable and robust and such that it satisfies Ofqual, the regulator, that this is a sufficiently reliable measure of progress.

Q235 **Chair:** Nick, you have already just reminded everybody that the Government is planning another consultation on primary assessment, baseline and so on. Do you think you should have carried out more thorough pilots or a consultation originally to avoid what happened last year and this next phase of consultation?

Nick Gibb: There was a long lead time. We started the curriculum review in early 2011 and the first SATs assessing the outcome of that new curriculum were sat in May 2016. So these issues do have a very long lead time and our view has always been that the existing curriculum, the existing SATs, did not have sufficiently high expectations of our young people when only 7% of young people who did not achieve the expected level in the 2015 SATs would have gone on to get five good GCSEs in 2015. This is the concern that we had, that the expectations were too low.

I am not saying we were in a hurry, because five years is a long time in terms of getting the system in place, but we did not want to wait another year and the work had been done; it was a thorough piece of work. We had some superb specialists advising us, the expert panel, and so on, a very wide consultation throughout 2012 and part of 2013, schools had a long lead time of a year or more—if they had followed the informal consultations they would have had even longer, but they certainly had a year—and then two years of teaching leading to the SATs. The test materials for those SATs were available in June 2015, so nearly a year before they were to be sat in May 2016, and the assessment frameworks were ready in September 2015. So I think things were in place. There were issues. For example, we had the breach of the security of the grammar/punctuation test, and we had multiple messages that were coming out from the STA that were designed to be helpful to this system but which, I think, served to overload the system with communications and that was part of the issue that led to the impression that this year's SATs were not under control.

Q236 **Chair:** Do you think the changes within the STA, as we heard earlier, will



HOUSE OF COMMONS

help deliver the kinds of outcomes you are expecting from now on?

Nick Gibb: Yes, I do. I was the Minister who ordered the root-and-branch review of the STA; I was unhappy about what had happened over the breach. It was a human failure, but it was a human failure that should not have happened, in my judgment. I reacted to that swiftly. As a Minister, this is the last thing you want to happen on your watch and I took the view that the key thing was to act swiftly, to take a decision about whether to use the grammar/punctuation key stage 1 test that year, and then to order the root-and-branch review. I am confident, now, that the measures have been put in place to ensure that we have a more robust and reliable Standards and Testing Agency than we had before.

Q237 **Chair:** Do you have any changes in mind for 2018-19?

Nick Gibb: The Secretary of State gave a clear commitment in her statement in October that there will be no change, no new tests introduced, before 2018-19. She also made it clear that we will be introducing the multiplication tables check in 2018-19. In terms of the cycle, as we have heard, there is a two or three-year cycle as you prepare new tests and we take on board the feedback we have had on the reading and the other SATs that were sat this year. Those will lead to reforms in the system as we go forward.

Q238 **Chair:** How will you reassure teachers that they will feel comfortable about the level of consultation and involvement and so on in this process?

Nick Gibb: We have gone to huge efforts to consult widely. We consulted widely during the preparation of the national curriculum, both informally and formally. We consult on the frameworks and we are going to be consulting again on the primary assessments.

What the Secretary of State and I want is stability in the system and that is why we have taken our time before making some decisions upon which to consult, because we want the outcome of that consultation to be a settled, statutory assessment system in primary schools that will last for the long term.

Q239 **Chair:** Finally, we heard about training earlier: training moderators and, indeed, actual training of teachers in the context of classroom assessment, pupil assessment. What are your thoughts about the requirements there?

Nick Gibb: There are two issues. One is the moderation. We have improved the training of local authority moderators. There does need to be consistency between different moderators across the country and that has, I think, improved significantly.

In terms of teachers, the teachers' standards that were produced after Sally Collier's review of the teaching standards had a key element to them, which is assessment. To be able to become a newly qualified



teacher, to get through initial teacher training, you have to fulfil those teachers' standards at the level at which you are in your career.

Q240 **Lilian Greenwood:** Chair, could I seek a quick point of clarification?

Chair: You are welcome.

Lilian Greenwood: There is a statistic you have quoted a couple of times during today's session; I think I have it. Only 7% of young people who did not achieve the required standard in their SATs went on to achieve five GCSEs. Is that right?

Nick Gibb: I have the wording in front of me. In 2015, only 7% of pupils achieving below level 4 at primary school went on to achieve five good passes, including English and Maths, at GCSE.

Q241 **Lilian Greenwood:** Of the pupils who did achieve level 4 or above, what proportion went on to achieve?

Nick Gibb: I feared you were going to ask that and I am looking behind me. There is a figure—it is not 100%, I can tell you that—and it is a figure that I regard as also too low, but if we do not get this number in time, I will write to you with that figure, because I think it is important for your evidence.

Q242 **Lilian Greenwood:** Chair, can I very briefly follow up? I completely understand the concern if only 7% of pupils who do not achieve level 4 go on to get decent GCSEs. Can you just explain to me why you felt that prompted the change to primary assessment rather than a focus on ensuring that all children reached level 4, or whatever is the equivalent? Do you see what I mean? Why not focus on those children who were not reaching the standard, rather than change the whole system?

Nick Gibb: Because of the other figure that I am trying to establish. The view of the Government at the time, of which I was a member, was that the general level of expectation in the curriculum was not sufficiently high, nor was it on a par with those education systems around the world that were significantly higher in the PISA league tables. It was to address that that we instituted the curriculum reform in 2011 and that is what led to the new primary curriculum, which I think is of a very high standard now. I am very comfortable with the work that was produced by the expert panel and the results that came out of the consultation on that curriculum; I think it is a curriculum we can be proud of. It is a step up, a significant step up, from where we have been and that is why 2016, the first year of assessment of that curriculum, was always going to be, and has proved to be, a challenging year. However, we now have a period of stability; schools are now used to that curriculum and they are certainly used to the assessment based on that curriculum. Going forward, I think we will have a period of stability.

Lilian Greenwood: Thanks.

Chair: Thank you, Lilian.



HOUSE OF COMMONS

I am now going to have another go at thanking you. Thank you very much for coming this morning. We have had a good dialogue.

Nick Gibb: Thank you very much.