



OFFICIAL REPORT
AITHISG OIFIGEIL

Education and Skills Committee

Wednesday 30 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 30 January 2019

CONTENTS

Col.

SCOTTISH NATIONAL STANDARDISED ASSESSMENTS INQUIRY 1

EDUCATION AND SKILLS COMMITTEE

4th 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:

Jackie Brock (Children in Scotland)

Professor Andy Hargreaves (Boston College and University of Ottawa)

Sue Palmer (Upstart Scotland)

CLERK TO THE COMMITTEE

Roz Thomson

LOCATION

The Robert Burns Room (CR1)

Scottish Parliament

Education and Skills Committee

Wednesday 30 January 2019

[The Convener opened the meeting at 10:00]

Scottish National Standardised Assessments Inquiry

The Convener (Clare Adamson): Good morning, and welcome to the fourth meeting in 2019 of the Education and Skills Committee. I remind everyone to turn their mobile phones to silent, so that they do not disrupt the meeting. We have received apologies from Tavish Scott and Ross Greer.

Our first agenda item is our inquiry into the Scottish national standardised assessments. We have two panels of witnesses today. For the first, I welcome Professor Andy Hargreaves, who is a research professor with Boston College and a visiting professor with the University of Ottawa. Professor Hargreaves, will you briefly outline your international experience as it relates to the inquiry?

Professor Andy Hargreaves (Boston College and University of Ottawa): Thank you, madam convener—is that how I should address you?

The Convener: You can address me however you like—“convener” is fine.

Professor Hargreaves: Okay.

Thank you for inviting me to present evidence to this very important committee at what is a crucial time in Scottish education, as you think about how best to forge a way forward with an assessment strategy that will benefit all students in Scottish education.

I began adult life as a teacher and then became a researcher. I worked in universities in England and then, in 1987, I moved to Canada, where I set up the international centre for educational change in Toronto. In the past 15 years, I have worked at Boston College, which is not in Boston and is not a college—it is 100m outside Boston and is a university. It is famous for the international maths and science studies that are administered from it, although I am not directly connected with that. My family and I have just moved back to Canada, where I am a citizen, as well as being a United Kingdom citizen, and I am connected with the University of Ottawa.

On my international experience, I have done research in a number of countries on educational reform and change, systemically and in terms of

the impact on teachers and the teaching profession. That has been across a range of countries but not too many, including Singapore, the United States, the UK and Canada—that is probably about it. I also do advisory work with Governments, sometimes on an occasional basis and sometimes on a more sustained basis. For several years, I was one of six advisers to Premier Kathleen Wynne, who was the premier of Ontario province, which has a population of 13 million people, until May, when she was deposed in an election. I have also been proud to be one of 10 international advisers to the Scottish Government over the past few years.

I have been engaged in Organisation for Economic Co-operation and Development reviews of various countries. Members will probably know about the one that took place in Scotland, which involved a team of four. Just prior to that, I was involved in a review in Wales, which is dealing with similar issues to those in Scotland. Some time before that, I was involved in a review of leadership strategies in Finland.

I am not really known as a measurement specialist, so if you ask me anything about technical items, design or validity and reliability tests, my answers will be extremely disappointing. However, I deal with work on change in schools, school systems and societies, and assessment comes on to my radar a lot, as it has a connection to everything else. What I am concerned about, and what I can best help you with, is how assessment is interconnected, in benign and less benign ways, with other parts of the improvement agenda.

Liz Smith (Mid Scotland and Fife) (Con): Thank you for providing us with your international experience, in which we are extremely interested. You have outlined your experience in different countries and you mentioned the OECD review here, in which we were given a set of six criteria to which we should adhere if we are to have effective attainment.

In the light of your final comment, will you provide us with examples from your international experience of where schools have improved outcomes for young people? Perhaps you could also relate that to the standardised assessments that those schools have used, although I know that you cannot go into the technical details. It would be helpful to the committee to hear where Scotland could learn lessons from the international experience.

Professor Hargreaves: The first thing to say is that it is important to learn and not to copy. With teachers or countries, I always advise that you should never look at one model and copy it, because no model will have everything that you want. However, if you look at a number of models,

you are empowered to learn from them what is most relevant to you or your country.

One country that many people go to is Finland. I am a great fan of Finland, which is one of the happiest countries on earth and is where I would live if I had to live anywhere else. It is a nation that values learning immensely and that has very low achievement gaps. Statistically, there is almost an accidental relationship between family background and achievement in Finland. Because it does well on overall performance and on equity, it is of high interest internationally.

In the system-wide sense, the assessment system in Finland up to secondary school leaving is based on samples rather than a census. Many people, including me from time to time, are extremely interested in the idea of a sample as a way of preventing people from teaching to the test or gaming the system. The benefit of the approach in Finland, which we can learn from, is that most assessment is directed towards improving learning and is done, chosen and developed within schools, with some collaboration within and across municipalities, which are the equivalent of our local authorities.

However, there is a difficulty with transposing the Finnish model to other places, which the six advisers in Ontario have considered in the past few months in conducting an assessment review for the province. We seriously considered the arguments for having a sample versus a census. However, the difficulty with that is that Finland is not very diverse, although it may become increasingly so over time. If a country is diverse and has wider inequities, as Scotland has—it is not unusual in that sense—there is a need to identify which populations are in greatest need.

For instance, in Ontario, the most persuasive argument that I heard as one of the advisers about the need for a census rather than a sample was from one of my Caribbean Canadian colleagues, who felt that there is neglect in Ontario—there is—of historically black Canadians. They are not recent immigrants and refugees, who get a lot of attention; they are black Canadians, sometimes going back to the times of slavery and the underground railroad. They are one of the most vulnerable groups in terms of disadvantage. My colleague felt that having data that enables the identification of exactly when and where such groups are being overlooked is essential to equity.

I have begun with an example that looks really promising and that uses a sample, but I am persuaded that, in cases where there is great inequity and increasing diversity, some kind of census can be more beneficial. In looking at countries that, unlike Finland, use large-scale standardised assessments, you first have to disconnect the term “large-scale” from the term

“standardised”. I have looked through the committee’s documents, and I have seen the point that teachers everywhere use standardised assessments, but they are just not on a large scale. This school might use a different assessment from that school or another school. There are good standardised assessments that have been reliability and validity tested in literacy, mathematics and so on.

The issue is whether large-scale standardised assessments can bring about authentic improvement. I can give many examples of such assessments bringing about improvement that is not authentic. Improvements that were documented numerically in the United States and in England were roundly denounced by the statistical societies of both countries as being statistically impossible without in some way faking or fabricating results or the practices that lead to results.

Ontario uses mid-stakes rather than high-stakes testing, so it provides one of the best examples that we might consider. A high-stakes approach means that assessments provide the power to intervene and to punish—for example, to remove a headteacher from a school or to close a school and reopen it as another kind of school. Ontario does not use such sanctions and provides a lot of support, but it has mid-stakes, which we probably have to pay attention to here. Knowledge of the results and their patterns can lead some school district directors to experience pressure from central Government to exert undue pressure on their schools to improve their results over relatively short periods. That creates all the negative impacts of large-scale assessments that we know of. Even in Ontario, the mid-stakes rather than the high-stakes approach produces some negative consequences. Influenced by Scotland, we spent time in the Ontario review trying to figure out ways to maintain a large-scale assessment without the negative impacts.

Liz Smith: That is extremely helpful. It is fair to say that a dilemma that was flagged up in the two previous evidence sessions on attainment is that the tests that might be used to foster better learning for an individual child might differ slightly from those that might be used to spot problems in the education system. In what you said about your international experience, you seemed to make a similar point.

We must grapple with the fact that we want not only to raise attainment among the youngsters who are involved but to use the testing to identify schools or local authorities that need more support. Will you comment on that dilemma in Scottish education?

Professor Hargreaves: That is the biggest dilemma. Some people think that the dilemma

involves learning versus accountability, and where there is a close connection to things such as parental choice of school, publication of the results and so on, that is a big dilemma.

For professionals, however, the dilemma is between supporting the teacher with information that will help them to help their students more effectively and the need for people who cannot know all their students but who are responsible for them—such as a headteacher of a large school, a new headteacher who wants to know where their school is so that she or he can help to lead it, or a local authority director—to have system-wide data that enables them to see where everybody is and intervene to provide support if people are falling behind. The biggest dilemma involves not accountability but the need for the system to know where it is and not be thrashing around in the dark, especially if the system is larger.

10:15

What we recommended in Ontario, which has not been implemented because of a change of government, even though it was accepted by the previous government and the other main party—in other words, it was accepted by two of the three parties—was the creation of a kind of firewall between the standardised assessments and the individual diagnostic assessments that are done in school. Just like you, we do not have total confidence in the approach. We are on the front edge, and we are in somewhat uncertain territory.

Five years ago, systems around the world were in denial of the fact that large-scale standardised assessments had negative consequences for students' learning and wellbeing and for the teaching profession that was responsible for them. That denial is disappearing very quickly everywhere, and Scotland is at the head of that. We are all starting to own the problem and to ask how we can gather large-scale information at the same time as providing teachers with good support diagnostically and formatively. The Ontario answer was to create a firewall and to say that the large-scale assessment agency should collect the results, which everybody will get to see, because they know where they are, about 10 months later. That information will be useless to the teacher from the point of view of giving feedback to their children. At the same time, lots of support will be provided through other kinds of instruments and processes to help teachers with assessment for learning.

I think that the solution that is being tried here is different. It involves asking how we use large-scale assessments to inform teachers' professional judgment. Local authorities will have knowledge of their schools, but they will not be able to make comparisons between one another

on the basis of the test results. That will be done on the basis of teachers' professional judgment, part of which is informed by the test results.

Scotland is at the leading edge in that regard. It is good that you are watching the world, but the world is watching you. Figuring out how to make this a success over the next three years—it is possible that it might not be—and to be a learning Government as much as an improving Government is the key challenge for the Scottish Government.

Johann Lamont (Glasgow) (Lab): On the question of purpose, I am not sure whether you think that it is necessary for the one test to do the two things that you have identified. Would another solution be to have a standardised test that informed the nature of the system and diagnostic testing that supported the child, which would not have to be standardised? Is it the standardised bit that matters?

The committee has had some discussion about purpose. You will probably be aware that the 2011 OECD review took the view that there should be one clear purpose for assessment and that the process is complicated if there is more than one purpose. We are now in the situation in which the Scottish Government says that the standardised assessments process is both a national survey and a diagnostic test. Do you think that that confuses the issue?

Professor Hargreaves: There is a general principle that many, but not all, people accept, whereby data that is collected for one purpose should not be used for another, but that does not mean that data should not be collected for two purposes.

Johann Lamont: The point is that there is a lack of clarity about what the purpose is. The OECD suggests that there should be one purpose. We can argue about why there has been a change, but there has been a shift from a position in which standardised assessments were just about having an understanding of what was happening across the system to one in which it is argued that standardised assessments are also of benefit to the child. Does that have an impact on the way in which the test might be structured?

Professor Hargreaves: Everything that you say is fair. The OECD is saying that the prime purpose of assessment—the first directive for assessment, if you like—is to support learning. There are four message systems in schooling: pedagogy, curriculum, assessment, and the broad area of care and support for the child and their development. Assessment is one of those message systems, and having a deliberate strategy that can develop teachers' expertise in

assessment to support their students' learning should always be the prime directive.

At the same time, there is a need to align the assessments with curriculum for excellence and the national improvement framework, which form a Venn diagram. Both are important, but they are sometimes in tension with one another, so we have to be careful and think about which is the moon and which is the sun, to ensure that one part of the Venn diagram does not eclipse the other. Curriculum for excellence would recede into the background as the national improvement framework took over. As advisors to the Scottish Government, we always urge it to remain vigilant that it keeps the focus on both.

The OECD recommended that there should be alignment with curriculum for excellence and with progress in curriculum for excellence; therefore, we also need to know whether progress is being made. At the same time, the OECD proposed that the assessments be subject to teacher judgment. As advisers, we have recommended that approach to the Government at every point, and that has been accepted, as people will know if they follow our recommendations in the public media.

Large-scale assessments do not have a direct impact on other decisions but are mediated through teachers' professional judgments. The theory of change that is going on here is that, if there is any aggregation at any point—which there is—it is an attempt to create consistency in teachers' professional judgment. Their judgment is really important but, as we all know, there is a clear understanding that all individual judgments are flawed. We are all subject to unconscious bias, and we all tend to prefer people who remind us of ourselves. Getting consistency of judgment means that a student, at whatever stage they are at in any class, will get a reasonably equal and professional response from the teachers who deal with them.

The theory of change is crucial, and it is different from the situation in Ontario. It involves the buffer of teachers' professional judgment between the large-scale assessments that the kids take on a screen and what teachers do with their children in the classroom. That is the theory of change, and the challenge is to make it work. It is not a case of an assessment being developed for one purpose and then being used for another—it is more complicated than that.

Johann Lamont: That is precisely what is currently happening. One has become the other, perhaps in order to persuade people that the assessments are a good idea.

There is perhaps international evidence on one of the issues that has emerged, which is the issue of consistency when pupils take the tests. We

were told by the advisers to the Scottish Government that pupils can take the test at any point during the year. For example, in primary 1, a pupil could take the test at any stage between the age of four and a half and the age of six. Is that valid, or do you take the view—as some of the members of last week's panel did—that, in order for the findings to be informative and valid at a national level, there has to be some consistency in both the stage in the year at which the tests are taken and the circumstances in which the tests are taken? We have heard anecdotal evidence that some teachers prepare the kids for the tests, whereas others do not. Maybe such factors do not matter, but do you have a view on the validity of a test that is not consistently applied?

Professor Hargreaves: As any assessment system, including this one, unfolds, it will contain risks, and knowing what the risks are—you just mentioned one of the most serious risks—is really important. Any and every system of collecting and aggregating data about a child is imperfect.

I remember the first test that I ever took, at the age of seven—you might remember the first test that you ever took. I was called up to the headteacher's desk to do a reading test. I was in P3. I can remember the last word that I could pronounce and the first and only word that I could not pronounce, when the test stopped. The last word that I could pronounce was "pneumonia". I had to give the meaning of it, which, frankly, was not bad for a seven-year-old. The first word that I could not pronounce—I still cannot pronounce it—was "phthisis". It is beyond me why they had a test for a seven-year-old that listed, in successive order, two words about pulmonary wasting diseases; however, I felt that the test was important.

Until 10 years ago, when the governors of my former school sent me class lists from the time when I was at the school, I did not know for sure that the test was used to decide who went into the A stream and who went into the B stream. I had the same class lists for the children at 11 years old, which were almost identical; the evidence from the time shows that only about 2 per cent of children transferred streams between those ages. Then the lists showed which secondary schools they went to: 70 per cent of the A stream went to grammar schools and zero per cent of the B stream did; they went to vocational secondary modern schools. That was all decided at the age of seven. We know that those tests were flawed and that the 11-plus was flawed. When the 11-plus was abolished—or replaced with teachers' judgment—we found that the results of the selection according to teachers and headteachers' judgment had more social class bias than the results of an objective test.

The first thing that I will reaffirm is that, if you are looking for a nirvana of the perfectly consistent way of making judgments or doing tests, you will be disappointed. They will all be imperfect to different degrees and in different ways. We should avoid treating teachers' judgment as individual, autonomous judgment. In the teaching profession, we need collective autonomy, not individual autonomy—we have argued about that here. That means that we might have more autonomy from the bureaucracy but less autonomy from each other. By looking at the ways in which we make judgments together and moderate them, over time, we will create some consistency. The data can help teachers to do that. However, the data will always be imperfect, depending on whether a student is sick on the day, whether they are tired and whether they take the test at the end of the week or at the end of the day rather than at the beginning of the day.

You have outlined the risk associated with the tests, and, for me, the biggest risk is not that what you describe might happen accidentally but that it might happen systemically. The risk is that, if there is undue pressure from the Scottish Government or from local authorities to drive results up in a short period of time, in order to demonstrate success within a period of taking on leadership or before an election, that pressure will and does lead teachers to do strange but utterly predictable things.

If I were cynically advising a school now, I would say that, if it wanted to show improvement in its results over three years, it should, first, introduce a test without any preparation or professional development so that, in the first year, the students would do badly and the school would have an artificial low for its baseline. Following some professional development, everybody would do better in the test, so there would be the appearance of an improvement over time. Secondly, in the first or second year, the school should test all the children early in the year, when they are younger. A couple of years later, it should test them all at the end of the year, when they have had a bit more practice and preparation and have learned a bit more. The school would then get better results over time.

10:30

Across the world, where truly high-stakes tests are used and punitive consequences can follow, such practices go on. Technically, you cannot alter that much, although it is a good thing to allow children to take the test at different times because of things such as student anxiety and unreadiness and the possibility of a dramatic event obviating the validity of the result. You can deal with those imperfections by creating a culture of assessment

and improvement in which everybody is genuinely focused on improvement, which includes accepting those moments when they are unsuccessful and they need to identify a different way of moving forward.

Johann Lamont: I sat in a class of 45, and we were tested literally every week. We were sat at desks from 1st to 45th, so we knew if we were the 45th person in the class not only because of the mark that we got but because of where we physically sat in the classroom. I therefore know the challenges around apparently objective tests, and I am aware of the assumptions that a teacher brings into a classroom.

Is there a danger that we are reinforcing bias in apparently objective testing? For example, if a test is trying to assess capacity in language, literacy and numeracy, is there a danger that we are reinforcing what children bring into the classroom in terms of the words that they know? It is not that they cannot read but that there is less richness in the language that they hear at home or in their community, yet we are saying that that says something about their literacy. For some of the test questions, whether a pupil gets them right can be a question of whether somebody has told them what a word means as opposed to a measure of their capacity to decode it and say what the word is. That matters because of what you have said. People have been conscious of their bias and they are trying to deal with it. Do you accept that a theoretically objective test that is actually reinforcing bias can do a lot of harm?

Professor Hargreaves: Absolutely, and there is a lot of evidence to support what you say. In a high-stakes or even a mid-stakes scenario—when the first test is in primary 3, for example—kids start rehearsing the words in kindergarten. Rehearsing those words from the first moment they enter school is geared towards preparing them for the test, not so much because of the existence of the test but because of the stakes that are attached to it in terms of the school's improvement record, the pressure that is placed on it and the interventions that can be made.

There is no way to resolve that situation within a census test other than by lowering the stakes from high stakes to mid or low stakes, so that there is not a culture of fear or anxiety or a feeling that a school must always demonstrate improvement or there will be unwanted consequences. You need to build a culture within the teaching profession—among the headteachers and in the regional improvement collaboratives—in which all leaders clearly understand that the purpose is to learn and to find ways to keep moving forward, never to create a culture of fear or anxiety that will lead people to contrive the results.

Johann Lamont: I am making a slightly different point, which I ask you to reflect on. It is not about what is taught or what is practised; it is about what a child brings into the classroom. A child can be very competent and able—they can know how to read to the expected level for their age and stage—but there are certain words that they will not know because they have not come across that vocabulary at home. You talked about diversity in Canada. Is there a danger that the testing reflects children as competent readers because they have had access to a particular experience outside the classroom, which has given them the vocabulary to understand and respond to a particular question? How do we take such bias out of tests? Have you looked at whether some of what comes out of Scotland's testing regime is the result of bias rather than pupils having the expected skill?

Professor Hargreaves: I took the P1 test yesterday—apparently, I did quite well, although I did not find all the questions easy—so I have some direct experience of it as an adult. All tests, but particularly those that involve words, are prone to cultural bias. In Ontario, we found that questions that involved appetisers on a menu were totally outside the experience of children who live in poverty.

There are three ways to deal with that issue in a test. One is to continually review and modify—people should never feel that a test should not be subject to review and improvement. A second approach involves accommodations, which can be offered not just to children who have legally identified needs that bring mandatory and statutory supports but to all children who struggle with an aspect of their learning. As the committee knows, 50 per cent of young people in Finland will have been identified as having a special need by the time they finish school. That does not mean that they have a medical condition; it just means that they are struggling with learning in some way.

The third approach is to consider the genuine importance of having an array of assessment measures and data, of which such tests are simply one part. Teachers' judgment must have primacy all the time. If the situation does not get there or starts to deviate from there, there is the serious possibility that the great experiment will have failed.

I ask you to own the problem. Two things are needed: knowledge to support the child, wherever they are, and knowledge to support the system, so that you know the system. Just as you are responsible for Scotland's people, so the head of a local authority is responsible for all the children in their authority. Those two things are needed, but they involve a dilemma, which I ask you to own. You must seek the best way forward to resolve

that and not favour one aspect over the other or deny the dilemma.

Iain Gray (East Lothian) (Lab): That comment gets to the heart of what the committee is struggling with. I am not entirely clear what your judgment is. We are asking ourselves whether the SNSAs—such as the test that you did yesterday—can provide a teacher with the capacity to improve the learning strategies that they pursue with an individual child to improve their learning while at the same time providing system-wide information about what the system is doing. Can the test provide both those things with validity?

Professor Hargreaves: We should remember that the test should be considered to be one thing out of all the data that a group of teachers has—I do not like to think of an individual teacher, because all professions are collective; if people cannot share their expertise, they should not be in any profession. The test is part of the data and should not prevail over all the other data that informs teachers' judgment. They might use other reading assessments if they are searching for other reading skills that the test does not cover.

As I see it, the test is largely about comprehension. It reflects a worldwide movement to understand what people see in a narrative. It does not test—as far as I can see—the creation of ideas or pupils' generation of their own sentence constructions. To test that, other kinds of test or knowledge would be needed, including knowledge of the child. The test will provide some information about some things that are important for teachers, parents and Scottish education, but by no means will it provide all the information that is required.

Iain Gray: That is a very powerful argument, but my question was about whether the test can provide data at an individual diagnostic level that can be used in the classroom at the same time as providing system-wide information at school, local authority and, in particular, national level. If the answer is that it can provide only one part of that data, do you agree with what some of our previous witnesses have said, which is that it would have made sense to keep the Scottish survey of literacy and numeracy data alongside the national assessments to enrich the data that is available at a system level?

Professor Hargreaves: I will tell you what I have seen at the individual teacher level, which you have probably seen, too. Have you taken the test?

Iain Gray: Yes.

The Convener: Yes.

Dr Alasdair Allan (Na h-Eileanan an Iar) (SNP): I was too scared.

Professor Hargreaves: How did you do?

Iain Gray: Fine.

Professor Hargreaves: So you will have seen the individual report cards that go back. Frankly, I would need a reading specialist or an early childhood specialist to enable me to say what worth or value that information would have to a classroom teacher. According to feedback from the Educational Institute of Scotland, which you will probably have received in your testimony, in the first year at least, teachers get value from such feedback and it helps them to identify some of the ways in which they can support their children. Of course, not all teachers feed in their views through the EIS—teachers' views are also gathered in other ways—but some teachers think that the test information contributes to the kind of feedback that is useful for their students.

To address the other half of your question about the skills that are identified by the test, it will feed information into teachers' judgments about how a system overall is moving or not moving over time and how sub-groups within that system are moving or not moving over time.

On the test itself—I am repeating everything that you know—although there is national knowledge of the test, it is not possible for the nation to intervene in a particular school or a class that is taught by a particular teacher because of that school or class's performance in the test.

Again, we come back to the issue of how to deal with teacher judgment. I might be anticipating a question that will be asked later. Professional development should not be seen only or mainly as training courses on how to do the test. That is part of what professional development is, but the research on professional development in the UK and the US shows clearly that the best professional development is on-going, embedded in the profession, collaborative and seen as directly related to the learning. If the leaders of your schools and local authorities continuously bring together their teachers to look at what is happening as regards the judgments that are made, based on all the data that they receive, that will create consistency between the individual feedback and the national-level trends.

Iain Gray: You have made very clear how much importance you attach to teachers' judgment—you said that primacy must be given to teachers' judgment—and you have obviously reviewed a lot of the evidence that the committee has received on the issue. It is fair to say that the evidence from teachers—as individuals and collectively through the EIS—contains the significant judgment that the tests do not provide useful information in the classroom for learning and teaching strategies. Should that ring an alarm bell for the committee?

10:45

Professor Hargreaves: That should be a warning, and it should prompt the Scottish Government to work with the Australian Council for Educational Research on what the tests contain. If the breakdown of skills is not seen as valuable or useful, teachers must be able to say collectively what skills and competences should be represented in the tests. The evidence is not a reason to do away with the tests, but it is a reason to ask what kind of test is most valid for the skills that are important for CFE.

Iain Gray: When the test designers gave evidence, I asked them how teachers had been involved from the start in the early design of the tests. The designers could not say that teachers had had any involvement. Was that a mistake?

Professor Hargreaves: Will you repeat the last part of what you said?

Iain Gray: When the test designers gave evidence, I asked about input from teachers in the initial design of the tests. There was none. Was that a mistake?

Professor Hargreaves: The design of most tests involves teacher participation. The danger is that, because there has been participation, that is it for ever—a test is seen as having been validity and reliability tested, as being able to be moved at any time to any country and as lasting in perpetuity. Teachers need to feel continuously involved in all the assessments that inform their judgments. It is important to have one-time involvement at the beginning, but it is also important to have a continuous feedback loop.

The Convener: I am conscious of the time, and a number of members have still to come in.

Rona Mackay (Strathkelvin and Bearsden) (SNP): I will continue the same line of questioning. Are the tests compatible with play-based learning? You will be aware that a body of opinion thinks that they are not and thinks that our children are being tested too much. What are your thoughts?

Professor Hargreaves: First, a clear philosophy and stance is needed on what early childhood education should look like up to and including P1. Debates are raging about the subject, and people who are sitting behind me know more about that and have even stronger views on it than I do.

Rona Mackay: I just wondered what your opinion is.

Professor Hargreaves: Play is an extremely important part of childhood. Clear evidence is emerging that young children are spending too much time looking at screens and not enough time engaged in other things; they are also spending

too much time indoors and not enough time outdoors.

Privileged parents read to their children from a young age. Some children will master a large vocabulary and a range of words from a young age, while other children will not—there are huge disparities. That is a fairly strong predictor of all kinds of indicators of later success or otherwise, including rates of imprisonment and employment and going to what Scotland thinks of as positive destinations.

An equal society such as Finland, where more people subscribe to public libraries than in any other nation in the world, can afford to have a philosophy for early childhood that is predominantly about free play. In an unequal society that has huge disparities in access to language at home, for example, it is important to consider on the ground of equity some forms of more structured play—I have seen that in Ontario—that provide ways of engaging with numbers or number sense, for instance. That engagement is still playful and enjoyable, but it is structured to help children who have less behind them to progress when they come to school, so that they have the same chance as all the other children.

Rona Mackay: I understand what you are saying, but I am talking about whether the tests are compatible with play-based learning. Can the two co-exist happily? Are tests at a very early stage necessary and do they provide value? To go back to Iain Gray's point, what value can we get from tests at such an early age?

Professor Hargreaves: The test is not a test of everything but is a test of literacy—and not even all literacy, because it is primarily a test of comprehension and reading. If developing reading to a certain degree is important in your curriculum, the tests will have some value.

On whether the experience of the test itself is incompatible with a play-based environment, I would say that the test does not come with bells and whistles, although it probably could. Apparently, the reason for that is a lack of broadband width in some of your schools. If you had more broadband width, you could have fancier tests that were even more playful and enjoyable.

My grandchildren—and possibly some members' children—will sometimes learn maths and other things by playing games on the computer, as well as through physical play with objects. Although I am broadly not in favour of a lot of technology in early childhood, a bit of familiarisation with technology in the classroom where possible, so that when children take the test it is not the first time that they have faced that technology, would make it seem less like an

extraneous event and more like a continuous part of classroom learning.

Rona Mackay: Do you see the tests as high, medium or low stakes?

Professor Hargreaves: The test is meant to be low stakes and is at risk of becoming medium stakes, but it is not at all high stakes.

Jenny Gilruth (Mid Fife and Glenrothes) (SNP): Good morning, professor. The 2011 OECD review advised that policy makers can

“reduce distortion and strategic behaviour by increasing teacher involvement and buy-in from an early stage”.

The SSLN arguably did not do that: it was a tool for Government and did not empower teachers. Historically, in Scotland at least, ownership of data in schools seems to have lain with headteachers and deputy heads. Do you have any examples of training teachers to engage with assessment data in a meaningful way? You have already alluded to building a culture of improvement, perhaps through regional collaboratives. Are there any other examples that we could learn from to empower teachers?

Professor Hargreaves: You have asked two related questions. The first question was about training on assessment for learning, which is typically given the lowest priority of all the four message systems that I have described, which are curriculum, pedagogy, assessment and care for young people. In Ontario, we are facing the same question that you face in Scotland. One of our recommendations was that more attention should be given to continuous learning of assessment and assessment for learning, in the classroom context.

In Ontario, there has been some success, because of the stability of government over a period of time. You can get stability of government in three ways. One way is not to have a democracy; for example, Singapore does not have a democracy as we would understand it and so has complete stability of government. Another form of stability is when one party is in control for a long time, as happened in Ontario for 15 years. Finally, we can get such stability through cross-party agreement and consensus that education is above political infighting—that is pretty much what there is in Finland. In that respect, I urge you not to be like Singapore but perhaps to be a little more like Finland.

I have forgotten what the first part of your question was about.

Jenny Gilruth: It was about the usefulness of the SSLN, comparatively, for teachers.

Professor Hargreaves: Over 15 years, Ontario has successfully built a very strong culture of

collaborative inquiry, whereby teachers will routinely inquire into problems of practice together, within their school. As part of that process, they will consider all kinds of data, including test data.

I will give a clear example. We have worked with a seventh of all the school districts, on and off, for 10 years. Ten years ago, when the stakes were higher in assessment, the focus was almost solely on literacy and numeracy, and there were consequences if results did not progress. Schools would identify what they called marker students, who were students whose scores were just below the acceptable point of proficiency. Here, that would probably be the level of progression that a student was supposed to be at in CFE. To get their schools up to a good score, school heads would have charts on their walls—we took photos of them. Proficiency was number 3, and the charts showed the percentage of students at 3, as well as the percentage at 2.9, 2.8 and 2.7. Teachers would give a disproportionate amount of their attention to the students at 2.7, 2.8 and 2.9. When they said, “What about the 1s and the 2s?”, they were directly advised to forget about the 1s and the 2s and to concentrate on the 2.7s, 2.8s and 2.9s. That was 10 years ago.

Now, Ontario has broader goals that are much more like those for CFE. Literacy and numeracy are still there, but wellbeing is now a goal, and equity is now defined as inclusion, which means that children must be able to see themselves in the curriculum. Teachers are now addressing the broad range of their children’s learning, including literacy and numeracy. They now focus on what they call mystery students or students of wonder. A student of wonder is a wonderful student who is struggling with a particular aspect of their learning, and teachers in the school, who work together collaboratively, wonder why. The school will bring together the teacher who teaches them now, the teachers who used to teach them, the special education support teacher, the language specialist, a school counsellor and a speech therapist. They will bring together 12 or 13 teachers to look at that student of wonder and to work out how to advance their learning. They will use all the data that they can collect, which will include things such as photographs of the student’s work taken on an iPhone and made available for everybody to see. They will use numerical data, test score data and diagnostic tests, as well as all the other information that teachers use to inform their judgments over time.

The ministry has a very good website that collects lots of materials and instruments that can be used, but the main thing is that the province now has a very good way of mobilising knowledge and moving it around within and between schools. The districts—for several years, at least—worked very well together in taking collective responsibility

not just for their own success but for one another’s success. The collaboration that took place at school level was replicated, to some degree, at the district level.

That deals with the first part of your question, and it almost covers the second part of it, too.

Jenny Gilruth: That was very helpful—thank you.

I want to ask about equity. In a previous evidence session, we heard from Professor Sue Ellis from the University of Strathclyde, who spoke about what had happened prior to the introduction of standardised assessments, when groups of children had been removed from class. She argued that that was unfair and unequal, because it created an unlevel playing field, as it were, by singling out children. With the SNSAs, do you think that there is an opportunity to level the playing field and to stop that?

11:00

Professor Hargreaves: The issue of exclusion is always controversial. One of the regrettable things that happen in Ontario education is that if, for example, a refugee from Syria who speaks almost no English arrives on the day or in the week of the test, the school has to decide whether to enter them for the test. For the child to have to do it can be humiliating, because they might sit there for over an hour trying to make sense of a language in front of them that they do not know. Alternatively, the school will exclude them, they will score a zero, and the school will therefore get a zero. The more refugees there are, or the more students there are with post-traumatic stress in the school, the greater the risk of zero scores. It is an impossible dilemma for teachers when a mid-stakes test that has dramatic significance can be taken by the children on only one day.

That can be got around by making the test less dramatic—by incorporating it and making it feel like it is part of the curriculum. I know that we are talking about the large-scale standardised assessment, but if other kinds of assessment also go on, children will learn that assessment is part of learning. If there are peer assessments and self-assessments children will understand that they do not do only learning, and that there is a thing called “assessment” that is part of their learning all the time.

If the test were to be modified so that it could be spoken as well as read, if necessary—the existing P1 test is partly, but not wholly, that way—and if the supports were available to accommodate modification of the test so that people with learning differences could access and express what they know in different ways, which is a resource question, there would be greater inclusion.

Dr Allan: Among the many interesting things that you mentioned was the idea of high, low and medium-stakes tests. It is fairly clear that the relevance of assessments depends partly on how closely related they are to the curriculum within which people work. I am keen to hear a bit more about how the assessments have fitted in with, or have been helpful to, the curriculum in Scotland.

Professor Hargreaves: I have not seen other assessments; I have seen only the P1 assessment. I know that all the activity and interest relate to that at the moment.

The Convener: For clarification, I point out that there is an issue relating to the P1 assessment, which the Government is dealing with, but the committee is interested in testing at all levels throughout the curriculum. Those are the terms of our inquiry.

Professor Hargreaves: The short answer is that I have not seen the other assessments, but I have seen the P1 assessment, because I realised that that is what is on the radar.

The P1 literacy assessment is basically an assessment of reading comprehension. It should be consistent with the literacy strategy. Curriculum for excellence is about many things other than the acquisition of literacy, so the test needs simply not to be inconsistent with those other things or to interfere with them. Other ways of judging the emotional and social development of children should also be a very important part of how teachers assess how kids are progressing.

Dr Allan: How we prepare teachers is related to that and to your work in Ontario. You have said that the Ministry of Education in Ontario should

“Implement professional learning and development for educators at all levels of the education system ... in concert with the roll out of the new ... assessment system”.

How would you translate that advice for Scotland? What analogous advice might you offer?

Professor Hargreaves: First, as the recommendations in the committee’s review pointed out, Heriot-Watt University has run a reasonably well-regarded training programme—assessment 1.01, if you like—which covers management of the basics of assessment, how to understand it, digital competence, developing digital competence among the children, and the significance of judgments that are made at different times. That is professional development as we typically understand it.

As important as that—perhaps it is even more important, once you have started moving—is professional development for middle-level teacher leaders in schools, for heads and deputy heads, and for local authority staff. That is in order to create a culture of assessment for learning and

assessment as learning. When that is the case, I can go into a school and see how the children are continuously reflecting on what they do, how they are setting goals for themselves and making judgments about each other’s work as well as their own, and how teachers are helping them to do that. Teachers and children will understand that assessment is not only part of learning, but a form of learning. That does not happen automatically; it needs to have conscious attention paid to it.

If that culture is developed effectively throughout a system, when an assessment instrument or device comes in people can figure it out in a strongly collaborative culture and it can be integrated into the understanding of learning, teaching and assessment that runs throughout school. The priority is learning and making shared judgments about learning—rather than there being just individuals’ judgments—in which there is consistency.

Dr Allan: I am interested in what it might take to create that culture. Earlier, half—or perhaps only a quarter—in jest, you mentioned the importance of consensus, including political consensus on some issues. Is there more that we could be doing to try to create that consensus within or outside the world of politics?

Professor Hargreaves: I hope that in one respect, at least, the Scottish Government can be different from Westminster, which is better able, on the issue that we cannot name, to articulate what it does not agree on than to articulate what it agrees on. Everything pivots not on the technicalities of the test, but on teacher judgment. If you can attain cross-party agreement, you can be world leaders in knowing where your country is going and how to help all your teachers to help your children to learn.

You will know from the OECD report that after devolution the National Assembly for Wales’s first act, almost, was to abolish standardised tests. That was a way of saying, “We’re going to do this differently from how we were doing it under Westminster.” Wales replaced the standardised tests with teacher judgments. Those judgments were somewhat moderated, but not in a very disciplined way, and the result was chaos and inflation of grades. Nobody wants to say that they are doing less well this year than they did last year, so there was improvement all the time until it could go no further. Wales was very clear that it wanted to get rid of standardised tests, but was much less clear about how it would create consistency in teacher judgment. The secret to moving Scotland forward will be in finding ways to support that quest, even if people differ on the best way to do it.

The Convener: The final question—I hope—is from Mr Mundell.

Oliver Mundell (Dumfriesshire) (Con): I have a couple of questions. From your experience of testing, does placing pupils in rank order, or deciding at a very early stage where they sit relative to their peers, not inevitably lead to bias in terms of the strategies that teachers use and how they teach in the classroom? If pupils are taught in sets or according to their ability—as in the example that you used—does that not add to existing differences and not focus on getting every pupil to where they can go?

Professor Hargreaves: I will check back with you, but I think that you are drawing our attention to the fact that there is no longer be an 11-plus examination and that we no longer put kids into streams, but put them into groups. In secondary schools, we put pupils into sets and—sometimes—into streams. The OECD data is clear that higher-performing countries select by ability later in the child's life, and lower-performing countries do so earlier. The countries with higher equity select later, and countries with lower equity select earlier.

Oliver Mundell: Does that mean that there is a danger in introducing, for pupils who are aged between four and six, a diagnostic test that starts to focus on individual interventions? As my colleague Johann Lamont said, pupils of that age might have ability but not knowledge. Should we not give them the chance to catch up and adjust to being in a more formal classroom setting?

Professor Hargreaves: There is a risk, but it is not inherent in the test. If we were to go into a school classroom, we might see four or five reading groups, which will be named after birds, planets or whatever, but which could be categorised as “fast”, “quite fast”, “in the middle”, “a bit slow” and “very slow”. The reading groups will work at different levels, and the kids will be able to pick up fairly quickly which group is which.

One of the purposes of a diagnostic test is to group kids in their learning so that they can be instructed in the most effective way. Teachers cannot respond to individuals all the time; sometimes they work with the whole class, occasionally they work with individuals and usually they work with smaller groups.

A good area of research on the issue is in co-operative learning. Sometimes children will be grouped by the same ability and sometimes they will be deliberately grouped by different ability. I do not mean that it is done randomly; a group could include someone who is a bit further ahead and someone who is a bit further behind, and they will work with each other at different levels.

Oliver Mundell: My question is: is there is a higher risk in testing children and segregating them based on their ability at the early-years

phase? I do not know what we would call it, but at that time children have not been given the chance for things to balance out a little bit in a formal education setting. Is there a bigger risk in diagnostic testing being used to decide how to teach children at that age?

11:15

Professor Hargreaves: It all depends on the culture of the school—I am starting to sound like a broken record. All teachers assess early. That is part of their judgment about things such as whether a child needs a bit of a push or whether the teacher should hold back, and whether a fight will break out or whether the teacher should let the children work their way through it. Those are all judgments and assessments of what teachers know about children in particular and in general from the evidence and their experience.

All of us always make early assessments. We might assess that a child has difficulty in forming relationships with other children and that we need to do something about that. First of all, we need to watch and wait a little bit—but not for too long—before we intervene. The same is true in respect of language. Whether an assessment is informal or formal, it is important to make it. One cannot possibly teach effectively without coming to judgments about and assessments of children from the very beginning.

Oliver Mundell: Is the risk of the teacher's judgment being imperfect at that early stage greater than the risk of the test producing a false positive or false negative? Having looked at the test and seen examples of its being done with children in schools, I think that there will be some children—perhaps those who have not had the same experience at home as others—for whom the tests will inevitably produce a result that does not give an accurate indication of their ability, for the reasons that you have outlined. At that early stage, is the risk of poor teacher judgment greater than the risks that are associated with testing?

Professor Hargreaves: When a doctor looks at a brain scan, the scan does not speak to the doctor automatically; it has to be interpreted by the doctor individually, and perhaps collectively by others. I have not had a brain scan but, six months ago, I fell off the Appalachian trail and broke my ankle in two places. I now have a plate down the right-hand side of my leg, which had to have 42 staples in it. Part of the plate is having difficulty healing. When I went back to see the surgeon, I saw the resident. He is more junior than the surgeon, and he did not seem to be certain about what the problem was, but he gave advice nonetheless. I asked, “Have you ever seen this before?” He said “No,” so I said, “Perhaps we could have someone in who has.” I was reminding

him that he works in a collective—not an individual—profession.

The next person who came in was the orthopaedic surgeon who sawed through my ankle—by the way, they are all men in orthopaedics, because they think that it is like being in the basement with tools, plugs and everything else. He looked at my leg, and I said, “Have you seen this before?” He said, “Not quite like this,” but it could have been this or that. He said, “Can I take a photograph of it, which I’ll send to dermatology?” The photo went off to dermatology.

At that point, they had the original X-rays and a photo—because we have iPhones—of the ankle. Three people had been consulted as well as me, because I am treated seriously as a patient. I make sure that I put my occupation at the bottom of every email before we connect so that I am taken seriously. If I were a plumber, they probably would not take me seriously.

Through that mix, we came to a judgment together about how to proceed, even though we were still not exactly sure and were trying to figure out what was best.

All judgment is imperfect, including a judgment that is based on a photograph, an X-ray or whatever it might be, as it depends on our collective ability to interpret. If we have a culture in which we teach people that X-rays are gospel and tell us what to do and in which the data drives us, we are in serious trouble. However, if we have a leadership culture in which we drive the data and the data does not drive us, and in which what matters is how we make sense of the data, including being critical of it, we have a chance of progressing.

Oliver Mundell: I have another, slightly different question. You talked about improving teacher judgment, for example. Would you start with standardised assessments, or are there other things that can be done in teacher training to teach people about bias issues and to enhance their ability to spot and identify different literacy problems? Are the assessments the best way to encourage a collaborative culture and help people to understand where other people are?

Professor Hargreaves: Exactly as you have said, there are many ways to improve our judgments, whatever our field, through referring to our collective knowledge and to outside knowledge that is somewhat independent of what we have among us.

The history of what we are looking at is important. Speaking as an adviser, it is public knowledge that the initial position was to have a high-stakes standardised test in the Scottish education system. As advisers, the advice that we

offered—whether you like it or not; the nature of advice is that you can ignore it—was that having a high-stakes, large-scale standardised test would have all kinds of negative impacts on teaching and learning. However, your Government feels that, in an unequal society, large-scale information is needed to guide it on where best to provide support and intervention.

There is now meant to be a lower-stakes assessment that is one of the things that informs teacher judgment. The main way in which we will figure out how the system is moving is through the aggregated data on teacher judgments.

That is the art and science of how we are trying to get beyond, on the one hand, a high-stakes, large-scale standardised test with utterly predictable and pervasive negative consequences and, on the other hand, no standardised testing at all, which leaves us unsure and unclear about the consistency of teacher judgment across schools and local authorities.

That is the dilemma and puzzle. As an adviser and somebody who has come to love Scotland—I courted my wife in Scotland a lot—I hope that you can help us to help you to figure out the best approach.

Oliver Mundell: By going for a compromise between the two approaches, might we end up losing the benefits of both? Have the advisers considered that?

Professor Hargreaves: We do not see it as a compromise; we see it as a third way that is between and beyond the two alternatives that the world has previously dealt with.

The Convener: Professor Hargreaves, thank you very much for your attendance this morning. We really appreciate your taking the time to come along.

11:23

Meeting suspended.

11:28

On resuming—

The Convener: With us on our second panel are Sue Palmer, who is the chairperson of Upstart Scotland, and Jackie Brock, who is the chief executive officer of Children in Scotland.

Jenny Gilruth: I will start with a question about the previous Scottish survey of literacy and numeracy. The Children in Scotland submission states:

“We believe evidence from SSLN and National Qualifications provided enough evidence to highlight and track attainment and the attainment gap at a national level”.

Do our witnesses recognise the limitations at local and school levels of the SSLN with regard to tracking pupil progress and informing teachers?

Jackie Brock (Children in Scotland): It is interesting that you are asking about the SSLN's limitations. Perhaps I could start by talking about its potential. I know that you have heard about that in previous meetings, too.

The first year of SSLN's reporting nationally on numeracy showed us that we were doing really well in terms of the ability of children in the early years of primary to add, subtract and do basic multiplication and division—essentially, teachers were doing really well in teaching basic numeracy to children. What was appalling was that children in primary 4 and beyond were not able to apply that knowledge to more sophisticated concepts. The situation with fractions was the evidence for that view.

That finding enabled us to understand teachers' needs for development in numeracy: they were good at teaching basic numeracy, but their ability to transfer that to enable children to apply it in more sophisticated ways needed more attention. That was the case across Scotland; it is not the case that some pockets of the country were doing better and some were doing more poorly.

11:30

The evidence on what was happening in primary 4 also helped us to unpack what was going wrong in the later years, and the implications for Parliament's and our aspiration that pupils do well in science, technology, engineering and mathematics subjects.

The Government then—this was probably anticipated—put in place a huge range of professional development that could be applied by every teacher in their teaching of numeracy. That demonstrated the lack of ownership that was mentioned in the session with Professor Hargreaves, and which I agree exists. However, the opportunity was lost to use evidence about how relevant the SSLN was to teaching in the classroom, in my view.

The following year, no one was surprised that the SSLN again showed that teachers were doing really well at getting children, whatever their background, up to scratch on basic literacy concepts—the basic comprehension in P1 and love of reading that the committee has heard about. However, across Scotland, in respect of applying literacy there was a huge gap in children's ability to talk articulately about what they were learning, particularly among boys.

There is what I see as a shared dilemma. We dismissed that evidence and did not follow through

on what it told us at national level in ways that could have improved and sustained performance, which is really important. In Scotland, we do not have a wealth of assessment data and information on follow-through at individual, school, local authority and Government levels.

I will say two things about that. No local authority chose to enhance the sample of the SSLN. What does that say? Also, the committee has lost an opportunity for consistent tracking at national level using the SSLN or other national information. You could have had an annual report based on evidence of improvement, and you could have been homing in on where we need to go to improve our education, based on real data, by addressing the individual literacy and numeracy needs of children. Scotland has lost that.

Sue Palmer (Upstart Scotland): One of the great strengths of the SSLN was that it did not cover P1. I fear that the results that Jackie Brock was talking about might have their roots in early years. I am here mainly because I am very much opposed to standardised testing of children at the age of five, other than general developmental testing. I am opposed to specific testing of literacy and numeracy at that age.

As Oliver Mundell pointed out in his question, if we focus too hard on children at an early age, many will not do well and will spend the rest of their lives playing catch-up. If we do as all of mainland Europe does and leave specific teaching of literacy and numeracy skills until children are six or seven, we would have an opportunity to create the level playing field that we have been talking about.

At that early stage, we would focus on elements including speaking and listening, which are hugely important and foundational throughout education, and not just for literacy. We would focus on self-regulation, which is children's capacity to control their behaviour and settle in a classroom; on social and communication skills, which are similarly important; on focus and control of their attention; and on their need to deal with complex information.

All those skills are foundational. If we were to concentrate on them at the early level that straddles nursery and P1, rather than homing in too soon on specific literacy and numeracy skills, maybe we would create a better foundation and there would not be so much of a fall-off at P4. That would be a strength of not assessing at that particular stage in children's lives.

If we build our education system on a shaky foundation because we are too busy doing the three Rs when there are other, more important things that we should be doing, we might look

good in the short term, but there will not be good long-term implications.

Jenny Gilruth: I want to go back to points that Andy Hargreaves made in the previous evidence session. He was keen to highlight assessment-for-learning methodologies, with which most Scottish teachers will be pretty au fait. He spoke about collaboration, about shared understanding, and about developing in schools a culture in which assessment is embedded in learning and teaching, so that it is not a high-stakes issue. In fact, he argued that SNSAs are not at all high-stakes tests.

I note that Sue Palmer said in her submission that

“SNSA is recognised by the public and media as a key factor of a high-stakes policy.”

Why is Professor Hargreaves wrong?

Sue Palmer: I do not think that Professor Hargreaves is wrong, at all; I said that SNSA is politically “a high-stakes policy”. That will affect public perceptions of it, which will affect what goes on in schools. If people feel under pressure to improve results, it is more likely that there will be the unintended consequences and behaviour that are often described as being related to testing.

I am sorry. What was your first question?

Jenny Gilruth: I also asked about assessment for learning.

Sue Palmer: Professor Hargreaves talked about Ontario. In Ontario, there is a developmental test at the equivalent of P1—the children are aged five, going on six. The early development instrument is used across Canada; the kindergarten teacher does the assessment. She looks at a checklist that covers social competence, physical health and wellbeing, emotional maturity, language and cognitive development, communication skills and general knowledge. Through that, teachers get a great deal of information about the sorts of developmental factors that are really important at that age. If we want to create a background for professional judgment, such a system could very well enhance professional knowledge.

If we focus just on literacy and numeracy, they become what is salient, and literacy and numeracy skills will tend to dominate what people do in the classroom. That will have the inevitable effect of grouping children, which was mentioned earlier.

There can be some sort of testing. I think that Professor Lindsay Paterson mentioned last week that the Netherlands has a developmental test at P1 age. Germany has a very good developmental test that helps to inform how the teachers work. It does not say that the three Rs are being done; it

looks at development. Results depend on what we look at, and that influences what we value and discuss and base our professional judgment on.

Professor Paterson talked about SNSA—I am sorry, but I say “sensa”, because that is what teachers call it; I can never remember all the letters. He said that we have based SNSA on the curriculum. We have not; we have based it on the benchmarks, and the benchmarks for P1 are extrapolated from the experiences and outcomes. That extrapolation is quite distorting. There are 54 benchmarks for literacy, 22 of which relate to speaking and listening—that is nowhere near enough; speaking and listening are big things—and 32 of which relate to specific literacy skills.

I disagree with Andy Hargreaves. I, too, have been given a demonstration of the P1 SNSA, and I would say that it covers a lot more than comprehension. It covers phonological awareness, word building, letter and word recognition and so on. Of the 54 benchmarks, the test covers about 10, which seems to me to be completely distorting the curriculum. The existence of the benchmarks, even without the test, would distort teachers’ impressions of the experiences and outcomes.

The original experiences and outcomes use words such as “explore”, “play”, “discover”, “choose” and “develop”—they are major verbs. Once we drill down and turn those words into specific tasks, we move away from a holistic developmental approach to early education, which is what curriculum for excellence is about, to a drilled-down, skills-based approach. If teachers look, as I suspect they will, at the benchmarks rather than at the experiences and outcomes, that will affect whether they achieve curriculum for excellence levels assessments.

Jenny Gilruth: I want to consider not what is happening in Ontario, but what is happening in Fife, which is where my constituency is. The Durham University centre for evaluation and monitoring’s assessments will be brought back in in Fife. Arguably, that is due to the politicisation of the SNSAs that you alluded to at the start of your answer. That will cost Fife Council up to £100,000, and because more than half of primary 1 pupils have not sat the baseline PIPs—performance indicators in primary schools—tests, we cannot just shift back to using the previous assessments. Instead, the Durham assessments will be used alongside the SNSAs, which will potentially double the assessment load on pupils. As a former teacher, I am appalled by that. Were Children in Scotland and Upstart Scotland against the Durham assessments?

Sue Palmer: We were. I am against specific skills-based assessments of literacy and numeracy skills. I am not against developmental assessments and checklists, which look at

children's development in a more holistic way and can inform what interventions might be needed for individual children. If we test literacy and numeracy skills, numeracy and literacy will be what is done in the classroom. I am also absolutely opposed to other sorts of specific assessment.

Jackie Brock: Children in Scotland is not opposed to any diagnostic formative assessments for children of any age throughout Scotland's education system. We are, however, opposed to standardised assessment when it is used to measure and shape individual children's performance and individual teaching strategies, for all the reasons that Andy Hargreaves set out in the previous session and, critically, because of Professor Louise Hayward's point last week about feedback.

There are pressures on politicians, local authorities, individual teachers and children in relation to how high stakes the tests are and to the semantics around the issue. If freedom of information requests are used to measure individual schools and, therefore, individual teachers, and to shape performance, and if that gets into the press, we will have a huge problem in how we consider Scottish education.

Critically, the tests will shape behaviours, and Children in Scotland's members are not satisfied, because the assurances that the Scottish Government made have changed. The Government has shifted the approach that is being taken to SNSAs, which is very welcome, but unfortunately the die has been cast in respect of how the tests will be used. The Scottish Government might say that the tests will be used in a certain way at the moment, but if the latest programme for international student assessment results, for example, show that there is a problem, more pressure will be applied to local systems and the Scottish Government to reveal more about what we know. There is a real danger that the information that will be formed, judged and used from SNSAs will become distorted, and it will be out of the Government's hands.

Jenny Gilruth: I appreciate what Sue Palmer said about being against the Durham assessments, but if we follow what Fife has done—it got rid of SNSAs and returned to that system—children can be removed from class in groups. I made the point to Andy Hargreaves that Professor Sue Ellis had previously raised about equity and singling out individuals and removing them from class. Surely, the SNSAs give us the opportunity to stop such things from happening and create a level playing field for all children.

11:45

Sue Palmer: I do not see how. The point about early level is that it is a stage in children's development when there is massive variation in what they can do in, for example, literacy and numeracy. It has been pointed out that that variation can be to do with their previous experience, including the richness of their experience at home and their family background. It is also to do with individual genetic predisposition. To put it simply, when learning to read, some children click later than others do.

I adore curriculum for excellence because it tries—especially at early level—to nudge the Scottish system away from going in heavy on the three 3Rs as early as P1. That is a developmentally appropriate stage—much more like what you would see in northern Europe. Unfortunately, it has never really taken off, because we are still stuck in the cultural habit of starting the three Rs early. It horrifies me that, just as we are beginning to see some schools starting to move towards play-based pedagogy—developmentally appropriate pedagogy—in P1, the introduction of the SNSA will kill that in its tracks. The SNSA firmly puts the focus back on saying, "Get on with the literacy and numeracy skills. Crack on with it now."

Jenny Gilruth: Are you saying that the SNSA will stop play-based learning from happening? That is not my understanding.

Sue Palmer: Yes. The two are inconsistent. I am not saying that you cannot be playful in your learning and put elements of play-based learning into a classroom in which you have groups working on literacy and numeracy skills. Those groupings will have to happen, if you are trying to address literacy and numeracy skills this early, and I have seen them every school that I go into. You can have such a hotch-potch, but if you are trying to provide a genuinely developmentally appropriate stage, testing will skew that away from being relationship centred and play based.

Jenny Gilruth: I visit schools regularly in my capacity as an MSP, and I was in a classroom not that long ago. It is certainly not my experience that that is what happens in our schools, so what is your evidence base for that assertion?

Sue Palmer: My evidence base is the same as yours. Every school that I know has reading groups.

Jenny Gilruth: They do not have any play-based learning.

Sue Palmer: Oh, no—I did not say that they do not have any play-based learning. I said that you can have some play-based learning and reading groups. However, the very fact that you have

reading groups indicates that it is not early childhood education based on development and on supporting every child at their individual developmental level. That is the ethos of a kindergarten; that is the ethos that you see in kindergartens in Finland and Germany. They do not say, “Oh, well, we’ve got a literacy standard, so everybody’s got to work to that standard.” They say, “No, we support the child at the stage it’s at, and we create a supportive, literacy-rich environment. We pay particular attention to things like speaking and listening. We are looking at how well children are learning to focus attention.” All those other things are going on as well. Indeed, in the Scandinavian countries, a great deal of emphasis is placed on self-directed outdoor play, which, as has been mentioned, is disappearing from children’s lives.

When we started Upstart Scotland—it was before the tests began that we got talking about it—it was nothing to do with numeracy and literacy. We were interested in reinstating play in children’s lives and having a ring-fenced period when that became very important.

It is not that you cannot have playful activities or games. You can, and you can turn those into lessons on how to recognise words or on sound/symbol recognition. However, that aims at a standard rather than at a genuine play-based environment in which children are gently supported at whatever level they are at.

Rona Mackay: I am interested in what you said about not being opposed to assessments, depending on whether they are for developmental purposes. For instance, if guidelines went out to schools and local authorities to say that such tests were not to be used for streaming children or as a benchmark for their future learning, would you be content if monitoring were to be done to ensure that that was not happening? I am interested to know what evidence you have that it is happening. Would that allay your fears, or do you dislike the entire nature of the test?

Sue Palmer: I am not sure that it would allay my fears, because I am not sure how easy it would be for teachers to do that. As we have said, once we have tests, the things that are on a test become salient, which then affects the way that we teach. If we were trying to teach P1, it would be very difficult to cover the specific skills that are in the test without grouping children. There are 25 children in a classroom, and it takes a lot of sitting down and helping them to understand such concepts. We have to keep repeating them, particularly for the less able groups. It is very time consuming, and therefore the grouping helps a lot. If we are aiming to concentrate on specific literacy and numeracy skills in the early level, I do not see how teachers can avoid using groups.

Rona Mackay: I do not have a teaching background, but would it not be possible to have that information so that the results of the test are noted but children are not streamed or grouped, and to leave seeing how much they have progressed until they are at a later level?

Sue Palmer: What happens in an early-level classroom is affected not so much by the results of the test as by its very existence.

Rona Mackay: But you are in favour of developmental tests. Would that not do the same thing?

Sue Palmer: No. We are interested in developmental tests that show the overall, holistic development of children. The EDI measure that I described earlier, which is used in Ontario and across Canada and Australia, has been piloted in East Lothian and validated for Scotland on the basis of that pilot. However, it never reached parliamentary level but stopped at civil service level. It was done roughly around the time that the idea of introducing standardised tests of literacy and numeracy came in.

Rona Mackay: I come back to my original question. If guidelines were to be put in place to ensure that tests were not used for the purposes that you do not believe they should be used for, surely that would be better from your side?

Sue Palmer: I have said that I do not think that guidelines can work in such circumstances.

Jackie Brock: In previous evidence sessions the committee talked a lot about the purpose and range of assessments. We need to be mindful of the amount of guidelines on teaching practice that is out there. I suggest thinking it all through again. You are obviously greatly exercised about the purpose of assessments. We need to look back at Scotland’s very strong legacy of thinking about assessment for learning and the points that Professor Hargreaves made about culture. Nationally and locally, we have had remarkable cross-party and political agreement on what we want to have on assessment. In 2005, the “Assessment is for Learning” guidelines stressed the importance of teacher judgment, supported by a range of assessment tools, which would be decided on locally.

Teacher judgment and moderation are critical of that. We all recognise the understandable propensity for bias. We all understand that, as professionals, teachers want to be able to check in with their peers and get support, so that they can support the progress and improvement of their pupils; of course they want to do that.

A huge amount of development has taken place since 2005; some of the committee members probably benefited from the professional training

and fantastic developments that went on in Scotland. All those principles were later reinforced under the 2011 “Building the Curriculum” guidance and there was a strong amount of reinforced pressure around moderation.

On the purpose of assessment and guidelines, and what we are actually doing with the information, it is interesting but frankly disappointing that we are not hearing about the thriving moderation that is going on in Scotland. Where are the moderation and discussion at school level? What are we hearing about the thinking on assessments in our schools? What successes do we have and how are we building on that improvement? What moderation are we hearing about at a thematic level?

We hear about a lot of amazing work that is being done around STEM at the school cluster level. Teachers in STEM know that they need to check out and work on standards in order to improve, and that is happening at a cluster level. However, there is a failure of confidence in the system—at local authority and national level—about whether it is actually good enough. That was the genesis of the SNSA.

Within all this, we have had a settled political, national and professional understanding of the purpose of assessment. We also have a legitimate, important and powerful requirement in our education system to remove inequality. For some reason, we have decided that we do not believe in valuing and strengthening teacher judgment and moderation, strengthening assessment, and building on our learning strategies. We have decided that we do not believe in all that; instead, SNSAs are the way forward in removing inequity.

We have heard powerful arguments about why that might be. However, it seems that we are now lurching towards a new way of looking at standardised assessment, but a huge range of international evidence suggests that that will not work in a high-stakes environment. We have heard that the timing of the tests cannot be standardised and that the information will not be known in a standardised way, either at the national level or between local authorities. I therefore worry what the guidelines on the use of the tests are for and how they will be used. How will teachers be trained and their development supported? How will the guidelines all of a sudden reveal clarity about how they can use that information to improve their teaching strategies?

The committee’s inquiry offers an opportunity to go back to basics around assessment and to think really carefully about what standardised assessments—as opposed to the measures that we have been using for some time—could offer.

Liz Smith: You make an interesting and powerful argument about international evidence.

If a school is not doing as well as it could be and requires more support, or if a particular local authority has not performed very well in the past, the big issue that troubles local authorities, many politicians and certainly many parents is the question of what kind of data we need to help those schools to do better, so that we can raise attainment. Scotland has not been doing as well as it might on a lot of the international measurements, which is a worry. We are therefore trying to use that data to improve things. I am interested in your views on that.

12:00

Jackie Brock: I hope that everything in our submission and in what we are saying today shows that Children in Scotland’s members and staff absolutely want to improve performance. It is good at the moment, but it must get better, and there are areas of some decline.

Information about qualifications and PISA is important. They begin to help with the fractions issue that I spoke about earlier and performance in mathematics and STEM. We could use information about where we are going wrong with some of the qualifications to unpick issues further down the chain so that we can say, for example, that we are not getting things right with regard to applying some basic concepts of numeracy to mathematical concepts at a later stage. We are not using the information that we already have.

We also have benchmarking within Scotland. A huge amount of money has gone into supporting schools to cluster with other schools that have similar socioeconomic characteristics. Therefore, we can consider why certain schools are performing better or worse than others that have similar characteristics, and that enables us to learn from those that are doing well.

With regard to primary schools, there has been a myth—in my view—that nothing can help us to compare schools in order to help us home in on poorly performing schools. However, there is plenty of information at local authority level because 31 of the 32 local authorities have bought into standardised assessment. I am sorry, but it is impossible for me to find credible the assertion that any local authority director of education does not know how well or badly their schools are doing, and, therefore, where they need to home in on with regard to support for certain schools—also down to year level—to do better.

A real issue that I do not think has been touched on sufficiently, if I may say so, is the question of what we, in Scotland, can do with the evidence to improve performance in relation to children. Again,

that is a legitimate concern of Government and the reason why it claimed, initially at least, to have introduced SNSA was that it wanted a tool to consider how to improve performance. That is legitimate. I disagree about the means, but there is plenty of information to suggest that what we need to be concerned about is the apparently inconsistent way in which we are improving performance across Scotland at the local level.

Johann Lamont: There has clearly been a shift in what the purpose of the assessment is. That started off as getting information across Scotland, but the assessment became a diagnostic thing. Which purpose would be the better one? Could SNSA testing fulfil either of them?

Sue Palmer: I wrote down the word “purpose” when Jackie Brock was speaking because the issue for us in terms of the primary 1 test is that we have the wrong purpose. The purpose for assessment at the early level should be children’s holistic development. The purpose of SNSA testing is to assess children against specific standards in literacy and numeracy. The two aims are at odds with each other. If development is being assessed, that is a holistic process that takes in things such as social competence, physical health and wellbeing, emotional maturity, language and cognitive development, and communication skills and general knowledge—it does not concern itself with specific literacy and numeracy skills. For me, as far as the early level is concerned, we have just got the wrong instrument—it is just not appropriate.

Johann Lamont: What would you say to the person who says that you cannot change what you do not know?

Sue Palmer: I would hope that we would be using that developmental information to help to improve educational outcomes, because we would know things about children’s development.

There are issues other than the background that you mentioned. We know that we need to provide a literary-rich environment, plenty of stories, lots of opportunities for songs and rhymes and so on, but there might also be issues with speech and language difficulty. If we pick that up, we can try to help with it. There can be issues with phonological awareness—perhaps children do not hear rhymes—that would mean that you might want to consider audiometric testing. Some children might need other physical check-ups, such as a visual check-up. That is the sort of thing that is regularly done in Germany when children are five—a physical and cognitive assessment that will help to ensure that the right sort of support for each individual child is put in place, if necessary.

Johann Lamont: I will play devil’s advocate once again. In my professional life, I have heard

the kind of characterisation that I could call the dismissive shrug: “They come from such and such a place, so we can’t expect any better.” However, my sense is that, in order to address inequality, we need rigour, and it could be argued that the standardised assessments offer the rigour that was not there before. How do we address that question for families, schools and teachers who are anxious about young people who are already disadvantaged when they come in the door? If we do not have rigour around understanding through assessment, how do we know whether those young people are being treated as seriously as children in other schools, that they are getting the same opportunities, and that there is the same kind of rigour around their learning rather than simply a low level of expectation, which is part of the characterisation around the debate? How do you respond to what is one of the most compelling arguments, which is that the choice is between rigour and treating every child with respect and therefore testing their understanding and ability versus something that is nice and warm but indefinable and can disadvantage some children?

Sue Palmer: If we were doing genuine developmental testing—which we are not doing at the moment—we would be applying the sort of rigour that is appropriate to that age group, and it would be very rigorous. In most of the world—including the whole of mainland Europe—children of that age would not even be at school, let alone being tested on the three Rs.

We have had a very early school starting age for 150 years and we have a cultural attachment to it, which means that we have assumed that children crack on with literacy and numeracy from P1. Some children will be fine with literacy and numeracy in P1, and we should support and encourage them, but some children will not have the foggiest, and they will need a different sort of support and encouragement—and, I hope, a very rich environment in which to make progress. That will ensure that there is a much more level playing field when specific instruction in schools begins.

It is in no way not rigorous to consider children’s development instead of saying, “Let’s just get on and aim at standards.” The point at which standards kick in is what is significant. International evidence on when other countries carry out standardised assessment shows that most countries do not carry out national standardised assessment before the age of 10. In Singapore, children do not start school until they are six. Previously, Singapore tested children at that age, but it has just abandoned that and it will not do any testing until after the age of eight, because it has realised that that changes the ethos of early years education in a way that is not productive for the children.

There are lots of different sorts of rigour. If you talk to specialists in early childhood education, you will understand that they are very rigorous indeed. However, that does not look the same as sitting down and doing the three Rs.

Johann Lamont: Children in Scotland is opposed to standardised testing at every level. We can see the argument in relation to the early years, but what is the argument against later testing?

Jackie Brock: Our response on the standardised assessment was in the context of how it had initially been proposed in the national improvement framework, which looked at ways to judge the performance of schools and local authorities and how that information would be used in relation to poorly performing systems. We were concerned because there is well-documented evidence about the distorting behaviours that come about as a result of such high-stakes testing.

We stress that we understand the purpose of assessment and the need to look at ways in which local systems, local authorities and schools work together to moderate performance and to make sure that the approach is robust. That is not about just sitting around having coffee and saying, “Oh, look at these results.”; it is about having a challenging approach to how we can demonstrate at the cluster level or, as I said, at the subject level—or whatever—that there is improvement. There is a problem in that teachers are finding that robust approach difficult. I do not know whether it is at the head or the subject specialist level where there may not be sufficiently robust professional development going on.

In one of your evidence sessions, you talked about the tests covering one tenth of the curriculum’s requirements on literacy and numeracy. We might revert simply to using that information. I suggest that that has the potential to distort all other efforts on literacy and numeracy.

Johann Lamont asked about the purpose of assessment. It is really important that I highlight what children and young people have said. In our work for the General Teaching Council for Scotland, we worked with 591 children and young people aged five to 18. In a moment, I will quote what a few of them said.

When you are reflecting on the purposes of education, it is really encouraging to reflect that the Scottish guidance on the assessment is for learning approach and “curriculum for excellence: building the curriculum 5: a framework for assessment” very much reflect what children and young people say that they want. Positive relationships are, of course, key to helping them to develop and learn. Specifically, children and

young people want to be able to focus on what they did well, what they did not do so well, and what the next steps are for their work. They want positive short-term learning goals and assessments that they can reflect on and discuss regularly one to one or in groups. They do not want assessments that are essentially memory tests; they do not feel that they are helpful to their learning, development and progress.

What do children and young people want? One young person said:

“If I make a mistake they explain what I did wrong and help me to understand for next time.”

Another said:

“They help us focus on what we do best and make us learn more about what we don’t know.”

I know that Johann Lamont has talked about children with additional support needs. Of course there is potentially greater variability for a whole range of children and a whole range of needs that may be additional and the extent to which some assessments can be modified, adapted and tailored for the individual needs of children, including those with additional support needs, those who are care experienced—I know that you have a significant interest in them—and those with particular health needs, including those with mental health conditions. They need a tailored approach. They need teacher judgment that is backed up by tests and assessments that can be modified and shaped to ensure that the teacher is getting it right in supporting the child’s learning and—this is critical—their progress on to the next levels.

We can make our report findings fully available to you. I make a plea that the voice of children and young people, which echoes national guidance, be reflected when you are reflecting and making recommendations on the purposes of assessment.

Johann Lamont: Thank you very much for that. I was going to ask a final question, but it has gone out of my head. Perhaps I can come back in when I remember it.

The Convener: Okay. Jackie Brock has talked about high-stakes testing. If I understood you correctly—I may not have picked you up correctly—31 out of the 32 authorities use Durham tests and cognitive ability tests. Why are they not considered to be high-stakes tests?

12:15

Jackie Brock: I do not know whether you are a parent. Did you know that those assessments were happening?

The Convener: No, I did not—but I do now, so the genie is out of the bag.

Jackie Brock: Indeed. That is an interesting expression. The genie is out of the bag. I understand the bureaucratic definition of high stakes, mid stakes and low stakes, but when the genie is out of the bag and parents have information that can help them to say where their child is, which the local press, councillors, ministers and the committee can also use, we have reached a high-stakes position, have we not?

Professor Hargreaves talked about this: if we are really clear about the purposes of assessment and about translating those purposes into the daily experience of children, which we can report to children and their parents and, in time, to the media, we can help to mitigate the impact of the high-stakes nature of the tests. I do not think that the discussion about SNSAs has been helpful so far, because the genie is out of the bag—or even the bottle.

I hope that the committee can dampen down some of the concerns about the authenticity of how standardised assessments will be used and how they will help teacher judgment. There is a long way to go before that will feel credible. If we have an honest conversation about how teacher judgments are being used to think about the progress of individual children and about how schools, local authorities and the Government are performing in terms of investing where they need to, that could lead to a healthier conversation, but I worry that, if we focus only on the results of SNSAs, we will lose a huge opportunity for us all to understand the importance of improving performance.

Having worked in the Scottish Government and seen the maelstrom of panic and concern that arises from the annual publication of data—frankly, the media and politicians all collude in distorting the really good work that is being done in schools—I feel that we need to be extremely cautious about the impact of high-stakes testing and assessment and how we use those results nationally.

Sue Palmer: That point about the genie being out of the bottle is particularly significant when it comes to P1, because the ratcheting up of parental anxiety impacts on the children. Within a year of the announcement that we would be testing primary 1 children, workbooks on how to help your child with P1 literacy and P1 numeracy had already appeared in the bookshops. As soon as people get wind of what is in the tablet-based tests, I dare say that there will be apps. That makes what is happening in P1 very high stakes, which is why something like a developmental checklist that the teacher goes through is much less distorting than a process that is linked to testing throughout the school system and which is

highly specific to particular literacy and numeracy skills.

The Convener: We have heard a lot of evidence about how helpful the testing that was done previously was, and some local authorities, such as East Renfrewshire Council and Fife Council, have reverted back to using that. Have we poisoned the water hole as regards what the perception of that testing will be in future?

Sue Palmer: I think that the Government has raised the whole question. I recently did a piece for *Sceptical Scot* in which I said that I hoped that the debate about P1 testing would start a national conversation about what is relevant at that early level and whether we should be thinking about getting on with the three Rs or whether we should be considering a different sort of approach. It could be that we have revealed that the water is poisoned.

The Convener: Johann Lamont has a quick supplementary.

Johann Lamont: There has been a lot of argument in the debate, some of it heated. The argument for SNSAs that gave me most pause was when it was said, probably both by the Government and in political debate, “If you had a child with special educational needs, you would want to know. The SNSAs are a means by which we can know that, and we would be putting young people at risk if we did not have rigorous assessment.” You can understand how compelling that argument is to anyone who previously thought that testing is not the best use of a teacher’s time. What is your response to that serious statement that the tests ensure that we identify young people with additional support needs early and can therefore meet those needs?

Sue Palmer: In many cases, we create some of the additional support needs by focusing on specific skills at a very early age. I worked with dyslexic children for a long time and, in many cases, it was clear when they came to me that it had started with an auditory or visual issue or something like that but, because they were being asked to do sound or symbol recognition that they could not do, there was an emotional overlay, which then grew. Then they felt the stigma of being in a remedial group; we do not call them that now, but they were in a special group doing special work—at a previous committee meeting, Sue Ellis spoke about the “walk of shame”. Children develop more problems as a result of being asked to perform tasks for which they are not developmentally ready. That creates the additional needs.

We need developmental checklists and assessment, both to inform policy and to direct funding to particular areas of need, and so that, by

becoming familiar with the sorts of things that developmental assessment covers, teachers' judgment about the children is better and, when they are worried about a child, they know who to refer them to for the best diagnostic tests. That is how it works in Finland, and there are far fewer children with special educational needs there; many of them are picked up through teacher judgment, proper diagnostic tests on the individual child and provision of a support package, so that, by the time the child starts school, the problem has been sorted out, rather than an emotional overlay being built on top of everything.

Iain Gray: Upstart Scotland's submission refers to the Australian national assessment program: literacy and numeracy—NAPLAN—and says that tests that were similarly labelled low stakes were introduced, but the information was then used in a high-stakes way and is now acknowledged to have had the "unintended consequences" of that kind of testing. Can Sue Palmer enlarge on that a little? Also, was that the fear that Jackie Brock was describing when she talked about information becoming available through FOI requests or otherwise?

Sue Palmer: The genie-out-of-the-bottle argument is very much at the back of that. Once national standardised testing is carried out, it is public knowledge and of great interest to the public. Parents become anxious, teachers are anxious to ensure that their classes get through the tests and schools worry about their results. The NAPLAN tests do not begin until year 3. However, as I said, the early development instrument is being used in Australia as well as Canada. Interestingly, its results correlated rather well with the year 3 results on NAPLAN, so a developmental check is good at predicting what will happen by year 3, as well as other stages.

Jackie Brock: To go back to the fractions argument and a couple of others, it is absolutely right that the public, the media and Parliament are engaged in a debate about how to improve teaching and learning in order to improve the outcomes for our children, and that can only lead to a deeper conversation. With the SSLN, rather than blaming and wagging our fingers at individual schools, teachers or children from a particular part of the country, we were saying that we had a systemic challenge, and then we could set out how to address it, along with a range of things that families and others could do to help us. We could have helped, in a very high-stakes way, to deepen our understanding of how to improve and to move the conversation on. The SSLN findings did not show that our teachers were rubbish and that our children were pretty rubbish, too, because they could not do sums—it was a systemic issue with the application of basic skills.

I have no problem with that discussion, as we would all benefit from a better-informed high-stakes discussion about how to improve Scotland's education. However, I want to resist the well-documented impacts on individual schools, neighbourhoods and types of children with particular needs as a result of league tables or some fancy way of presenting the information when children appear not to be performing well based on the SNSAs or the Durham University assessments, which, as we all know, are very narrow tools. I am not saying that they are necessarily the wrong tools, but basing high-stakes judgments on very narrow tools in isolation can lead only to distorting factors and poor consequences for our children's prospects.

The Convener: As members have no more questions, that concludes our session. I thank Sue Palmer and Jackie Brock for their evidence.

12:27

Meeting continued in private until 12:34.

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