

# The use and misuse of assessment data in test-based accountability systems

37th IAEA Annual Conference

Manila, 23-28 October 2011,

Peter Hermans, Jan Wiegers

Cito, Arnhem, The Netherlands

## Abstract

Over the last two decades we have seen a global change in the governance of public education in the aftermath of the introduction of a new, market-driven educational policy in the early 80s. The Dutch government is increasingly involved in decision-making that directly affects teaching and learning.

Important characteristics of these changes are:

- the quality of education is defined in terms of student outcomes;
- the introduction of (national) standards-based curricula and instruction;
- data driven, test-based accountability
- the introduction of national tests in primary, vocational and higher education.

Performance- and test-based accountability systems tend to focus on students and schools failing to meet required levels of performance. There is a growing concern among test developers about the quality of high stakes decisions in this context that are, in many cases, based on the results of a single exam or test. Although the field of educational testing has emphasised over and over that judgements about student performance, teacher effectiveness or the quality of schools should be based on multiple measures, this principle seems to be largely ignored by the media and by policy makers. In this paper we will explore ways in which Cito tries to limit the misuse of assessment data in order to maintain the credibility of performance-based accountability.

There are three different approaches of accountability in education: through strict regulations, through (occupational) standards and through a data-driven (also referred to as a test-based or measurement-supported) system. In the first variant, schools behave in accordance with the rules and are accountable to the government. The second variant requires schools to set their own best effort requirements in order to meet standards. In this scenario, they are accountable to the rest of the world of education. In the third variant, the education system is responsible for the performance of its pupils and it is accountable to society. Most education systems

consist of a combination of the three, but in the past 30 years, the introduction of test-based accountability systems has led increasingly to educational reform. Following other social tasks which are financed by public funds, the idea that performance, productivity and value will improve through systematic public accountability is also gaining ground in the education system.

Test-based accountability systems are controversial because a relatively high number of tests must be taken in order to gather sufficient information. Test-based accountability is less popular, especially with parents and teachers; according to public opinion there is already too much testing. In addition, it is difficult to demonstrate the effectiveness of test-based accountability. There are many different ways of structuring these kinds of systems, from single test-based systems to extensive tracking systems with multiple tests per year over a number of successive years. Comparative research into the effectiveness of these systems is difficult to conduct, but hardly any supportive research is being carried out into the effectiveness of each of the systems.

The key common features of test-based accountability systems are:

- For a number of subjects, pupils' performance levels are established at different points in their school career (or at different ages);
- The descriptions of these performance levels will be set by laws or regulations, obliging schools to make efforts to enhance pupils' performance so that as many pupils as possible reach the desired level;
- The performances of pupils are measured through standardised tests
- The results obtained by schools are made public.

It appears as if accountability systems are especially focused on managers and directors. After all, they are the ones in charge of steering quality improvement, but ultimately the teachers are the ones who must achieve the quality improvement. However, especially in the Netherlands, a test-based accountability system is made by and made for teachers. They are the ones who provide the input for the system, and they must transform the outcomes of the analyses into concrete changes in their daily profession.

Test-based accountability systems can only have a positive effect on the quality of education when results are systematically collected and analysed, and when the decisions that are made based on this information are geared towards improvement and expansion of the curriculum. This requires an organisation which acts rationally and that is what is still missing in many schools.

Test-based accountability only works when administration and management respond to disappointing results by first investing in increasing the effectiveness of teaching and learning, which is in fact the only decisive factor that can easily be influenced. When the system is mainly driven by sanctions and teachers do not receive sufficient support in implementing the required improvements - whilst demands are raised or the sanctions become more severe - all sorts of negative effects arise and the value of the system is reduced to a useless ritual. It is useless, because ultimately the results are no longer a proper representation of the real level of education. A full test-based accountability system encompasses not only standards, tests and reports, but also effective teaching and learning, tools and generally accepted, widely implemented sanctions.

Test scores form the basis of test-based accountability systems, but the question is whether they are an adequate reflection of the pupils' actual performance level and whether it is advisable to compare and rank schools. For instance, the image may be distorted by differences in the size of the test population and because the pupils' motivation to perform well in the test may differ from one school to another school. In addition, the way in which pupils' results are aggregated to an indicator on a school level can have a negative effect on the effectiveness of a test-based accountability system.

### The situation in the Netherlands.

In the Netherlands, the education inspectorate is the pivot of the accountability system and test results are part of the assessment framework used by the inspectorate.

Nearly all Dutch primary schools use a pupil tracking system. The most extensive tracking systems, such as the Cito LOVS, consist of two tests per year (mid and end) for a large number of components of mathematics and language in the years 3, 4, 6 and 8.

At the end of primary education, almost all pupils take a national test; more than 80% of primary schools use the Cito Eindtoets Basisonderwijs (End of Primary Education Test) for this. The Dutch education inspectorate uses the data from this test to identify weaker schools, along with information from annual reports, complaint records and other sources of information. When the pupils' scores for the Eindtoets Basisonderwijs (marked taking into account social and economic background variables) measure a falling trend over a number of years or remain below the national average, a school will be considered 'at risk' and the inspectorate will conduct a supplementary investigation based on data from the tracking system from the five previous academic years and qualitative evaluations.

When the results are half a standard deviation below the national average of a group of comparable schools and when the quality of education has been assessed as insufficient during inspection visit, a school will be labelled as 'weak'. Schools which have been labelled by the inspectorate as weak or very weak will have increased supervision, where the interventions intended to lead to improvement are intensively monitored by the inspectorate. If this does not result in an improvement, it could ultimately lead to the discontinuation of the government funding and the closure of the school.

In secondary education in the Netherlands, there has been much less talk of systematic test-based accountability. In case of a lack of sufficient standardised tests, examination results are the key indicators used by schools in the frame of accountability. A test-based accountability system will be created at this level shortly, by extending the existing pupil tracking systems and introducing reference levels for language and mathematics and corresponding tests.

Test results therefore play a decisive role in the accountability system of the Dutch education system, but are not the only data on which schools are assessed. However, as a developer of systems and tests, the following two questions are important for Cito when setting up and implementing a test-based accountability system:

- Do the decision-makers have sufficient expertise to make decisions based on the test results?
- Do test results offer sufficient information to validate the decisions?

Cito and the quality assurance of the test-based accountability system

The quality of a test-based accountability system will benefit from having several measurement points and from monitoring pupils' progression over a longer period of time. Cito's tracking systems contain tests halfway through and at the end of a number of successive years. Furthermore there are other measurement tools available to create the full picture of pupils' performance.

The items that are used in the Cito test programmes are regularly refreshed in order to prevent score inflation. For the high stakes test programmes, such as the Eindtoets Basisonderwijs (End of Primary Education Test) and the tests taken as final exams, new tasks are used every year.

The possibility of working with ICT is used increasingly. Wherever possible, tests are computer-based, and if such tests are not available, schools will have the opportunity to enter their scores and to conduct their own analyses of the available data.

#### Information at all levels

To stimulate the responsible use of test results when making decisions, Cito has chosen to fully and actively inform all those involved. This is done not only through publications and information on the internet, but also through face-to-face information, discussions and dialogue. In addition, Cito responds quickly and adequately, through the media if necessary, when it appears that test results are used in an improper, incomplete or careless manner in the decision-making process.

#### Support

Cito provides different types of reports, directed towards target groups and adapted to the way in which the report will be used. In addition, schools will be given the tools to create custom-made analyses, not only to deal with specific issues, but also to enable them to assess the development over a number of years and to draw conclusions about the effectiveness of the teaching and learning, the teachers' skills, the methods used and the curriculum.

#### Education and training

School leaders and especially teachers must learn how to use a test-based accountability system effectively in order to increase the benefits of education. Without training, the data is not used fully, because only a small part of the possible analyses are actually conducted. The possibilities of the system are not used to their full potential by schools in order to assess

developments over a number of years. Using the data without training is in fact similar to incompetent use, because it is easy to draw the wrong conclusions from the analyses. Citowill increase the effectiveness of education and training by adjusting the system to the wishes and options of the schools. In addition, teachers and schools will be more closely involved in the further development and extension of the system.

All things considered, those involved would benefit on *all* levels from a working test-based accountability system, not just members of leadership and school leaders. Politicians and other policy-makers use the aggregated national data for administrative decisions and compare those with the results of international test programmes. They are completely dependent on analyses and conclusions that have been made by others and, at best, allow experts to advise them about their quality.

The directors and management mainly use the results to compare the quality of their school to other schools and for the vertical accountability to the inspectorate and the government.

Teachers must understand the role the system plays in the quality control of their teaching; they must know how the results of analyses can be translated to improvement. Parents must know what the indicators mean when choosing a school for their child.

A number of features of test-based accountability which are important for the success of the system can be distinguished.

The different interests of the various parties can lead to a mixture of too many purposes. It is not inconceivable that the data offered by a test-based accountability system could be used by different parties for different purposes, but when these purposes are mixed there is an increasing risk that the system as a whole will be corrupted. Test developers should be more proactive in clarifying the use(s) of a test or a system and should openly resist the improper use of tests and test data.

When the interested parties lack the necessary expertise to make full use of the possibilities that a test-based accountability system offers, this may lead to a number of negative effects. The tendency then is to pay close attention to the pupils with borderline scores, at the cost of pupils performing well and those performing badly. Test developers must not only deliver the instruments, but also enable the user to use these tools properly.

The effectiveness of a test-based accountability system is under pressure due to the tendency by all concerned to absolutise the aggregated and often merely summarised results.

International surveys almost exclusively look at the place in the ranking order and pay much less attention to other, more detailed outcomes. Test developers must be more independent, especially in their reports, and allow scientific integrity to prevail over commercial interest.

However, it is particularly the lack of understanding about the accuracy of the tools which undermines the effectiveness of test-based accountability systems, but this is something the users of the system can hardly be blamed for when the developers of the test take no responsibility for this decisive feature of the system.

A successful test-based accountability system enables schools to systematically follow the development of their pupils and to report the results and the effectiveness of the decisions made based on the results. Whether it works depends not only on the system, but especially on the way in which it is used.