



SUMMARY REPORT SICI WORKSHOP 5-6 MARCH UTRECHT

How schools are judged is an important part of the work of inspectors. The assessment can have a big impact on school boards, school leaders, teachers as well as on students. The aspect of "judgement" is also one of the most challenging parts of the work of inspectors, because the assessment that they carry out must meet high quality standards. It should always be reliable and fair so that all schools are treated equally. In this SICI Workshop we dived deeply into this matter.



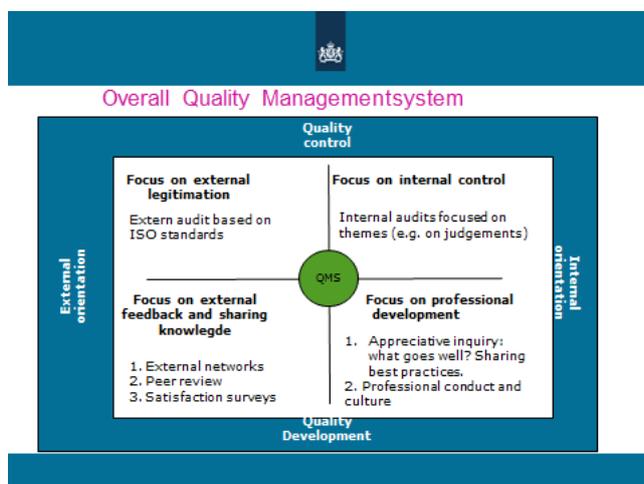
This report summarizes each part of the programme.

Index

1. Presentation Internal quality assurance on judgements by the Dutch inspectorate	3
2. Lost in control	3
3. Workshops.....	4
a. Basque Country	4
b. Ireland.....	4
c. Lower Saxony	5
d. Scotland	6
e. Sweden.....	6
4. Interview with governors and school principals.....	7
5. Conclusions of the workshops	7
6. Discussing statements.....	7
7. Dutch inspectorate evaluating judgements.....	8
8. Ofsted's research into inspector's reliability.....	9
9. Professional judgement: A behavioral problem for regulatory governance.....	10

1. Presentation Internal quality assurance on judgements by the Dutch inspectorate-Rutger Meijer, Herman Franssen

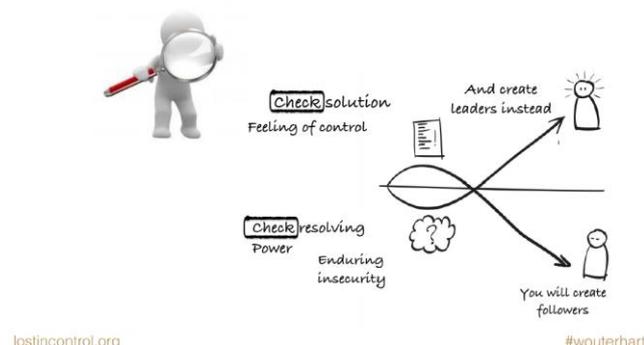
Rutger explains the new Dutch framework and the reasons for this innovation. The Dutch inspectorate now focusses on school boards and encourages improvement of educational quality at all schools. Besides heavy supervision on schools that are inadequate or very weak, the Dutch encourage adequate schools to become better, by challenging them to choose own ambitions and grading them as 'Good' or 'Excellent'. The current framework has much less descriptive, more holistic standards than the former one, which puts more weight on the expertise and professionalism of inspectors. Because the criteria for the judgements inadequate, adequate and good (per standard and for the school as a whole) are less detailed, the internal quality assurance on judgements shifted more to training, case studies and consultations. Besides that, the Dutch have internal and external audits, process descriptions and instruments and proof reading of reports. In this phase in time the Dutch want to invest more in professional culture with consultations, dialogues and reflection, and more teamwork. [Link presentation](#)



2. Lost in control by Wouter Hart

Did hierarchy take over the spirit of leadership? And bureaucracy the art of professionalism? Organizations are [Lost in Control](#), when serving the system has become more important than serving its original purpose. When protocols, policy and KPI's took over customer focus, ownership, and common sense.

Key note presenter and author of Lost shows us that if we want our schools to take responsibility and find their own solutions to improve the quality, we should give the problems and not the solutions. Rules and strictly defined ways to improve may cause a culture of irresponsibility and followers instead of leaders. So, we must at least think about what we write down in our framework, reports and what we give as feedback. Forms and rules are needed to serve, but are not a goal in itself. Giving control means creating leaders. [Link presentation](#)

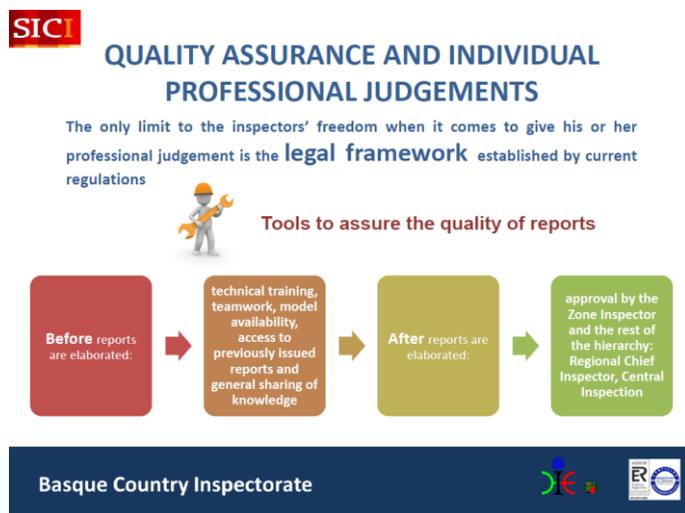


3. Workshops

The elements of internal quality assurance that were named in almost all workshops are set out in presentation 5 (Conclusions of the workshops by Bert Lichtenberg). This only summarizes the exceptional elements that were presented.

a. Basque Country

Basque Country does not grade or judge schools but has a well worked out quality management system, with standardized processes and kpi's (key performance indicators). The process and result indicators lead to corrective improvement actions. The Basques have worked out all of their processes (about 40) and each process has an evaluation team. The focus lies very much on the content of the reports, as there are no gradings. The Basques have a platform with all the reports in which inspectors can search on subject, school type, etc etc, and publication of good practices. The (regional) zone manager has to give each report his/her seal of approval. [Link presentation](#)

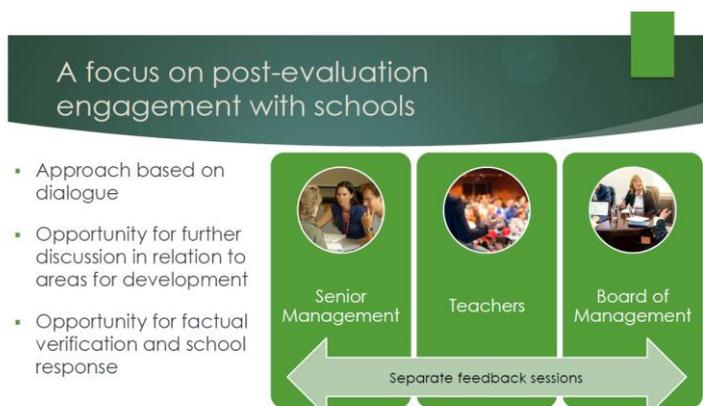


b. Ireland

In Ireland quality assurance is meant to be part of the culture of the organization. The aim is to create continual reflection and professional dialogue. At the moment, the Irish inspection uses a quality assurance framework which consists of professional practice standards or benchmarks, administrated inspection processes and procedures and tools for measuring inspection impact. This is the first time a quality framework is used in their system and schools really appreciate the framework.

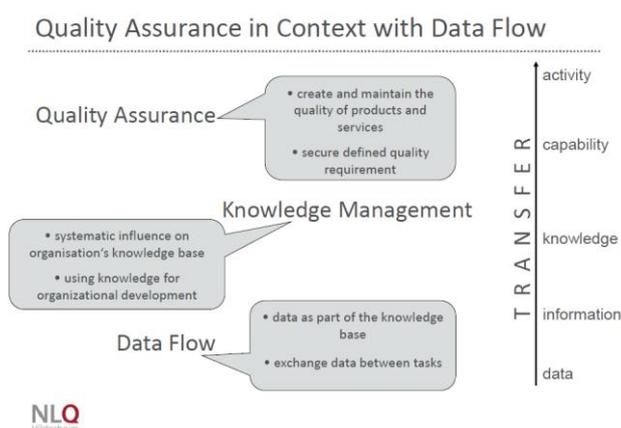
The colleagues of Ireland state that it is difficult to ensure that inspectors make 100% reliable, fair and effective judgements but they do try to do so with a number of mechanisms. A notable example is the post-evaluation engagement with schools. This approach is based on dialogue and consists separate feedback sessions at the school (senior management, teachers, board of management). It provides the opportunity for further discussion in relation to areas for development. The inspectors try to address all the comments and questions during that day.

The colleagues of Ireland note that the development of quality assurance is an ongoing process. Their quality framework is not a finalized product. They are open for feedback. And they present plans for the future. A compliance check by reviewing a sample of inspection files is one of the suggestions for further improvement. [Link presentation](#)



c. Lower Saxony

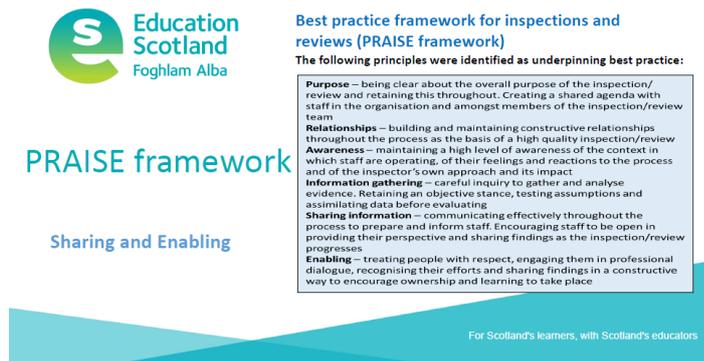
In Lower Saxony all instruments are developed by inspectors in cooperation with scientific department of NLQ, external scientists, teachers and school boards. They pilot the instruments in schools before adaption. They also do frequent video trainings and an inspector has to do one calibration lesson per inspection. The self-evaluations are analysed and compared to the findings of the inspectors. The inspectorate works with a modular approach, offering schools different kinds of modules to choose from. They also find accordance between inspectors important, which is assessed per item. After every step of the inspection there is an evaluation moment. Parents, school teachers and school management participate in that evaluation. [Link presentation](#)



d. Scotland

In Scotland, the frameworks used by HM Inspectors during school inspections were developed in collaboration with external stakeholders. These are used by schools, centres and local authorities as tools to support self-evaluation (How good is our school?). In this way, there is a clear understanding, widespread use and strong ownership of the frameworks. The framework has detailed descriptions of indicators on a six point scale.

Besides the framework the Scottish Inspectorate works with a code of conduct called PRAISE. The inspectors pay a lot of attention to evidence gathering during the inspection process. They have continuous dialogue in the inspection teams about the gathered information. The gathered information is also shared with the school during the inspection. [Link presentation](#)



Education Scotland
Foghlam Alba

PRAISE framework

Sharing and Enabling

Best practice framework for inspections and reviews (PRAISE framework)

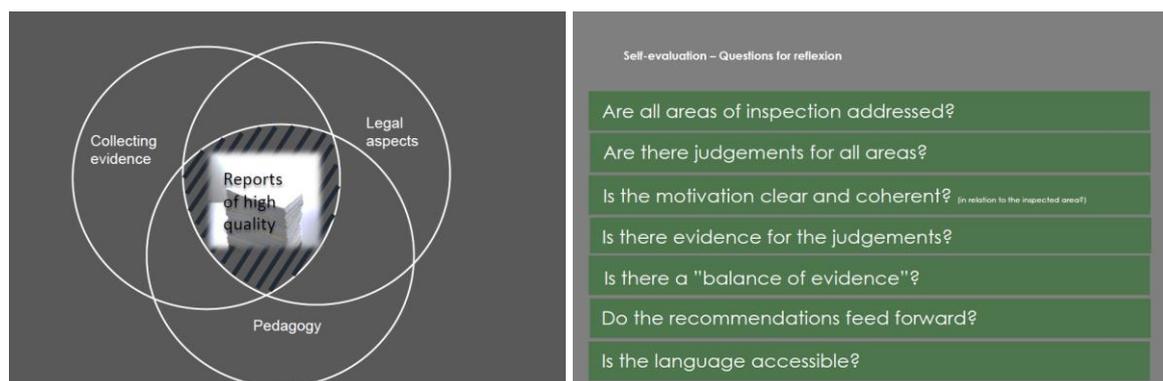
The following principles were identified as underpinning best practice:

- Purpose** – being clear about the overall purpose of the inspection/ review and retaining this throughout. Creating a shared agenda with staff in the organisation and amongst members of the inspection/ review team
- Relationships** – building and maintaining constructive relationships throughout the process as the basis of a high quality inspection/ review
- Awareness** – maintaining a high level of awareness of the context in which staff are operating, of their feelings and reactions to the process and of the inspector's own approach and its impact
- Information gathering** – careful inquiry to gather and analyse evidence. Retaining an objective stance, testing assumptions and assimilating data before evaluating
- Sharing information** – communicating effectively throughout the process to prepare and inform staff. Encouraging staff to be open in providing their perspective and sharing findings as the inspection/ review progresses
- Enabling** – treating people with respect, engaging them in professional dialogue, recognising their efforts and sharing findings in a constructive way to encourage ownership and learning to take place

For Scotland's learners, with Scotland's educators

e. Sweden

The Swedish put a lot of effort into evidence gathering. Inspectors make objective and factual memos with evidence which is sent to the school for approval before writing the reports and making the judgements. Inspectors are trained not to jump to conclusions, but to stay objective during observations and discussions. Only if there is enough balanced evidence inspectors can start with judgements and conclusions. There are three pillars that reflect on an inspection: collecting evidence, pedagogy and law. There is also a lot of comparison between regions and their judgements. [Link presentation](#)



Collecting evidence

Legal aspects

Reports of high quality

Pedagogy

Self-evaluation – Questions for reflexion

- Are all areas of inspection addressed?
- Are there judgements for all areas?
- Is the motivation clear and coherent? (in relation to the inspected area?)
- Is there evidence for the judgements?
- Is there a "balance of evidence"?
- Do the recommendations feed forward?
- Is the language accessible?

4. Interview with governors and school principals

Governors and principals were mostly positive about the way the Dutch inspectors carry out the inspections nowadays. They said they could openly discuss their challenges in a constructive way. The governors are pleased that they have to account for their responsibility for the educational quality of their schools, and both principals and governors feel it's only natural to be held accountable for what they are doing with public money. They also appreciated the joint classroom observation (learning on the spot). They did experience our descriptive tailor made inspection plan as 'low trust' and see our new framework as a step in the right direction. One governor says it would be better to focus even more on the processes (like quality assurance) of the schools, and less on the results. Another principal didn't agree with the short period of time we gave for improving. He feels he needs to take a shortcut, while development needs more time. [Link presentation](#)

5. Conclusions of the workshops by Bert Lichtenberg

Bert first states that there are cultural and political and differences in system of education and the position of the inspectorates. Therefore copying another method/system is difficult, it's about learning and using parts that are desirable and possible.

The elements of quality assurance we all seem to find important are a good training programme for new inspectors (variation 5 month – 2 years), structurally developing knowledge and skills of inspectors, teamwork and cooperation (always 2 inspectors, consideration, intervision, peer-to-peer), evidence gathering and triangulation and feedback of external stakeholders. Inspectorates also value transparency of our framework and processes, before, during and after the inspection, reflection/research on our data and ISO certification.

The underlying mission for all the effort put into internal quality assurance is that we should have impact on the quality of education. [Link presentation](#)

QUALITY ASSURANCE

- ◉ It is not only Plan-Do-Check-Act.
- ◉ Impact inspectorate on and for others means: Quality Assurance for process, judgements, reports and conduct.
- ◉ But on what level?
 - Different inspectors.
 - Different school situations and development.
 - Different context.
- ◉ Most important outcome: we must have impact on the quality of education.

6. Discussing statements

Wibo presented four statements. We discussed these statements in groups of around six participants and then voted with green and red cards. Some participants were asked to explain their votes to the whole group.

The statements that were presented led to some interesting discussions. A large part of the participants agreed that our judgements should be reliable and according to the rules. Maximizing impact should be reached by other means than being flexible in judgements, like the way inspectors communicate our feedback (both in words as on paper). One participant says that if

inspectors have to judge schools inadequate too often, we should reconsider the criteria, rather than judging too softly.

A large part of the participants agree on giving and publishing judgements/gradings as they are an important means to have impact, although some disagree. Although schools can sometimes be less open about the challenges they meet, knowing that we will grade them at the end of the day, many participants find it more important that schools are held accountable.

Cases are known when negative judgements caused schools to further deteriorate (because parents and teachers might choose better schools), but some participants react that this happens only on the long term. If the school recovers quickly (within one year) and the grading is then changed back to adequate, not much harm will be done, and schools usually improve significantly.

Some inspectorates do not give an overall judgement, only judgements on certain standards/quality areas, or no judgements at all, partly because of this potential effect. [Link presentation](#)



7. Dutch inspectorate evaluating judgements-Aster Dijkgraaf

The ways the Dutch measure the inter rater reliability and the effect/impact of judgements are:

- Satisfaction surveys
- Teacher, school leader, board panels
- Effectiveness study
- Quantitative study of given judgements
- Fairness and reliability studies

With the quantitative study of given judgements, the Dutch look into the discrimination of the standards, patterns in judgements, regional differences and inspection team differences.

The core of fairness is that judgements are not affected by which inspection team visits the school or board, and that schools or boards with the same quality are given the same judgements. The Fairness and reliability study of inspector's judgements, looks into where we set the pass mark, which underlying criteria within a standard are easier and which are harder to judge, and what the considerations are for a judgement. Normally the inspector assesses a standard as a whole, but in the fairness studies we ask the inspectors to also assess underlying criteria and provide further motivations for the assessment. This leads to clarifying clarification (instruments) or improvement of the standards, further training, etc.

For example, this study showed which motivation inspectors used to pass an inadequate criterion within a standard (when giving a positive judgement):

1. The criterion is not important enough to judge the standard as insufficient
2. Other criteria compensated enough
3. The difficult circumstances of the board or the school justify deviating
4. It is still insufficient but the improvement has started

The Dutch inspectorate agreed that the first two motivations were valid, but that the latter two weren't.

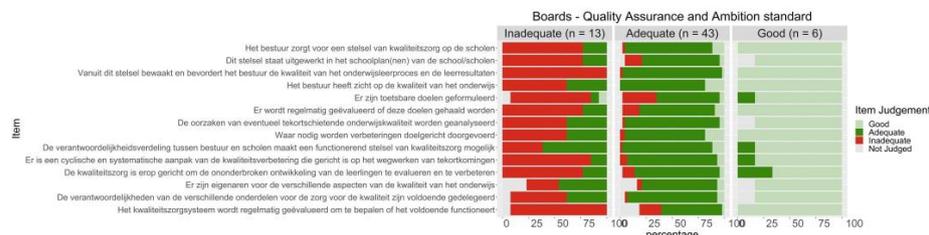
The Dutch research department gives the following guidelines for high quality of judgement:

- Discussion of judgement before finalizing
- Consensus is reached about any differences in interpretation of the standards
- Rotation within teams of inspectors
- Inspectors regularly join another team
- Regular training including case discussion
- Regular evaluation studies & sharing results with inspectors
- Management cares for inter rater reliability and reports on the status
- Reviewers of reports
- Assessment committee for checking reports and judgement of very weak schools

The fairness study also identifies criteria that are more difficult to achieve for schools and boards than other criteria. This provides information for the inter-rater consultation of inspectors. [Link presentation](#)



Analysis of response patterns



8. Ofsted's research into inspector's reliability-Alan Passingham

Ofsted has done several studies on reliability of judgements. Alan explains:

1. Reliability of short inspections (2017) – focus too broad?
2. Lesson observation (2019) – focus too narrow?
3. College reliability study (2020) – the right balance? (still to be started)

For the first study two inspectors inspected the same school at the same time without any form of cooperation/communication. Of course only one inspector could lead the discussions. Of the 24 valid inspections 92% of inspectors arrived at the same judgements, and 100% same overview of the school as a whole. Because of the complexity, small sample size and heavy burden, Alan does not recommend other inspectorates this type of study.

Behaviour indicators were scored more strongly than those for teaching and curriculum



The second study focused on how valid and reliable the use of lesson observation is in supporting judgements on the quality of education. Paired inspectors observed the same lesson to assess indicators spread over three domains: curriculum, teaching and behaviour. This study concluded that the reliability in primary schools was higher than in secondary education and much higher than in colleges. The observation length (15 or 30 minutes) didn't seem to make a difference. Lesson observations are only a part of the judgements that are given (triangulation). This is why Ofsted calls this study a narrow one.

The last study focusses on whether groups of college inspectors reach reliable judgements about the quality of education when using the deep dive method of inspection. This methods involves discussion on the intent of the curriculum with senior leaders, followed by activities that look at curriculum implementation and triangulation of the evidence collected results in a quality of education judgement. The two stage approach which will provide qualitative and quantitative evidence. This study has still to be carried out.

Some drivers for reliability suggested by Ofsted because of these studies are:

- Structure (ensuring that the framework and handbook are valid)
- Triangulation of evidence from different activities essential
- Inspector collaboration
- Inspector training
- Post-visit quality assurance procedures

However

- There remains a small trade-off between validity and reliability
- 'Perfect reliability' difficult to attain for complex, real-world judgements
- Subjective judgements vs test score validity

[Link presentation](#)

9. Professional judgement: A behavioral problem for regulatory governance-Thomas Schillemans

Thomas shows us how important it is to be aware of our human processes when giving judgements. Our mood, how much we like or trust a person or organisation, previous inspections, prejudices, etc can affect our judgements. Professionals can also become over-confident and human minds are 'lazy machines for jumping to conclusions'. Psychological experiments show that our judgements are feeble, but that accountability does significantly improve judgements.

Good judgements are made when we are in the reflective mode; which means that we are controlled, using full effort, taking time (slow) and being conscious. In contrary to the automatic mode in which we are more uncontrolled, effortless, reacting on emotions, fast and unconscious.

He shows us a few things that can be done to make sure we judge the right way:

- Checks and balances
- Professional norms, training, teams and routines
- We know 'repeat players' are better ... but also over-confident
- Timely and targeted feedback
- Power of anticipation
- Confidence + self-reflection
- "Respect for the game"

[Link presentation](#)

SO?

Regulatory professionals all have impossible – inhuman - jobs