SICI General Assembly 2019
Keynote speech | SICI strategy 2020–24
Amanda Spielman, Her Majesty's Chief Inspector, England

At this session of this year's SICI General Assembly, we thought it would be useful to discuss some of the challenges we face as inspectorates, and how we can address them.

Education and schooling have never been more prominent in the eyes of the public, and therefore policy, than they are today. Education is high up the policy agenda in many countries. Schools matter, and the public and policymakers know it.

This makes our work ever more important, but also means we face more scrutiny than ever before. I am going to talk in my introduction about seven areas that often come up in my discussions with other inspectorates. They are:

- Transparency and independence
- Inspection or improvement
- High or low stakes accountability
- The substance of education
- Meaningful inclusion
- The value of human judgement

Transparency and independence

Greater public scrutiny comes with the demand for transparency in our work. This is a good thing. As public servants, we have to be accountable, just as we hold schools to account. What we do has to be open to scrutiny. We are all of course accountable, though exactly who to and how much independence we have depends on our national context.

Even where we are more independent, we know that it is not helpful to school systems if inspectorates and education ministers pull in different directions. But as

inspectorates, we must maintain our independence, and be seen to do so. Otherwise we lose credibility and are seen merely as the enforcer of changing government policies.

Different inspectorates experience this tension between independence and alignment in different ways. Being part of SICI can help us draw on our common and different experiences to help us find the right balance.

Greater scrutiny also means that our role as inspectorates is more likely to be part of the political debate in our countries. And if others are thinking about our role in the education system, so should we. While we obviously never have full control over this, we do have acknowledged expertise.

It is valuable for us as inspectorates, and for SICI as our international association, to explore the role from our different perspectives. That gives us clear messages to take back to our policymakers and public.

Inspection or improvement

A second tension that exists in many inspection systems is between inspection and improvement. What is the role of the inspectorate? To judge and hold schools accountable for the quality of education they provide, or to be an agency for improvement, or a combination of both?

In England, government policy separates judgement and improvement work, but we do nonetheless need to make sure that our work is valuable to and valued by the schools we inspect, as well as by parents and government. Our guiding principle is therefore to be a force for improvement through intelligent, responsible and focused inspection and regulation.

We do that by concentrating on the quality of professional dialogue during inspection, and by providing feedback and reports that schools and other agencies can use for improvement work. Our inspections also prompt others to take action.

This approach keeps our work distinct from the work of improvement agencies, so as to minimise confusion and duplication.

Of course, other approaches exist and are equally valid. In some countries, inspection is seen as more of a school support service. In others, inspectorates are tasked with combining accountability and support. This of course relies on having

multi-skilled people as inspectors, since the knowledge and skills required for good inspection and good school support are necessarily the same. A support model probably also requires governments to make less high-stakes use of inspection than is the case in England.

But no model is perfect. In ours, which has a clear divide between inspection and improvement, we do at times run into the limitations of our ways of working. While most of the schools we judge to be less than good do improve, a small number have remained stuck at less than good since 2005, going through inspection after inspection without measurable improvement. These schools have been through many improvement initiatives, from the Department for Education, from local authorities and from other agencies, but none of it has worked, or not for long.

For these schools, we are looking at whether a more diagnostic approach is needed alongside our standard inspection model. We hope that this may provide a better foundation for well-directed improvement work.

Whatever model of inspection we each work with, we should not be held directly responsible for the improvement of the schools we inspect. This idea can often confuse people: the National Audit Office here did not much like the fact that we do not attempt to attribute improvements in schools to our actions. Aren't we supposed to be a force for improvement in the education system?

Yes, of course, but for an inspectorate to be accountable for improvement in the schools would introduce a perverse incentive. We would essentially be marking our own homework. The temptation would be to make sure that schools found to be falling short at one inspection did indeed improve at the next. And it wouldn't be hard to achieve.

High- or low-stakes accountability

One area in which we see big differences between SICI members is the nature of accountability in our national systems. In England, inspection can have consequences for schools. We grade schools on a four-point scale and our reports are published. Parents use them to inform school choice. As schools are essentially funded per pupil, parental choice affects the school's income. Then government uses our grades to decide when to intervene, with additional support or by moving the worst-performing schools from local government control to an academy trust.

We are only one of two main pillars of accountability in England, with national assessments at the end of primary and secondary school forming the other main component. School-level exam results are published, and local and national press frequently turns them into so-called 'league tables'.

High-stakes accountability and high-stakes inspection are always controversial. Although parents hold Ofsted and our judgements in high esteem, headteachers are more critical, and complain of the pressure this model applies. The evidence about high- and low-stakes models of inspection is quite limited, although what international research there is tends to suggest that greater 'accountability pressure' leads to both greater positive impact on pupil outcomes and more negative unintended consequences such as teacher workload.

It is probably a decision to be made in the light of national policy priorities. But it might be worth doing more detailed study of this trade-off, and might be a fruitful area of work for SICI?

The substance of education

We live in an era of big data, in which algorithmic and data-driven models affect more and more parts of our lives and work. In education, data has of course become a strong driver in recent decades. None of us ignores PISA, for example. PISA results have driven education policy in a number of countries.

German colleagues will certainly remember the so-called 'Pisa-Shock' after the publication of the 2001 PISA results, where the reported performance of the German education system was well below what had been expected by policymakers and the public. Many states in Europe and elsewhere have reacted to PISA with educational reforms.

In national systems too, many countries are considering increasing the use of pupil assessment data as a primary means of assessing education systems and providers. The new Flemish government, for example, is introducing national tests in languages and mathematics for the first time.

This leads us to a discussion about the extent to which inspection should be test- and data-driven. In our own system, a move began 15 years ago towards a greater reliance on data on inspection. Inspection frameworks came to rely more and more heavily on pupil data, especially the progress pupils made over time. This of course

has some value. Attainment is important, and we want pupils to progress successfully from one stage of education to the next.

There were, however, a number of major problems with this approach.

Because national assessment only takes place at the end of the second and final years of primary education and after the fifth year of secondary school, schools make up for the lack of data in other years with internal testing and data collation.

In many schools, teachers were asked to generate large amounts of progress data on their pupils every couple of months. This caused a big spike in workload, with little or no gain for pupil learning. This data was also quite problematic, as progress models typically assumed every pupil should make linear progress in a subject over time, which of course is not necessarily the case. The reliability and validity of internal tests is also often doubtful, for many reasons.

This dubious data was then presented to and scrutinised by inspectors. And – what a surprise – the internal data inevitably showed better pupil progress than the national test data.

A whole data generation industry developed. More schools started teaching narrowly to tests, and for longer, and slicing untested areas out of their curriculum. Pupils, especially the most disadvantaged, were often directed into low-level qualifications in which it was easier for them to obtain high grades. This in turn inflated school attainment profiles.

Another problem was that in primary school, national assessment only covers English and mathematics, so most schools taught these separately in the morning and lumped all the rest into cross-curricular topic work in the afternoons. Secondary curriculums often became impoverished, because schools stuck tightly to the final exam syllabus, frequently simply repeating topics year by year with scant regard for the sequence of content or the need to provide building the blocks of knowledge and skills that pupils need to progress.

English schools have unusually wide freedom in choosing curriculum and planning teaching, and haven't always used it well. I have been honest about this publicly: the curriculum and quality of education was suffering for many children.

That is why we decided to develop our new framework, putting far less emphasis on data, but instead looking above all at the quality of the curriculum that schools adopt

or develop and teach. This has been widely welcomed by teachers, school leaders and parents, who largely agree that we are now looking at what matters. You could say that we have restored the second pillar of our accountability system.

This is an important message for us as inspectorates. What children learn matters, and we should not lose sight of this, even in more data-driven systems.

Meaningful inclusion

This leads us to one of the major areas of discussion in many of our education systems, which is of course the inclusion of children with special educational needs and disabilities. Of course, we want our schools to be inclusive. But that does not mean just placing all children in mainstream schools. Sometimes we see these children getting a substandard education in classes or groups taught by the least qualified staff, while better teachers teach the others. That is not real inclusion.

As far possible, all children should have a broad and rich curriculum that does not underestimate their ability to learn. And we should not fall into the trap of attributing all learning difficulties to the child. Mild and even some more serious learning difficulties can sometimes simply be the result of a low-grade curriculum and weak teaching early on. In particular, many reading difficulties can be prevented or addressed by better teaching that draws on the rich evidence base on reading instruction.

But at the same time, there are children for whom specialist provision really is the most suitable option, for a variety of reasons. We must be able to have honest conversations about these children. A dogmatic approach does not help them: indeed, it can sometimes force them to wait far too long before getting the help they really need.

So there is a real challenge for inspectorates: how to encourage schools to be inclusive, without accidentally creating harmful consequences.

The value of human judgement

Of course, one criticism that has been made of our new framework is that it must be less reliable, because it is harder to assess quality of education than it is to make judgements based on data. And in some ways that is true. Assessing the quality of curriculum, or teaching, or leadership inevitably relies more on human judgement than looking at a set of numbers. This may be seen as reducing reliability. But to

take that view is problematic for several reasons. Critics of the framework overestimate the reliability and validity of testing. But, more importantly, they underestimate the value of expert human judgement.

Okomentoval(a): [EB1]: We haven't mentioned critics, just criticism.

Okomentoval(a): [EW2R1]: Yes agree with the change

Quality of education is not a simple, one-dimensional construct. To attempt to reduce it to this, for example by equating it too closely to outcomes on national tests, means losing sight of the substance of education and so reducing the validity of our judgements.

Of course, human judgement is not perfectly reliable, but then neither are tests! In fact, human judgement is the only way we can reach a view on true quality. Our inspectors are our greatest resource. Their expertise is what makes inspection unique and a large part of what makes our methodologies valid.

This means that, as inspectorates, we must work to get the best possible people to inspect for us. Recruitment and retention really matter, although they are a challenge for many of us. Of course, we want our human judgements to be as valid and reliable as possible. At Ofsted, there are a number of ways we do this.

The first is having valid and reliable frameworks, with clear inspection handbooks to guide inspectors on how to apply them. We have already heard from Daniel about how we now develop our frameworks.

The second is to give thorough, high-quality training to all our inspectors. This is a rigorous programme that teaches inspectors the theories behind the framework and the research that underpins it, as well as the practice and methodology of conducting inspections. We refresh this training regularly and introduce new training whenever we find something new that needs addressing.

The third element is quality assurance. We have a rigorous QA system for all our inspections, drawing on our most senior inspectors, through which particularly difficult inspections can be escalated.

The final piece in the puzzle is evaluation. We evaluate all our frameworks very thoroughly, both for the quality of implementation and for their impact on schools.

One thing we don't do in our inspections, although a lot of you do, is use indicators and scales to come to our judgements, such as example rating scales of teacher behaviours or of quality in the classroom. While we appreciate that these can have high levels of reliability, they are problematic in a high-stakes system like ours,

because any scale used would quickly become the default performance management system for schools across the country. These are the types of unintended consequences that we have to watch for as we develop our frameworks and methodologies.

Rating scales may also reduce the discretion of inspectors to explore issues they encounter on inspection, which in our view is an important part of creating valid systems. But of course we always want to learn more about your inspection models and methodologies. We are always looking to improve our practices, and being part of SICI helps us do that.

And we do have some evidence that our inspections are reliable. In a parallel inspection study of 24 primary inspections, we found that two inspectors reached the same judgement in 23 out of 24 cases. We continue to monitor validity and reliability of what we do through our evaluation work.

This focus on the importance of human judgement is not to say that data isn't useful. Under our new framework, we still look at the pupil outcomes in national exams. Without these qualifications, the routes to further study are closed off. A curriculum that leads to poor attainment cannot be effective.

Like many of you, we also use data for our risk-assessed inspection scheduling. Indeed, we use an artificial intelligence algorithm to help us find schools that were last rated as good or better but may be declining in effectiveness. While in the past we used pre-determined thresholds on performance data, we now use 'supervised machine learning' for the first stage of decision-making, before applying the expertise of our regional inspector teams to make the final calls. Using technology such as AI has great potential to inform expert judgement, and we hope that SICI can play a role in exploring the uses of such technology for inspection.

So we try to find the right balance between inspection validity and reliability. And of course, there is always room from improvement, and we continue to learn from other SICI members as we grapple with similar issues and concerns.

Learning from others

At this forum, we are meeting as school inspectorates. But this is only one part of our work. We also inspect early years providers, post-16 education, adult education and children's social care, as well as initial teacher training. This breadth has proven to be extremely valuable, because there is so much we can learn from other sectors.

Our new deep-dive methodology draws in part on the experience of looking at topics in depth during our inspections of children's social care providers. Working with inspectors who look at early years and social care also helps us to understand how these sectors interact with what happens in schools, and so deepens the knowledge and skills of our school inspectors.

We also work with other inspectorates. On our inspections of provision for special educational needs and disabilities, we collaborate with the Care Quality Commission, which inspects healthcare in England. Again, this allows us to learn and bring a richer perspective to what we do. There is so much we can learn and exchange with our colleagues, at home and internationally. The joint projects we have been part of have enriched us all.

To meet all these various challenges, be they the call for transparency, the tension between reliability and validity, making sure we look at what really matters, and choosing the appropriate inspection model, we need to make sure we are evidence-informed.

This is particularly important in education, because unfortunately we can be taken in by fads masquerading as research and by the slick salesmen peddling them. Too often, people in education embrace the latest shiny new idea or technology without any real idea of whether there is any evidence to support it. Sometimes, they almost intentionally ignore past experience that suggests that the shiny new thing isn't likely to have much value. This is something we can and should guard against as inspectorates. For example, in the 18th century, a medical inspector might well have praised the use of a popular cure for headaches – bloodletting, using leeches[1]. In medicine we take a rather more evidence-based approach these days. We should make sure we do the same in education, and don't just end up promoting the latest fashions. Evidence underlies what we do here in developing and evaluating our frameworks. I believe this reinforces our credibility. It is something we have put a lot of effort into in recent years, and we are continuing to expand the research programme. And we'd very much like to work internationally in this area where it's appropriate.

And of course, the challenges I have outlined are only a snapshot of the many different issues we see in the system. I am sure you will all have your own perspective. So let us use the next hour productively. Let us discuss the challenges we all face as inspectorates and, especially, let us discuss the solutions we have found.

---

[1] Leech = une sangsue (French) or ein Blutegel (German)