# How to define and measure output/outcomes of schools?

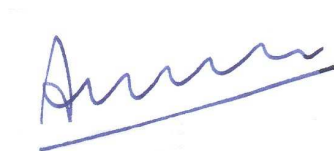Using student achievement data for school inspections

Strategic report
on the SICI Conference in Amsterdam,
May 18 - 19, 2010

Utrecht, 24 January 2010

## Preface

It is our pleasure to present to you the strategic report on the May 2010 SICI Conference in Amsterdam, The Netherlands. The Standing International Conference of Inspectorates (SICI) is the association of European Inspectorates of Education, established in 1996. The impressive rate of participation in the conference may largely be attributed to the subject of this conference: 'How to define and measure output / outcome of schools? - Using student achievement data for school inspections.' Clearly, these questions are very much alive for inspectorates of education all over Europe, even though circumstances obviously vary greatly from one country to the other. Notwithstanding these differences, major common insights emerged form the conference. First and foremost, the conference concluded that calculating school output based on student achievement is difficult, but not impossible. In addition, common agreement was reached on a definition of the use of output data by inspectorates. In my opinion, these results are a very good starting point for further fruitful and mutually supportive exchange of views and approaches between SICI-members concerning these issues.

The conference was a joint effort of the SICI Inspection Academy (SIA) and the Netherlands' Inspectorate of Education. We very much appreciated the vivid interest in the conference and the constructive exchanges that took place in Amsterdam. This strategic report on the major outcomes aims to support SICI members and other interested parties in developing and implementing an inspection system which uses the nationally available data in the best way possible. In addition, SICI hopes this report will help to motivate individual countries to increase the quality and quantity of student achievement data available to inspectorates of education.

Annette Roeters, Senior Chief Inspector of the Netherlands Inspectorate of Education

Gonnie van Amelsvoort
Mineke Laman
Bruno Vreeburg
Inge de Wolf

Utrecht, the Netherlands, January 2011

# TABLE OF CONTENTS

# Summary

The interest among SICI members in the subject of using student achievement data for school inspections turned out to be vivid. The participation rate was high: 85 participants from 23 SICI-countries were present and representatives from ten countries presented the approach of their inspectorates. In the workshop-discussions, there were frequent references to the input by the keynote speakers, indicating that their contributions were helpful.

Overall conclusion of this thematic conference on the use of student achievement data in school assessments was that calculating school output based on student achievement is difficult, but not impossible. Agreement was also reached on a general definition for the use of output data by inspectorates: 'Inspectorates strive for reliable and fair output indicators of all types of results of all students: a measure which is practical as well as recognizable for schools.'

Participants agreed that inspectorates should aim to implement fair and reliable assessments of schools. Output of schools was seen as one of the main quality areas making up school assessment, amongst other important areas such as school climate, education processes and management. Student achievement was considered a crucial element of school output, implying that schools should incorporate student achievement data in their quality assurance systems. In addition, the importance of effective communication was stressed: inspectorates should communicate clearly about results with all the stakeholders.

Some inspectorates use the 'value added' approach, which reflects the schools' contribution to student results. This approach was appreciated, but it is not always suitable / feasible for other countries, since it requires the availability of quite a lot of reliable data and research capacity. What's more, these issues are obviously not purely technical: phenomena like central exams, national tests and related centralised data-collection are important political issues.

As to the use of data about outcomes (longer term results such as labour market performance, contribution to social cohesion etc) the general feeling seemed to be that it would be useful to include indicators of such measures in quality assessment. In fact, the discussions built on the insights that emerged from the 2010 Copenhagen conference on social cohesion in the context of citizenship. However, it was clear that there is still a long way to go in developing suitable approaches towards this goal.

***The way ahead***
The conference identified quite a few areas for further enquiry, research and development. Consequently, participants saw many possibilities for further cooperation and mutual support. SICI/SIA could support this process, not only by publishing the strategic report on its website, but also for example by starting a digital community of practice on the subject of defining and measuring output and outcome of schools. Overall objective is for inspectorates to use data on student achievement and output and outcome of schools as an essential component of fair and reliable assessments of schools. Judgements of schools should not be based on the coincidental composition of their student population but on their contribution to student achievement.

# 1 Introduction

## 1.1 Objectives of the conference

The overall purpose of the conference was to exchange knowledge and experiences in using data of student achievements (attainment, progress, social performance) for school inspections. Presentations and discussions focused on two key questions:

a)      Which information on student achievement is available for school inspections?
b)      How can this information be used for inspections?

Efforts were made to answer these questions for countries with standardized information available on student achievement data (test scores, exams, drop out rates), as well as for countries where these student data have to be collected and assessed during an inspection visit.

This strategic report aims to summarise the main outcomes of the conference, thus supporting inspectorates in developing and implementing school supervision systems which take due account of student achievement data. In addition, the report should support national and political discussions about data-collection on student achievement.

## 1.2 Programme and participation

The programme of the conference included various lectures by researchers, presentations by staff of inspectorates and workshops on social performance, strategic behaviour and fair comparisons. The full programme is attached to this report (annex 1).
All but six out of 29 SICI-members were represented at the conference. In addition, there were some representatives of inspectorates which are not SICI members (Lower Saxony, Berlin Brandenburg and Serbia). Thus, European coverage was substantial. Ten members presented current practices in their own countries with respect to the measurement of school output / outcomes and the use of student achievement data for inspections. These members were Flanders, England, Wales, North Rhine-Westphalia, Romania, The Netherlands, Scotland, Berlin-Brandenburg, Czech Republic, Sweden and France.

## 1.3 Terminology and definitions

Terminology and interpretations related to output and outcomes of schools are not unambiguous. Therefore, some definitions were proposed to the conference and these will also be adhered to in this report. The current section specifies some of the essential terminology.

*Student performance indicators*

There is a wide range of student performance indicators. These may be used to calculate school output or school outcomes. Important distinctions to be made, are:

  **output:** short term results in terms of
  - o  effectiveness, such as:
    - ▪ achievement of students and pupils;
    - ▪ results or the achievement of the school: the gross and net school effect and the value added as an estimation of the net school effect.
  - o  efficiency results, such as:
    - ▪ number of students who reach the objectives;
    - ▪ attendance rates and drop-out;
    - ▪ the time it takes to obtain results.
  **outcome:** longer term results such as success on the labour market.

If we look at these terms in detail, the following specifications may be given.

*Output in terms of effectiveness*

**a. Student level: achievement of students**

'Achievement' refers to the performance and the results of the students in general. There are several domains of achievement:

- Cognitive attainment like examination results, results on (standardized) tests, grades and certificates of qualifications. Cognitive attainments are very often results with reference to national standards.
- Social performance of students:
  - o Social behaviour, social results or social output. Student performance in social areas is another component of student achievement. How do students contribute to the school community and to society at large?. The results of citizenship education are social results.
  - o Healthy behaviour: do students adopt a healthy lifestyle?

**b. School-level: gross and net school effect and value added**

It is possible to calculate the results or the achievement of the school from the examination-results of students on standardized tests. This may be done in various ways, for example by calculating the mean examination results, or the percentage of certificates acquired. These are indications of the gross school effect.

Value added (or the school net effect) may be used as a better way to estimate the contribution of the school to student achievement. With value added measures, the characteristics of the studentpopulation are taken into account when calculating the school results in terms of student achievement. These characteristics influence the results but they are beyond control of the school. The most important characteristics to be taken into account are:

1. prior attainment of the student;
2. background characteristics of the students like ethnicity, social background or social deprivation;
3. contextual factors, such as the percentages of students with social deprivation or the mean starting level.

***Output in terms of efficiency results or internal efficiency: the school career***

Indicators frequently used for measuring internal efficiency of a school are:
1. the number of students who reach the objectives of the school compared with the total number of students;
2. attendance and drop-out rates;
3. the time it takes to obtain results and the prevalence or absence of delays.

***Outcomes***

Outcomes are longer terms results, also called external efficiency results. Indicators of outcome are for example: success on the labour market, success in the next stage of education or in society in general.

***Other types of output and outcomes***

Other examples of indicators for school output are evaluations of student satisfaction and of satisfaction of other stakeholders. In quality assurance systems these indicators are included in the results box. This type of school output is not part of the results of the learning process, but it certainly is part of the output of school.

# 2        Current practice

## 2.1        Evidence from SICI members

**Variation (availability of data, methods, objectives)**
There is remarkable variation in the data and methods applied and the objectives to be achieved by inspectorates in using student achievement data. Here we mention just the main characteristics of some of the many interesting examples that were presented at the conference.

In Wales for example, education policy strongly emphasizes working towards better learning outcomes. Provision of education and skills in Wales was devolved in 1999 to be the responsibility of the Welsh Assembly Government (WAG). Estyn, the Welsh inspectorate, has strong links with the Department of Children, Education, Lifelong Learning and Skills within WAG. There is general agreement among stakeholders to share data on education providers, including a wide range of analysis of primary and secondary school performance data. Over the last three years, considerable progress has been made. The inspectorate groups schools into so-called 'families' of schools with similar characteristics, such as free school meals, prior performance and so on. Comparisons can therefore be made not only with national averages but also with averages of all schools in the same 'family' (cf. figure 1).

**Figure 1**
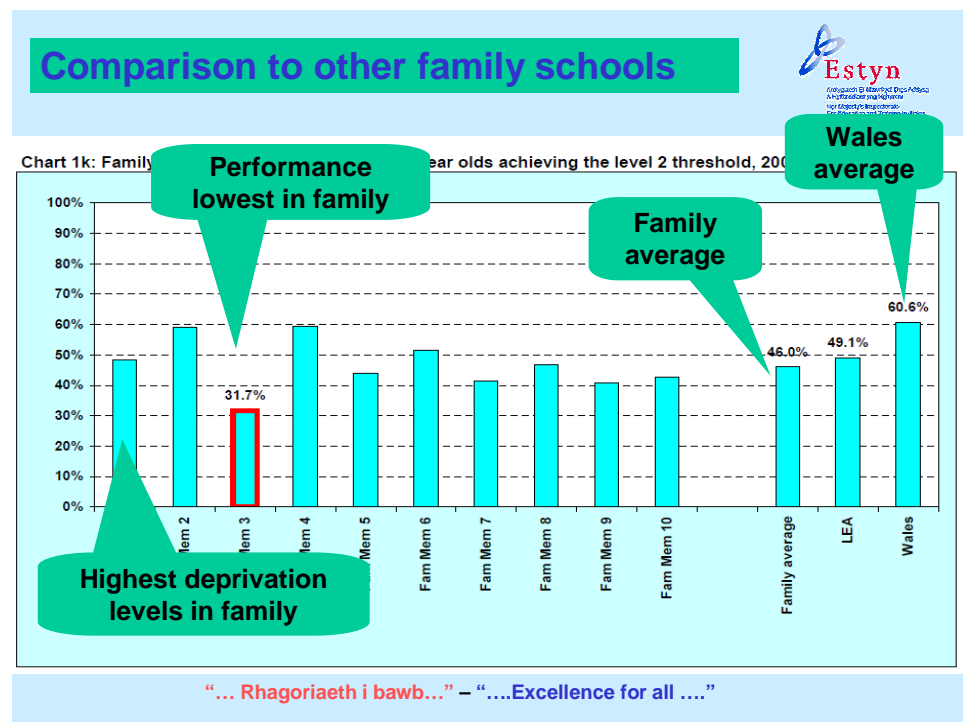


According to the national Education Act (2004) in the Czech Republic, the Czech School Inspectorate is required to 'determine and assess the conditions, course and results of education in accordance with relevant school educational programmes'.

Since there are no national programmes for testing of pupils in the key stages and no national school-leaving examinations, the inspectorate developed its own methods of data collection to make an approximation of student achievement per school.

Examples of criteria monitored by Czech inspectors to evaluate results are:
- number of students and failure in education programmes (drop outs);
- transfer of students to a higher educational level;
- transfer between education programmes;
- changes in education leading towards development of key competencies.

In France, national data of good quality on student achievement are available, but the inspectorate does not use them to make comparisons or to analyse or judge the performance of schools. This is largely due to the fact that national policy notions as to what the role of the inspectorate should be do not allow for such comparative judgements to be made by the inspectorate.

Flanders operates in a context similar to the Czech Republic, given the fact that there are no central or public exams and there is no central testing. It developed the CIPO (Context – Input – Process – Output) model (figure 2), which is basically a datawarehouse drawing amongst others on financial data.

**Figure 2**
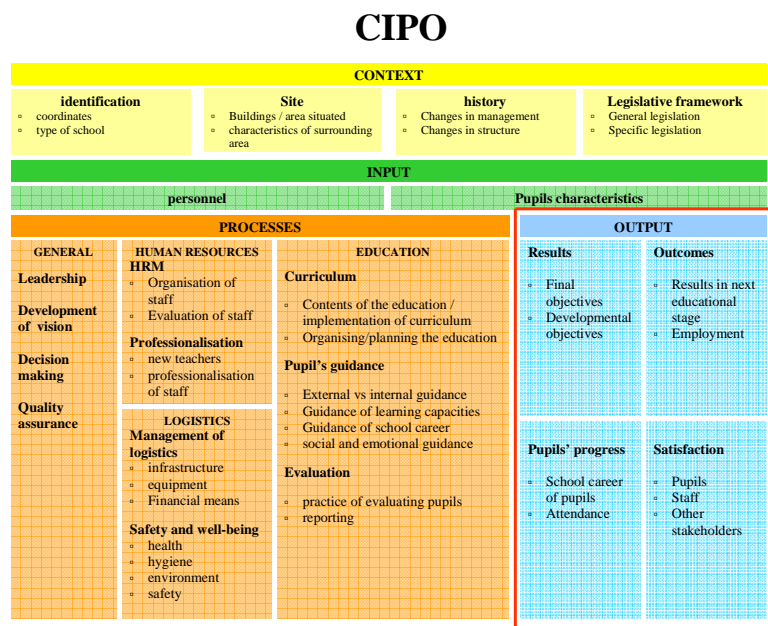


**Risk analysis (identification of weak schools)**
Risk analysis is used by quite a few SICI members, the CIPO model form Flanders being a case in point. Other examples are Rumania / Aracip (risk index), England / Ofsted, the Netherlands' inspectorate and Sweden. Sweden developed SALSA (http://salsa.artisan.se), a tool for local context analysis. Using SALSA, the Swedish inspectorate came up with the following scenarios for the situation of schools with respect to student achievement.

Scenario 1: Schools in which virtually all students reach appropriate in all subjects in grades 3, 5 and 9. Assessment: These schools do not receive any negative ratings or comments on the evaluation check-list. They are instead provided with a general assessment that describes the school's positive performance. This is because we use an assessment based on deviation.

Scenario 2: Schools where more than a negligible proportion of students fail to achieve the objectives of all the current subjects and have several complaints about the activities, in which at least one report includes an area that has a direct impact on results. Assessment: Students are not given a chance to reach the goals to achieve in all subjects. A written warning can be issued in conjunction with the formal assessment report. These are linked to the defects that directly affect the achievement of objectives.  E.g. "It is especially critical that ..."   "It is not acceptable ...".

Scenario 3: Schools where more than a negligible proportion of students are failing to achieve the goals in the current subjects, but the affecting factors are assessed as functioning. Assessment: "Despite well-functioning activity and special support, the school reported a high proportion of students failing to achieve goals in all subjects." (The positive aspects of the school can never change the fact that the goals are not achieved).



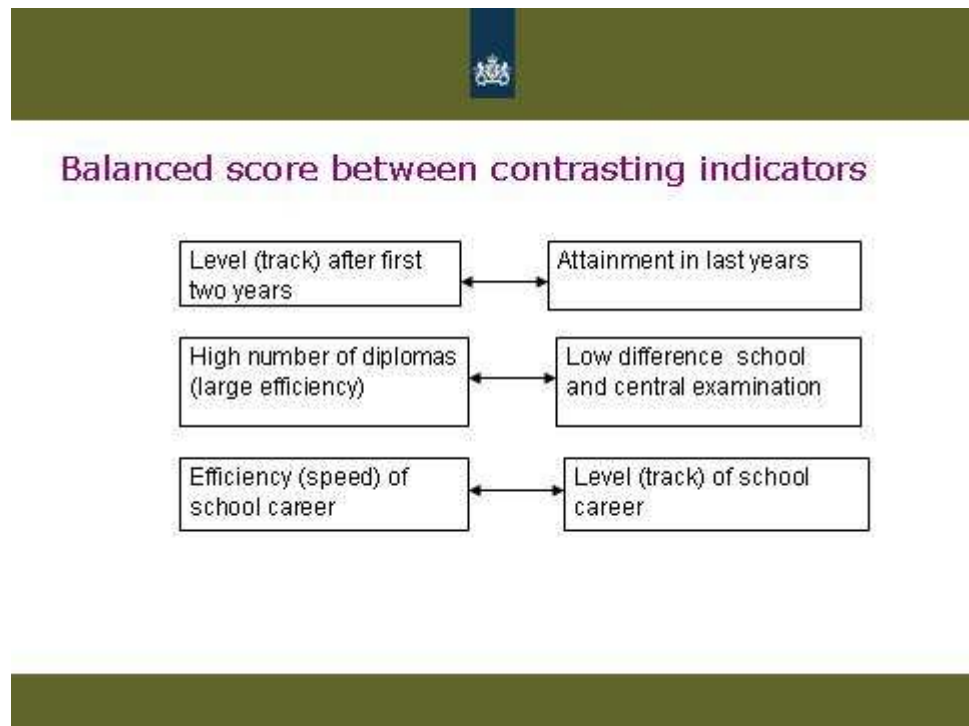In the Netherlands, a balanced score between contrasting scores is used for assessing school results in secondary education (figure 3). This method uses the following indicators:

1. Result of the first and second year (primary level).
2. Efficiency result of last 2-4 years of secondary education.
3. Results of central examination (national tests).
4. Difference between the marks of school examination and central examination.

In the near future, an outcome indicator will be included: successful placement in the next school.

**Figure 3**

## Balanced score between contrasting indicators

| | |
|---|---|
| Level (track) after first two years | Attainment in last years |
| High number of diplomas (large efficiency) | Low difference school and central examination |
| Efficiency (speed) of school career | Level (track) of school career |

Ofsted, the British inspectorate, uses seven key judgements to assess outcomes:

1. how well pupils achieve and enjoy their learning - achievement
2. the extent to which pupils feel safe
3. pupils' behaviour
4. the extent to which pupils adopt healthy lifestyles
5. how well pupils contribute to the school and the wider community
6. how well pupils develop workplace and other skills that will contribute to their future economic well-being – taking account of attendance
7. pupils' spiritual, moral, social and cultural development.

National comparators for these outcomes :

| Data | Associated judgements |
|---|---|
| Absence | Pupils' attendance |
| Exclusions | Pupils' behaviour |
| Take up of school meals | The extent to which pupils adopt healthy lifestyles |
| Participation in sport | The extent to which pupils adopt healthy lifestyles |
| Young people who are not in education, employment or training | The extent to which pupils develop workplace and other skills that will contribute to their future economic well-being |

Inspectors should compare the school's data with the national picture, but:

1. the data must be used carefully
2. it provides a starting point for discussion with the school – particular when there is a large difference between the school's data and the national picture
3. the data must be set alongside all the other evidence gathered by inspectors.
4. these are important principles in all inspections by Ofsted

The website www.raiseonline.org provides interactive analysis of school and pupil performance data. RAISE stands for Reporting and Analysis for Improvement through School Self-Evaluation.

**Use of output data in supervision and judgement**
In several countries, inspectorates use student achievement data as a component in their judgement of schools. Cases in point are Sweden, the countries of the United Kingdom and the Netherlands. In the Netherlands, this judgement is public (published on the website of the inspectorate). Inspectorates taking this approach do aim to develop methods to take into account factors such as the context of the school, notably the socio-economic background of the pupils.

**Feedback oriented practices**
Some inspectorates very consciously decide not to use output data in their (public) judgement of schools. In Berlin – Brandenburg for example, the achievement data are used for inspections in two ways. The results of the school are compared to state-wide results for the same track of school, but have no impact on the evaluative part of the report. However, the inspectorate does pay attention to whether the school has analyzed the feedback systematically and whether the school has derived measures for future school and instructional development.
Description of school results at the national level
Many SICI-members use their analyses of student achievement data and school output in the descriptions they make of school results at the national level.

### 2.2 Evidence from research

Roel Bosker made a convincing case arguing that student achievement is influenced by many factors, with the school accounting for only 15 percent of the total influence (cf. figure 3). The value added model he developed is based amongst others on this notion.

**Figure 3 Decomposing the original contribution of the secondary school**
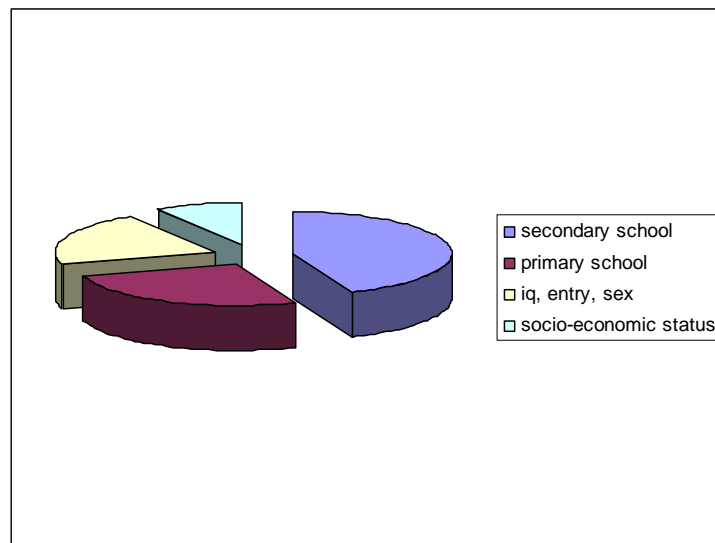


Analyses made with the value added model led Bosker to conclude that it does not matter much which technique we use to get at the value added by a school. 'Good' schools according to one method are also 'good' schools  according to another method; schools that are 'good' one year are mostly also 'good' the next year; schools that are 'good' in one subject are mostly also 'good' in another subject, and schools that are good' in the cognitive domain are mostly also 'good' in affective domains. But also, it is an illusion to assume that we might ever get at the true value added by a school - at best, we will only be able to reach an approximation.

Professor W. Bos of the Technical University Dordtmund addressed the question of background characteristics of pupils and their influence on learning achievement. In doing so, he built on Bourdieus theory on social, economic and cultural capital. Not only socio-economic deprivation has an impact on learning achievement but also social skills and cultural baggage pupils are provided with at home. Professor Bos developed an instrument which gathers information from pupils about such characteristics. This way, a much more accurate picture of the background of pupils may be obtained. The discussion in the workshop focused on how to use this instrument in practice. How should one collect such data on a larger scale? More information may be found in: W. Bos a.o.: Zur Konstruktion von Sozialindizes; ein Beitrag zur Analyse sozial-räumlicher Benachteiligung von Schulen als Voraussetzung für qualitative Schulentwicklung. Bundesministerium für Bildung und Forschung, Berlin 2010 (http://www.bmbf.de/publikationen/index.php).

Koji Miyamoto from OECD argued convincingly that education plays a significant role in improving health and social cohesion by raising competencies. However, having better information and cognitive skills is not enough. Social and emotional skills

empower individuals to better mobilise available information and cognitive skills so that they are more capable of preventing and coping with health challenges and promoting social cohesion. Education can contribute to raising such capabilities not only by facilitating the acquisition of these skills, but also by developing habits, norms and ethos of healthy lifestyles and active citizenship. Learning also takes place in the family and the community. Both are important environments in which children develop critical competencies. The difficulty is to ensure that the various environments are coherent and consistent. Government can play an indispensible role by promoting policy coherence and providing the right incentives for stakeholders to invest in the right resources. In this way, education can make a significant contribution to social progress.

## 2.3        Concluding remarks

From the large variety of indicators for school results used by SICI-members for various purposes, some general characteristics do emerge. In the first place, usually more than just one indicator is used. Three major groups of indicators may be distinguished:

- ➢ indicators based on final test results (effectiveness);
- ➢ efficiency indicators: school drop out rates number of students who pass the exams, survival rates;
- ➢ outcome indicators.

As to the norms used to assess achievement and output data: it is clear that these are very often relative norms, such as comparing results of schools to the national mean, or to the mean of comparable schools. At several points during the conference the question came up as to whether absolute norms are possible. And perhaps an even more urgent dilemma is whether we should actually wish to have absolute norms at all.  Several SICI-members reported recent national discussions on this issue that caused considerable political upheaval. The Netherlands recently actually introduced national reference levels for the Dutch language and maths. The public debate on this national measure continues.

The SICI members using student achievement data as a component in their judgement of schools do notice very clearly that this approach tends to foster manipulative actions by schools and institutions. Some examples of such phenomena will be expounded in chapter 3 of this report.

# 3      What works and what does not?

## 3.1      Effectiveness of using standardised data

In general, SICI-members using student achievement for school inspections are positive about the effectiveness of this approach. Analysis and benchmarking of standardised data allow inspectors to build a detailed picture of how a school is performing, whether this performance is improving over time and where the school fits within a national context. It also allows inspectorates to make a more effective selection of schools that need special attention as opposed to schools that are performing well. The latter do not need frequent inspection visits, which saves time and staff capacity for schools as well as inspectorates. In addition, schools increasingly use the results of analysis and benchmarking by inspectorates to assess their own effectiveness and to draw conclusions as to actions they should take to improve their performance.

Notwithstanding these encouraging results, it is clear that using student achievement data for judgement of school results also causes unintended side-effects which are less positive. The following section of this chapter provides some examples.

## 3.2      Unintended side-effects

Using student achievement data for assessing school performance often also causes unintended reactions by schools or students. Manipulative strategies by schools or students to manipulate the data may prove to be a problem. Unexpected visits to Dutch primary schools at the time the national test was administered showed a variety of strategies, such as:

  a. Manipulation of test administration conditions: 5.5% of the schools turned out not to keep to guidelines, for instance by allowing the use of scrap-paper, by not administering the test on the correct date or by giving away answers.
  b. Strategic filling in of questionnaires in which schools account for the administration of the test. In some cases the number of students in 8th grade did not equal the number of students that took the test + the number of students that did not take the test. In other cases, the percentage of students for which background correction was requested did not match general school population data.
  c. excluding the weakest students or not accounting for individual students (reasons given: students were ill, participated in practical training or received learning support).
  d. Strategic use of capacity tests of which is commonly known that they usually result in lower capacity scores for pupils; the (relatively low) cognitive outcomes of these pupils may in these cases be more in line with the expectations from the capacity test.

A well-known example of strategic behaviour from Dutch secondary education is the difference in results between school exams and central exams, with students structurally scoring higher on school exams than they do in the central exam. In weakly performing schools, this difference tends to be bigger.

Although these examples are taken from the Dutch situation, similar cases from other countries were mentioned in the discussion in the workshop on this theme. Portugal for instance mentioned the problem of rankings published by newspapers. The rankings are based on test results of public as well as private schools.  In the private schools however, 25% of the pupils do not participate in the national test. This causes the rankings to present an unrealistically positive picture of private schools. Parents like to see rankings, but they are not interested in the way the rankings were established. Good teachers apply for positions in good schools. For weakly performing schools, it is hard to get good teachers and difficult to improve. It is not easy to break through such a vicious circle.

## 3.3        Outcome indicators

Inspectorates make various efforts to take a broader perspective than just looking at student achievement and school results in the narrow sense. HMIe, the Scottish inspection for instance, does this by
> observing classes and young people within the school environment
> talking to learners, staff, parents and other school partners
> analysing documents and looking at learner work
> applying professional judgement within the inspection team to enable attainment in tests and exams for example and wider achievement relating to social outcomes to complement and support each other.

## 3.4        Education for pupils with special needs

At the conference, limited time was available to discuss the possibilities for using student achievement data in education for pupils with special needs. This section uses the case of the Netherlands only because detailed information about the experience of other SICI-members on this issue is not yet available at the time of finalizing this strategic report.

Schools for special education in the Netherlands do not have a long history of using pupils' achievement data for reflecting upon their added value or their own quality. There are several reasons for this, but an important factor is certainly the lack of sufficiently standardised tests for the heterogeneous population in special education. The range of tests being used in regular education were not qualified for assessing the – usually - quite slow progression in skills of the 'special needs' population.

In 2009 – 2010, the Dutch Central Institute for Test Development (CITO) has started the development of valid and reliable tests for the assessment of (growth in) outcomes for pupils with special needs. The tests are expected to become available in the near future. Thus, schools will gain insight in the outcomes of pupils and educational achievement over time. In addition, schools will then be able to relate progress to peer-group, school and region with all the benchmarking possibilities.

*Educational developmental goals*
Apart from the above-mentioned developments, the Netherlands' inspectorate will establish an outcome-oriented approach in assessing the quality of schools for special education. From the year 2011, the inspectorate will request schools to define an 'educational developmental goal' (EDG) for all pupils when entering special education. Defining such an EDG has already been done since 2004 in primary schools for the 'inclusive pupils' (pupils with special needs, but enrolled in regular primary schools). An EDG, which assumes specified levels of achievement necessary

for streaming into a next educational setting, should open the way to goal-oriented education.

When defining an EDG, it is necessary to have insight into the capacities or IQ of the pupils. However, in special education all pupils are tested before they receive an indication for special education. One of the tests included is an intelligence test, so schools can use the test results in defining the EDG for each pupil. Taking into account the IQ and the diagnosed disorders of the individual pupil, as well as his or her protective and impeding factors, schools should be able to decide upon an ambitious educational and developmental goal for their newly admitted pupils.

By formulating the goals that the pupils should have reached when leaving school, it is also possible for teachers to deduce the goals that the pupils should have reached within a single school year and subsequently for the next months and even weeks. This leads to a 'planning approach', instead of the 'following approach' which was practised in the past. Schools get more control over the progression in curriculum that they need to accomplish within a certain period of time or at the end of a year. The inspectorate gains insight into pupils' learning outcomes by analyzing the EDGs and comparing them to the outcomes of the pupils when they leave school. Of course schools can change the EDGs, if the goals turn out to be too ambitious, or possibly too pessimistic. Obviously, working with the EDGs needs to be elaborated, and there may be some pitfalls or statistical problems. However, one of the good things is that regular schools working with the EDGs have accepted this planning approach. Most of these schools acknowledge that the it helps to make education more effective.

### 3.5        Lessons learned

A major question during the conference was: can we actually use student achievement data for school inspections? The lectures and presentations showed that the answer to this question was affirmative. However, more complicated questions arise, such as the issue of how exactly to define output/outcomes of schools? Participants agreed that we need output indicators of all types of results of all students, indicators which are reliable and fair, measures which are practical as well as recognizable for schools.

As to the use of data about outcomes (longer term results such as labour market performance, contribution to social cohesion etc) the general feeling seemed to be that it would be useful to include such measures in quality assessment. In fact, the discussions built on the insights that emerged from the 2010 Copenhagen conference on social cohesion in the context of citizenship. However, it was clear that there is still a long way to go in developing suitable approaches towards this goal.

Manipulative behaviour by schools with respect to achievement data is hard to prove (unexpected visits may prove effective). But data and literature provide strong evidence that it does indeed exist and comes in different forms and shapes. Inspectorates should at least be aware of this phenomenon because it leads to undesired differences in assessments between schools. It also might have undesired consequences for students. There are measures that can be taken to prevent this behaviour from distorting data patterns, such as assigning fictive (low) test scores to students that are excluded from the test, looking for unexpected outcomes, trends, high percentages of students repeating a year, and mobility in school population.

Looking for trends across different years of performance – for the same group of students as they progress through education and for different groups of students across years but at the same stage – could also be helpful in determining 'true' performance.

# 4          Results of the conference

## 4.1          Shared principles

The overall conclusion of the conference was that student achievement data should be a main component of measures that try to assess the quality of the schools' results of output. Calculating school output based on student achievement is difficult, but not impossible. Agreement was also reached on a general definition for the use of output data by inspectorates: 'Inspectorates should strive for ***reliable*** and ***fair*** output indicators of ***all types of results*** of ***all students***: a measure which is ***practical*** as well as ***recognizable*** for schools.'

Elaborating upon this definition, the conference came up with detailed specifications. ***Reliability*** of indicators may be optimised in several ways. In the first place, it is important to minimise measurement errors, by using (administrative) data of good quality, and by not relying on one single indicator but on a set of indicators (over more than one year). Data should be as specific as possible, meaning that measurement should be done at the lowest level. If at all possible, data collection should take place at the individual level. In  addition, attention is needed for specific measurement problems: ceiling effects, student mobility, etc. Academic partnerships and advice are crucial;  many of the inspectorates represented at the meeting cooperate with scientists in improving their use of output data.

***Fairness*** of indicators means that their measurement should result in an assessment which expresses the achievement (or added value) of the school. Results depend on characteristics of the student population and context. We have to take this into account. The value added approach used by some inspectorates is very appropriate from the point of view of fairness, but it is not suitable / feasible for every country, as it requires the availability of quite a lot of reliable data and research capacity. What is more, these issues are obviously not purely technical: phenomena like central exams, national tests and related centralised data-collection are important political issues.

In order for it to be fair, assessment of schools should also take into account manipulative behaviour.

***All types of results:*** Assessment and judgement of schools should ideally be based on a balanced model with a (smart) combination of more than one output indicator: effectiveness results, efficiency, outcome, etc.; test scores and examination results of a large range of subjects as well as efficiency results such as low school dropout rate and numbers of students repeating a year should be used. Results obtained by students in the next stage of their education are an example of outcome-information that may improve the quality of the assessment. As to the use of other data about outcomes (longer term results such as labour market performance, contribution to social cohesion etc) the general feeling was that it would be useful to include such measures. At the same time however, it was clear that there is still a long way to go in developing suitable approaches towards this goal. Another case in point is the well-being of the students: ideally, it should be included, but how to measure it? To conclude this discussion about types of data: using a combination of indicators also helps to counter the impact of strategic behaviour.

***All students:*** An absolute level of skills is needed (basic skills for every one). Such reference levels are important from the point of view of the stakeholders: a certain minimum level should be attained - every child must be able to read and write. Thus, there should not only be relative norms, but absolute ones as well. This also means that achievement of children with special needs or specific groups needs separate attention and that differences in results between groups should be taken into account. Separate norms should be available for every group.

***Practical:*** Indicators selected should preferably fit into the systems of data collection already used by schools. That way, the schools can use the indicators in their quality system for improvement. We must keep in mind that there is a difference between the school perspective and the perspective of the inspectorate.

First and foremost, for indicators and assessments to be ***recognizable*** by schools they should be as simple as possible and build on output measures that are used by schools themselves. It is essential that they be recognized by schools with weak student populations as well as by schools with a strong student population, and by parents as well as students, teachers, etc. Therefore it is useful to have (a panel of) stakeholders agree upon the indicators to be used.

All in all, participants agreed that inspectorates should aim to implement high quality assessments of schools. Student achievement was considered a crucial element in assessment of schools, implying that schools should incorporate student achievement in their quality assurance systems. Finally, the importance of effective communication was stressed: inspectorates should communicate their findings clearly to all stakeholders.

### 4.2       Areas of further enquiry

The conference identified quite a few areas for further enquiry, research and development. The major ones are the following:

a) Characteristics of learners: for which factors should we correct our results? Quality for which group is at stake? (differences between coungtries).
b) Presentation of results: what are effective ways to communicate with stakeholders?
c) School improvement: what are the most successful interventions of inspection in improving results? How to use output/outcomes in this?
d) Lack of data: how to deal with data restrictions (for example: no national test)? Countries lacking national data collection systems may find this strategic report a useful input in their national policy discussions.
e) Developing suitable approaches towards including data about outcomes (longer term results such as labour market performance, contribution to social cohesion, student well-being etc) in quality assessment.

### 4.3       Possibilities for further cooperation and mutual support

Participants saw many possibilities for further cooperation and mutual support. SICI/SIA can do much to support this process, amongst others by publishing this strategic report and all other inputs to the conference on its website. Other useful follow-up could consist of: a) further work on theory, glossary and principles; b) bilateral exchanges and cooperation in developing measures; c) further cooperation with scientists, including statisticians.

### 4.4 SIA community of practice or expert network

In addition to the support activities mentioned in the preceding section, SICI/SIA will consider facilitating cooperation amongst its members on issues raised during the conference by starting a digital community of practice on the subject of defining and measuring output and outcome of schools.

A useful activity of such a community of practice could be to develop country-profiles on themes related to the use of achievement data in assessment of schools. The country-profiles could be based on a format developed by a small working-group composed of representatives from three to four SICI-members. Themes / working groups that might be of interest are for example:

a) Working group 'Assessment of schools by using student achievement data'
b) Working group 'Development of alternative measures in case of lack of data at national level'
c) Working group 'Calculation on large data-sets / development of added value measures'.

## Annex 1  Programme of the conference

| MONDAY, MAY 17th, 2010 | |
|---|---|
| Arrival of participants | |

| TUESDAY, MAY 18th, 2010 | |
|---|---|
| 9:00 | Arrival and registration |
| 9:30 | Welcome and Introduction : *Chair: Gonnie van Amelsvoort,* IvhO and EC SICI<br>*Annette Roeters,* Senior Chief Inspector of the Netherlands Inspectorate of Education<br>*Graham Donaldson,* President of SICI<br>*Bruno Vreebrug/Inge de Wolf,* Netherlands Inspectorate of Education:<br>Introduction: aims and outlines |
| | Keynote *Roel Bosker* (University of Groningen)<br>'Practical and methodological issues in using student achievement data for accountability and school inspections' |
| 11:00 | Coffee break |
| 11:30 | Two examples of using student achievement data for inspections:<br>*René Vanotterdijk*: Assessing student achievement in school inspections in Flanders<br>*David Humphries*: Assessing student and school achievement; the use of Raise Online by Ofsted inspectors in England |
| 12.30 | Similarities and differences . Introduction to workshops<br>*Bruno Vreeburg / Inge de Wolf* |
| 12:45 | Lunch |

| 13:45 | Workshop 1:<br>Experiences with measuring output and outcomes of schools. Different countries present their methods.<br><br>1. Wales (*Mererid Stone, Sioned Moffet*)<br>2. North Rhine-Westphalia (*Klaus-Georg Wey*)<br>3. Romania (*Serban Iosifescu*)<br>4. The Netherlands (*Klaas Bos, Bruno Vreeburg*)<br>Presentations and discussion | Workshop 2:<br>Experiences with measuring output and outcomes of schools. Different countries present their methods.<br><br>1. Scotland (*Sheila Page*)<br>2. Berlin Brandenburg (*Holger Gärtner*)<br>3. Czech Republik (*Petr Drabek*)<br>4. Sweden (*Per Ingvar de la Motte*)<br>Presentations and discussion |
|---|---|---|
| 16.15 | Break | |
| 16:45 | Summary and feedback from the workshops<br>What kind of similarities and differences are there?<br>Which basic principles do we have?<br>What kind of problems in assessing school outcome do we meet and how can we solve these problems?<br>How can inspectorates support each other in these matters?<br>*Bruno Vreeburg/Inge de Wolf* | |
| 17:15 | End of work session | |
| 18.30 | Bus tour, sightseeing | |
| 19:30 | Official Dinner, hosted by the Netherlands Inspectorate of Education<br>Restaurant Jaap Hannis, Borneosteiger 2, 1019 KM  Amsterdam,<br>tel. +31 20 4189690 | |

| Wednesday, May 19th, 2010   Chair: *Gonnie van Amelsvoort* | | | |
|---|---|---|---|
| 8.45 | Coffee/tea | | |
| 9:00 | Basic principles of using student achievement and introduction workshops<br>*Bruno Vreeburg/Inge de Wolf* | | |
| 9:30 | Workshop 1:<br>Social performance;The role of education in social outcomes – including adolescents' social and health related behaviours<br><br>*Koji Miyamoto* (OECD) | Workshop 2:<br>Strategic behaviour; Unintended consequences of test-based school inspections; theory and examples of strategic behaviour<br><br>*Melanie Ehren* (Twente University) & *Machteld Swanborn* (Netherlands Inspectorate) | Workshop 3:<br>Fair comparisons; Comparing the unequal. A school based index to describe student's background characteristics<br><br>*Wilfried Bos* & *Heike Wendt* (University of Dortmund) |
| 11:15 | Coffee break | | |
| 11:45 | Lessons from the (morning) workshops<br>social performances<br>strategic behaviour<br>fair comparisons<br>*Bruno Vreeburg/Inge de Wolf* | | |
| 12:15 | Final speeches:  conclusions from the conference; challenges for the inspectorates and SICI<br>*Annette Roeters, Graham Donaldson* | | |
| 12:45 | Lunch | | |
| Afternoon | Departure of participants | | |

## Annex 2   List of participants

| familyname | first name | country |
|---|---|---|
| Amelsvoort, van | Gonnie | Netherlands |
| Andersen | Anders | Denmark |
| Andersson | Jonna | Sweden |
| Arnouts | Kristien | Flanders |
| Baak | Peter | Netherlands |
| Baltussen | Miriam | Netherlands |
| Balvers | Maarten | Netherlands |
| Bojanic | Milan | Serbia |
| Bos | Wilfried | Germany, North Rhine - Westphalia |
| Bosker | Roelof | Netherlands |
| Braunová | Jarmila | Slovak Republic |
| Broek d'Obrenan, van den | Vic | Netherlands |
| Bulder | Bert | Netherlands |
| Camilleri | Kenneth | Malta |
| Capric | Gordana | Serbia |
| Cvetkovic | Gordana | Serbia |
| De La Motte | Per Ingvar | Sweden |
| Diedrich | Martina | Germany, Hamburg |
| Dijkstra | Anne Bert | Netherlands |
| Donaldson | Graham | Scotland |
| Drábek | Petr | Czech Republic |
| Duriacova | Tatiana | Slovak Republic |
| Dzida | Branislava | Serbia |
| Ewa | Hallberg | Sweden |
| Gärtner | Holger | Germany, Berlin/Brandenburg |
| Groot-Wilken | Bernd | Germany, North Rhine - Westphalia |
| Henkens | Leon | Netherlands |
| Hornboell | Rikke Juul | Denmark |
| Humphries | David | England |
| Iosifescu | Serban | Romania |
| Kartawidjaja | Heidi | Netherlands |
| Keating | Yvonne | Ireland |
| Key | Tim | England |
| Kjems | Bo | Denmark |
| Klaver | Liesbeth | Netherlands |
| Kooijman | Jakeline | Netherlands |
| Kordes - De Vaal | Joke | Netherlands |
| Korduliaková | Anna | Slovak Republic |
| Kovac Cerovic | Tunde | Serbia |
| Kozelová | Beata | Slovak Republic |
| Laman | Mineke | Netherlands |
| Lambrechts | Dirk | Flanders |
| Lemmer | Beate-Maria | Germany - Hessen |
| Livin | Franck | Flanders |
| Lukackova | Zuzana | Slovak Republic |
| Mair | Rudolf | Austria |
| Melichárek | Kamil | Czech Republic |

| | | |
|---|---|---|
| Micallef | Joan | Malta |
| Moffett | Sioned Mair | Wales |
| Müller | Sabine | Germany, NRW |
| Myamoto | Koji | France |
| Noivo Roque | Carlos | Portugal |
| Onna van | Maria | Netherlands |
| Page | Sheila | Scotland |
| Pagrach | Karel | Netherlands |
| Quax- e/v van Cauwenberghe | Carla | Netherlands |
| Reezigt | Gerry | Netherlands |
| Rijkers | Jan | Netherlands |
| Roeters | Annette | Netherlands |
| Rydin | Anna | Sweden |
| Schaapman | Eef | Netherlands |
| Schouten | Rob | Netherlands |
| Schreier | Bernd | Germany, Hessen |
| Sommer | Norbert | Germany, Lower Saxony |
| Stone | Mererid | Wales |
| Swanborn | Machteld | Netherlands |
| Thurner | Josef | Austria |
| Vaicekauskiene | Vaiva | Lithuania |
| Vanotterdijk | René | Flanders |
| Verbauwhede | Yvan | Flanders |
| Verkroost | Jos | Netherlands |
| Vikydalová | Adriana | Slovak Republic |
| Vosse | Agnes | Netherlands |
| Vreeburg | Bruno | Netherlands |
| Weir | Tony | Ireland |
| Wendt | Heike | Germany, NRW |
| Wetzels | Jo | Netherlands |
| Wey | Klaus-Georg | Germany, NRW |
| Willems | Fons | Netherlands |
| Woitalla | Eric Peter | Germany, Hessen |
| Wolf, de | Inge | Netherlands |

## Annex 3  Abbreviations

| | |
|---|---|
| ARACIP | Romanian Agency for Quality Assurance in Pre-University Education |
| CIPO | Context - Input – Proces – Output |
| EDG | Educational Developmental Goal |
| HMIe | Her Majesty's Inspectorate of education |
| ISQ | Institut für Schulqualität der Länder Berlin und Brandenburg |
| OECD | Organisation for Economic Cooperation and Development |
| Ofsted | Office for Standards in Education, Children's Services and Skills (UK) |
| RAISE | RAISE stands for Reporting and Analysis for Improvement through School Self-Evaluation |
| SALSA | (tool for local context analysis, Sweden) |
| SIA | SICI Inspection Academy |
| SICI | Standing International Conference of Inspectorates |