



OFFICIAL REPORT
AITHISG OIFIGEIL

Education and Skills Committee

Wednesday 23 January 2019

Session 5



The Scottish Parliament
Pàrlamaid na h-Alba

© Parliamentary copyright. Scottish Parliamentary Corporate Body

Information on the Scottish Parliament's copyright policy can be found on the website - www.parliament.scot or by contacting Public Information on 0131 348 5000

Wednesday 23 January 2019

CONTENTS

	Col.
DECISION ON TAKING BUSINESS IN PRIVATE	1
SCOTTISH NATIONAL STANDARDISED ASSESSMENTS INQUIRY	2
PUBLIC PETITION	41
Free Instrumental Music Services (PE1694)	41
ATTAINMENT AND ACHIEVEMENT OF SCHOOL-AGED CHILDREN EXPERIENCING POVERTY	42

EDUCATION AND SKILLS COMMITTEE

3rd Meeting 2019, Session 5

CONVENER

*Clare Adamson (Motherwell and Wishaw) (SNP)

DEPUTY CONVENER

*Johann Lamont (Glasgow) (Lab)

COMMITTEE MEMBERS

Dr Alasdair Allan (Na h-Eileanan an Iar) (SNP)

*Jenny Gilruth (Mid Fife and Glenrothes) (SNP)

*Iain Gray (East Lothian) (Lab)

*Ross Greer (West Scotland) (Green)

*Gordon MacDonald (Edinburgh Pentlands) (SNP)

*Rona Mackay (Strathkelvin and Bearsden) (SNP)

*Oliver Mundell (Dumfriesshire) (Con)

*Tavish Scott (Shetland Islands) (LD)

*Liz Smith (Mid Scotland and Fife) (Con)

*attended

THE FOLLOWING ALSO PARTICIPATED:

Dr Keir Bloomer (Royal Society of Edinburgh)

Professor Louise Hayward (University of Glasgow)

Professor Lindsay Paterson (University of Edinburgh)

Gil Paterson (Clydebank and Milngavie) (SNP) (Committee Substitute)

CLERK TO THE COMMITTEE

Roz Thomson

LOCATION

The Robert Burns Room (CR1)

Scottish Parliament

Education and Skills Committee

Wednesday 23 January 2019

[The Convener opened the meeting at 10:00]

Decision on Taking Business in Private

The Convener (Clare Adamson): Welcome to the third meeting of the Education and Skills Committee in 2019. I remind everyone to turn mobile phones and other devices to silent, to prevent interference with the broadcasting equipment. We have received apologies from Dr Alasdair Allan and we welcome Gil Paterson, who is attending as substitute.

The first item of business is to decide whether to discuss our work programme in private at the next meeting. Are members content to do so?

Members indicated agreement.

Scottish National Standardised Assessments Inquiry

10:00

The Convener: Item 2 is the second week of the Scottish national standardised assessments inquiry. I welcome Dr Keir Bloomer, who is convener of the education committee of the Royal Society of Edinburgh; Professor Louise Hayward, who is professor of educational assessment and innovation at the University of Glasgow, based in the School of Education in the College of Social Sciences; and Professor Lindsay Paterson, who is professor of educational policy in the School of Social and Political Science at the University of Edinburgh.

I thank you all for coming along to participate, and ask you for your brief perspectives with regard to the SNSAs and the move away from the Scottish survey of literacy and numeracy.

Dr Keir Bloomer (Royal Society of Edinburgh): The Royal Society of Edinburgh has no objection in principle to standardised testing. It is concerned about the fact that we have, if anything, too little information and data about how the education system in Scotland operates, particularly in the whole of the primary phase and the early part of the secondary phase. The society broadly welcomes gaining more information through the introduction of the assessment. It does not necessarily say that it welcomes every aspect of what has subsequently taken place, but the principle is perfectly all right.

Our major concern about what has happened is that the purpose of the assessments has become less certain as time has passed. We were fairly clear at the outset that the main purpose was to monitor performance of the system, which we welcome. Since then, the emphasis has been placed on the diagnostic capacity of the tests and their ability to help teachers to help individual pupils. I know that the convener wants us to be brief at the outset, so I will not go into detail at the moment, but we are much less persuaded that the tests work effectively in that role. There are concerns about the way in which the tests are being used, but we think that the tests have the capacity to supply information that is of value and that has not been available hitherto.

We are puzzled by the abandonment of the Scottish survey of literacy and numeracy. It is unfortunate that there has been no continuity in the information that has been made available in the past. We had a previous assessment called the Scottish survey of achievement, which ran for five years; a short interval and then the SSLN ran for six years; abandonment; and now we have a

third system. A sample survey, which is what the SSLN was, is not incompatible with universal assessment of the kind that is provided by the new SNSAs. We do not see what the rationale was for abandoning the SSLN, because it would be perfectly possible to run both systems in parallel.

Professor Louise Hayward (University of Glasgow): It is important to remember that both the SNSA and the SSLN are simply different ways of collecting evidence, and both sit within the national improvement framework.

It is important to see any part of the system in the context of the whole. Tests and surveys are different ways of collecting evidence—the important thing about them is their purpose—and the different ways of collecting evidence relate effectively to different purposes. Once there is clarity about the purpose, and we know what we want to find out, the second-order question is what the best ways are of finding that out. That is an encouragement to come back to what the central purposes are. The idea of having information from tests that supports teachers' professional judgment is an entirely appropriate approach. The issue is, however, that we have to decide what matters. If it is the curriculum for excellence, our assessment system should reflect all that matters in CFE. We have to find ways of gauging how much and how well children are learning in relation to all those processes.

I turn to the move away from the SSLN. Surveys can provide very helpful information. If the purpose of such information is to give, at a national level, feedback on how the system is progressing, survey evidence is a very good way of doing so. It can provide evidence at that level without having the unintended consequences that other ways of collecting evidence can have, such as narrowing the curriculum or encouraging teachers to teach to particular parts of it.

Our central focus should be purpose, and then we should decide how best to collect information.

Professor Lindsay Paterson (University of Edinburgh): Thank you for the invitation to appear before the committee this morning.

Collecting data that is neutral and reliable is always better than not having it. Whatever their faults may be, the new tests—or assessments, as we are supposed to call them—are more reliable, neutral, objective and independent of bias than anything that we have previously had in Scottish education in recent decades. I say that because all of us who teach—I include here university teachers as much as any other sector of schooling—are unavoidably subject to bias, which is sometimes unconscious. We know that when we do not allow students' essays to be marked anonymously, there will be bias—for example,

against women, whose assessment is more accurate when it is done anonymously. That is an illustration of the bias that all teachers inevitably have. The bias of school teachers—which I should emphasise is no less than and no greater than that of university teachers—is evident from the previous survey, which was the Scottish survey of achievement. Year after year, it was systematically shown that when teachers assessed children they tended towards optimism—sometimes very great optimism—in comparison with the results of objective tests conducted as part of that survey. For me, the principal attraction of the new Scottish national standardised assessments is that they provide neutral, objective information that guards against bias. From the history of examinations, we know that guarding against bias has been one of the major means by which equality of opportunity has been improved—for example, for women; in Scotland, for Catholics; and, more recently, for ethnic minorities.

Secondly, I agree completely with what has been said about the abandonment of the SSLN: the two surveys could have run in parallel. The great advantage of a survey is that it can ask for a much wider range and much deeper kinds of information. Incidentally, I agree that the design of the Scottish survey of literacy and numeracy was not adequate for some of those purposes. The older one—the Scottish survey of achievement—was better. One way in which it was better was that it could provide a national picture, as has been said. The cabinet secretary's legitimate complaint was that the SSLN could not tell us where things were happening—where they were getting better or worse—which was a design feature of the SSLN but not of the SSA. The SSA's design allowed us to say, anonymously, that a particular school was doing better for certain reasons, such as homework practice, discipline or school uniform. In other words, it is possible to design a survey that gives us a national picture, and also council-level and school-level ones. Both could be done.

The Convener: Thank you. We move to questions from committee members.

Tavish Scott (Shetland Islands) (LD): Perhaps I could start with a question about purpose. Professor Hayward, in your very helpful submission, for which I am grateful, you say:

“These three main purposes interact in any national assessment system. Any action taken in one area will have an impact on the other areas.”

You said earlier that there was no clarity on the purpose of the standardised assessments. Should there be clarity and what should the purpose be?

Professor Hayward: I would argue that there is greater clarity, in that the national assessments are there to provide one part of the information

profile on an individual child. Lindsay Paterson pointed to the advantage of the reliability of the items in the survey. The danger is that that compromises their validity. These assessments are able to give information on only certain important aspects of the curriculum, but not all of it. For example, we might get information on punctuation or spelling, but writing is more than that. To get information on what matters, we have to depend on teachers' professional judgement, an approach that is central to Scottish policy. It is teachers who work day to day with the young people and who collect the information. The system should support teachers, so that we can build and enhance the dependability of their professional judgement.

Tavish Scott: So, there is more than one purpose. It is not just about supporting teacher judgments, it is also about the whole school system and understanding what that is doing and how it is performing. Is it fair to say that there are at least two purposes?

Professor Hayward: There are three main purposes, but we need to recognise that no one part of the process will be able to address all of those purposes. That is why we have a national improvement framework that draws evidence from a range of sources that are linked to a range of different purposes.

Tavish Scott: Does the panel believe that having three purposes is appropriate?

Professor Paterson: In the absence of any other source of information of the kind that we have already referred to in the case of surveys, it could lead to confusion. It would be possible to design what would unfortunately be a very cumbersome system, where all the SNSA results would be supplemented by the full range of kinds of information that one might collect in a well-designed survey. However, that would impose such burdens on teachers and schools as to make it unmanageable.

To take a specific example, the home language of the child is known through returns that schools give in the school census. That is already a difficult thing for the school to establish. If, in addition, they had to establish, for example, the education level of the parents and their occupation, or matters to do with the size of the family, the living arrangements or whether it is a single-parent family, that would be a ridiculous burden. That is not what schools are for. That is why sample surveys can give you deeper information. Therefore, although in principle you could design an SNSA-type thing that would cover all the purposes, I do not think that you could do so in practice.

Professor Hayward: I was not saying that.

Tavish Scott: I quite understand that. In your view, the Government, or those who are promoting the best motives behind standardised assessments, need to be very clear about what the purpose is. Has that purpose been established? Keir Bloomer, in your opening remarks you suggested that it had been changed.

Dr Bloomer: The emphasis has clearly changed. It was on national monitoring at the outset, and it is now more on the diagnostic capability. One has to recognise that the diagnostic value of the tests is limited. They have some strengths. They can monitor the same pupils over time, which we were not able to do through the sample surveys, because the same pupils did not figure in successive runs of the survey. We now have what are described as long scales, which stretch through from primary 1 to secondary 3, and it is possible to monitor how the individual pupil has progressed up the scale. That is valuable information, and researchers will be able to make something of it in the future.

10:15

On the other hand, the assessment looks at a restricted area of the curriculum once every three years. Therefore, as far as the individual is concerned, a minimal amount of information will be available at any given time. Although the information that is available in the print-out—it will be available to parents and teachers—has more value than it has been given credit for, it is still restricted. It is a standard description of—for example—what performing at band 6 means, and not much more than that.

Some teachers have raised with me another source of difficulty. One of the features of the assessment is that it is adaptive; as the child goes through it, depending on how he or she is getting on, they will be fed more difficult or less difficult questions. Therefore, in order for the teacher to work out what the outcome of the assessment says about the child, the teacher needs to follow the child's path through the questions. In the feedback, that is not easy to do. There is also the issue that children can get to the same banding as a result of taking different paths, so different interpretations might attach to that. There are complications in the nature of the feedback. Teachers would need to be aware of the strengths and weaknesses of the assessment in order to get what is of value out of it.

Tavish Scott: Thank you, that is very fair.

Professor Hayward, you have done some helpful international comparative work, which has been supplied to the committee. Correct me if I am wrong, but in your international comparisons I cannot find any other country that does P1 testing.

Am I missing something about how other education systems around the world look at what is happening or assess how children aged four and five are doing?

Professor Hayward: I am not sure that the question was asked, so we may not have the evidence.

Tavish Scott: That is fair enough.

Professor Hayward: There are countries that would use tests.

Tavish Scott: Even at that young age or at that early a stage of school?

Professor Hayward: Countries that test at that young age are few in number and they tend to be countries where there is a strong tradition of testing throughout the system. The purpose of the tests is important. It is really important to know what young people bring, what they are able to do, what they know, what they understand and how they feel about learning. It is important to gather information about young people as they come into the system. How best to do that is a matter for debate.

Professor Paterson: The Netherlands tests from the beginning. I got the information from the Organisation for Economic Co-operation and Development report on testing in the Netherlands in 2014, which goes from year 1 to year 8. Year 1 pupils are about five years old, so it is much the same as here. In years 1 and 2, pupils are tested on elementary mathematical things, which we call ordering; language; and orientation in space and time. It is possible to do it. Let me remind you that the Netherlands is a very high-performing country in, for example, the programme for international student assessment tests.

The arguments about play-based learning—which we may come on to later—are never confined to the age of five. They are usually thought of as relating to the period from the ages of three to seven. If you include that range, then there are many countries that start testing at the equivalent of either our P2 or our P3, depending on whether they start at the age of five or six. For example, Denmark and Australia do the same. I agree that it is unusual; it is more common not to start testing until the age of about eight. Nevertheless, there are perfectly respectable countries that we like to emulate and that in many respects are doing better than us, which start from an earlier age.

Tavish Scott: I will ask one final question about purpose. Professor Paterson, given how much we have chopped and changed in Scotland, there will be some argument for continuity. However, if testing were to remain the same, how long would it take Scotland to work out what was genuinely

happening in our schools? With regard to the point that you made in your opening remarks about the whole-school experience—how long would it take us to know?

Professor Paterson: I would give the same answer to any question about educational reform of any kind: it would take at least a decade to know what was happening.

Iain Gray (East Lothian) (Lab): I have questions of clarification for each panel member. The first is for Dr Bloomer and follows up his answer to one of Mr Scott's questions.

You said that, when the SNSAs were first proposed, their purpose was not clear and the emphasis seemed to be on getting a national picture but, latterly, the emphasis has been more on diagnostic use, which I presume means teachers using the information to plan a pupil's individual learning and teaching strategy. However, in your answer to Mr Scott, you implied that you feel—or the RSE feels—that the tests are not particularly effective for that purpose. Is that fair?

Dr Bloomer: Not entirely. The assessments have strengths and weaknesses—for example, I said that the ability to track the pupil over time is a strength that we have not had in the past. However, the amount of feedback on the individual from any one test is limited, which is obviously a weakness. Teachers need to become skilled users of the information that is available, and a degree of professional development has been made available with that purpose in mind. However, my overall conclusion is that the form of assessment does not yield a wide range of valuable information. It is not without value, but it is limited.

Iain Gray: I will move on to the other purpose, which is to get a national picture. The Government's stated core objective is to close the attainment gap. You made an interesting point that the SSLN applied across all schools in Scotland, but the SNSAs do not take place in the independent sector. Will you elaborate on the impact of that on the measurement of the attainment gap?

Dr Bloomer: That is relevant to the attainment gap. There is no reason in principle why the new national assessments should not take place in independent schools, although whether the Government could or would wish to oblige independent schools to use them is another matter. That dimension was present with the SSLN and is now absent.

The SSLN had information about family background and surveyed teacher views, so there was a richness to the information, although I accept Lindsay Paterson's point that the SSLN's predecessor—the Scottish survey of

achievement—was probably a better test still than the SSLN. We have lost quite a lot of contextual information, which is valuable in trying to narrow the attainment gap.

Iain Gray: We have removed from the data a cohort that, in general, is likely to be at the more privileged end of the spectrum. Is that correct?

Dr Bloomer: Absolutely—yes.

Iain Gray: My next question is for Professor Hayward. The cabinet secretary has said that a lot of the impetus for the change came from the Organization for Economic Co-operation and Development's comments on the availability of data in the Scottish education system. However, the University of Glasgow's submission says that the shift away from the sample approach was based on

"A misinterpretation of the recommendations of the OECD report."

Will you expand on that point?

Professor Hayward: No one voice ever causes a shift in policy direction.

Iain Gray: That is very much the core evidence that has been presented to us. It is the only evidence that has been presented to us and Parliament as an evidential researched reason for making the change so, in this case, it is the only voice.

Professor Hayward: I included the quote from the OECD report in my submission. The OECD is clear that it does not mean that, by necessity, one particular path must be followed. It was open to a wider debate to think around the issues.

The question goes back to purpose. What do people want to know and what use will they make of the evidence? The word "data" sounds hard and impersonal, but there is an advantage in having some objectivity.

On the other hand, the central purpose is to improve children's life experiences, so the issue is about the way in which we collect evidence and who will use that evidence and for what purpose. One grows flowers not by weighing them but by creating the circumstances in which they develop. One feeds them, looks after them and helps them to grow.

Closing the attainment gap is shorthand for improving the life chances of all young people in Scotland, and we have to ask ourselves serious questions about how best we do that. The focus therefore has to be on the action that is taken in relation to the evidence that we have, rather than all our attention being on the evidence.

In my submission, I listed all the areas in which evidence is collected. Our system has to operate

at all levels, and there is information that national policy makers need in order to think about policy development and the action that they will take to enhance the direction of policy. However, do they need all the information right the way through the system? Is it the case that the teacher in the classroom needs the evidence about every individual child, the school's headteacher needs the evidence about the dependability of the professional judgment of every teacher in the school and the local authority needs other information? The model is layered, and all the layers have to work for the system to operate effectively; otherwise, we move into a world in which we collect so much information that we cannot use it.

Iain Gray: That is why, in your evidence, you said:

"A view emerged that the OECD had recommended the introduction of standardised assessment",

which is a

"misinterpretation of the recommendations".

In the terms that you just described, they are much broader. Is that fair?

Professor Hayward: That is fair. The OECD also argued that we should look at the range of sources of evidence that we had available and relate them back to the purposes that we intended them to serve.

Iain Gray: That is helpful.

Professor Paterson, I will ask you not so much about your evidence as about previous comments that you made on the introduction of SNSA back in 2017, when the policy was first being described. You said that the varied local approaches to SNSA

"cannot give a valid national picture"

and that, therefore,

"the whole exercise is a waste of time".

Those are quite strong words.

As recently as this time last year, you said:

"Scotland has no reliable method of monitoring the performance of schools in literacy and numeracy for the first time in almost 60 years",

which you described as a "woefully inadequate" situation. Those are strong words—stronger, perhaps, than the evidence that you have given this morning. Do you still hold those views to be correct?

10:30

Professor Paterson: I was being diplomatic in my submission to the committee.

Yes—I still hold those views. To start with the second quote, it was about the situation concerning evidence. The context is that I was discussing the demise of almost all surveys of school students, including those of leavers or any other group. The only survey that remains is the programme for international student assessment, which is inadequate for most purposes; it is only for pupils who are aged 15 and so on.

I referred to 60 years, but we could even say that we need to go back nearly 80 years, because Scotland pioneered the use of good-quality surveys to understand the progress of people through education systems. From that came a series of things, including the Scottish school leavers survey, various surveys of primary school children, the SSLN, the SSA, predecessors to that, and the assessment of achievement programme. All of them have gone and are no longer used.

We do not now have the kinds of information that we had 20 years ago, for example, when the Parliament was established. We cannot monitor and it is impossible to know reliably whether we are closing the attainment gap, because we do not collect valid data. The Scottish index of multiple deprivation—the area thing—is not valid as a measure of social inequality.

I therefore hold strongly to the view that I expressed. I suppose that I feel strongly about it because my job is to do research, so perhaps you can discount my strength of feeling, because the situation means that I lack opportunities to do research.

On the first quote, which was about the use of the proposed SNSAs, the question is still very much open. I have been somewhat reassured by the approach that has been taken by the contractor that is doing the surveys—the Australian Council for Educational Research. The details and rigour of its approach as submitted to the committee and in its first annual report, and in information that Reform Scotland kindly helped me to get from freedom of information requests, show that it is trying to produce standard and reliable information that can be interpreted in the same way across Scotland.

However, there are still major worries. One is that we will not know when the child is tested. If we consider a child in primary 1, the difference between testing them when they arrive in September and just before they leave in June is about one sixth of the child's development up to that point. That is an enormous amount of child development at such a young age. We could allow for that, statistically, in appropriately technical ways, if we knew when the child was tested but, as far as I understand it, that information will not be collected. Maybe I am wrong—I hope that I am.

That information is needed to enable us to standardise the test results and make sense of them at a national level. There are other circumstances that we will not necessarily know, such as the context in which the testing takes place. Some schools do it all at the same time, almost like an exam, as the Educational Institute of Scotland has pointed out. Others do it much more informally. Through teachers and parents, I hear of many schools in which testing is essentially integrated into the classroom environment.

A scientific study that was aware of such variation would collect information about the context and conditions in which testing was taking place. It can be standardised, so my original comment might be wrong, but I am still somewhat pessimistic about it at the moment.

Iain Gray: If we have no reliable method of monitoring the performance of schools nationally, what about the other purpose that we have mentioned this morning—the diagnostic purpose of planning individual learning strategies? How do you feel about the strength of the SNSAs for that?

Professor Paterson: I agree that problems have been identified already. However, one valuable way in which the SNSAs could contribute to that purpose is through what we might call calibration of teacher judgments. I referred earlier to the unavoidable bias that all teachers have. One way in which teachers can try to improve their judgments and correct for bias is by looking at objective data and comparing their judgments with its results. That is what other professionals do all the time, and teachers should do it. In that sense, although the SNSA's measures look only at part of what a child can do, they are valuable.

Good secondary schools do such comparisons every year when the Scottish Qualifications Authority exam results come in. They sit and look at the results and compare them with the forecasts that they made for the students who took the exams, in order to improve their forecasts and, in turn, their teaching. That is how I hope that the tests will be used, but it is not clear that they will be integrated into programmes of teacher development in the thorough way that would be required to achieve that.

Ross Greer (West Scotland) (Green): Professor Hayward, in your very useful submission you say that assessment systems have three main purposes, one of which is

“to hold people to account.”

Will you talk about how the SNSAs do that? Forgive the daft laddie question, but who is it that the SNSAs are holding to account?

Professor Hayward: I think that what I intended to say is that any system serves a range of purposes, and just now in Scotland, at a national level, the question is whether the system is performing as well as we would like it to perform. The evidence that is available in that regard comes from the national improvement framework.

Let me go back to something that I said earlier. The issue to do with putting too much emphasis on the SNSA is that what the SNSA looks at is very narrow compared with curriculum for excellence, with its four purposes and its vision of what it is to be an educated Scot—we want people to be successful learners, to be able to contribute, and so on. The SNSA can give us reliable information on a very small part of two areas of our broad curriculum. The suggestion that, from those two small areas of the curriculum, we can then generalise on the education system as a whole leads us to ask questions.

We have to be very clear about the purpose. If we want to ask questions about how much and how well young people are developing, we have to do that across the curriculum, and we can do that only by basing our reflections on the evidence that we get from teachers' dependable judgment. Over time, we have to work to make sure that that judgment becomes more and more dependable.

However, there are other ways—that is, other than testing—in which that is done in Scotland. For example, Education Scotland runs professional moderation activities whereby, just as Lindsay Paterson described in the context of SQA, teachers come together to look at examples of pupils' work and consider them against the national benchmarks in order to develop and share an understanding that informs their professional judgment, so that we build professional judgment that is more consistent across every school in the country.

No system is perfect. We look to develop approaches that will give us sufficiently dependable information to allow good-quality action to be taken to support young people's learning.

Ross Greer: That was useful. I suppose that what I was getting at in my question is the concern, which teaching unions and a number of individual teachers have raised, that SNSA data might be used to judge teachers' performance. Is that an appropriate use of the data? Should it be used by a headteacher or local authority as evidence of a teacher's performance, given that class-level data can be disaggregated?

Professor Hayward: No. That is the short answer.

Assessment is very simple; there are two world views of it. There is the world view that says that

assessment is about ways of gathering evidence to inform learning, so the focus is on learning and improving learning. Alternatively, there is the world view that says that assessment is about judgment and categorisation. Those two world views sit uneasily together. In the real world, they mesh to a certain extent but, ultimately, the focus has to be on learning. If it is on judgment, we get into all kinds of perverse behaviour. If teachers believe that they will be judged by evidence that comes from one test, they will, naturally, teach to that test and spend more time on that part of the curriculum.

Standardised assessment gives teachers one important source of evidence, which they can use to inform what action they take to support children's learning, but it covers only a small number of areas. We would not want standardised assessments to cover all areas of the curriculum, because then we would work on nothing else. The focus should be on learning, not on assessment.

Jenny Gilruth (Mid Fife and Glenrothes) (SNP): I have a brief supplementary to Ross Greer's questions. I was quite taken by Professor Hayward's comment about learning needing to be the principal concern in what we are doing here. I am interested in the panel's views on whether learning was the principal concern under the SSLN.

Professor Hayward: I completely agree that the Scottish survey of achievement was a better survey than the SSLN.

Jenny Gilruth: To give you some context, and by way of background, I should say that I was a teacher. Children were removed from my classes to provide sample groups, as was the case previously, but the data was never shared with me as a classroom teacher. There is a disconnect in the teaching profession, more generally, between what happened previously and what we are seeking to achieve through the SNSA. In the past, data was held in the hands of headteachers and deputy heads; in my experience, the SSLN was not used to empower the teaching profession. I am interested in hearing the panel's views on that.

Professor Hayward: The items within the SSA and SSLN tests were designed and constructed by teachers across the country. Courses were run that were designed to help people to use the information that came from the SSLN.

However, Jenny Gilruth has put her finger on the crucial issue, which is that some people had access to the information and others did not. That was simply not good enough. I used to tease that, if we had called the SSA "save Scotland from accountability" rather than the Scottish survey of achievement, it would have attracted a great deal more interest. This issue is crucial. We need to be

clear about the purpose of such surveys. Information from surveys can provide very helpful information for classroom teachers but, if the information does not reach them, we will miss a significant opportunity.

Another interesting thing about the SSA was that local authorities, in addition to having access to information from the national survey, had the opportunity to ask for a boosted sample within a particular local authority, which would give them information at a local authority level. Technically, there is nothing to suggest that a headteacher, for example, did not take that information from a survey to use in a school or for a teacher to use in a classroom. With the SNSA, the norming studies provide the opportunity to develop a survey approach that could build some of the advantages that I have described into our system if that is our purpose.

Professor Paterson: No teacher would have been given the results for the individual children who were assessed in the surveys for exactly the same reason as with any survey: the normal ethical requirement of any survey is that all survey responses are confidential. If I were to conduct a survey and give to anyone apart from the respondent the responses that a respondent gave, I would be severely disciplined and, ultimately, could be sacked by the university. An absolutely fundamental principle of surveys is that only the survey contractor, who is sworn to confidentiality, and the individual respondent know what the individual respondent has replied to the survey.

10:45

The reason why local authorities could get access to that level of information in the SSA just as they can get access to, for example, the information in the Scottish household survey is that the level of aggregation—that is, the number of people who are involved in the sample at the local authority level—is such that there is no risk of any individual identity being compromised. I doubt whether that could be done at the level of the school, and it certainly could not be done at the level of the classroom. It might be argued that that is an advantage of the SNSA. The contractual situation with the SNSA is different, and it is intended that the teacher knows what the results of each child's test are. That is how it is designed; no one is in any doubt about that. However, a survey could not and should not do that kind of thing.

More positively, we might ask how the survey might have been useful to teachers, and there are two ways. Louise Hayward has mentioned one of those, which is that the overall national report was useful to teachers in the same way as it was useful to Government, politicians and so on. The

other way—it was a good thing about the SSLN, which was developed after the SSA—is that the people who were running the SSLN would pick out those test items that children were not doing well with and would use them as the basis for professional development sessions for teachers. That was extremely good practice. For example, if they found that children were not good at telling the time, they would use the mistakes that children made in their answers to the questions about telling the time to advise teachers on how they could teach that better. That was a great idea, and it shows how a survey can be used. Of course, that information was totally anonymised, because it was aggregated across the whole country—it was not about children in one teacher's classroom; it was about children across Scotland.

A survey can be used in that way, but it cannot address the questions that the testing of individual children can—that is not the purpose of a survey.

Dr Bloomer: In my view, it is a mistake to assume that a survey—or, come to that, a system of universal assessment—that says something about how the system as a whole is performing has nothing to do with learning. Learning in the system will improve if we know more about how we are doing and whether we are progressing or moving backwards. Although the connection is less direct than it is in the case of the feedback that is given to teachers about the individual's performance, survey information of that kind is still a valuable contribution towards improvement.

At present, there is a kind of orthodoxy in Scottish education that nothing influences the quality of provision other than the quality of teaching. That is not true. There are lots of other factors, such as the curriculum and the nature of education policy, that influence the way in which the system is performing and, therefore, the experience of the individual. So, we require to have that kind of information. The sample surveys that we used to have fulfilled a very important function, and it is not clear that we have that kind of information available any longer. At any rate, it is not available in the same depth as before.

In a couple of years' time, all of you will be vocally expressing views about whether the attainment gap has narrowed. It is probably possible to predict each individual's views on that matter. However, you will be basing what you say on what is, at the present moment, remarkably thin evidence.

Oliver Mundell (Dumfriesshire) (Con): My question leads on from that point, Dr Bloomer, but I also want to go back to some of your comments about the adaptive nature of the tests and the feedback that you have received.

Given all the variables—the adaptive element, some of the accessibility features that have been built in, the variable timescale for an individual to complete the test, the different testing circumstances, the different timescales for carrying out the tests and so on—do you think that we can consider the tests to be standardised at all?

Dr Bloomer: Clearly, they are not fully standardised. Lindsay Paterson has talked about the issue of timing, and I think that it is relatively common for schools to have a set pattern of timing.

For example, a school that I visited recently had, last year, carried out almost all its tests in May and had come to the perfectly reasonable conclusion that a primary school will get little value out of testing primary 7 pupils in May—the school will not have any opportunity to make use of the feedback that it gets—and so had decided to carry out all the tests for primary 7 pupils in November. You can see the reason for that. However, if that is a common phenomenon—and I think that it is—it sits ill with the idea that every pupil has been tested at the point at which they were judged by the teacher to be ready.

There are many such circumstances that mean that the circumstances of testing for the individual are likely to vary quite widely across Scotland, which will have a clear effect on the overall outcomes and whether we can fairly compare what is happening in one place with what is happening in another—not that we have the opportunity to make that comparison, but, if we did, those variations would make it less than valid.

Oliver Mundell: I also wonder whether you feel that it is odd not to have road tested some of the testing models with teachers. We heard last week that teachers were consulted only in passing on the design of the tests, particularly at primary 1 level. If the tests are designed to help with teacher judgment, would you have expected teachers to have been asked about those tests before they were implemented?

Dr Bloomer: I will make an initial point on teacher judgment. One effect of the tests is that it may assist teachers in relating their own judgment to national expectations and standards. That is quite helpful in itself.

The tests were the subject of some previous road testing, although I cannot offer the committee a view on whether that was done to an adequate extent. There is always a tension, in policy making and implementation, between taking time to get it right and getting on with the job. If anything, the tendency in recent years has been to accelerate timescales, which means that less is done to perfect the instrument before starting off.

However, to be fair, that is not a criticism that I have heard much canvassed by teachers.

Oliver Mundell: Professor Paterson, let us turn to the points that you made about teacher judgment. Other people might not think this, but I consider myself to be an optimist. I would hope that teachers at the early stages of education would be optimistic in considering a child's ability, because we know that there is less variation in ability than there is in attainment. Do you think that focusing on those narrow aspects and considering solely current attainment is enough, or do you think that teacher judgment would pick up what the child is capable of at that stage but a standardised assessment would not?

Professor Paterson: Whether we distinguish between potential and the point that someone has reached is an interesting question. Ultimately, I do not think that it is possible to distinguish between so-called formative and substantive judgment. As Louise Hayward has said, those two things always happen. In order to know what it is best to help a child with at age 5, one needs to know in a summative way what they already know, and that is a judgment. We cannot get away from a judgment as a precursor to helping a child to progress.

Judgment is not a bad thing; it is intrinsic to good teaching. A teacher must be optimistic that they can take a child forward. However, in order to be optimistic, they need reliable evidence. There is no point in being optimistic on the basis of fallible evidence or wishful thinking, because that does not help at all. That goes back to the common accusation that children ultimately suffer if they are praised for trivial things. For example, according to Professor Carol Dweck of Stanford University, California, who proposes the idea of a growth mindset, children should be praised only for effort, because it is effort that will improve what they do, and they should not be praised for doing trivial things that do not require effort at the stage that they are at, which will vary according to age.

That just suggests to me that being optimistic is a necessary part of being an effective teacher. However, being optimistic also requires that one is realistic about the limitations of one's judgment. To be optimistic, one must be able to listen to judgment that is independent of one as a teacher. It is only on that basis that one can reliably act; otherwise, one is potentially living an illusion of what the child can do and what one can help them to do.

Oliver Mundell: Going back to my point about the variability of the test, do you think that the test gives the teacher enough information for them to compare with their own judgment? I have heard from teachers in my constituency who worry about a child listening to something rather than reading

it. They also believe that a child might be more engaged with the test if they were shown two picture cards instead of having to sit at a computer, where they are not necessarily very focused. Do you think that those are valid points about the design of the tests?

Professor Paterson: They are, indeed, valid points, and they are the kind of things that the improvement framework of the whole testing regime has built into it. My understanding is that it was always expected that people would try to learn from the experience of the tests, particularly over the first few years, and build in improvement. That is happening this year and is already documented in, for example, the ACER submission.

On your point that some children do better in reading than in listening—or vice versa—I note that the tests, apart from in primary 1, assess both listening and reading as well as writing. That means that a teacher could choose to give greater attention to one aspect of the test than to another, depending on their feeling about what the child would respond best to. That is a good example of how the tests—even though they inevitably assess only certain aspects of the curriculum, as has been said—are already sufficiently rich to allow the kind of distinction that you mention to be made.

Johann Lamont (Glasgow) (Lab): Specifically on what we test at each stage, in the presentation on testing, we were told that a primary 1 child would have to choose to press a button to hear the word but that, in the assessment information that is given to the teacher, no distinction is made between a child having pressed the button to hear the word and a child having decoded it and read it themselves. What value is there in a test that does not make that distinction and does not tell the teacher what level a child is operating at?

Professor Paterson: That, too, is a telling question. In principle, I agree that a teacher would want to know that information. Of course, whether it matters is an empirical question. The teacher would need to have information about whether the child had responded to the written form or the aural form, and they would then have to see whether one gave a better assessment of the child's overall ability in language. With that information, the teacher could make the decision; it might turn out to make no difference or to make an enormous difference.

That is not about the existence of tests but about their design, and such an improvement in the design seems to me, in principle, to be desirable. Of course, to make the improvement valid and reliable, there would have to be a lot of replication of items—we would have to give some children only aural tests and some only written

tests so that we could compare their performance. That would have to be done as part of an experimental add-on, as it were, to the annual testing. It would be a deliberate add-on to improve the quality of the whole testing regime.

Johann Lamont: The point is that the test does not show whether the child pressed the button.

Professor Paterson: It could.

Johann Lamont: But it does not show that.

Professor Paterson: No.

Johann Lamont: The information that we get about two children is that they can both read a word, when, in fact, one child needs to hear it and the other does not. That is pretty important information.

Professor Paterson: I agree.

Johann Lamont: Does that mean that there is a danger that what looks like standardised testing that gives full information is actually not that? I have been told by very committed teachers—not teachers who resist or repel all boarders, but teachers who really want to do their best—that it takes 50 hours of teacher time for a primary 1 class to do the testing but the information is not particularly valuable. Are you concerned that that is teachers' experience of the tests? My sense is that the testing seeks to be objective but cannot be taken out of the context in which it operates. Do you agree?

Professor Paterson: I agree entirely that, if it could be shown that whether a child responds to a written version or to an aural version is important, it would then be important for the test to allow the teacher to distinguish between the two and for the reporting to allow that to happen.

Johann Lamont: Maybe I am missing something, but it seems self-evident that it is important that the teacher knows. It might be that the child can read and decode the word but presses the button just to reassure themselves while another child cannot decode the word but knows that pressing the button will help them.

Professor Paterson: Yes.

11:00

Johann Lamont: Surely, it is self-evident that two different skill sets are being assessed. Teachers might be able to make that assessment anyway, without a standardised test, so maybe we are digging ourselves in on a point that is not very important. However, it strikes me that a test that is presented to teachers as being rigorous is not—in my view—particularly rigorous, because it conflates two groups of children or it gives us less

information than might be identified by the teacher working with the child in a classroom.

Professor Paterson: Nothing is self-evident. Any claim that something is or is not the case needs to be tested by evidence. If we were to set up an experiment in which we compared children's responses to the different stimuli—aural and written—we could find that the distinction was so important that the two things would have to be reported separately, exactly as you suggest. However, it could also be the case that one predicts the other so reliably that we do not need two separate versions. It is an empirical question that requires evidence in order to be satisfied. If it turned out that the two things are sufficiently independent—that they need to be reported separately—the evidence would say that that is the case, as you suggest.

Johann Lamont: Do you share my concern that there appears to be no evidence that the question of whether the two different approaches actually matter has even been asked? The standardised assessment is supposed to be based on evidence, so I presume that we could identify evidence that shows that it does not make any difference whether the question is asked in this way or in that way or whether the child has access to hearing the word.

Professor Paterson: That is a constructive recommendation that the committee might make. The point of debates such as this, and of the Government's new inquiry on P1 testing, is to come up with constructive ways of improving quality in the design of and reporting through the system. That does not damn the system; it points out the ways in which the system might improve. The committee might make recommendations to see whether the system can be improved.

Johann Lamont: If the basic work was not done before the tests were put in place, that might call into question some of the assertions that are made about the benefits of the testing.

Professor Paterson: That is in the past. I agree entirely that the points that you make are very important, but, to improve for the future and to move forward, evidence relating to those important points would allow improvement of the system. Such debate does not damn the system and does not mean that we should not have the system, but it might produce reasonable ways in which we could collect evidence to see whether the system could be improved, as the committee recommends.

The Convener: Dr Bloomer spoke earlier about information for teachers and the pathways that a child would follow. Is Johann Lamont's point an example of something that would be dealt with by the pathways, or am I missing something? Are

other factors involved in the different pathways that allow a child to achieve a particular level?

Dr Bloomer: The instance that Johann Lamont is referring to could be an example of where children would be taken with different pathways. The notion of the pathways is that the test responds to what it gets back from the young person and, to put it crudely, makes things easier or more difficult. The pathway also has a facility built in that allows a child to listen to the question as opposed to reading the question. Those are examples of the pathway in action. The particular example depends very much on what the question is designed to test. It is conceivable that there could be a question that is concerned with comprehension, in respect of which it is not terribly important whether the child gets the question aurally or by reading it. However, self-evidently, if the question is trying to assess the individual's ability to read, the question of whether the child read the question or had it read to them is critically important. I cannot imagine that such an obvious failing is built into the system.

Johann Lamont: I will give the example that we saw. The question asked which word sounds like another word. If the word was "pie", the question asked which of three other words sounds like "pie". Is there a difference between hearing that question and seeing it?

Dr Bloomer: If that is the example, I am obliged to agree with you.

Oliver Mundell: There are multiple similar examples of questions in which there is a choice between looking at pictures and reading words. Seeing and identifying something and reading words are different skills.

I return to Professor Paterson's point about the evidence base. Do you need bespoke evidence and trials for such tests, or can you look to other educational research? There is plenty of existing research on how learning happens that looks at different skills, including different techniques for reading. Again, someone who has a wide vocabulary and can see a whole word and identify it is different from someone who is able to decode or read new words. There is plenty of evidence on how those different skills work. Can that be used to inform how tests are designed?

Professor Paterson: Yes, it can—and should—be. Professor Sue Ellis described some of the ways in which that can happen. She probably knows about that body of research better than anybody else in Scotland.

I agree about that, which very much supports Johann Lamont's point. What would be required is well-designed research into how that operates, but that well-designed research has probably already been done—if not in Scotland, it will certainly have

been done in other places that have similar cultures and education systems, such as England. You would not want to reinvent the wheel. Scotland is terribly bad at not learning from elsewhere. We should certainly learn from research elsewhere to inform the kind of questions that Johann Lamont has been asking.

Oliver Mundell: I know that you want to look to the future and not go back when it comes to introducing new things and taking forward new policies, but do you not think that those questions about what the evidence base is and whether what we are doing matches up with educational evidence should be asked before any new educational policy is introduced?

Professor Paterson: Yes, I completely agree. We should, for the future and without going back, be looking far more at the evidence. I say that not just because an academic will ask people to pay attention to academic research. This is not just about academic research; it is about the accumulated wisdom of the professionals in the system, which is often very well articulated by bodies such as the General Teaching Council for Scotland and the Educational Institute of Scotland. I think that evidence should be much more a part of the policy formation cycle. After all, that was one of the aspirations 20 years ago, when the standing orders of this place were constructed. It would be nice if that was done more than by just attaching the necessary consultative memorandum to bills.

Liz Smith (Mid Scotland and Fife) (Con): My very direct question is this: in the light of what you have said to the committee, do you believe that greater standardisation would be helpful and that schools should be asked to undertake the tests at a specific point in the year, or should we be slightly more open-ended about that?

Professor Paterson: There are two types of answer to that question. If I was answering as a researcher or as someone working in the Government's statistical service or something similar, I would say that there has to be as much standardisation as possible, or, failing that, that there has to be the collection of sufficient information to allow an estimate of the effects of not standardising, for example, the precise date when children are tested. That would be the researcher's or civil servant's answer. Of course, I completely recognise that that cannot be the political answer to the question.

This goes back to Keir Bloomer's earlier point that the purpose of the tests has shifted. In so far as the emphasis is placed much more firmly on the diagnostic value of the tests, it would be impossible in the circumstances that have come about in the past two years to require that the tests take place at a standard time of the year, and that

poses a real dilemma. I think that researchers who tried to insist on standardisation would be flying in the face of the political reality of it being impossible to have a standardised week in May or whenever. They would be failing to pay attention to how things happen in the real world.

My compromise would be the caveat that I expressed in answering the first question. We cannot hold the tests on a single week in May or November, but we can collect information that would allow us to take account of the possible effects of maturation on, for example, the difference between the autumn and spring of primary 1.

Liz Smith: Your point about the dilemma is a very important one. I think that a parent is interested in two things: how their child is getting on—what progress he or she is making at school—and how well the school is doing. It seems to me that, at the moment, we have relatively good information on how well a child is doing, and I think that we all more or less agree that the new tests are designed to provide more information to teachers on that basis. However, the new tests are also designed to provide more information to local authorities and the Scottish Government to enable them to assess how well schools are doing and therefore to pinpoint areas of concern. If there are schools and/or local authorities where educational standards, year on year, are not as high as they should be, we need to find the relevant data on that. If we do not do so, it is very difficult to help underperforming local authorities or schools to improve.

What specific additional data do we need, or how can we better interpret the existing data, so that we can find weaknesses in the system and therefore help schools that are underperforming?

Professor Hayward: There is a real danger of overgeneralising if we say that an instrument that is designed to collect information on very specific aspects of reading, writing and numeracy can be generalised to provide information on the quality of a school. That is one issue.

A second issue is that standardised assessment is one way of collecting information and can be an important and helpful source of evidence to inform a broader judgment. The tension is always in not using such evidence in a way that can have unintended consequences for other activities. For example, if the test is taken in a particular week of the year, an atmosphere starts to develop around it and it starts to attract stakes that no one wants it to have. We have anecdotal evidence of that happening in our system in certain circumstances. Everything takes place in a context.

Thirdly, we want to ensure that the consequences that follow from the use of any assessment are positive.

A final issue is that assessment is not the only part of the education system. It is the responsibility of education authorities to ensure that the quality and standard of education in schools are appropriate. We therefore have quality assurance officers and school inspectors—we have lots of sources of information that come together to give a picture of performance in a school.

It is about recognising that we have multiple sources of evidence in the system and ensuring that, when we ask key questions, we draw on a range of sources of evidence to give us dependable answers.

Liz Smith: In that case, do you think that there is work to be done on the school inspection process to enhance that qualitative judgment?

Professor Hayward: The school inspection system is one part of our national improvement framework that is a way of gathering evidence on what happens in schools, and local authorities have their own quality assurance processes. We have a national self-evaluation system that is moderated by critical friends. We have a great deal of evidence in the system and, if we focus on only one tiny element, we risk ending up with a less-dependable judgment than we might have had if we had paid attention to the range of sources of evidence that are available to us.

Liz Smith: You make an interesting point. Let me explain what I am trying to get at. If there are variable standards across local authorities—and particularly within local authorities, where some schools might have improved their performance over time—one of the most important trends to measure involves measuring a school against itself. How do we identify that? How do we get a satisfactory measure for a local authority director of education or Scottish Government minister if there are concerns about the flatlining of performance in a local authority area? How do we drill down into the results—as you say, those are in the national improvement framework—to help local authorities to improve what they are doing?

11:15

Professor Hayward: Going back to the earlier conversation about the interrelationship between research, policy and practice in the system, the truth is that, when such issues arise, we often do not know why, so we have to ask further questions about what is going on in the particular establishment that is leading to that situation. Such situations are a trigger to seek further evidence that will lead to action. It is about seeing it at a whole-system level and thinking about what

evidence we need to collect that will give us the best-quality information that is likely to lead to improvement.

Interestingly, the research evidence suggests that, in most circumstances, the differences between schools are largely explained by socioeconomic circumstances. The most significant differences lie within schools.

Liz Smith: Dr Bloomer, the Royal Society of Edinburgh's report on the curriculum for excellence and how to measure it pointed to quite a few gaps in the information and research that we can use to draw conclusions. To pick up on Professor Paterson's point that we are not very good at learning from international comparisons, I note that that report also pointed to international evidence. Is there a need for additional information in Scotland to improve our efforts to close the attainment gap, or is it a matter of interpreting the data that we already have?

Dr Bloomer: The Royal Society of Edinburgh believes that Scottish education is relatively data poor and that we need more information, particularly at stages below the senior phase in secondary education. I think that all of us at this end of the table hope that the work that the committee is engaged in will make a contribution to improving information gathering in the system although, no doubt, some parts of that are beyond the remit that the committee has taken on.

We are now involved in only one international survey. In my view, it was a mistake to abandon the other two, and I hope that at some point that will be reversed, because we need more information about how we compare with other countries. Although PISA is an excellent survey, it operates at age 15, so it tells us nothing about what is happening at the stages of the education system that we are already most ignorant about. As I say, I suspect that that is not the kind of issue that the committee is immediately concerned with, but you are concerned with the assessment regime and therefore, by implication, you are concerned with whether we would benefit from reinstating something like the SSA or the SSLN. I am not entitled to speak for my colleagues, but I rather think that the three of us believe that that would be a good thing to do.

Whether or not that happens, I am sure that we all think that it is important to be clear about what information the national standardised assessments are supposed to generate. It is of course possible to use a single assessment to generate information of more than one kind, although, in doing so, you have to be careful that one purpose does not compromise the other. Therefore, it may not be necessary to say that the assessments serve only one purpose, but it is necessary to be clear about the hierarchy of

purposes. Either the assessments are designed to monitor the performance of the system, in which case what they generate by way of diagnostic information is secondary, or they are a tool to assist teachers to aid individual young people and to refocus teaching so as to benefit from what they learn about how the whole class is getting on, in which case the assessments' role as a source of evidence about the performance of the system as a whole is secondary. We need to know which it is and act accordingly.

If the assessments are primarily to generate information about the system, they need to be able to fulfil that purpose, which points us in the direction of greater standardisation of approaches. If their purpose is diagnostic, that will not be important. It is a question of clarity about objectives first of all, and the rest follows on from that.

Professor Paterson: May I come in again on the question about individual schools? Keir Bloomer described graphically the distinction between assessing the system as a whole at the national and possibly also the local authority level and the other purposes that the information can be put to. You have asked us what a local authority director of education could do with knowledge about individual schools on that basis.

There is a workable model called contextual value added. Unfortunately, it has been moved away from in England, but it operated until about four years ago. There were two components. One was that, in looking at a school, you would look at what it adds to children's learning. If you take the end of secondary school as an example, it is not about the average number of highers that pupils in the school get but about the progress that the school has enabled children to make towards highers. That is the basis of some of the contextual admissions decisions that universities are making. That is one thing, and it is about the progress that children make at primary as well as secondary school.

The contextual bit of that method in England was about also taking account of the social circumstances that children live in. We sometimes think of parental social class or parental education as background variables that we will allow for once, but they should not be that. If someone's parents can help them because their education is advanced, that will continue to be of help throughout. The child who has well-educated parents is more likely to make more progress between, say, P1 and P4 than the child whose parents are not so well educated. That is why the contextual bit is important.

After a lot of argument between the mid-1990s and the middle part of the following decade, a system was put in place that, by and large,

commanded a lot of consensus in England. I cannot remember exactly when it was put in place, but it was at some time in the previous decade, and the system ran until a few years ago. It certainly ran right through the period of the coalition Government, and some of the policy decisions were taken under the previous Labour Government. It worked quite well. It was not perfect, but it allowed school-level information to be generated while also taking account of the complexities of children's learning in terms of both their progress and their family and other circumstances. There might be some possibility of using the SNSAs in that way.

I finish by noting that school-level information is bound to find its way into the public domain whether we want it to or not, because of freedom of information. It would be far better to prepare for that by addressing the questions that you have raised.

Tavish Scott: I very much agree with that last point, but that is a different subject altogether.

My question follows on from Liz Smith's line of questioning. The Government says in its submission to the committee that the achievement of curriculum for excellence levels return is a replacement for the SSLN. Do you believe that it is a replacement for that? We are not quite sure that it works, because it is still badged as being experimental, even after three years. What is its role? What is it there for?

Professor Paterson: I do not think that it is an adequate substitute for the SSLN, for two major reasons. First, the assessment of where children have reached is made according to teacher judgments. We have already talked about the unreliability of those. Secondly, it is not an adequate substitute for a completely different reason, which is to do with the measurement of social circumstances. Actually, the SSLN suffered from that, too. We need much better measures of social circumstances. I think that the committee has addressed that before, but it comes up over and over again.

We know that two thirds of children who are living in poverty are not in the 20 per cent most deprived neighbourhoods. Your constituency probably has no deprived neighbourhoods, but that does not mean that it has no deprived families. There are other ways in which the annual December report is inadequate, but those are the two major ways.

Tavish Scott: Yes. That supports the contention that we should revisit the SSLN, but with some enhancements and some careful, creative thought about how it should properly work.

Professor Paterson: Yes, and we have good models for that. The growing up in Scotland survey, which is an excellent survey that traces children through their lives, contains really good, sensitive measures. I am not saying that we should replicate that every year, as it would be too expensive to do that. However, the experience that ScotCen and particularly Paul Bradshaw, who is the director of the survey, have built up over the 15 years since the survey was established would be very useful in helping to strengthen the evidence that you are talking about.

Tavish Scott: You have all said that all of us in politics are basing our arguments on closing the attainment gap on some pretty thin evidence, if we are where we are. Would an enhanced SSLN help politicians of all political persuasions with a genuinely difficult issue? Is there some purpose in it in that sense?

Professor Paterson: Yes, absolutely.

Professor Hayward: It is designed to serve that purpose.

Tavish Scott: That raises the question of why we took it away, but you have answered that question already.

Rona Mackay (Strathkelvin and Bearsden) (SNP): I would like to go back to the purpose of the tests and some comments that Dr Bloomer made a few moments ago. If I was explaining standardised assessments to a constituent and I said that they are there to monitor the performance of the system, my constituent would be surprised and confused. They think that the assessments exist to monitor their child's performance. Has something been lost in translation when we have been trying to get all this over to the public? I am not suggesting that it is in any way your responsibility, but people are confused and the general perception is that they are essentially diagnostic tests.

Dr Bloomer: The message has changed. As a result, parents have been persuaded that the primary purpose is diagnostic. That was certainly not the advertised primary purpose at the outset.

Professor Paterson: There is a brief comment on that in the National Parent Forum of Scotland submission that puts it succinctly, and I agree.

Professor Hayward: Policy should also be susceptible to development in the light of evidence. I do not know whether we would all agree, but I would argue that the shift to using the tests to lower the stakes and have them as part of the repertoire on which a teacher can draw is a positive move.

Professor Paterson: However, it leaves a gap.

Professor Hayward: Yes, but that is why we are talking about how that gap might be addressed in a way that would not have the potential unintended consequences that there would be if the policy stayed as it was.

Rona Mackay: You believe that the public could do with some clarification about the purpose of the tests.

Professor Paterson: When parents start getting report cards that incorporate the results of the tests, the misunderstanding will go away. In fact, it will then be difficult for the Government or anybody else to go back. Once parents start getting the scale that has already been published on the Education Scotland website and in the ACER submission, they will wonder why they did not get such detailed information previously.

Teachers might then face quite different problems with how to explain the sort of thing that Professor Hayward has been talking about, which is that the child's progress is about more than just the result of a test.

Rona Mackay: Professor Hayward, you said that the system is a layered model, coming from the local authority down. Is that working in practice, or is it too early to tell?

Professor Hayward: As with any complex system, there are parts that work very well and parts that work less well. Learning from evidence is as important at the level of the system as it is at the level of the child. We need to make sure that we have good-quality evidence that will allow us to reflect on that question and then allow us to realign policy.

Going back to the question that came up earlier, there is the idea of research to inform. There is research that, along with other sources of dependable professional evidence from teachers in classrooms, school inspectors and a whole series of others, should inform any new development. It is also about research to align. Once we have the vision of what we want to achieve, we need to keep an eye on what is happening as that is developing so that we can make sure that we stay consistent to the ideas of the vision. The history of every country that I have worked with internationally is that they start out with clear and coherent visions of what they want to achieve and, over time, divergence happens. As we do not go into the system to better understand why the gaps are beginning to emerge, it continues to develop until we get to a point when a new innovation has to come in.

We need to change the model. We need a vision for what we want to achieve, and we must use research evidence as we develop the model to make sure that it remains consistent with that

vision. We also need to feed the evidence from that back into developments in practice and policy.

11:30

Dr Bloomer: Although I agree with Lindsay Paterson that parents will be clearer about this once they begin to receive test feedback in school reports, I am not sure that they will necessarily all be well equipped to interpret what they are told.

In relation to each test—for example, the reading test—they will be told in which of the 12 bands their child is considered to sit. They will be offered a standard, pre-written paragraph of three or four lines that tells them something about the band. Each of those descriptors starts with the words

“Learners in this band are typically able to”

and continues with something such as

“read a wide range of straightforward texts”.

It says that, typically, a child who falls into band 6, for example, is able to do this but perhaps not that.

Whether the child fits the stereotype of the band descriptor is another matter and, as we have already discussed, a child can be assessed as being in band 6 by answering a different set of questions from those answered by somebody else who ends up as being considered to be in band 6. A different mix of skills might emerge from the answers that they give.

The descriptor adds information to the parents’ understanding, but there are limitations to the nature of the information that it adds.

Professor Paterson: Children will get different questions, but if the design of the tests has been done adequately and scientifically, the tests will address the same underlying skills.

Most people are aware that, if they go to their doctor and a blood pressure test is taken that shows something unusual, the doctor will almost certainly not—and should not—rely on that one assessment. The person has probably gone there in some apprehension—perhaps they travelled by ScotRail and they are late—and there might be other issues, so the doctor will repeat the test.

We all know about the essential randomness of things, yet it is not being conveyed—this is a big failing of the public discourse on the issue—that all assessment is subject to random error. There have been detailed studies of that in England, which have found that the degree of random error has diminished since the national curriculum assessments in England were first introduced 20 years ago. However, there is still an inevitable amount of random error, and we have some way to go.

That was the purpose of the so-called reliability measures in the new standardised assessments. They are pretty high, but they are not perfect and a degree of misclassification will go on. That is not because anybody is doing the tests badly, the teachers are failing to understand them or anything like that; it is intrinsic to the nature of measurement that an element of error is introduced.

There needs to be a public education programme about that, which is difficult as it involves acknowledging that random mistakes are made—it is not that there are deliberate biases. There will be a great challenge in educating parents on what to do with the results and, sadly, I do not currently see any programme from any agency that intends to educate parents about that.

Jenny Gilruth: At the start of the meeting, Lindsay Paterson alluded to bias and objectivity, and you just mentioned that again. You said that no teacher is objective and, when I was teaching, we used to be able to identify when pupils came from a certain primary school in the city, because it used to inflate grades. We knew that that happened in the system.

At a previous evidence session, Professor Sue Ellis made the point that the SNSAs could challenge unethical and biased approaches to assessment, whereby, for example, children are removed from class and put in different groups. Does the panel agree with the assertion that the SNSAs could potentially stop that kind of thing from happening?

Professor Paterson: Yes, they could if they help to induce a mindset among everybody involved that, if you are going to get properly reliable evidence, you have to adhere to standardised conditions in the same way as any scientist or doctor would do to get reliable evidence. You cannot, as it were, fix the results by fixing the conditions under which the results are obtained. So, yes, that would be a really good thing.

Dr Bloomer: I agree.

Jenny Gilruth: Professor Hayward, you gave an example earlier of moderation and quality assurance at Education Scotland, with teachers working collaboratively to get a better understanding of standards. Do the SNSAs offer the same opportunities for teachers to work collaboratively to get a better understanding of CFE levels? Lindsay Paterson talked about the accumulated wisdom of the profession. Could there be an opportunity to improve that as a result of the SNSAs?

Professor Hayward: It comes back to my earlier point that the SNSAs give you information about very limited areas. For example, one

assumes that the purpose of being a teacher in the classroom—as you were—is to help children to become better readers, for example. In that context, the SNSAs will give you information on aspects of that. However, as a teacher, you know that motivation, whether a child believes that they can read and whether they see reading as being important are all crucial factors in whether a child will make progress in reading.

It is about bringing all that information together and living with that complexity. I would argue that parents also want to know what they can do to help their child next, what their child is moving on to and what the most important things are for them to focus on. The SNSAs can play a role within that broader picture, but it is the quality of the teacher, their understanding of the curriculum and their ability to generate tasks and experiences for young people that will allow those young people to develop as positively as they can. It is about the teacher's ability to discern progress and focus on what happens next in learning. It is a complex picture and we have to learn to live with and support that complexity if we are really concerned about improving the life chances of every child in Scotland.

Johann Lamont: Some of the questions that I had about what Professor Paterson called neutral and reliable data might have already been answered. However, given that the test can be applied at any point between a child being four and a half and six years old—and, as we were advised at the demonstrations, it can be taken either with a lot of support and practice or with no practice—is it fair to say that that will distort the information that the classroom teacher gets?

Professor Paterson: If the purpose is, as it now appears to be, to give the teacher diagnostic information about how to help the child make further progress, I would say that the risk is not too great, because the teacher has already taken into account the fact that they have chosen to test that child at age six—perhaps in the summer of P1, for example—rather than earlier, so that would not be a problem.

Where it is a problem—as I said in answer to Liz Smith's question—is in trying to aggregate the results to make interpretations about the system as a whole, the local authority or the school. If that is happening to an extent that we do not know about, it comes close to invalidating the results when they are aggregated to those levels.

Johann Lamont: The other thing that I am interested in is how much importance this process has within the system. I will give an example. When I was still a classroom teacher, I might have had to assess an S1 English class in October because there was going to be a parents' evening. I would give the parents an initial idea of how their

kids were doing in respect of their progress, behaviour, homework and effort. I would want to give all the kids As because they were really enthusiastic and keen, they had come in to a new school and they were doing their very best, but the headteacher told me that I could give only 20 per cent of them As because, after all, by the time they got to higher, only 20 per cent of them would be able to compete.

However, by giving a child an A and recognising what they are trying to do, you are keeping them engaged in school, so it is entirely valid for a professional to say, "I want to keep these wee people enthusiastic—I am not going to tell them now, 'By the way, you're not going to get a higher.'"

Do you accept that that is part of the assessment? Perhaps objective testing allows the teacher to know both what they want and aspire to for the child and what they want for themselves against the testing. Do you think that that is valid?

Secondly, we talk about not teaching to the test. If it could be established that support staff in schools have been taken away from children with additional support needs to manage the process, which would disproportionately impact on schools with disproportionately high numbers of children with additional support needs, would that matter? Is that a judgment on the effectiveness of the policy of a standardised assessment? I have heard anecdotally that, in a primary school with a lot of children with additional support needs, the support staff are being taken away to run the system. Is that not another form of distortion, just as teaching to the test is?

Professor Paterson: That is a serious failure in so many respects that the committee does not need me to point out. It completely contradicts the idea that the purpose of the test is to inform the teacher's judgment. The teacher cannot, as it were, subcontract their judgment. They have to hone their judgment on the test that they, as a teacher, administer. What you describe is not a consequence of the test; it is a consequence of school management and local authority management.

Johann Lamont: It might be a consequence of the compulsory nature of the test in a school that does not have the resources to do anything other than manage it in that way.

Professor Paterson: It might be a consequence of the ways in which the tests are implemented by the Government as well as by the school and local authority, but it is not a consequence of testing as such. It is a consequence of the context of the testing.

I return to Johann Lamont's point about the headteacher with his—or perhaps it was her,

although I suspect not—normal distribution in mind. What was said was nonsense and should never happen. Clearly, we should never constrain people by completely non-evidence-based standards. That is the point. Giving As to everyone to encourage them is fine, but it does not produce a judgment. It is a form of exhortation—it is what the team coach does at the beginning of a football match or something similar. It has nothing to do with actual performance. After the match, the team coach would presumably want to say that one person did well and another did not do well and did not try hard enough. The point is that that would be based on evidence.

If the system of national assessments encouraged a greater respect for evidence in making judgments across the system of Scottish education as a whole, that would be a good thing. People would no longer get mixed up between exhortation and assessment.

Johann Lamont: Would it be valid in assessing the benefit of standardised assessments to ask schools what the consequence has been for their routine processes? I am troubled by the fact that we were told during demonstrations that a child could basically be tutored in how to do the test and could have any number of chances to practise it before they did it. That must distort what is happening in the classroom, in terms of time.

Professor Paterson: What you describe is part of the practice sessions that children would have. I do not think that it is part of the assessment itself.

Johann Lamont: If there is not a standardised test, self-evidently, a child does not have to practise the test before they do it. Some schools might make the judgment that standardised tests should be done in the way that the previous survey was—pupils go and do it, then they come back to the classroom, and it does not have any immediate impact on them as individual learners.

Professor Paterson: Teaching to the test is a bad thing only if the test is bad and is not a valid assessment of the content of the curriculum. Given that there is going to be lots of teaching to the test, we had better make sure that the tests are valid and actually assess what is in the curriculum.

For example, in primary 1, we expect children to tell the time from analogue, not digital, devices. If that is a reasonable thing to have in the curriculum, it is a reasonable thing to ask children to do. It is not unreasonable at all to ask them to look at an analogue clock. In primary 1, it might be unreasonable to ask them to look at Roman numerals on an analogue clock, but that is not the point—the task is about interpreting the position of the hands.

The mantra about teaching to the test is overused. Sometimes, teaching to the test can be a good discipline that forces people to think. After all, we expect people studying higher mathematics to have been taught to the test to the extent that they are learning how to perform mathematical operations.

In primary school, it is true that the tests assess only certain aspects of attainment. In some respects, however, those aspects are fundamental to any other progress being made. Unless a child can do the elementary operations of arithmetic, they will never make progress in any other aspect not only of maths but of science and many social sciences, too. Although it might seem narrow to check that the child can add, subtract, multiply and divide mentally as well as on paper, those skills are the basis for the child flourishing in later life. Teaching to the test is not necessarily a bad thing—it depends on what the test does.

11:45

Professor Hayward: I agree with that. Ms Lamont raises interesting issues about the relationship. It focuses learning. If an English teacher wants to encourage someone to learn, a system that asks them to put a label on that learning is not necessarily the most helpful way to do it. The issue for the teacher is what the child can do now; what their understanding is of how the child relates to progression in the learning journey from the time they walk into the school until the time they are likely to leave; and how they might support the child to make progress in that journey, which is absolutely crucial.

It is interesting that it is written into law in Norway that a letter or number cannot be put against a child's name before they are 12—it is illegal to do so. In that context, there is a recognition that using letters or numbers, which are shorthand symbols for professionals and can be intended to communicate with people externally, can have a negative effect on the self-esteem and confidence of the very young people whom we want to support most effectively. There is sometimes a confusion between criterion referencing—looking at the child's progress and development in relation to a criterion—and norm referencing, by which we look at the 20 per cent who can do something, and so on.

I make a plea not just for better understanding about standardised testing but for better understanding as a society about assessment's potential to enhance learning and its challenges for trying to achieve a society in which every child makes good progress.

Dr Bloomer: Louise Hayward's point about norm referencing and criterion referencing is

interesting. If we want a well-rounded and comprehensive picture of how a young person is developing intellectually, we should ask the teacher—that has always been true and it remains so. Very few classroom teachers would have any difficulty in giving some kind of norm referencing of all the children in their class off the top of their heads, whether for reading, arithmetic or whatever, particularly if they operate in primary school, where they spend more or less the whole week with the child.

How that would relate to how children elsewhere in the country are performing is an entirely different matter. If we want a criterion-referenced assessment, we probably should not go to the class teacher. The information from standardised assessment will be more helpful—with regard to the limited part of the curriculum that it covers, at any rate. In recent years, we have become much more interested in how teachers' judgment correlates with a more objective notion of expectations and standards, hence the emphasis that has been placed on moderation, which we talked about earlier. The new assessments provide teachers with a tool that will help them to do some of that, which is a valuable contribution.

Ross Greer: I return to the issues that were raised about the comparability of the data and Johann Lamont's point that some children in primary 1 take the test at the age of four and a half and some take it at the age of six, which is a significant difference. Did I pick up Professor Paterson correctly as saying that the aggregate group-level data at that stage would be invalidated if that variability was not recognised?

Professor Paterson: As a simple headline, yes, I would say that it would invalidate the data—it is too big a variation at that age. I have students whose ages vary by more than that who are doing their final honours exam, and we do not apply an age adjustment. Clearly, the ages vary. However, at that very young age, one could not draw valid inferences if one just had the test result with no measure of progress on the basis of it. Incidentally, that is an argument for having baseline testing in primary 1, because it would allow a measurement of progress in the later stages of primary and would take account of that.

I apologise for introducing too many caveats. The answer to your question is yes, it would invalidate the data.

Ross Greer: In your experience, is there a sufficient level of data literacy in local authorities and schools to recognise and compensate for that?

Professor Paterson: No, there is not. It is demonstrable that local authorities do not have that statistical expertise. However, it must be said

that the vast majority of Scottish teachers do not have that expertise either. Remember that one can do a primary teaching degree with a C in what is now called national 5 applications of mathematics—the equivalent of what those of us of a certain age would call arithmetic O grade or a standard grade pass. That is not enough to understand the complexities of statistical sampling and measures of reliability.

What is more, you might think that that would be part of the teacher education programmes, but the committee heard evidence from some student teachers last year that they get no more mathematics in their undergraduate programmes than they took with them from school. They get courses on the teaching of maths, but they are not taught any more maths. A typical primary teaching graduate emerges as a primary teacher with no more than application of mathematics national 5, which is not nearly enough. That is why I say that there is not enough expertise to allow the evidence to be interpreted in schools.

Ross Greer: Do the other witnesses share that opinion?

Dr Bloomer: Yes.

Professor Hayward: There might be some variance across the different teacher education institutions.

Professor Paterson: No, the evidence produced for the committee meeting to which I referred included a paper from the Scottish Government that examined the amount of time in a typical four-year programme that is devoted to certain activities, one of which was mathematics. There was variance, but none of it was more than a few hours a week. The students did not even get to the level of higher mathematics.

Dr Bloomer: There is variance from one student teacher to another, because they come in with varying levels of expertise in mathematics.

Professor Paterson: Yes.

Dr Bloomer: Placing increasing importance on teachers interpreting evidence has implications for initial teacher education, which, so far, have largely not been considered.

Professor Paterson: Finland is a place that is often—and rightly—admired. One of the questions that is asked is why Finland does so well when it does not have national testing until the end of primary school. It has often been said that that is to do with the quality of teacher education in Finland. If we look into what that means in detail, we see, for example, that in Finland about 15 per cent of primary school teachers have enough of a mathematics component in their degree to have a mathematics qualification—they would satisfy our

requirements to teach mathematics in secondary school.

If we had that, it would mean that, on average, every primary school would have at least one person who was qualified to a level that was equivalent to a mathematics honours degree. That does not mean that every teacher would have to do that, but we would want every school to have someone who could interpret the evidence and share that interpretation with their colleagues. The same is true of other specialisms in the Finnish curriculum, such as foreign languages.

Professor Hayward: The only thing that I would add to what Lindsay Paterson said on assessment literacy is that it is about assessment in its broader sense. It is about not only interpreting statistical evidence but the broad picture of how assessment relates to the curriculum and pedagogy and the skills that are needed.

I do not know what kind of induction programmes there are for members who come to work in the Parliament, but, in this context, it is about the extent to which people are supported in carrying out the roles that society is asking them to carry out and ensuring that that support is there in all the layers throughout our system.

Ross Greer: I want to move up a layer from schools to the local authority level. There is a challenge for teachers in that such data literacy is just one of many skills that would be desirable in a teacher.

At local authority level, there is an opportunity to create posts and recruit people with the specific skills for them, but there is some evidence that local authorities no longer have the quality improvement staff who have that level of understanding. Have you picked up on the fact that the introduction of SNSAs—with the need for local authority staff with that level of data literacy—has come at a time when local authorities have lost the staff who had the relevant skills?

Dr Bloomer: That is unquestionably the case. Local authorities have a declining capacity to offer support to schools. As long as local authorities remain an important tier of organisation within the system, that is decidedly unfortunate.

Professor Hayward: The idea of building capacity in the system—which is, fundamentally, what we are talking about—might vary from authority to authority, depending on their size. The other issue is about seeing those skills and competences as part of being a professional teacher. It is about not just initial teacher education but making sure that there are opportunities throughout a teacher's professional career for them to develop, hone and enhance their skills in those areas.

The Convener: Thank you. This has been a very long session. We thank Dr Bloomer, Professor Hayward and Professor Paterson very much for attending the committee today and for their submissions, which have been highly valued by members. Our next evidence session on Scottish national standardised assessments will be on 30 January.

11:56

Meeting suspended.

12:02

*On resuming—***Public Petition****Free Instrumental Music Services (PE1694)**

The Convener: We move to agenda item 3. Petition PE1694, on free instrumental music services, was referred to us by the Public Petitions Committee in the course of our inquiry into instrumental music tuition. I put on the record my sincere thanks to all who signed the petition; the music education partnership group, which raised the issue with the committee during our evidence session; and all those who gave evidence as part of our inquiry including the young people who were involved in our deliberations, who made very powerful contributions.

We published our inquiry report yesterday and we hope to hold a committee debate in the chamber in the near future, involving members from across the Parliament. At a future committee meeting, we will consider the responses of the Scottish Government and the Convention of Scottish Local Authorities to our recommendations. In addition, we will consider how the “What’s Going On Now?” research aligns with the committee’s findings.

Our papers suggest that we close the petition at this stage. Alternatively, we could leave it open until we have received correspondence from the Scottish Government and COSLA regarding our report. I look for guidance from committee members on their preference between keeping the petition open and closing it today.

Tavish Scott: Might it be courteous to keep it open until such time as the debate has happened and the Government and other bodies have also responded to the committee’s report? While I was at home yesterday, it struck me that a couple of people had mentioned the committee’s report—that does not happen every day—which suggests that it has struck a chord with people. Given that the petition is entirely relevant to the work that we have just done, it might be courteous to act in that way.

The Convener: As no other member is minded to add to that, is the committee content to leave the petition open for deliberation at a future meeting?

*Members indicated agreement.***Attainment and Achievement of School-aged Children Experiencing Poverty**

12:04

The Convener: The final agenda item to be taken in public is consideration of a further response from COSLA to our report on attainment and achievement of school-aged children experiencing poverty.

Last year, the committee considered responses from the Scottish Government and Education Scotland and an initial response from COSLA. However, the substantive response from COSLA covers issues that the committee acknowledges would take longer to analyse and therefore to respond to.

I observe that the Government responded to the committee’s recommendation that all local authorities be surveyed on charges that are made for core education and how they contribute to the cost of the school day. The Government said that it would pursue such charges with COSLA. However, COSLA’s recent submission does not mention the committee’s recommendations in that area.

It is suggested that the committee writes to the Scottish Government, copying in COSLA, to seek clarification of which organisation is taking forward the work that stems from the recommendation on the cost of the school day, and details of the specific work that is planned. Are members content for us to do so?

Members indicated agreement.

The Convener: As members have no other observations on the response from COSLA, that concludes our public session for this week. We will move into private session.

12:05

Meeting continued in private until 12:33.

This is the final edition of the *Official Report* of this meeting. It is part of the Scottish Parliament *Official Report* archive and has been sent for legal deposit.

Published in Edinburgh by the Scottish Parliamentary Corporate Body, the Scottish Parliament, Edinburgh, EH99 1SP

All documents are available on
the Scottish Parliament website at:

www.parliament.scot

Information on non-endorsed print suppliers
is available here:

www.parliament.scot/documents

For information on the Scottish Parliament contact
Public Information on:

Telephone: 0131 348 5000

Textphone: 0800 092 7100

Email: sp.info@parliament.scot



The Scottish Parliament
Pàrlamaid na h-Alba