OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes

COUNTRY BACKGROUND REPORT FOR THE NETHERLANDS

June 2012

Jaap Scheerens, Melanie Ehren, Peter Sleegers and Renske de Leeuw
University of Twente, the Netherlands

TABLE OF CONTENTS

LIST OF ACRONYMS AND GLOSSARY OF TERMS

International acronyms

EU	European Union
EURYDICE	Information on Education Systems an Policies in 33 European countries
IEA	International Association for the Evaluation of Educational Achievement
INES	International Indicators of Educational systems
OECD	Organization for Economic Co-operation and Development
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
TALIS	Teaching and Learning International Survey
TIMSS	Trends in International Mathematics and Science Study

Dutch Community acronyms

CBS	National Bureau of Statistics	Centraal Bureau voor de Statistiek
CCE	Center for consultation and Expertise	Centrum voor Consultatie en Expertise
CEB	Primary Education Evaluation Commission	Commissie Evaluatie Basisonderwijs
CEI	Committee Educational Innovation	Comite Onderwijsinnovatie
Cito	Central Institute for Test Development	Centraal Instituut voor Toetsontwikkeling
CPE	Committee on Program Evaluation	Comite voor de evaluatie van het programma
COOL	Cohort survey school careers	Cohort Onderzoek Onderwijsloopbanen
CVE	College for Exams	College Voor Examens
DUO	Service Design Education	Dienst Uitvoering Onderwijs
HAVO	Senior general secondary education	Hoger Algemeen Voortgezet Onderwijs
HBO	Higher professional education	Hoger beroeps onderwijs
ICO	Intercultura Education	Intercultureel Onderwijs
IRT	Item Respons Theory	Item respons theorie
IVBO	Individualised pre-vocational education	Individueel Voorbereidend Middelbaar Beroepsonderwijs
LVS	Pupil monitoring systems	Leerling Volg Systeem
MAVO	Junior general secondary education	Middelbaar Algemeen Voortgezet Onderwijs
MBO	Secondary vocational education	Middelbaar Beroepsonderwijs
MIS	Management Information System	Management informatie systeem
NWO	Foundation for Scientific Research	Nederlandse Organisatie voor Wetenschappelijk Onderzoek
N-term	Norm term	Norm term

PO	Primary education	Primair onderwijs
PPON	Periodic assessment project	Peridodiek Peilings Onderzoek
PRIMA	Primary Education and Special Education Cohort Studies	Primair onderwijs en special onderwijs cohort studies
RCEC	Research Center for Examinations and Certification	Onderzoeksinstituut voor examens en certificering
ROA	Research Centre for Education and the Labour Market	Onderzoeksinstituut voor onderwijs en de arbeidsmarkt
SBL	Sick Leave Bank	Belangenvereniging voor docenten
SES	Social Economical Status	Sociaal Economische Status
SIS	School-leavers Information System	Informatie systeem over schoolverlaters
SLO	Foundation for Curriculum Development	Stichting Leerplanontwikkeling Nederland
SLVO-cohort	School and origin by pupils in secondary education, cohort 1982	Schoolloopbaan en herkomst van leerlingen bij het voorgezet onderwijs, cohort 1982
SMVO-cohort	School and origin by pupils in secondary education, cohort 1977	Schoolloopbaan en herkomst van leerlingen bij het voorgezet onderwijs, cohort 1977
SVO	Foundation for Educational Research	Stichting voor Onderzoek van het Onderwijs
VAVO	Adult general secondary education	Voortgezet algemeen volwassenenonderwijs
VMBO	Pre-vocational education	Voorbereidend Middelbaar Beroepsonderwijs
VO	Secondary education	Voortgezet onderwijs
VOCL cohort		Voortgezet Onderwijs Cohort Leerlingen
VWO	Pre-university education	Voorbereidend Wetenschappelijk Onderwijs
WEC	Act of expertise centers	Wet op de expertisecentra
WO	University education	Wetenschappelijk onderwijs
WOT	The Act on the Supervision of Education	Wet op het onderwijstoezicht
WPO	The Act on Primary Education	Wet op het primair onderwijs
WVO	The Act on Secondary Education	Wet op het voortgezet onderwijs
WMS	The Act on Participation in Schools	Wet medezeggenschap op scholen

Acknowledgements

EXECUTIVE SUMMARY

Context and history of the evaluation system

A key feature of the educational system in the Netherlands is the principle of freedom of education. This principle, established in the constitution, implies that, when certain basic requirements are met, there is freedom to establish a school and, secondly, that parents are free to choose a school for their child. Freedom of education can be seen as the historic background for the particular patterns of central and de-central elements in educational governance. Two thirds of the schools in the Netherlands are government dependent private schools; these schools are to a large degree still organized according to religious denomination. Representative bodies of these types of schools consisted of school governor's organizations and teacher and parent organizations.. More recently this structure has been partly secularized into central councils for all major educational sectors; the ones for primary and secondary education (the PO and VO Council) being the most important for this report. These bodies have a strong influence on educational policy.

Another important implication of the freedom of education and the importance of educational organizations representing actors in the school field is the traditional autonomy of Dutch schools. This autonomy has been particularly marked in the domain of pedagogy and educational content. During the last three decades school autonomy has also grown in areas like financial management (the introduction of block grants and lump sump financing) and personnel policy, and a continued effort is being made to deregulate and to decrease central administrative pressure. Currently, according to publications from the OECD, the Netherlands is one of the countries with the highest degree of school autonomy in the world.

When it comes to positioning and describing evaluation and assessment, these historically developed structural and institutional arrangements are of great relevance. Types of evaluation range from being embedded in centralistic control measures to internal "formative" assessment procedures that are seen as instrumental to the development of individual students and schools.

The history of evaluation and assessment in the Netherlands goes back to the 1970s, when there was a temporary upsurge in more government driven educational innovation policy. This so called "constructive educational policy" was lead by the social democratic Minister of Education, Van Kemenade; it was characterized by a somewhat centralistic orientation, combined with a rational planning orientation. Large scale innovations were planned as experiments, and scientific evaluations were to point out the viability of these innovations. In many ways this approach did not go well with the traditional autonomy of schools, and neither did it sit well with the intermediary structures, including the denominationally organized educational support organizations who had to, more or less, manage these innovations. The efforts to scientifically

evaluate these programs largely failed due to unclear organizational positioning of the evaluation researchers and particularly due to strong resistance from teachers and schools.

During the 1980s educational policy changed to a more incremental system wide development orientation and was matched by an evaluation approach that left the model of program evaluation, turning to a more "systemic" evaluation approach, based on key data streams and continuous monitoring. In this period, major instruments for system level evaluation were developed: the periodic assessment project (PPON), the cohort studies in primary and secondary education as well as a gradual development of policy relevant educational statistics and educational indicators, the latter strongly stimulated by the active participation of the Netherlands in the OECD indicators projects and in initiatives from the EU, particularly EURYDICE.

In the wake of these developments in system level evaluation, systematic student evaluation and school evaluations (in that order) were gradually developing. Despite of the large autonomy of schools, the Netherlands traditionally have had central examinations at the end of secondary schools. In primary schools, a school leavers test, the so called Cito test, is being used as a basis for supporting the choice of a secondary school track since 1976. During the 1980s, school inspection was structured and shaped in an empirical analytic way. Partly based on knowledge from school effectiveness research, and also partly following similar orientations of Her Majesties' Inspectorate in the United Kingdom, an elaborate supervision framework was developed, and applied in school inspections.

Currently educational evaluation and assessment is being aligned to newer models of school governance, with slightly changed roles for national government, intermediary organizations and autonomous schools, giving rise to new forms of evaluation and assessment of schools such as "risk based inspection", "windows for accountability" and a larger focus on the use of assessment data in schools to improve student achievement (under the heading of "result oriented work"). The newer models of school governance also include clearer attainment targets and benchmarks for achievement of students in basic school subjects, which are described in so called Quality Agendas and Action Plans of the Minister of Education. Recently, new draft laws have been prepared requiring all primary schools to administer one central school leavers test and all primary and secondary schools to implement a pupil monitoring system (LVS). These new policy emphases underline the importance of summative and formative student assessment, and instrumental feedback to improve teaching and learning.

Demarcation of evaluation and expertise in evaluation technology

All instances of evaluation, assessment and appraisal addressed in this report confirm two basic elements of systematic evaluation: they involve structured empirical data and allow for an evaluative judgment. In the conceptual introduction of Chapter 2, three major evaluation functions

are distinguished: certification, accountability and improvement/organizational learning. When crossing this dimension (functions) with two other dimensions, namely type of data and aggregation level, a taxonomy of evaluation types can be drawn up, featuring a total of 14 evaluation types. Practically all of these types are used in the Netherlands, which means that the scope and application of evaluation and monitoring is quite broad. Among this broad range of evaluation approaches, teacher appraisal is one of the areas that lacks a structured systematic approach. Teacher appraisal in the Netherlands belongs to the discretion of the competent authorities of schools, i.e. school boards and municipalities. Just a few monitors exist on how, and to what extent schools fulfill this evaluative function. The fact that this domain is not penetrated by external organizations, not even the Inspectorate, can be seen as one the purest features (or toughest strongholds, when one takes a more critical perspective) of professional autonomy within autonomous schools.

Expertise in evaluation technology is well developed in the Netherlands. The test development company Cito has an international reputation in advanced applications of educational testing. Next, a range of research institutes and university departments exists that have the research technical skills to carry out various forms of educational evaluation. In the past, the development of evaluation has been stimulated by national expert committees, such as the Committee for Program Evaluation, and the research school (network of universities) on educational research, ICO. Last but not least the systematic approaches of the Dutch inspectorate have often been cited as exemplary by other educational inspectorates in Europe.

System evaluation

After the developments concerning the evaluation of national innovatory programs and the gradual move to a monitoring type of systemic evaluation, as described in the above, a number of stable data collection procedures were established. These are the periodic assessment project, the cohort studies, the gradual development of national educational statistics to a set of "system indicators" (Dutch: *bestel indicatoren*) and the various data collections by the Inspectorate. In addition, the Netherlands takes frequent part in international assessment studies, like TIMSS and PISA. Finally, a large number of smaller scale evaluation studies are being contracted out by the Ministry of Education. It should be noted that system level evaluation partly depends on information that is primarily collected for purposes of pupil assessment or school evaluation. Examples are examination results, aggregate data based on the Cito primary school leavers' test, and aggregate information based on the inspection of individual schools.

Several reporting frameworks have been created, in which sub sets of these data are synthesized, annually. These are the publications Trends and Key Figures, and the Inspectorate's annual report.

School evaluation

The main instruments for school evaluation are: school self evaluation and school inspection. School self evaluation is one of the instruments for the quality policy of a school. This quality policy has a legal basis as schools are required to produce various documents, such a school plan and school prospectus, in which they describe their quality policy and its results. In addition the quality of school self evaluations and quality care as a whole is monitored and assessed by the Inspectorate. Early 2000, two new support organizations (called Q5 and Q Primair) were established with the purpose of stimulating school self evaluation practices in primary and secondary schools. In addition, the Educational Supervision Act of 2002 introduced 'proportional inspections' as a means to motivate schools to implement self-evaluations. Proportional inspections were to use self-evaluations of schools to determine the intensity with which schools were to be inspected. From the late 1990s onwards school inspection became more systematic and guided by explicit supervision frameworks in which quality aspects and quality indicators were defined. The Inspectorate also issued quality cards, in which a school's functioning was rated on a number of indicators. School quality cards were made publicly available to support school choice. At about 2005 new concepts on educational and school governance gave a new impetus to both school self evaluation and proportional school inspection. The policy white paper 'Educational Governance' (Parliamentary year 2004-2005) outlined new governance relationships, which were intended to give more autonomy and responsibility to schools, and to diminish administrative burdens. The school boards' responsibility for educational quality was underlined, urging for a clear delineation of internal supervision and governance of schools,. In correspondence with these changes, the role of internal supervision and horizontal accountability by schools was underlined and distinguished from external supervision and vertical accountability through school inspection. The adapted governance arrangements as well as budget cuts,, called for a new interpretation of proportional inspection, which is aligned to the stronger positioning of school boards and their responsibility for horizontal accountability. Risk based inspection was implemented in 2008 and consists of an initial screening of all schools, based on a relatively limited set of information sources, (among which educational achievement indicators), on the basis of which one of two inspection arrangements is chosen. An arrangement can be: basic (no risks for the quality of education), or adapted (weak or very weak quality). Next, more intensive supervision and improvement stimulation occurs for the schools that are classified as (very) weak. Apart from receiving support, weak schools are also urged to improve by the threat of sanctions, which may go as far as holding back the complete budgetary funding of the school. If no risks are detected, schools are inspected less frequently, yet at least once every four years. With the introduction of risk based inspection a shift occurred in the kind of information that was required from schools for proportional inspection, which initially was expected to depend on school self evaluations. In the 2012 version of the Educational Supervision Act the idea of using information from school self evaluations was abandoned, and instead "publicly available accountability information" (e.g. on

outcomes and the financial situation of the school) was to be used as a basis for proportional inspection.

School inspection and school evaluation in general are likely to benefit from value added performance measurements, which can be based on existing and prospective instruments for student evaluation (see Chapter 6), and are currently explored in pilot projects.

Internal supervision and horizontal accountability is currently supported by a new procedure in which the VO Council and (very recently) the PO Council cooperate with the central data unit of the Ministry of Education (DUO) and with the Inspectorate of Education in "Windows for Accountability". Through this procedure schools obtain core statistical information on their own functioning from DUO and are supported to create school based indicators on, for example, parent satisfaction with the school. This development might be seen as a more structured and externally supported stimulation of school self evaluation, as compared to the more autonomous arrangements of earlier periods, which had somewhat disappointing results.

Teacher appraisal

In the Netherlands, the evaluation or appraisal of individual teachers belongs to the jurisdiction of the Competent Authority of the school, the school board, or the municipality.
Although the central role of teachers is fully recognized in current educational policy, there is no external teacher appraisal. Effective teaching is an important issue in school inspection, but it does not regard the functioning of individual teachers.
Few evaluative studies are available in which the way school boards and school leaders appraise teachers is described and evaluated.

Student assessment

Centrally specifying educational objectives and testing them by means of national tests is a theme that is not uncontroversial, given the principle of freedom of education and the traditional strong autonomy of schools in the Netherlands. Despite of the sensitivity of this issue the Netherlands has a central examination at the end of secondary education. For a long time educational attainment targets (Dutch: eindtermen) were only described in rather general terms. A fairly recent development is the formation of somewhat more specific "reference levels", or benchmarks. An important step is also the plan to implement a national school leavers test, by 2014, and to make pupil monitoring systems in primary and secondary schools mandatory.

The three most important instruments for student assessment in the Netherlands are: the Cito school leavers test at the end of the primary school period, the secondary school examinations,

which consist of a central and school-based part, and pupil monitoring systems in primary and secondary schools, the most important of which are also developed and supported by Cito.

The Quality Agendas and Action Plans to improve student achievement and achievement orientation of schools stimulate the use of formative and summative student assessments. A consequence of these policy plans is expected to be an increase in formative use of achievement tests, which are part of pupil monitoring systems, to diagnose and improve student learning and to improve the achievement orientations of teachers and schools. One could say that it is particularly at this micro level of teaching and learning that the improvement potential of assessment is at stake in a very concrete way. Experiences so far are promising, but also point at strong needs for professional development and external support to teachers, in order for them to learn how to work effectively with information from tests.

Responsibilities for evaluation and assessment

System level evaluation is mostly controlled by the Ministry of Education. Cohort studies are a joint venture of the Ministry of Education, the Central Bureau for Statistics and the organization for scientific research (NWO).
In school evaluation, autonomous schools have an important say. The recent legislation on "Good Education, Good Governance", has underlined the responsibilities of the school board, for quality enhancement as such, and internal supervision in particular. School level external supervision is the responsibility of the Inspectorate of Education. The educational organizations, united in the VO and PO Council have a supportive role in stimulating internal school supervision and horizontal accountability.
Teacher appraisal fully belongs to the discretion of the competent authorities of schools, i.e. school boards and municipalities.
Apart from the central examinations, student assessment belongs to the jurisdiction of schools. Instruments like the Cito school leavers test at primary level and the pupil monitoring systems at primary and secondary level are purchased by schools. Although the application of these instruments has become (as in the case of monitoring systems), or is becoming mandatory (as is the case of the primary school leavers' test), schools still decide about the particular instrument they want to use.

Implementation, appreciation and use of evaluation and assessment

Implementation of evaluation and assessment procedures in the Netherlands has sometimes hampered because of a lack of cooperation from schools in data collection procedures. This lack of cooperation first occurred in the 1970s when program evaluations were implemented and is also indicated by reoccurring problems for the Netherlands in obtaining sufficient response rates in international studies. For this reason, the Netherlands was excluded from the international

reporting on PISA 2000 and the first wave of the TALIS study (about the functioning of teachers). School autonomy and a general weariness of administrative burden might explain this phenomenon. Still, a large number of (autonomous) schools (85%) have purchased important student assessment instruments like the Cito school leavers test and pupil monitoring systems.

With respect to the implementation of school self evaluation a mixed picture emerges. It is the impression that schools generally own school self evaluation instruments, including administrative systems. Yet, the proportion of schools which, according to the Inspectorate, have a well-functioning internal system of quality assurance is not increasing at a level that was expected.

Systematic information on schools' appreciation of evaluation procedures is only available with respect to school inspection. Generally schools are satisfied with the work of the Inspectorate. An internal review by the Inspectorate pointed out that the recent risk based inspection is successfully being implemented and has shown results in the sense of a diminishing number of very weak schools.

The notion of evaluation and assessment stimulating the improvement of teaching and learning works differently for evaluation procedures at system, school, teacher and student level. In a general sense all types of evaluation and assessment, both summative and formative, are ultimately meant to improve educational achievement through improved teaching and learning. Feedback loops and improvement mechanisms will differ, however, both in length and in the role of different actors in using evaluative information for improvement purposes.

With respect to the use of system level evaluations there is only fragmented and anecdotal evidence available. The availability of periodic synthetic publications such as the annual report of the Inspectorate and the publications on Trends and Key Figures must be seen as an important condition for facilitating the use of system level evaluations. Since the reporting of the Parliamentary Committee "Dijsselbloem" in 2008, public interest in the position of the Netherlands on international assessment tests, such as PISA, seems to have grown, and has been the object of some debate in the press.

Research studies point out that the extent to which schools implement self evaluation procedures and use self-evaluation results for school improvement is often superficial and problematic.

Similarly, recent and ongoing studies into schools using student achievement data to improve teaching and learning (which is motivated by Departmental action plans) point out that teachers often lack required skills and expertise to make optimum use of these data. The good news is that these practices can be considered as touching the core of what evaluation and assessment can do for improving teaching and learning, and that current improvement and evaluation policies in the Netherlands are addressing this very core.

Policy initiatives

With respect to system level evaluation the Netherlands has a broad range of procedures and instruments available and the continuation of these procedures and instruments seems to be guaranteed.

At the school level, evaluation procedures are being aligned to new governance arrangements, which may be more efficient, include less administrative burden for schools and which may provide more effective support.

Finally, the most important recent policy initiative is probably the current orientation and action planning with respect to educational quality, including the stimulation of achievement oriented work by schools. Among others this is a strong stimulant of the formative use of results from achievement testing and pupil monitoring by teachers and schools.

CHAPTER 1: THE SCHOOL SYSTEM

In this chapter the structure of schools and the responsibilities of different administrative levels for different domains of educational functioning will be described. Specific attention is given to the place of high stakes testing and examinations. In a final section some historically grounded features and recent trends, relevant to the application of evaluation, appraisal and assessment, will be touched upon.

1.1 The structure of the Dutch school system

Education is compulsory from the age of 5 to the age of 16; but pupils can (voluntary) enter primary education at the age of 4.
The Dutch education system is divided into three levels: primary, secondary and tertiary education. These levels include the following types of education:

- Primary education:
- primary education for children aged from 4 to 12 years
- special primary education for children (aged 3 to 12) who require special educational treatment; for older pupils in this category, there is also special secondary education

Primary schools in the Netherlands cater for children from four to twelve years of age. They are usually arranged into eight year groups.
Children in need of special care can attend special schools.

- Secondary education:
- secondary education caters for pupils between 12 and 18 years.

Secondary education is divided into:
- pre-vocational education (VMBO) and individualised pre-vocational education (IVBO), 12 - 16 years
- junior general secondary education (MAVO), 12 - 16 years (the MAVO was abolished in 1999 and integrated into pre-vocational education)
- senior general secondary education (HAVO), 12 - 17 years
- pre-university education (VWO), 12 - 18 years

Evidently the system of secondary education is strongly stratified. After primary education, pupils move to one of the type of schools described above. Selection is informed by teachers' advice and achievement levels within primary education. The brightest students attend HAVO and VWO,

while less proficient students go to VMBO schools. From 1993 onwards, secondary schools shared a common curriculum during the first three years of (the so called "*Basisvorming*" or basic general education). The introduction of such a curriculum could be seen as an attempt to introduce comprehensive schooling. Despite this attempt most secondary schools stream their students in the first three years of schooling and, in doing so, foreshadow the tracking of students at the upper secondary level. In this respect the attempt to introduce comprehensive schooling was not successful. The failure of the "*Basisvorming*" is documented in the report of the Parliamentary Inquiry Committee Educational Innovations, in 2008.

At upper secondary level, different educational tracks include HAVO, VWO and a vocational track of upper secondary education (see below).

- secondary vocational education (previously divided into senior secondary vocational education (MBO) and apprenticeship training), 16 to 20 years, divided into four levels of training:
  (1) training to assistant level, 6 months - 1 year
  (2) basic vocational training, 2 - 3 years
  (3) professional training, 2 - 4 years
  (4) middle-management training, 3 - 4 years, or specialist training, 1 - 2 years

Level 2 (basic vocational training) is deemed to be equivalent to a basic qualification: the minimum qualification that anyone should have on entering the labour market. Holders of a basic qualification are capable of carrying out relatively complex routines and standard procedures within their own field of work. Level 1 (Assistant level) is for those students who are not able to obtain a basic qualification, giving them the opportunity to obtain some sort of qualification nonetheless. Compared with workers with level 2 qualifications, assistants will carry out less complex procedures, usually requiring a less rapid response. Holders of level 3 qualifications (professionals) will have responsibilities over and above their own duties. They must be able to account for their actions to colleagues and monitor and supervise the implementation of standard procedures by others. They will also be capable of devising preparatory and supervisory procedures. The fourth level (middle management or specialist) requires non-job specific skills such as tactical and strategic thinking and involves responsibilities in keeping with such skills.

There are two learning pathways at each level.
- block or day release (equivalent to the old system of apprenticeship training, with practical in company training taking up at least 60% of the course);
- vocational training (equivalent to the old senior secondary vocational education, with practical in company training taking up between 20 and 60% of the course).

- Adult education

The purpose of adult education, unlike vocational education, is not to train students for a particular occupation but to provide a solid foundation for vocational and secondary education courses and to enable adults to participate in society (social and life skills).

The following courses are offered at various levels:
- adult general secondary education (VAVO)
- courses providing a broad basic education
- Dutch as a second language
- courses aimed at fostering self-reliance to improve general social skills

Adult general secondary education gives adults (18 years and older) a second chance to obtain MAVO, HAVO or VWO qualifications in one or more subjects.

Courses providing a broad basic education may, for example, have a qualifying level equivalent to completion of the first stage of secondary education. They are not intended to qualify students for the job market but to provide a basis for further education.

A key area of adult education is the teaching of Dutch as a second language. Courses of this kind are designed to bring the language skills of non-native speakers up to an acceptable level. Newcomers to the Netherlands are obliged by law to attend a social integration programme at a Regional Training Centre, during which they receive not only Dutch language lessons, but also training to help them to cope with the Dutch way of life.

Finally, there are courses of a general introductory nature that aim to give students the minimum language, numeracy and social skills necessary to get by.

- Tertiary education:

Higher education is divided into:
- higher professional education (HBO)
- university education (WO)
- open higher distance education (Open University)

Given the scope of this report, no further information will be given on tertiary education.
A scheme of the Dutch education system, excluding adult education, is presented below.

| WO | HBO | MBO |

Age 18

Part-time

| VWO | HAVO | VMBO | PRO | VSO |

| VO 1/2 |

Age 12

Full time

| BAO | SBAO/SO |

Age 4

VVE

Age 0

Compulsory education
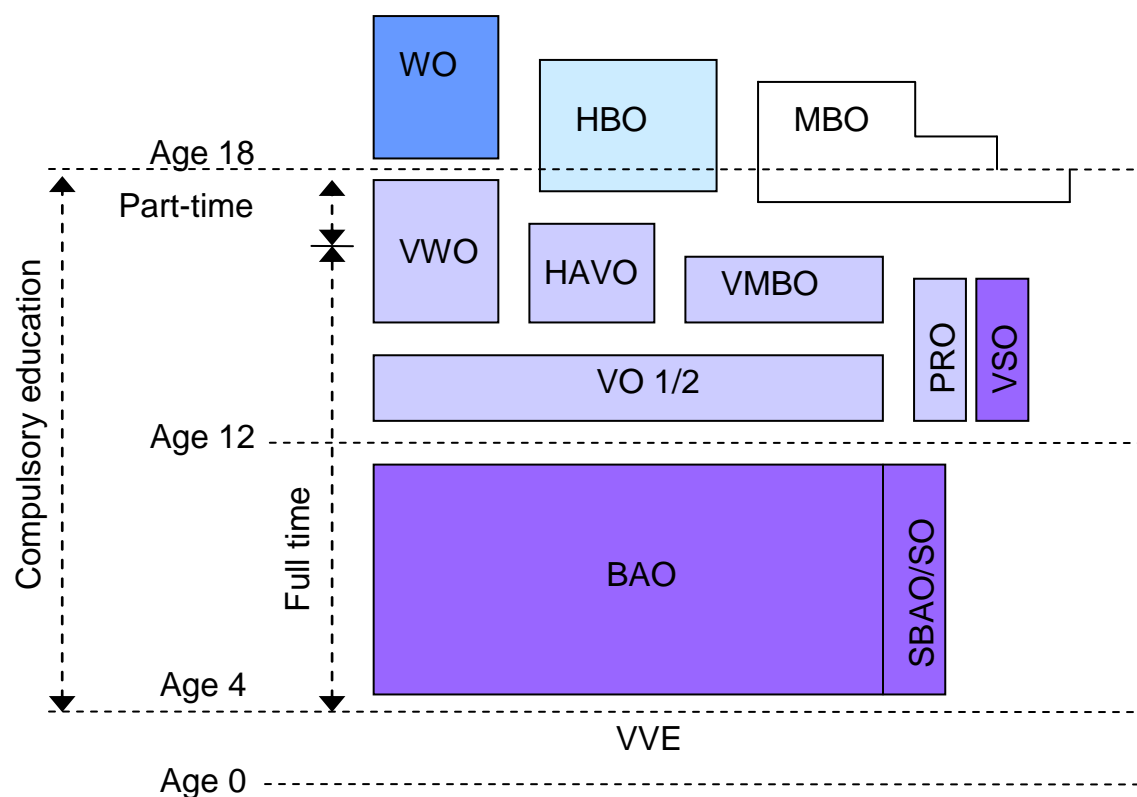
BAO	Mainstream primary education
HAVO	General secondary education
HBO	Higher professional education
MBO	Vocational education
PRO	Practical training
SO	Special education
VMBO	Pre-vocational secondary education
VVE	Early childhood education
VWO	Pre-university education
WO	University education

Figure 1.1: The Dutch school system

1.2   Number of primary and secondary schools

In 2010 the number of primary schools and, different types of, secondary schools are indicated in the table below.

Table 1.1: *Number of schools and students per school category*

School category	Number of schools in 2010	Number of students enrolled (number x 1.000) in 2010
Primary	6848	1.653,3
Secondary VMBO	646[1]	147,0
Secondary HAVO		151,1
Secondary VWO		164,8
MBO	118	525,4

1.3   High stakes testing and examination at primary and secondary level

Most primary schools (in 2012 this amounted to 85% of all primary schools) administer the "school leavers' test primary education" (*eindtoets basisonderwijs),* developed by Cito (the Central Institute of Test Development) at the end of primary education. The prime function of the "Cito test", as it is usually called, is to support teachers in advising students and parents on the most suitable track of secondary education. The strong stratification of secondary education in the Netherlands explains the high usage of the Cito test in primary schools. Research has shown that the test score closely matches the actual advice that students get by the school. Driessen (2011)[2] reports a correlation of .86 between test score and the advice by the school. This strong correlation applies to all student groups, regardless of their ethnic background. Studies from De Boer et al (2007) and Driessen et al. (2007) show no indications of migrant students in the Netherlands receiving a lower or higher advice for a track in secondary education, compared to native Dutch students. Gradually the Cito test, i.e. the school aggregate score, is also being used as a basis for school evaluation. School scores on the Cito test are used for School Inspection, and as a basis for external school evaluation by municipalities and school boards.

At the end of secondary education students are to take a set of final examinations in a number of subjects within a profile that the student has chosen. The final examination is divided into two parts: a school examination and a national examination. Dutch language is a compulsory subject in the national examination in all types of secondary education.
English language and some form of mathematics are compulsory elements in the national examination in pre-university and senior-general secondary education. Other compulsory subjects depend on the profiles (pre-university and senior-general secondary education) or type of vocational training the student has chosen. Schools are free to define the form of the school examination. Sometimes it is an aggregate of earlier accomplished tasks and partial tests,

---

[1] This it the total of VMBO, HAVO and VWO, school types that are often integrated. Source: http://www.stamos.nl/index.rfx?verb=showsectors

[2] Driessen G., (2011) Onderwijsadvisering van allochtone leerlingen. Radboud Universiteit, ITS http://www.onderwijsinspectie.nl/actueel/publicaties/Onderadvisering+van+allochtone+leerlingen.html

sometimes it is a specific testing event, designed by the school itself, or bought from an external examination institute. The Central Examination takes place at a fixed day, and is externally administered. For 2014 an arithmetic test will be added to the Central Examination.

A current trend is to weigh the Central Examination heavier than the School Examination in determining the passing or failure of a student. Students have to reach a minimum standard in the basic subjects on the Central Examination in order to pass. When school results on examinations are used for external school assessment, e.g. by the Inspectorate, the discrepancy between the results on the School and Central Examination is taken into consideration. It is one of the current quality standards to minimize this discrepancy.

1.4   Division of responsibilities

The Dutch school system is characterized by a large segment of government dependent private schools. About 2/3 of all primary and secondary schools are government dependent private schools; these schools often have a specific religious profile.. School governance and responsibilities for specific domains of functioning are largely similar for the public and (government dependent) private schools, and differences do not appear to be relevant for the subject of this report.

Decision-making and responsibilities for organizing education are too a large degree located at the local school level; schools in the Netherlands are among the most autonomous in the world. In 1998, 73% of all decisions in areas like instruction, planning and structures, resources and human resources were taken by the school, 4% was taken at the local level and 24% by the central level (source: Education at a Glance, OECD, 1998, p. 299). In 2008, the percentage of decisions taken by the school had risen to 94%, whereas only 6% of the decisions were taken at the central level (source: Education at a Glance, OECD, 2008, p. 488). School autonomy in the Netherlands is very high in all functional domains that were considered in the OECD study: organization of instruction, planning and structures personnel management and resources. For the subject of this report it is important to note that setting examinations is one of the areas in which the central level remains ultimately responsible, although the technical process of developing examinations is delegated to organizations like CvE, Cito and SLO (Foundation for Curriculum Development).

School autonomy and "freedom of education" is also manifested in the domain of curriculum development. At the central level so called "core objectives" are established. For example, for secondary education 58 core objectives, covering all subjects have been formulated. Schools are however responsible for deciding on how to implement these core objectives in their own school curriculum. In the recent past the "Education Council" (*Onderwijsraad)*, which is the central advisory committee, has made a case for more specific standards. As a result, so called "reference levels" were developed which indicate an emerging trend towards a more detailed description of

educational objectives; although this is a very sensitive issue in the Netherlands (see, for example, the most recent Advice of the Education Council, 2012).

Financing of general secondary schools was adapted in 1998, when very detailed regulations on financing of schools were replaced by block grants. The lump sum per school is determined on various ratios, all of them ultimately depending on the number of students. How schools allocate their budget however has to meet certain requirements , which are subject to accountancy control.

1.5   Conclusion: some specific features and relevant trends

School autonomy in the Netherlands can be seen as rooted in the principle of "freedom of education", which was regulated in the constitution in 1917. The consequence of this constitutional arrangement was the creation of equality between state schools and private denominational schools in for example the requirements to establish schools and the financing of schools. This arrangement also resulted in the institution of a new unofficial intermediary power structure of denominationally grounded organizations and pressure groups that represented teachers and school governors. Some analysts (e.g. Leune, 2007)[3] have described this structure as corporatist. Part of this intermediary structure is also a substantial educational support structure, which is also to some extent organized according to denomination. This support structure has a particular position in the national governance structure of education as it is controlled by educational organizations and individual schools and not by the central government, i.e. the Ministry of Education. During the last decade this intermediary structure of educational organizations has however been concentrated and secularized in the form of the PO, VO and MBO Councils. These councils have actually taken over the role of employers in their respective sectors. Additionally they also support schools in improving and innovating their education; some of these initiatives have also been placed in separate institutions and organizations, accommodating improvement of both primary and secondary schools (e.g. "*scholen aan zet*" – initiative to schools).

Major trends in educational governance in the Netherlands during the last twenty years are a shift towards more school autonomy, and more local decision-making, particularly in areas of finance and personnel policies and a transformation of educational organizations functioning at the intermediary level, who still have a lot of influence in educational policy. In areas like curriculum, accountability and quality control, however, there is a tendency to decrease the autonomy of schools and implement more centralized arrangements. Particularly in areas of defining and evaluating final outcomes of schools, central government has recently implemented additional legislation on performance standards/reference levels and centralized national testing. This latter trend is manifested in the Quality Agendas of the Ministry of Education, more prescription and

---

[3] Verstandig onderwijsbeleid, Antwerpen/Apeldoorn, Garant-Uitgevers N.V., 2007 (102 blz.) (ISBN 978-90-441-2110-0).

weight of Central examinations, and more specific "end terms" in the sense of reference levels. The Inspectorate of Education has been a more constant factor in these slowly changing organizational arrangements, holding its central role of supervising the quality of education, notwithstanding the fact that in the Netherlands the quality of education is seen as a responsibility of the individual school, i.e. the school board.

CHAPTER 2: THE FRAMEWORK FOR EVALUATION AND ASSESSMENT

2.1   Current approach

General Orientation

As a basis for an analysis of context, key factors and policy responses with regard to the evaluation system in The Netherlands, the conceptualization of educational monitoring, evaluation and assessment by Scheerens, Glas and Thomas (2003)[4] can be used as a starting point. This framework encompasses all evaluation, appraisal and assessment forms that are mentioned in the OECD guidelines for the review, and provides specific directions with respect to use and application. The conceptual framework is based on the distinction of three basic functions of evaluation (used as the general, overarching term to cover program evaluations, systemic monitoring, appraisal and assessment): *accreditation/certification*, *accountability* and *improvement*. As a second dimension, three basic data strategies are distinguished, test and assessment data, administrative data and educational statistics, and data based on systematic inquiry and review. Various levels of the educational system, the national system level, the school level, the teacher level and the student level are used as a third dimension (see Table 2.1).

Table 2.1: *Overview of M&E types; MIS means Management Information System*

Data Source	Test and assessment data			Administrative data; statistics			Systematic inquiry and review		
Function *Object*	**Account ability**	**Improve -ment**	**Accredi- tation**	**Account ability**	**Improve -ment**	**Accredi- tation**	**Account ability**	**Improve- ment**	**Accredi- tation**
System	Nat. Int. Assess- ment			MIS	MIS		Internat Review panels	Internat Review panels	
Program	Formative and summative evaluation of outcomes and processes using various data sources								
School	School Perf.- Report.	Test- based school self-eval.	School accreditat ion/ audits	School MIS	School MIS		Inspec- tion	Inspec- tion School Self Eval.	Quality Audits
Teacher	Assess- ment of competen cies		.	School MIS	School MIS		Inspec- tion	Inspec- tion	
Student		Student monito- ring syst.	Exams		School MIS			Mon. of behavior by teach.	

The different cell-entries are listed as distinct evaluation types in Table 2.2, which also gives a first impression of the coverage of these evaluation types in the Netherlands.

---

[4] Scheerens, J., Glas, C., & Thomas, S.M. (2003). *Educational Evaluation, Assessment and Monitoring, a Systemic Approach*. Lisse: Swets & Zeitlinger.

Table 2.2: *Evaluation types; coverage in the Dutch educational system*

Different types of educational evaluation	Present in the Netherlands' evaluation and assessment framework?
Test and assessment based types:	
1. national assessment programs	Yes
2. international assessment programs	Yes
3. school performance reporting	Yes
4. student monitoring systems	Yes
5. assessment-based school self-evaluation	To a limited degree
6. examinations	Yes
Two basic kinds of monitoring systems that depend on statistics and administrative data:	
7. system level Management Information Systems	Yes, educational statistics, Key Figures
8. school Management Information Systems	Used by unknown quantity of schools
The following forms depend on data from expert review and systematic inquiry:	
9. international review panels	No
10. school inspection/supervision	Yes
11. school self-evaluation, including teacher appraisal	Yes, without teacher appraisal
12. school audits	Only by Inspectorate
13. monitoring and evaluation as part of teaching	No quantitative information
Program evaluation is distinguished as a 14th form that may use various and mixed data strategies.	Mixed use of partial evaluations and system level monitoring studies

The overview in Table 2.2 shows that test and assessment based types of evaluation are well represented in the Netherlands at the system, school and individual student level. Noteworthy is the strong participation in international assessment programs by the IEA and OECD as well as participation in the OECD indicator project INES. Test and assessment based data are used for all evaluation functions: certification, accountability and improvement. Assessment-based school self evaluation is however a weak area which has seen relatively little improvement over the last years, despite of potentially strong and relevant data from student assessments.

Educational statistics, including descriptions of statistics on school careers of students, are used for monitoring at system level, and published in periodic publications on key figures and trends. Schools use (computerized) administration systems, which sometimes have the potential of being used as Management Information Systems. Incidental quality reviews (Visscher, 1998, Branderhorst, 2005) however indicate that actual use of those systems for reviewing educational quality is only done sporadically and is hampered by bugs and technical problems. More recent

quantitative overviews on the level of application of such systems by schools were not found, but an analysis of available systems shows an increase in their availability, particularly because existing and frequently used pupil monitoring systems also allow for analyses of trends in student achievement at the school board level. The quality of available systems has also been improved to enable a more stable use of such systems in schools.

The last international review of Dutch education was an OECD Review that took place in 1989, the final report was published under the title "Richness of the Uncompleted; Challenges facing Dutch education" (Ministry of Education, 1989). Since then no international reviews of Dutch education have taken place.

School inspection has a long tradition in the Netherlands. From the late nineteen eighties onward, inspection was structured by means of explicit Supervision Frameworks, consisting of standards and indicators, and supported by observation check-lists for school visits.

School self-evaluation is regulated in the Law on primary Education (art. 10 and 12), which requires schools to produce an annual report, a school plan and a school guide. Schools are required to report to parents on their goals and the (results of) educational processes in their school once every four years in their school plan and school guide. The annual reports of schools mostly include reports on financial indicators and do not provide indications on the educational quality in the school. See: http://www.schoolgids.org/schoolgids_wettelijke_eisen.html

At the turn of the century, school self evaluation was strongly stimulated by two dedicated, semi-independent bodies, founded by the Ministry of Education, known as *Q Primair* and *Q*5.* School self-evaluation is also stimulated by the Inspectorate, first of all as one of the quality aspects of schools that are monitored, on the other hand as part of the concept of *proportional inspection.* Proportional inspection was implemented to increase the efficiency of inspection, and implies that schools with adequate quality care and self evaluations will be inspected less frequently and intensively. The "rise and fall" of school self-evaluation is further documented in Chapter 4 on School Assessment.

The obvious "white spot" in the set of arrangements for evaluation and assessment in Dutch education is a lack of (external) teacher appraisal. Only very recently did the Inspectorate started reporting on the personnel policy of schools and how schools evaluate and improve the quality of their teachers, in addition to a school level assessment of the quality of teaching. Judging individual teachers belongs to the jurisdiction of the Competent Authorities of the schools, i.e. school boards and municipalities.

Little systematic information is available on how teachers use evaluative information to inform and improve their teaching. Older work by Janssens (1986) and quite recent work by Visscher and

Ehren (2011) indicates that teachers tend to have difficulty in applying systematic student evaluation and using the outcomes to improve teaching practice.[5]

Program evaluations have occurred in various forms and shapes during the last four decades. These evaluations have evolved from fairly big and well documented projects, to a large set of smaller studies and ongoing "monitors". The historical overview is presented in Chapter 3 on System Evaluation.

The framework presented in Table 2.1 can be used to obtain an impression of the completeness in coverage of evaluation provisions in a country, as has been done in Table 2.2. The framework however does not provide information on the coherence and efficiency of the entire set of arrangements. Currently, this set of arrangements cannot be seen as the result of a comprehensive design. Instead it has developed from various discrete and relatively independent backgrounds, partly driven by ideas on educational governance, rational planning, functional decentralization and subsidiarity, and partly driven by the availability of institutions (such as the Inspectorate) and technology; technology for educational testing in particular.

When looking at the current set of provisions some kind of integration can be discerned, as far as the use in aggregated form of student assessment data, and examination data is concerned. This data, originally collected at the individual level, is being used for school evaluation and system level evaluation as well. In the very recent plans for teacher appraisal, in a context of merit pay, value added student performance data are also envisaged to play a role at teacher level.[6]

A more specific overview of system level, school level, teacher level and student level evaluation and assessment
Below we provide a schematic overview of the major evaluation and assessment procedures at system, school, teacher and individual student level. The overview is a summary of the contents of the subsequent chapters of this report.

---

[5] Janssens, F.J.G. (1986) *De evaluatie praktijken van leerkrachten.* Rijks Universiteit Groningen, dissertation
[6] The plans for experiments with merit pay have recently, after the fall of the Cabinet, been abolished

System evaluation

Type of evaluation	Short description	Formal responsibility	Implementation and use
Policy & program evaluations	Evaluation of educational policies and programs	Minister of Education, in one occasion Parliament	Strong resistance from the field against early program evaluations. Little evidence on actual use.
PPON	Periodic national assessment primary schools	Central Test Agency, Cito	Relatively low profile.
Cohort studies	Achievement and attainment indicators of cohorts of primary and secondary school students	Joint responsibility of the Ministry of Education and the Foundation for Scientific research, NWO	Question marks with respect to use by education policy planners and schools
Annual report Inspectorate	Comprehensive report on the state of education	The Inspectorate of Education	Relatively high profile for policy use. Modest press coverage
Monitors	Partial effect and evaluation studies contracted out by the Ministry of education, some of them longitudinal (monitors)	The Ministry of Education	Extensive information, no clear evidence about synthesis and policy use
Key data, "trends in beeld"	Comprehensive annual reports containing key data and indicators	The Ministry of Education	Appear to have high potential for policy use, given active dissemination and user friendly formats
International indicators & international assessments	Dutch participation in IEA, EU and OECD studies	The Ministry of Education	Have obtained high profile in public debate on education concerning the quality of education

School assessment

Type of evaluation	Short description	Formal responsibility	Implementation and use
School self evaluation	Internal quality care by schools	Schools	Hampering implementation, substantial underutilization
School Inspection	Systematic school supervision, using structured formats and check-lists	Inspectorate of Education	No implementation problems, schools have a positive attitude to inspections
Quality cards	User friendly set of key indicators on school functioning to inform general public and parents. Recently reduced to an indication of the inspection regime a school has to follows, which is indicative on good or failing performance	Inspectorate of Education	Disappointing use by parents for purposes of school choice.
Windows for Accountability	Information dossiers on each school, consisting on centrally delivered quantitative indicators and qualitative indicators provided by schools	A new foundation resorting under the Councils for Primary and Secondary Education, as of 2012	No use and impact information available as yet.

Teacher appraisal

Type of Evaluation	Short description	Formal responsibility	Implementation and use
New initiative Inspectorate to appraise teachers	Inspection of personnel policy of schools and the quality of teaching in a school; classroom observations in a national sample of schools	The Inspectorate of Education	Results are published in the Annual Inspection Report
Within school teacher supervision	Individual teacher appraisal by school leadership and governance	The competent authorities of the school	No systematic information available

Student assessment

Type evaluatie	Short description	Formal responsibility	Implementation and use
examinations	Formal assessments at the end of secondary schools for purposes of individual certification	The Ministry of Education, with delegated responsibility to the CVE and Cito. Schools, monitored by the Inspectorate are responsible for the internal school examination	Implementation is obligatory. Use and application is straightforward.
Cito test	The Cito test is a school leaver test at primary school level, used by 85% of schools.	Schools are responsible for taking part. Cito takes care of technical aspects.	The test is used in supporting students' choice of a specific secondary school track. In aggregated form, use for school and system level evaluation.
Cito LVS	A pupil monitoring system for primary schools, all grades and broad coverage of subjects.	Schools are responsible for taking part; i.e. they buy into the system. Cito takes care of technical aspects.	Tests are used for didactic diagnosis and formative student assessment. In addition aggregated data are sometimes used for school self evaluation. Actual use by schools is still far from optimal.

The overview confirms the previous conclusions about the balance of evaluation procedures across the various levels of application (system, school, teacher, and student). At the system level, a broad range of evaluation procedures and instruments exists, including important efforts to synthesize some of the available school level information in the Annual Inspection Report and in the Key Figures and Trend publications. The use of cohort studies and smaller scale effect studies and monitors is less easy to capture.

There a vast number of activities and initiatives can be discerned at the school level. A constant factor is the systematic framework for school inspection, used by the Inspectorate, although the scope and focus of application has changed over time. While school self evaluation is still a formal requirement, see the earlier reference to the school guide and school plan that schools have to produce annually, and remains an important quality aspect in the supervision frameworks of the inspectorate, the hey-days of strong stimulation seem to be over (see Chapter 4). The focus of self-evaluation has changed from an overall evaluation of educational practices in the school to

analyzing and using data on the school and classroom level to improve instruction. Quality cards, in which the Inspectorate presented a summary of their assessment of individual schools, have also been reduced to just an overview of the inspection regime assigned to each schools. Perhaps the new initiative of "Windows for Accountability" will be able to replace school self evaluation and quality cards at the same time, but is still too early to say at present.

Appraisal of individual teachers is a practice that is hard to capture, as it resides under the responsibility of schools, and little systematic information on use and application is available. External appraisal of teachers does not exist in the Netherlands.

At the level of individual student assessment there are three well-established procedures in place: a formal examination system at the end of secondary school, a school leaver test at the end of primary school which is administered in 85% of the schools (which will be obligatory as of 2014) and an IRT based pupil monitoring system at the primary and secondary school level.

Main objectives and purposes of the Dutch system of educational evaluation and assessment are not officially stated in some kind of overall planning document. Instead, they may be inferred from the current practices that have been indicated above, and will be further documented in the remaining parts of this report. On a general level one could say that the three major functions discerned by Scheerens, Glas and Thomas, 2003, *certification, accountability* and *improvement* are all represented as goal areas.

The ambition to make evaluation instrumental to policy was perhaps most evidently present in the period of the large scale, centrally initiated, innovation programs and their evaluations in the 1970s. More recently, during the last five years, there is a new upsurge in the ambition for evidence based educational policy, and the climate for the use and application of student assessment data for educational policy making and improvement of educational practice has improved. All this will be documented further in the subsequent chapters.

Responsibilities for educational evaluation and assessment are about evenly divided across administrative levels. A fair amount is controlled by the central level, e.g. the Ministry of Education and the Inspectorate. Another essential part, student assessment is ultimately controlled by schools, although supported by external institutions, such as Cito. Schools are also responsible for their own quality control and are obliged to to draw up a plan for assessing quality, possibly by school self evaluation. Finally schools are responsible for teacher appraisal. Organizations at the intermediary level of education (PO and VO council) also control evaluation of schools to some extent, particularly through the recent program of "Windows for accountability" (see Chapter 4).

The different components of the set of arrangements for evaluation and assessment can be seen as loosely coupled. Synergy in the use of evaluation and assessment at the different evaluation levels (student, school and system) is enhanced through the use of aggregated student attainment and achievement data to inform school and system evaluation. The only area where the efficient use of

evaluation and assessment information is questionable is perhaps the large quantity and diversity of system oriented evaluation and monitoring projects; the added value of cohort studies is for example not obvious given the large number of other types of studies.

As far as the expertise for evaluation and assessment and knowledge management are concerned, the following summary statements are in place. Further details are given in other chapters, particularly chapter 3 on System Evaluation:
- the Netherlands has high expertise in test development, examination technology and psychometrics; most of that expertise is concentrated at Cito, but also in the RCEC centre, which is a structural collaboration between Cito and the department of educational measurement at the University of Twente;
- a number of specialized research institutes in educational research have expertise in educational evaluation research and monitoring as well;
- in the past, know-how on program evaluation was stimulated by the work of the interdepartmental Committee on Program Evaluation (CPE);
- the research school ICO has provided graduate training in evaluation research;
- the systematic approach of the Dutch Inspectorate had been a source of inspiration for other Inspectorates in Europe;
- at system level, the Knowledge Directorate of the Ministry of Education has a special role in knowledge management, and the stimulation of evidence based policy;
- active participation in OECD's INES project, on educational indicators has stimulated national work on educational statistics and indicators (e.g. the recent Trends publications);
- at school level there is increased attention for enabling and facilitating schools and teachers in learning to work with pupil monitoring systems, and interpreting data; this happens as part of the policy to stimulate result oriented work in primary and secondary schools.

Information technology plays an increasing role in the administration of the school leavers test at the end of primary school and in the use and application of pupil monitoring systems.

The use of evaluation and assessment results by schools, for purposes of improving school organization and teaching and learning is stimulated in various ways, but cannot be seen as a dedicated and explicit system wide *evaluation* policy. Yet, summative and formative use of tests by schools is an important part of the current *improvement oriented policy* of the Minister of Education, under the heading of Result Oriented Work. Also, recent plans for legislation on a national standardized test for all schools in primary and secondary education, requirements to implement a pupil monitoring system in primary education and inspection criteria to evaluate how school assess and monitor their students, highlight evaluation and assessment as an object of central concern. Most evaluation and assessment procedures that are based on data from schools and students, incorporate some kind of feedback, and reporting to schools. The most significant

and concrete efforts are the attempts to enhance the didactic use and applications of data from pupil monitoring systems, mentioned in previous sections.

2.2   Context

Over the last five years national educational policy has stated explicit standards and benchmarks to improve educational performance of students (and therefore also of schools and the system as a whole). This new policy is most evident from the Quality Agendas for primary and secondary schools, and the action programs aimed to improve student performance. One important facet of these policies is the aim to increase the achievement orientation of schools, and urge them to strive for excellence. Another important feature in recent educational policy is the ambition to make policy measures *evidence based.* This policy context has led to a tendency to give educational tests a stronger formal position and to give more weight to the central part of the examination as compared to the school examination (see Chapter 6). Also school inspection has become slightly more "high stakes" for schools, given the intensified monitoring of potentially failing schools and current mandate for the Inspection to install sanctions in schools that fail to comply to legislation or fail to provide sufficient educational quality for a long period of time. Yet, the funding of schools doesn't get stopped very readily.
Finally the formative use of student assessment is being stimulated by the policy to enhance schools' use of student achievement data to improve their education and outcomes of students.

Evaluation and assessment practices (practices rather than policies, since there are no explicit overarching evaluation and assessment policies)[7] in the Netherlands are somewhat aligned to the governance structure, characterized by strong school autonomy (see Chapter 1). The most important centralistic element is the examination at the end of secondary education. Next, school inspection also has a centralistic design, although it encounters constraints due to school autonomy, when it comes to the appraisal of individual teachers. "Summative" and formative assessment at primary level is controlled by schools, although the large majority of schools chooses the same tests, namely the "Cito test" and the Cito pupil monitoring system.

During the last two decades, the wide use of the Cito test and the pupil monitoring system by primary and secondary school have created larger acceptance in schools for the use of these tests.. Assessments and tests are however still not fully used for the purpose of data-driven teaching or school management. Even though tests are administered, the majority of schools and teachers only limitedly uses the results for improvement purposes.

Political forces that have impacted on the use and application of evaluation and assessment in Dutch education can be interpreted in the sense of the endorsement of the principles of New

---

[7] One might discern partial evaluation and assessment policies, however, like the intention to have a national standardized school leavers' test at the end of the primary school period by 2013.

Public Management, among others expressed in the so called Scheveningen Council (*Schevenings Beraad*, 1993). http://www.vosabb.nl/uploads/media/Lumpsum_retroperspectief_02.pdf Implication are, on the one hand, centralistic control of outcomes, while emphasizing at the same time that schools are responsible for their own quality, i.e. how they try to achieve these outcomes.

The framework for evaluation and assessment seems to be rather independent from evaluation arrangements in other social sectors. An exception is formed by certain aspects of school inspection. The ideas of proportional and risk oriented inspection have been developed in cross-departmental studies.

2.3   Initiatives and implementation

Recent intentions from the Minister of Education to make student assessment at primary level obligatory can be seen as the clearest recent example of strengthening the evaluation and assessment framework. Given the traditional strong degree of school autonomy this is a controversial issue. In a very recent Advice, the Educational Council (*Onderwijsraad*) has opposed to implementing one similar central test in primary education; according to their advice, schools should be able to choose a test that fits their curriculum and teaching best.

As pointed out before, a range of other characteristics of current educational policy (quality agendas etc.) are stimulating and enforcing the role of evaluation and assessments.

Major stakeholders in shaping evaluation and assessment policies are: the Ministry of Education, the "intermediary" organizations (VO and PO councils), teacher unions, and, more implicitly also the providers of evaluation procedures and technologies. The Netherlands has a tradition of negotiation among pressure groups and stakeholders in education. There are ample formal and informal channels for these stakeholders for interaction and debate.

The main difficulty in implementing empirical educational evaluations and assessments is a lack of cooperation from schools. This has been most overt and clear in the case of the early program evaluations in the 1970s and 1980s and in the great difficulties of obtaining a high enough response rate in international studies, like PISA. More covert is the lack of real use of data from school self evaluation and pupil monitoring in a large proportion of schools.

The policy priorities in the area of evaluation and assessment are a slight expansion of obligatory tests, a slight increase in the high stakes nature of inspection, and the stimulation of formative use of tests in primary schools as part of the policy of result orientation. One might say that the emphasis is not so much on expansion of evaluation and assessment procedures but at a wider and better use and implementation of results.

CHAPTER 3 SYSTEM EVALUATION

3.1 Current practices

In this chapter an overview will be presented of evaluation and assessment at the national level; describing the various types of system evaluation that have taken place in the Netherlands. Current practices will be placed in a historical perspective as shifts in educational policy have also implied different choices in evaluation and assessment at the national level.

3.1.1 Overall framework for system evaluation

Large scale policy evaluations in education were implemented during the 1970s, as part of the so called "constructive educational policy" of the social democratic government, in office during this period. Constructive educational policy was an attempt to centrally initiate educational innovations. Several experimental innovations were planned, which were assumed to be grounded in scientific evidence and which were subjected to independent external evaluation. The most well-known innovation program was the "Middle School Experiment", which was an attempt to introduce a comprehensive secondary school. Other innovation programs included "Open Schools" for adults with little initial schooling, integration of primary schools and nursery schools, and a dual form of vocational education (*Participatie Onderwijs).* In the same period several large cities had large scale innovation programs as well: Rotterdam (OSM), Utrecht (GEON) and Amsterdam (IpA), and some of these programs, particularly the ones in Rotterdam and Utrecht were subjected to program evaluation as well.

Analyses of the evaluation of the national innovation programs (Scheerens, 1983, 1993) pointed out that a lot of problems were experienced in implementing evaluation plans. This had to do with organizational aspects of the management of these programs, fuzzy program implementation, program innovation committees acting as political lobbyists for the programs in question and large resistance from schools and teachers towards external, quantitative evaluation research. Creemers and Hoeben (1985, p. 37) make the following observation, while looking back on this episode: "*Educational evaluators in the Netherlands have faced many difficulties during the last decade. While evaluating large scale innovations and curriculum development projects they were confronted with vague or unspecified objectives, with unclear and loose project organizations and with confusing responsibilities that were inherent in the implemented change strategies. Evaluation research was therefore caught between conflicting interests, could not be independent and gained results which had doubtful objectivity, validity or reliability. Most evaluation research was not even evaluative, but merely descriptive.*"

During the 1980s the ideals of constructive educational policy were replaced by a more piece meal approach, with a strong urge for economization and budget control. This was for example manifested by an increase in school size. During this period the emphasis in policy evaluation gradually shifted from program evaluation to a monitoring approach. This monitoring approach was enabled by the creation of new evaluation instruments such as periodic assessment in primary education (PPON, *Peridodiek Peilings Onderzoek)* and cohort studies in primary and lower secondary education.

In the early 1990s, educational policy was further shaped according to the principles of 'new public management", with a new emphasis on school autonomy, but combined with stronger accountability requirements. The well-known *Schevenings Beraad*of 1993 laid the foundations for this policy approach. This development was a strong boost for educational evaluation and assessment. Existing forms, like PPON and the cohort studies were enforced. Next, school inspection was intensified, and explicit "school supervision frameworks" were developed and implemented. At the school level, schools were obliged to produce annual reports and school plans, which were expected to motivate school self evaluation. Participation in international assessment studies of IEA and later PISA was stimulated, and the Netherlands took active part in OECD's education indicator project (INES).

The main emphases in educational policy for primary and lower secondary education during the 1980s and 1990s, some of them continuing into the first decade of the 21st century, were the policy to stimulate equity under titles like "Educational Priority Policy", which meant giving priority to schools with a sizeable low SES composition, the integration of nursery schools and primary schools, a national program to stimulate the use of information technology (the Informatics Simulation Plan), a program aimed at more inclusive primary education, with an attempt to diminish enrollment in special education at the same time (*Weer samen naar School),* shifts in the school structure of lower secondary education, leading up to the structure described in Chapter 1, continued policy to enhance equity (the so called *Achterstandsbeleid*), continued attempts to have at least some kind of integration of different school tracks during the first years of lower secondary education (the *Basisvorming*); reform of the didactic approaches in upper secondary education (the so called *Study House).* These policy emphases differed from the "top down" experiments in the 1970s in the sense that they were applied, not just to a small group of experimental schools, but across the board in all schools. Although, initiated from the centre, these policy initiatives were not implemented through detailed work programs, but limited to a set of policy priorities and budgetary conditions, while decisions on the actual implementation were left to the autonomous schools, assisted by the equally autonomous intermediary support organizations.

As far as the evaluation approach was concerned, all of these policy initiatives at some point were appraised by evaluation committees consisting of educational experts. The reform of upper secondary education was subjected to analysis and evaluation by a Parliamentary Advisory

Committee (the Committee *Dijsselbloem*). From a methodological perspective, without exception, the evaluation committees had to depend on fragmented sets of coincidental information. Only the Dijsselbloem Committee was able to use more systemic information from more structured sets of cohort studies, educational statistics and international assessment studies. None of these policy evaluations had a program evaluation design that could enable answering questions about outcomes being attributable to the policy programs in a straightforward way.

The last few years, national educational policy has taken a new turn in emphasizing quality and excellence, particularly in basic subjects. Detailed Quality Agenda's and Action Plans have been drawn up by the Ministry of Education, emphasizing professionalization of teachers, outcome control and incentives for excellent and efficient performance. Schools are urged to work in an achievement oriented way, among others, supported by formative student assessment. Explicit attainment standards have been formulated. An evidence based approach to policy initiation is propagated and supported by a new directorate, the "Knowledge Directorate", at the Ministry of Education. This new achievement orientation is accompanied by a renewed interest and expansion of existing forms of monitoring, assessment, evaluation and appraisal.

As far as responsibilities are concerned, the Ministry of Education has been and still is responsible for most system level evaluation procedures, in terms of initiation, finance and organization of the implementation. Procedures like PPON and the Cohort Studies are carried out rather independently from the Ministry. The Inspectorate of Education has a specific responsibility for monitoring the quality of education. The Inspection coordinated and partly carried out two of the major policy evaluations, namely the evaluation of primary education and the evaluation of the first phase of secondary education the "*Basisvorming*". Other evaluation committees resided directly under the Ministry. In addition the Inspectorate publishes a yearly report on the State of Education.

According to the Constitution (article 23, section 8, and in accordance with the Departmental Arrangement on School Inspection, April 22, 2003) the Inspectorate of Education publishes an annual report on the State of Education in the Netherlands. The report describes positive and negative trends in primary and secondary education and provides a summary of key indicators and facts in the inspection framework on for example the quality of teaching or quality assurance in schools. Societal trends and the information requests of the Department of Education partly inform the choice of additional topics in the report. The Inspectorate of Education uses the data on quality of schools collected in regular inspection visits of schools and additional relevant studies by external parties to write the annual report.

As stated before, the evaluation of the didactic innovation in upper secondary education, by the Committee Dijsselbloem, was initiated by Parliament.

Various organizations are involved in implementing and carrying out the various system level evaluations. The Netherlands have a rather elaborate structure of research institutes specialized in educational research. Some of these institutes are connected to universities, and some are private companies. Usually a number of research institutes compete for obtaining a contract to carry out evaluation projects. Next, Cito, as the institute for Test Development is responsible for the periodic assessment project (PPON), but Cito also takes part in the development of achievement tests in the Cohort Studies, which are carried out by a consortium of educational research institutes, Cito and the National Bureau of Statistics (CBS). The various Monitors are initiated by the Ministry and carried out by educational research institutes. Attainment indicators, "key data" and other relevant statistical data are developed by the Ministry of Education, currently this responsibility is placed in the Knowledge Directorate.

System level evaluations are articulated with respect to school evaluation and student assessment in the sense that the prime users of system evaluation are situated at the national level, the Ministry of Education, Parliament or the general public. From a technical perspective, school evaluation and student assessment data are likely to be used in aggregate form in system level evaluations.

3.1.2   *Procedures used in system evaluation*

In this section, each of the evaluation forms and instruments, mentioned in the introduction, will be briefly described.

3.1.2.1 Policy and program evaluations
Policy evaluations can be roughly subdivided in an early phase (1970s) when the evaluations tried to approach the design of program evaluations assessing the impact of experimental innovation programs, and a second phase of implementation (1980s and 1990s), when system wide policies were evaluated by committees in a more retroactive way. Over time the latter kind of policy evaluations could increasingly benefit from a range of system wide data collections, such as PPON, the cohort studies, information from school inspections, and data from international studies. Examples of program evaluations from the early phase are the evaluations of the Middle School and Open School Experiment (Scheerens, 1983). During the second phase, the major policy evaluations were the planning of the evaluation of the innovations in secondary education carried out by the Coordination Committee Evaluation Plan Secondary Education, final advice, 1987[8]), the evaluation of the creation of the "basic school"; report of the CEB (*Commissie Evaluatie Basisonderwijs,* 1994); the evaluation of the Informatics Stimulation Plan (CEI), the Final report of the Committee "Zegveld" (1988); the evaluation of the policy to counter inequality

---

[8] See Reaction by the Minister of Education:
http://resourcessgd.kb.nl/SGD/19861987/PDF/SGD_19861987_0004010.pdf

in primary education (*Achterstandsbeleid*); evaluation of the policy aimed at inclusive education *(Weer samen naar School,* Meijer and Peschar, 1995), evaluation of the integration in the first grades of lower secondary education, the evaluation committee "*Basisvormin*g", final report 1999[9] and the evaluation of the innovation during the second phase of secondary education by the Parliamentary Committee "*Dijsselbloem*".

As far as aims and standards are concerned, in a very general sense all of these policy evaluations tried to reconstruct aspects of policy implementation, and capture some of the intended outcomes. Re-occurring issues in these policy evaluations at primary and secondary level were: overall attainment levels, the overload in the curriculum, the position of disadvantaged students and the lack of integration of different educational tracks at lower secondary level. By way of illustration two major policy evaluations, the work of the CEB, and the Parliamentary Committee Educational Innovation, will be described in somewhat more detail.

Primary Education was evaluated between 1992 and 1995 as a result of an agreement in the Primary Education Law, to conduct a periodic evaluation of the implementation of the Law. The work of the committee was based on available studies, specifically designed short term research studies and consultations of key stakeholders. The committee identified problematic areas and designed solutions. The recommendations by the committee are summarized in the letter that the Minister sent to Parliament, together with the final report of the Committee http://resourcessgd.kb.nl/SGD/19931994/PDF/SGD_19931994_0002450.pdf

The following recommendations were made:
1. Counter the overload in the curriculum by concentrating on basics.
2. Stimulate a result oriented working culture in primary schools.
3. Initiate an action program to improve Dutch language education.
4. Stimulate instructional leadership in primary schools.
5. Create structural provisions to realize the ambitions of inclusive education.
6. Stimulate the implementation of systematic quality care in schools; among others by having Cito develop a school test that is aligned to the Periodic Assessment Tests (PPON).

The report of the "Parliamentary Committee Educational Innovation", usually indicated as "the Committee Dijsselbloem", after its chairman, was published in 2008. The Committee investigated innovations in Secondary education, including the "Basisvorming" (a structural reorganization of secondary education, introducing a new VMBO track), and the didactic innovation in upper secondary education which is generally referred to as the "Study House". The latter element appeared to be the trigger for having the parliamentary evaluation, as a lot of public discomfort

---

[9] Inspectie van het onderwijs (1999). *Werk aan de basis. Evaluatie van de basisvorming na vijf jaar.* Utrecht, september 1999. Also see the Commentary by the Education Council:
http://www.onderwijsraad.nl/upload/publicaties/428/documenten/advies_herzbasisvorm.pdf

was expressed, at the time, about the "new learning" (i.e. self regulated learning) propagated as part of the Study House. The Committee used hearings of (groups of) stakeholders and contracted out specific studies to collect data and draw evaluative conclusions.

With respect to the aims of the "Basisvorming", the Committee concluded that the overall aim of higher achievement levels was not attained (nor had attainment levels declined, for that matter). Next, the ambition of more flexibility and ease of transfer between different education tracks had not been realized. Concerning the didactic innovation in upper secondary education, the Committee noted a lot of confusion in the field and extensive resentment in schools regarding the perceived top down nature of the innovation. To the extent that the Committee was able to collect information on the actual implementation of "new learning", the results presented a rather mixed pattern. The Committee concluded that the scientific basis of the New Learning approach was not yet established.

Apart from looking at the implementation of the Innovations, the Committee also analyzed the general quality of education at secondary level. They collected, among others, views from participants and stakeholders about the quality of secondary education. The results were rather positive, in the sense that stakeholders gave relatively high marks for the quality of education. Next, the Committee analyzed the results of the Netherlands on international achievement tests. The position of the Netherlands has been consistently strong (see further information in a subsequent section of this report) on these tests, in all main subjects. Interestingly the Dijsselbloem Committee seemed to "talk down" the importance of these positive results, and, on the contrary emphasized a negative trend of slightly declining country average scores (see discussion further on in this report).

Recommendations of the Committee were, among others, to counterbalance the large autonomy of schools with a focus on the control of outputs of schools, to strengthen test and assessment provisions, to try and improve the flexibility and transparency within the tracked systems to enhance possibilities for students to move between tracks, and, as far as future educational innovation policy is concerned, to follow an evidence based approach.

In the context of this report it is interesting that the Dijsselbloem Committee, like the evaluation committee on primary education in 1995, recommended stronger and more systematic arrangements for educational evaluation and assessment, including new mandatory achievement tests.

3.1.2.2 Periodic National Assessment (PPON)

The Periodic National Assessment, PPON is being conducted since 1987. It consists of periodic assessments on a sample of primary schools, testing students in group 8, the last grade level of primary school, and sometimes also in group 4. The Ministry of Education contracted Cito to implement PPON. The aim of PPON is to inform the Ministry, and other educational organizations, on the attainment levels of students in primary education in the following subjects:

mathematics/arithmetic, language, world orientation (social science), English, musical education, physical education, traffic education and visual arts. PPON consists of achievement tests, administered to students and a school questionnaire to measure subject matter coverage, i.e. what has been taught in the respective subjects and other relevant school characteristics. Test taking is managed by test leaders, external to the school. The results of the periodic assessments are published in research reports, indicated as "Balances", which are distributed to the Ministry and to other organizations in the education sector. In addition, participating schools receive so called "school folders", in which the results of the school are summarized. General information on PPON is available on http://www.Cito.nl/onderzoek%20en%20wetenschap/onderzoek/ppon.aspx

From this web-site an overview of the results of the various waves of PPON in each subject is available. Reports on Dutch language, for example, have appeared in 2002 and 2005. Reports on reading skills were published in 2007 and 2008. Results in mathematics/arithmetic came out in 1999/2000, 2005 and 2012. The periodic administration of the PPON tests enables a comparison of achievement levels across time. Scheerens, Luyten and Van Ravens (2010, pp. 78-79), give the following overview of trends in the major subject matter areas:
"In Dutch language, no marked change in achievement was seen between 1988 and 1998, although there was a slight improvement in written skills between 1993 and 1998. Unfortunately, no comparable data is available for the period after 1998. While data is available for reading skills and comprehension, only very slight changes can be seen, both positive and negative.

It is numeracy and arithmetic in primary education which has drawn fiercest criticism in recent years, with many commentators lamenting the fact that many children seem unable to add up without the aid of a calculator. Twenty-one specific aspects of numeracy were monitored between 1987 and 2004. In fourteen aspects, performance remained consistent. The remaining seven aspects show two contradictory trends: an improvement in basic arithmetic (numbers and the relationships between numbers, percentages, mental arithmetic and estimation). A negative trend was observed in the so-called 'arithmetic processing' skills: addition, subtraction, multiplication and division and problem-solving requiring a combination of these processes. The test questions required students to show their problem solving strategy and how they arrive at the correct answer, making use of a standard algorithm. The decline in performance may be due to the fact that most students attempted to solve the problems by head rather than on paper.

The domain of 'world orientation' includes four subject areas: history, geography, general science and social awareness/citizenship. No significant changes were seen in the latter two subjects between 1991 and 2002. In geography, a decline in map-reading ability was noted over the period 1995 to 2001, while scores in other sub-disciplines such as topography and physical geography remained reasonably constant. In history, a slight decline in performance was noted in two of the four aspects: ancient history and temporal awareness or 'time-scaling' (which refers to a knowledge of major events and the order in which they happened).

In English at primary school level, there were minor changes in student performance between 1991-2006 in the areas of oral comprehension, reading and vocabulary."

The authors (ibid) conclude, based on these results, that very little change occurred in the attainment levels of students at the end of primary school in the major subject matter areas over a period of twenty years. The most recent outcomes of PPON, to be published in 2012, show improvement of attainment levels of students from groups 4 and 8 in arithmetic and Dutch language.

The PPON reports also include additional information on for example the comparative effectiveness of the main text-books used in schools, the influence of the socio-economic status and gender of the students, the age and experience of the teachers, and the average time that schools dedicate to teaching in different subject matter areas in question.

3.1.2.3 Cohort Studies
The idea to use cohort studies to inform educational policy evaluation (i.e. analyzing the educational attainment of age cohorts of students throughout their school career), in a context of educational policy evaluation originates from an advice by of the Professors Hofstee and Meijnen to the board of the Foundation for Educational Research (SVO), in the early 1980s. This advice was given in the context of a discussion on the use of evaluation designs for ongoing policy programs in education, such as the policy in primary schools to stimulate equity and the restructuring in secondary education. The authors concluded that quasi experimental designs were unfeasible and proposed cohort studies as a next best design.

Cohort studies in secondary education
Early cohort studies in education were carried out by the Central Bureau of Statistics (CBS) in 1977 (SMVO-cohort) and 1982 (SLVO-cohort). These studies had a more descriptive sociological orientation, in describing for example the relationship between socio economic background of the students and their educational attainment.

Basically the cohort studies consist of achievement testing of students in language and mathematics/arithmetic at various levels during the school career, recording data on students' progress throughout the school program, in terms of class repetition, drop out, transgression to another school type, and examination results, and the collection of school background data by means of questionnaires to parents, teachers and school heads. The first cohort study was the one for secondary education, the VOCL cohort study. As of yet, three subsequent cohort studies have been conducted, which are VOCL 1989, 1993 and 1999/2000. The year signifies the year in which students started their school career and the start of the cohort study.

Cohorts are based on representative samples of "regular" primary and secondary schools, excluding school for special education. Typically the sample of students is in the magnitude of 20.000 students (Kuyper and Van der Werf, 2007)

http://gion.gmw.eldoc.ub.rug.nl/root/2007/resultaten_vocl/?pFullItemRecord=ON

The test design of the 1999 VOCL cohort consists of an entrance test in the first school year, which resembles the Cito final test of primary education. At this stage an intelligence test is administered as well. Next, in the third school year an achievement test in mathematics and language is administered. Finally, examination results are registered of students in the cohort study.

The reports form the cohort studies provide detailed statistics on the achievement levels of students throughout their school career. Results are presented on grade repetition, drop out, percentage of students that obtain their diploma without delay, and movement of students across the different school tracks. At the secondary level, comparisons of these attainment statistics can be made across cohorts.

Next, the test results, compared across cohorts provide a picture on growth or decline of scores over time. Kuyper and Van der Werf (2007), for example, note slightly higher average scores on the third year test for the 1999 cohort, as compared to the 1989 and 1993 cohorts, particularly for mathematics. These authors found that, after controlling for background characteristics the differences between the cohorts are minimal. Attainment of students as a function of students' positions on school career ladders was higher for the 1999 cohort compared to the earlier cohorts.

http://gion.gmw.eldoc.ub.rug.nl/FILES/root/2007/resultaten_vocl/Resultaten_VOCL.pdf

Cohort studies in primary education

The first cohort studies in primary education are referred to as 'Primary Education and Special Education Cohort Studies' (Dutch abbreviation: PRIMA). These studies were implemented within the framework of the evaluation of the Educational Priority Policy (aimed at enhancing equity), in 1988 and 1992.

In a later stage, the primary school cohort studies were seen as having significance for educational research next to the original objective of policy evaluation. Subsequent waves of PRIMA data collection and reporting are presented in the table below, obtained from the Social and Cultural Planning Bureau, SCP, 2012.

http://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Primair_onderwijs_en_speciaal_onderwijs_cohortonderzoeken_PRIMA

PRIMA wave	Response
PRIMA 88	3,5% random sample + additional sample Educational Priority schools
	696 schools took part (56%)
PRIMA 94/95	789 schools; 14.432 pupils in group 2, 15.343 pupils in group 4, 14.041 pupils in group 6 and 16.054 pupils in group 8
PRIMA 96/97	622 schools; 18.113 pupils in group 2, 17.125 pupils in group 4, 14.879 pupils in group 6 en 13.847 pupils in group 8
PRIMA 98/99	602 schools; 16.830 pupils in group 2, 16.953 pupils in group 4, 14.992 pupils in group 6 en 14.182 pupils in group 8
PRIMA 00/01	653 schools; 16.181 pupils in group 2, 15.223 pupils in group 4, 14.945 pupils in group 6 and 13.888 pupils in group 8; 3610pupils from special education; [597 schools exclusive of special education]
PRIMA 02/03	600 schools: 420 schools in the representative sample and 180 schools in the additional sample of schools with many disadvantaged students; 15.681 students in group 2, 14.678 students in group 4, 14.106 students in group 6 en 14.006 students in group 8
PRIMA 04/05	600 schools: 420 schools in the representative sample and 180 schools in the additional sample of schools with many disadvantaged students; 16.060 pupils in group 2, 14.997 pupils in group 4, 13.998 pupils in group 6 en 13.847 pupils in group 8

The PRIMA-cohort studies include data at grade levels 2, 4, 6 and 8 of the primary school program, in the subjects language and arithmetic; similar to the VOCL cohort study, school background information was collected by means of questionnaires to parents, teachers and school heads. http://www.ru.nl/its/onderwijs/afgerond_onderzoek/vm/het_cohortonderzoek/

Integration of cohort studies; the COOL study
From 2007 onwards cohort studies in primary and secondary education have been integrated in the COOL (*Cohort Onderzoek Onderwijsloopbanen*) study.
http://www.cool5-18.nl/doelenopzetbo/#4; http://www.cool5-18.nl/scholenbo/

The COOL study comprises data collection in primary, secondary and vocational schools. In primary schools, data is collected at the level of groups 2, 5 and 8. In general secondary education data from students is collected in the third year, and in vocational education in year two. The subsequent waves of data collection for COOL have been planned until 2015. Student achievement is tested in language and mathematics/arithmetic. The achievement tests make use of

the Cito pupil monitoring system LOVS. In addition short questionnaires are administered to students.

The COOL study is presented as a data source for various user groups, such as the Ministry of Education, The Social and Cultural Planning Bureau and the Educational Council (*Onderwijsraad)*. Next, research groups can use the data base to carry out more in depth studies. The COOL study also comprises specific services to feed back information to schools. The study is jointly financed by the Ministry of Education and the Foundation for Scientific Research (NWO).

More recently, the pre-Cool Study has been started. This study, of which the first data collection has already taken place, will collect data from nursery schools and kindergarten.

3.1.2.4 The annual reports by the Inspectorate

The Dutch Inspectorate of Education is responsible for the inspection and review of schools and educational institutions:

- assessing the quality of education offered in schools
- reporting publicly on the quality of individual institutions
- reporting publicly on the educational system as a whole
- encouraging schools to maintain and improve the education they offer
- providing information for policy development
- supplying reliable information on education
- financial supervision and control

Since 2007, the Dutch Inspectorate of Education has carried out risk-based inspections of schools, assessing potential problems that could affect the quality of education. This system reduces the administrative burden as perceived by schools and makes inspections more efficient. Schools providing good education (no risks detected) and achieving good student outcomes are not scheduled for inspection visits, allowing the Inspectorate to focus on monitoring and improvement of potentially failing schools. More information: Risk-based Inspection as of 2009

According to the Constitution (article 23, section 8, and in accordance with the Departmental Arrangement on School Inspection, April 22, 2003) the Inspectorate of Education publishes an annual report on the State of Education in the Netherlands. Annual reports by the Inspectorate have been published since 1801.

The annual reports usually consist of two major parts: A general introductory chapter on the State of Education and the Supervision of the Inspection and a part in which the different educational sectors are separately described. In the most recent report on the 2009/2010 school year, a third part is added, dedicated to a number of specific themes.

English summary reports are available for [2008/2009](), [2006/2007](), [2005/2006]() and [2004/2005]().
By way of illustration a brief summary impression will be given of the Annual Report of 2011 and 2009/2010. The 2011 report is available online through:
[http://www.onderwijsinspectie.nl/binaries/content/assets/Onderwijsverslagen/2011/Onderwijsverslag+2009+2010+printversie.pdf]()

The Inspectorate of Education uses a supervision framework (which will be described in the next chapter on school assessment) , to assess the proportion of schools functioning above or below certain achievement standards. In the annual report of 2009/2010, strong and weak points of the school system were noted.

Favorable developments were:
- There are fewer weak and very weak schools in most sectors
- Pupils in primary education are doing a little better
- More student follow higher forms of education and this is why the general level of education is rising
- More students obtain a diploma, and there are fewer students who leave school without a qualification
- Schools and school boards are more and more aware of the importance of high attainment

Reasons for concern were:
- Students who are badly in need of good education are overrepresented in weak and very weak schools
- In secondary education lower marks are obtained on the central examination in the subjects mathematics, Dutch language and English than in previous years
- Schools and educational programs do not sufficiently guarantee the quality of diploma's
- There is a lot of variation in achievement levels between schools, which means that pupils are much better off in some schools as compared to others
- Students have the right to be taught by good teachers, but a part of the teachers is failing

As far as strategies for improvement are concerned, school boards and school leaders are expected to take on more responsibility, and more professionalization of teachers is required. The Inspectorate emphasizes a need to improve schools' and teachers' use of student achievement results to improve teaching, using clear standards and objectives, a structured teaching approach, and ample use of feedback from achievement testing. Finally a strong case is made for subject matter coverage and alignment of subject matter contents between school sectors.
The second part of the report provides analyses and evaluations of the respective school sectors: primary, secondary, special education, vocational and adult education and higher education. In the third part the following themes are addressed as separate chapters of the report: educational

attainment, the quality of educational governance, finance and legitimate spending of funds, teacher quality and the quality of education for students who need special care.

The annual inspection reports on the State of Education are quite comprehensive and detailed. The unique evaluative contribution that the inspectorate can offer are descriptions and evaluations of classroom teaching and school functioning that are generated through systematic school inspections, guided by explicit supervision frameworks, standards and indicators (see chapter 4). In addition, a wealth of data from other sources are used, such as examination results, school career data (based on the cohort studies), information on test results from PPON and the cohort studies, data from international assessment studies and from national educational research studies.

3.1.2.5 Monitors and small scale evaluative studies
Part of the empirical research studies that are contracted out by the Ministry of Education each year, have an evaluative orientation, in which partial aspects of educational policies, innovations or new legal arrangements are the object of investigation. The yearly report of the Ministry of Education of 2011, section Secondary Education, for example, provides an overview of studies "looking into the effectiveness and efficiency of policies", in which 31 ongoing or recently concluded projects are mentioned. A small part of these studies are assessing targeted phenomena at regular intervals, and could therefore be indicated as "monitors". Examples of such longitudinal studies from the list of 2011 evaluative projects in secondary education are: the second measurement of the study that looks into the mastery of "reference levels" (subject matter tied achievement standards), periodic research into school leaving in secondary education and the upstream monitor VMBO-HAVO and VMBO-MBO. By way of illustration the monitor on school leaving, described as the school-leaver survey is described as follows:
"Since the early nineties the Research Centre for Education and the Labour Market (ROA) has conducted research among recent school-leavers in most sectors of education in the Netherlands. The surveys among different groups of school-leavers have been strongly standardized, and the resulting data has been integrated into a single School-leavers Information System (SIS). This system is designed to function as a monitoring instrument for the transition from school to work, covering the full breadth of the Dutch education system. ROA bears the responsibility for the design, implementation and management of SIS" Source: http://www.roa.unimaas.nl/
At primary level a similar number of evaluation studies (31) are described in the same source. An example is the "Monitor on the implementation of the educational number", this is a unique code for each student. The monitor includes a study from 2007 to 2013. Nine more evaluation studies are listed in the domain of interest of the Directorate on Teachers of the Ministry of Education. An example is a study that runs from 2009 until 2012 on "Realization and effectiveness of remuneration measures that are part of the Action Plan for Teachers."

3.1.2.6 International Indicators and International Assessment Studies

The Netherlands has taken active part in OECD's education indicator project, the INES project, with its annual publication, Education at a Glance. Education at a Glance presents indicators on educational outcomes, the school careers and civic effects of schooling, educational finance and indicators on the school environment (school and system level indicators). To facilitate its use, the OECD also publishes "Education at a Glance Highlights", in which the most salient indicators of a particular edition are presented. See for example:`

http://www.oecd-ilibrary.org/education/education-at-a-glance-2011_eag_highlights-2011-en

The Ministry of Education usually prepares summaries in Dutch in which indicators that are of particular salience for the Netherlands are highlighted,

e.g. http://www.europa-nu.nl/id/visqndzi9av4/brief_regering_oeso_rapport_education_at

Results from Education at a Glance are also included in summary publications like Key Data and "Trends" of key figures on education, to be discussed further on.

The Netherlands has also been an active participant in international assessment studies. In the section below, cited from Scheerens et al. (2010), an overview and discussion of the Dutch international results is presented.

The Netherlands in TIMSS, PIRLS and PISA

Table 3.1 presents the Netherlands' most important results in the various TIMSS, PISA and PIRLS surveys to date. It shows the average score for each subject and the relative position of Dutch students compared to their international counterparts. The total number of participating countries is shown in brackets. The assessment scores are calculated in such a way as to ensure that the international average is always 500, with a standard deviation of 100. The scores shown for TIMSS 95 relate to primary Grade 4 and secondary Grade 8 students.

As Table 3.1 clearly demonstrates, Dutch students have consistently achieved scores which are (well) above the international average. The Netherlands generally show extremely good performance in these international studies, particularly when compared to other European countries. The Netherlands achieve a slightly lower position on the global ranking due to the exceptionally high scores achieved by countries such as Japan, Singapore, Korea and Taiwan. The Dutch students' scores for science are generally slightly lower than those for reading and mathematics.

Table 3.1: *The Netherlands' results in TIMSS, PIRLS and PISA surveys*

Survey	Mathematics		Science		Reading		Problem-solving	
	Score	Position	Score	Position	Score	Position	Score	Position
TIMSS 95 - 3	493	6 (24)	499	6 (24)				
TIMSS 95 - 4	577	5 (26)	557	6 (26)				
TIMSS 95 - 7	516	7 (39)	517	10 (39)				
TIMSS 95 - 8	541	9 (41)	560	6 (41)				
TIMSS 99 - 8	540	7 (39)	545	6 (39)				
TIMSS 03 - 4	540	6 (25)	525	10 (25)				
TIMSS 03 - 8	536	7 (45)	536	8 (45)				
TIMSS 07 - 4	535	9 (36)	523	17 (36)				
TIMSS Advanced 08-12	552							
PIRLS 01					554	2 (35)		
PIRLS 06					547	12 (45)		
PISA 00	564	1 (42)	529	6 (42)	532	3 (42)		
PISA 03	538	4 (40)	521	8 (40)	513	9 (40)	520	12 (40)
PISA 06	531	5 (57)	525	9 (57)	507	10 (57)		
PISA 09	526	11 (66)	522	11 (66)	508	10 (66)		

The figures presented in Table 3.1 seem to suggest a (slight) downward trend in mathematics, science and reading. However, the scores in successive surveys cannot be directly compared, since the calculation methods used vary from one year to the next. The most recent TIMSS, PIRLS and PISA reports do however include an analysis of the development in each country's scores (see Table 3.2).

Development over time: trends

Table 3.2 shows the change in the performance of Dutch students as revealed by the successive TIMSS, PIRLS and PISA assessments. The figures in the table refer to the difference between two measurements and the relevant standard errors. Most differences are statistically insignificant ($\alpha < 0.05$ in a two-tailed test) but there are three statistically significant results which indicate a (slight) decrease in performance. In fact, twelve of the fifteen reported differences indicate a decline. It would be inappropriate to conclude that the international comparative surveys provide evidence of worsening performance on the part of Dutch students, although they also do not suggest any improvement.

Little differences can be seen in the trends for each subject or age group. Insofar as there is any actual decline in learning achievement, it would seem to be at both the primary and secondary level. Recent criticism (in political circles and in the media) of the quality of Dutch education has

chiefly been directed at secondary education. However, Table 3.2 shows no statistically significant decline in this sector. In terms of primary education, the criticism is largely concerned with numeracy and mathematics. However, the figures also suggest a slight decline in literacy and reading skills. It is interesting to note that the indications of a decline in student achievement offered by the international surveys are not confirmed by the more detailed national "PPON" surveys, which have been conducted by Cito since 1987, as described in greater detail below. The biannual PRIMA cohort studies actually show an improvement in language and numeracy skills between 1994 and 2002 (Mulder, Roeleveld, Van der Veen & Vierke, 2005). The dataset used by PRIMA relies on two random samples of schools, the first of which is representative of the entire country. A supplementary sample is then drawn from those schools with a high proportion of 'disadvantaged' students. This makes it possible to accurately monitor the progress of all students whose parents have low educational attainment, regardless of ethnicity. Numeracy and literacy tests are given to students in the final kindergarten year and in Grades 0, 2, 4 and 6. Improvement was particularly noticeable among the younger students. In 2007, a new cohort study (COOL[5-18]) was introduced.

Table 3.2: *Trends in learning performance in the Netherlands*

	Mathematics		**Science**		**Reading**	
	Deviation	S.E.	Deviation	S.E.	Deviation	S.E.
TIMSS						
99-95 (sec. year 2)	11	9.5	3	9.1		
03-99 (sec. year 2)	-4	8.1	-9	7.6		
07-03 (Grade 4)	-5	3.0	-2	3.1		
03-95 (Grade 4)	**-9**	3.7	-5	3.5		
03-95 (sec. year 2)	7	7.3	-6	6.8		
07-95 (Grade 4)	**-14**	3.7	-7	4.0		
PIRLS						
01-06 (Grade 4)					**-7**	2.9
PISA						
06-03 (15-year-olds)	-7	4.3			-6	6.1

Deviations of statistical significance ($\alpha < 0.05$ in a dual symmetrical test) appear in bold type. How to read this table: the difference in score between 1999 and 1995 is presented in the second column.

It is possible that changes in learning performance are related to changes in the composition of the student population. The reduction in the number of students referred to special education would seem to be relevant in this respect. The proportion of students receiving special education dropped from 3.8% during the 1994-95 school year to 2.7% in 2009-2010. Inclusion in mainstream education is probably better for the cognitive development of the individual student. For statistical purposes, however, the inclusion in general assessments of students with learning difficulties (who

would have been referred to special education in the past) is likely to have a negative impact on the average scores".

The interpretation of the position of the Netherlands on these international tests, has given rise to quite a lot of debate. As noted in an earlier section, the Parliamentary Committee Dijsselbloem tended to play down the satisfactory achievements of the Netherlands, while emphasizing the significance of the slight downward trend, as indicated in Table 3.2. The report from which the above citations were taken (Scheerens et al., 2010) was met by a lot of disbelief in the media, as the general opinion in the Netherlands is that the quality of education is particularly low. A recent study by the Central Planning Bureau established that the position of Dutch students in primary and secondary education on international tests is not that splendid at the top of the score distribution. These results are used in a new emphasis on "excellence" in current educational policy, cf. Vermeer en Van der Steeg (2011). www.cpb.nl/.../cpb-achtergronddocument-onderwijsprestaties-nederland-perspectief.pdf

3.1.2.7 Publication of Key Figures and data Trends by the Ministry of Education
In 2011, the Ministry of Education, Culture and Sciences published the 15th version of "Key Figures", providing information over the period 2005 - 2010. Each edition of Key Figures publishes key data and statistics on education, culture and sciences over a period of five years. The latest English version of Key figures is the version on 2004-2008 (a more recent version, available in Dutch, is the one for 2006-2010). The chapter of the report titled "Education National" describes the major development in education and presents data on enrolments in education, main movements within the educational system, expenditures and the level of education of the population. The 2008 version of the report addressed four special themes: early school leavers, school size, non-subsidized education and lifelong learning. The chapter on "Education International" describes Dutch education from an international perspective. The chapter starts with the goals agreed upon within the European Union. Several common aspects, such as enrolment in education, student performance and the education level of the population, are used to present a picture of the international position of Dutch education, with special attention given to the European objectives.

http://english.minocw.nl/documenten/key%20figures%202004-2008.pdf

Next, the Key Figures editions provide detailed information on the specific educational sectors, on issues like numerical attainment figures, finance, schools and teachers. The contents of Key figures are quite descriptive. Evaluative comments primarily provide summaries of data, describing for example upward or downward trends in attainment, comparing the Dutch figures to the EU benchmarks, and comparing the Netherlands to other countries and international averages.

Parallel to the Key Figures publication, the Ministry annually publishes "Trends in the Picture" (Dutch: *Trends in Beeld).* This publication is shorter and is more explicitly evaluative, and analyses education in terms of ease of access (*toegankelijkheid)*, quality and efficiency. The

36

overall evaluative comment in the 2011 version of Trends is as follows: "Dutch education is achieving far above the international average on a number of facets. On a number of policy issues a positive trend can be noted. For example, the percentage of early school leavers has declined considerably, partly as a consequence of intensive policy. At the same time the percentage of people who have completed higher education has increased, as has the number of students that have obtained a starting qualification (although this is a more modest increase). On these latter indicators the Netherlands is approaching the international top five. Explicit attention is needed for the increase of the proportion of 15-year old students with low reading, mathematics and science skills between 2003 and 2009. Improvement is also indicated for an increase of high performing students in the beta subjects and the number of beta technology graduates" (Trends, 2011, Ministry of Education, Culture and Science, p. 5).

http://www.trendsinbeeld.minocw.nl/TrendsInBeeld_2011.pdf

http://www.wwwords.co.uk/pdf/freetoview.asp?j=eerj&vol=3&issue=1&year=2004&article=6_Benchmarking_Summary__EERJ_3_1_web

http://www.rcec.nl/en/

3.1.2.8 Other facets of procedures used in system evaluation

Criteria and standards

Most of the procedures that were described have used explicit evaluation criteria. Evaluations that include an assessment of student attainment over time (PPON, cohort studies) often use relative standards (e.g. whether students have made progress). European benchmarks, like the Lisbon objectives are used as more absolute attainment targets (Key Figures, Trends).

To the degree that policy agendas and programs are more and more expressed in terms of explicit attainment targets, as is the case in the current Policy Agendas and action programs at primary and secondary level, interpretation of the information from current evaluation procedures is becoming quite straightforward.

3.1.3   Competencies to evaluate the school system and to use evaluation results

Roles in the demand and supply of system evaluations are: (a) initiating and contracting evaluative studies, (b) the technical execution of data collection and analysis, including dissemination, and finally (c) the actual evaluative judgment and use of evaluative information. In the Netherlands the Ministry of Education and the Inspectorate are the main actors for the first aspect. Before contracts for some of the described procedures are given out quality advice is provided by the Foundation for Scientific Research (as in the case of some monitoring studies and the Cohort studies). In some cases expert groups have carried out ex ante evaluations and developed evaluation plans (as for example the CCE). Research institutions, Cito and the Inspectorate of

education generally function as the key agents in the execution of empirical studies, (aspect b). Organizations like the Central Bureau of Statistics and Planning Bureaus sometimes have an executive role as well.

Evaluative conclusions are usually reported as a result of the empirical studies by the actors mentioned under b, but may also be left to evaluation committees or boards. The Inspectorate of Education has a specific evaluation task that comprises all of the three aspects, although evaluative conclusions on major evaluations are likely to be drawn in correspondence with the Ministry of Education. The actual use of system evaluations for decision making and future policy planning is in the hands of the Minister of Education and (in some cases) Parliament.

According to international scientific standards, the technical competency in the Netherlands to carry out evaluative research should be considered as high. University based research groups have good international standing and Cito is one of the world's top institutes in test development and psychometrics. Among inspectorates in Europe the Dutch Inspection has a high standing; its systematic supervision procedures having functioned as examples for other countries. Specific expertise in program evaluation was stimulated through the work of the Committee Program Evaluation, CPE, which was operational from 1986 until 1991. The CPE considered program evaluation not just in the field of education but also in fields like health care and law. Among others this committee submitted rewards for outstanding evaluation research, and supported methodological and meta-analytic studies. Later on the Inter University Center for Educational Research ICO, a so called "Research School" (a recognized interuniversity center of excellence) was quite active in post doctoral courses in evaluation research and attracting international experts for conferences in seminars.

Expertise at the demand side, the Ministry of Education in particular, has been enforced by the creation of the Knowledge Directorate. The Knowledge Directorate has a special role as a clearing house of scientific knowledge and the stimulation of evidence based work in education.

Direct use of results from system evaluation by schools is explicitly aimed for in some of the procedures described, but not in all of them. Positive examples are the feedback from schools based on PPON and the Cohort studies. From experiences of ongoing research and development projects aimed at strengthening result oriented work in schools, it appears that schools and teachers often have difficulties and are in need of specific training to learn how to use feedback from achievement tests (Visscher and Ehren, 2011). Currently the emphasis on the use of data to improve teaching in primary and secondary schools is to be seen as a relevant stimulant to enhance competencies of school leaders and teachers to use test and other empirical data.

Empirical studies on the use of system evaluations are rather scarce. Analysts of the early program evaluation studies in the 1970s, Scheerens (1983) and Creemers and Hoeben (1985), noted the partial failure of the evaluation studies, because of strong resistance of schools and their support organizations to cooperate in evaluation studies and resulting in a lack of quality of these evaluation studies. A positive example of the use of system evaluation however occurred as a

result of the rather disappointing results of the Netherlands in the IEA Reading Literacy Study at primary school level (Postlethwaite and Ross, 1992). These outcomes were a major motive to facilitate a stimulation program of reading literacy education carried out by the Expert Center Dutch Language (Sijtsma, Aarnoutse and Verhoeven, 1999). An increase in reading results in the Netherlands in more recent international studies, might perhaps be partly attributed to this stimulation program.

Scheerens and Hendriks, (2004), collected information on the use of international indicators. They found that the knowledge of politicians, teachers and school heads about the existence of this kind of information was practically absent. Only policy planners at the Ministry of Education were familiar with international indicators.

http://www.wwwords.co.uk/pdf/freetoview.asp?j=eerj&vol=3&issue=1&year=2004&article=6_Benchmarking_Summary__EERJ_3_1_web

3.2   Implementation and use of system evaluation

There are no studies that have assessed the impact of system evaluation in the Netherlands. As the literature on the use of results from evaluation research shows (e.g. Weiss, 1997), impact and use hardly ever occur as a linear process. Instead, research "impacts in ripples, not in waves" (reference), and information trickles through to decision makers gradually. Moreover, use may be "conceptual", rather than "instrumental", and take shape as a gradual process of re-shaping of frames of reference. The most that can be said about the probable impact of system evaluation in the Netherlands is that conditions for the information to be used seem to be favorable: there is an abundance of actual and relevant evaluative information; major "carriers" of system level evaluative information are produced close to the main users (i.e. the Ministry of Education) in the form of the Inspectorate's Annual Report and the annual publications Key Figures and Trends. Monitors and smaller scale evaluation studies are also used to periodically inform the government, i.e. the Minister and Parliament. Cases in point are the various monitors on teacher policies. Finally, as is to be explained further below, current educational policy is increasingly being formulated in terms of measurable targets and standards, which can be seen as another favorable condition to the use and impact of system level evaluation and assessment.

The impact of system evaluation on educational practice in schools is likely to be limited, as system level evaluations are primarily conducted to inform national policy-making. Some procedures have built in feedback to schools, but there is no information on the use and impact on educational practice.

A major difficulty to implement system evaluation, and other kinds of evaluation as well, is a certain reluctance of schools to cooperate in data collection procedures. This has been a problem since the early program evaluations in the nineteen seventies, and, more recently, lead to the

Netherlands being excluded from the international reporting on OECD's TALIS study, because the minimum response rate was not reached. The same happened with the first wave of PISA in 2000. The resistance to collaborate in evaluation and assessment studies can be related to the large degree of autonomy of schools, which in its turn can be seen against the background of the traditional "freedom of education". National assessments and evaluations are likely to be associated with too much state interference. Although schools make ample use of tests, like the Cito test and the Cito pupil monitoring system in primary education, there is a lot of caution towards new legislation of making these tests obligatory. Despite the evidence based orientation of recent educational policy, the educational community remains very alert on too much testing. There is an ever-lasting debate on core educational objectives on the one hand, and autonomous schools being free to be responsive to all kinds of needs from the society at large to broaden the curriculum. The most recent advice from the Education Council (*Onderwijsraad*) to the Minister on this subject is an example of this debate. The council acknowledges the importance of norms and standards but, at the same time, warns against "measurement fixation".

http://www.nrc.nl/nieuws/2012/02/16/onderwijsraad-waarschuwt-voor-verplichtingen-op-scholen/

Another source of the reluctance of schools to cooperate with external research and data collection are 'schools' complaints of too many administrative reporting requirements to central government and their perception of research as being another administrative requirement. On this issue there is considerable support in national politics to be aware of too many centrally induced administrative burdens to schools.

Finally there is also an "anti-test" lobby in the education field. In primary education, this is manifested, for example, by the pressure group "Save the primary school".

3.3 Policy initiatives

From the description of procedures and instruments for system evaluation in education in the Netherlands it is obvious that during the last two decades procedures have been expanded and increased in scope and depth.

New policy initiatives to further system evaluation procedures are closely geared to current policies to gradually expand educational testing, and to monitor educational policies by means of explicit benchmarks. These policies are part of the Quality Agendas and Action Plans for better performance in primary and secondary education and the overall ambition to make educational policy more evidence based. A specific program in this line of policy making is the research program "Onderwijs Bewijs", which literary means "educational evidence", in which research groups can compete for longer term experimental research studies related to specific areas of educational improvement. Other programs, partly carried by the Organization for Scientific Research are dedicated to improvement in language and arithmetic/mathematics. Other policy initiatives are aimed at improving value-added measurement of educational performance,

stimulating school excellence, and experiments to enhance the efficiency (*arbeidsproductiviteit*) of school functioning.

http://www.vosabb.nl/werkgevers-in-onderwijs/archief/item/artikel/kwaliteitsagenda-primair-onderwijs/

http://www.vosabb.nl/werkgevers-in-onderwijs/archief/item/artikel/kwaliteitsagenda-voortgezet-onderwijs/

http://www.vosabb.nl/werkgevers-in-onderwijs/archief/item/artikel/kwaliteitsagenda-speciaal-onderwijs/

http://www.onderwijsregelingen.nl/overzicht_regelingen/regelingen/28

http://www.rijksoverheid.nl/nieuws/2012/02/16/registerleraar-nl-open.html

CHAPTER 4: SCHOOL EVALUATION

4.1   Current practice

4.1.1   Overall framework for school evaluation

Responsibilities of schools

The Act on Primary Education (WPO and WEC), the Act on Secondary Education (WVO) describe the objectives of education (attainment targets) and indicate conditions on how teaching should be structured and organised (content, quality, school plan, funding, school prospectus, complaints procedure). These acts also lay down rules governing the special needs support structure (special needs plan, consortia) and the position of staff, parents and pupils. Recent adaptations include requirements regarding the governance structure and the instruments schools need to have in place to ensure and evaluate educational quality.

According to these laws, primary and secondary schools are required to draw up a school plan, every four years, in which they describe their policy on educational quality and personnel matters (including a description of the instruments and protocols used for evaluating the quality of the school and for ensuring continuous high quality of school personnel). In addition, schools are required to have a school prospectus in which they annually provide information to parents and students on their goals, the achievement of students in the school, choices regarding teaching time, and other matters.

In summary, the responsibility for the quality of education and the pursuance of a quality policy oriented towards improvement is explicitly mentioned as a duty of the school itself, i.e. the school board. Three policy documents: a school plan, a school prospectus and an arrangement for complaints are the requirements that form the legal basis for the internal quality care of schools and school self evaluation in the Netherlands.

Recent requirements on "Good Education and Good Governance" in the Law on primary and secondary education set additional conditions on the governance structure of primary and secondary schools, which are expected to contribute to high educational quality, including student achievement outcomes.

"Good Education and Good Governance" underlines the responsibilities of the School Board for realizing good educational quality and urges for a clear delineation of horizontal and vertical accountability. Horizontal accountability refers to the school's duty to inform the local environment of the schools and various external stakeholders. Vertical accountability operates by way of external school inspection. A further important element in the new directions for

governance is the requirement to stimulate "voice" of parents and students. This is laid down in the law on participation (WMS).[10]

As of August 2010, schools are required to instate an internal supervisory board that will be charged with approving/authorizing the annual report of the school, supervising the extent to which school (boards) meet legal requirements, codes of good conduct and the financial management of the school. Schools are also required to meet minimum student achievement levels.

In the "Language and Numeracy Act", which was established in August 2010, additional requirements on attainment of students in primary and secondary education were set in literacy and arithmetic. These attainment benchmarks, or reference levels as they are usually called, further specify attainment targets and describe the knowledge and skills pupils are expected to have acquired in language and numeracy at different stages in their school career. The reference levels (Dutch: *referentieniveaus*) provide a general framework for the organisation of the curriculum by schools and teachers at all levels of primary, secondary, secondary vocational and special education. The reference levels were introduced to improve pupils' language and numeracy skills, and ensure continuity of learning across grade levels and school types (primary and secondary schools). Schools are responsible for deciding on how to apply the core objectives and benchmarks in their curriculum and instruction and are autonomous in designing their own curriculum and instruction, as long as they make sure that students meet the core objectives and benchmarks.. External standardized tests are used to evaluate the extent to which students are meeting the core objectives and attainment benchmarks.
Legislation is planned to assure that, as of 2014, primary schools will be required to administer a national student achievement test to assess whether they meet these achievement targets and performance standards.

Responsibilities of the Inspectorate of Education

The Inspectorate of Education is charged with the external control and evaluation of schools. The Education Inspectorate is an independent executive agency, under the Ministry of Education, Culture and Science. The Education Inspectorate performs its tasks on the basis of the provisions in the Educational Supervision Act. According to the Supervision Act of 2002, adapted in 2012, the Inspectorate performs its investigations guided by eight, later ten domains of quality and also guided by all regulations schools have to comply to on the basis of any educational law. The approach is risk-based, meaning that the investigation starts with a first screening on a limited number of quality domains and ends with a broader investigation when the risk analysis suggests that quality is insufficient. The Act is explicit on the interventions for failing schools. The Act also requires the Inspectorate to publish their assessments of schools.

---

[10] Source: Governance in het Onderwijs. Tweede Kamer Vergaderingen 2004-2005

The inspection framework contains a description of both the way schools are inspected and the indicators and standards used to evaluate schools. These standards are derived from the aspects of quality described in the Supervision Act. The standards include the legal requirements as prescribed in the Act on Primary Education (WPO and WEC), the Act on Secondary Education (WPO and WVO) and the Act on 'Good Education, Good Governance', and include additional quality standards describing the output as well as the process and prerequisites for teaching and learning (e.g. curriculum, school climate, monitoring of students) of schools. These quality standards were developed in close cooperation with schools and their stakeholders to ensure maximum support for school inspections of these standards. The revision of the Supervision Act in 2012 also assigned the evaluation of the quality of teaching personnel, the quality of governance of schools and financial compliance of schools to the Inspectorate of Education. According to the Supervision Act, the inspectorate has to consult relevant stakeholders in education on the standards of the framework.

Compliance with legal requirements (Dutch: deugdelijkheid) and quality standards

Article 23 of the Constitution dates back from 1917 and states that schools with a religious denomination are entitled to receive funding on the same grounds as public schools. In return, the education of the denominative schools had to be as "sound" and in compliance with legal regulations as the education in public schools. Through the years, this aspect of "soundness" of education has taken on different meanings and - as intended by the legislator - lead to different criteria in consecutive educational laws (Mentink, 2012). At first, these compliance requirements kept well away from the freedom of education (which was then conceptualized as: the choice of curriculum and materials, the appointment of teachers and -in current times- the freedom of admission of pupils). These minimum requirements of "soundness" regarded the competence and morality of teachers, compulsory subjects and amount of teaching time, and were the basis for funding.

In the Quality Law of 1998, quality of education became conceptualized next to the "soundness", in the sense of compliance to basic legal requirements, of education. The competent authority has to ensure educational quality in the school which should be outlined in a School plan and includes both legal requirements and self-chosen goals. Quality is seen as broader than the "soundness" of education: all actions that are taken to attain the goals of the school, partly because of legal requirements, partly because the particular responsibility and circumstances of a school, which may amount to a certain pedagogical vision or didactic approach. The Inspectorate inspects the degree to which a school meets legal requirements and meets minimum student achievement results, not the specific choices a school makes.

In addition the inspectorate monitors the level of education at system level. In the period between 1998-2002 the inspectorate became gradually more focused on the quality of individual schools. The Supervision Act of 2002 states that the inspectorate "judges the quality of individual schools with respect to legal requirements and other quality aspects". The quality aspects consist of the results of education (learning results and progress in development) and the process of education

(teaching time, pedagogical climate, pupil care and guidance a.o.). Next to the control and evaluative tasks, the inspectorate was attributed a stimulating task: to stimulate the quality of education. In this period, a clear statement was made that the quality requirements were not funding requirements, meaning that schools were not obliged to live up to these norms individually. Yet, when a certain amount of the most crucial quality indicators are judged as 'insufficient', a school is judged as 'very weak' or 'weak'. In 2012 the Education Council stated that the fluid boundary between "soundness" aspects (i.e. compliance with legal aspects) and quality aspects is defensible because of the right of pupils to good education, given the growing importance of education in society (Education Council, 2012, p. 56).

Quality assurance policy and proportionate / risk based supervision

In the law on the Supervision of Education (WOT, as determined in 2002), validation of the quality assurance policy of the school was seen as a major part of the inspection process and the degree of external school inspection would be proportionate to the extent in which schools succeed in appropriate forms of quality care and school self-evaluation.
This approach of proportional supervision is in line with a general orientation in Inspection and Supervision across societal sectors in the Netherlands to enhance efficiency in supervision. http://www.inspectieloket.nl/Images/A5-folder-VT-aug09_310809_en_tcm296-264134.pdf

In the period after 2004, proportionate supervision was developed in order to reduce administrative burden for schools. Proportional inspection refers to an inspection method where the inspectorate evaluates only those schools and those quality aspects that are at risk. At first, this was performed on the basis of school self-evaluations, but as these proved to be less informative for the inspectorate (no substitute for inspectorate's instruments) risk based inspection was introduced as a way to enhance efficiency in school inspection. Since 2008, for risk based inspection, the inspectorate relies on compulsory accountability information of schools. Around 2010, the performance of schools with respect to the reference levels also became a quality aspect. In 2012 the supervision framework was broadened to include the policy for assuring teacher competence, as an additional quality aspect.

The Educational Supervision Act also states that the Inspectorate is authorized to promote the quality of the school. Supervision should stimulate the implementation of excellent quality care as well as a permanent improvement of quality.

Recently, policy changes described as "Good Education, Good Governance" also motivated additional changes in school inspection methods, involving specific measures to stimulate very weak schools, increased discretionary powers with respect to sanctioning mismanagement of schools and addressing school boards (as the competent authority of schools) in the inspection of schools (instead of the school principal).

Responsibilities of DUO, Windows for Accountability and 'Schools have the initiative'

With the increasing importance of accountability information, not only for government but also for pupils, parents and the broader public, means have been sought to diminish administrative burdens for schools. An agency (DUO) that originally merely had the function to fund schools has been redesigned to act as an 'Information Window' for schools (just "one window"). Part of the data on student performance is centered here, as all students have their own 'civil service number'. The administrative data of schools can be used to evaluate students' school careers across the school system. DUO also receives all other accountability information of schools.

These data are used by the inspectorate on the one hand and by a new intermediate organization, representing schools in primary and secondary education (Windows for Accountability, Dutch: *Vensters voor Verantwoording*). Windows for Accountability also uses other quality information (e.g. from student satisfaction surveys from the Students Organizations) and publishes this as benchmark information for schools. Part of this benchmarking information is available for the broader public.

Windows for Accountability has an agreement with the Inspectorate of Education on the usage of this information for school inspections; inspection information can also be published in the 'Windows for Accountability'.

Next to providing accountability information, intermediary organizations have taken a role with respect to quality improvement.in the sense of assisting schools that have been judged as weak to develop into good schools and stimulating schools to attain policy goals formulated in the Action plans. This is carried out under the heading of 'Schools have the initiative', (Scholen aan zet).

Conclusion

In conclusion it can be stated that assessing school quality in the Netherlands had best be seen as the interplay of two main responsible agents, autonomous schools and the Inspectorate of Education. The idea of proportional supervision is an attempt to regulate mutual roles and responsibilities. Interestingly, the intermediary organizational structure in Dutch education (see chapter 1) has recently been added as a third actor as the councils for Primary and Secondary Education have initiated a new foundation (developing an online website 'Windows for Accountability' providing standardized information on benchmark indicators) to provide schools with benchmark information to use in their horizontal accountability to stakeholders (e.g. student satisfaction surveys). A recent agreement between the Inspectorate of Education and the council of secondary schools enables the Inspectorate to use this information in their supervision of secondary schools. Vice versa, the Inspectorate provides data that schools can use (e.g. normed student achievement data). Schools in primary education have just started to provide benchmark information and will most likely eventually be included in this arrangement.
See: http://www.venstersvoorverantwoording.nl/nl-NL/pages/38/Samenwerkende-partijen.aspx
Further details on these procedures are presented in subsequent sections.

Consequently, the main procedures for school assessment are: school self-evaluation, horizontal supervision and school inspection (consisting of systematic school supervision and the publication of "quality cards", which are currently indicated as "supervision cards").

4.1.2   School evaluation procedures

4.1.2.1 School inspection
Within the framework of Risk Based Inspection, the Inspectorate of Education uses annual risk analyses to target inspection visits to potentially failing schools. In these analyses, called primary detection, information is collected on possible risks of low educational quality in all schools, such as student achievement results on standardized tests, and financial reports of schools, complaints of parents and news items in the media. Since the implementation of the 'Good Education, Good Governance' Act, school boards are responsible for providing such information.

De Wolf and Verkroost (2011)[11], discuss the three pillars of the new inspection approach: next to risk analysis, (see above), these include holding the competent authorities of schools, i.e. the school boards, responsible and a broader gamma of intervention measures (including financial sanctions).

The results of students in the final grade 8 of primary education (corrected for the socio-economic background of students) on the national standardized Cito-test and students' results on the national examination in secondary education are the primary indicators in the early warning analysis. The Inspectorate considers the student results to be a good output measure of the educational quality of schools on the inspection standards. It should be noted that the risk analysis (using test scores) is not used to evaluate outcomes of the school. The test scores are only used as a first indicator of potential risks of low educational quality in schools. The quality of schools is evaluated during additional inspection visits in case the primary detection shows potential risks. Also, after this data-based phase, an expert analysis is performed by the team of inspectors on additional data (not for schools that show no risks). Then it is decided if a quality investigation will be performed. During these visits, school inspectors observe classes, interview the principal, teachers, parents and students and analyze additional documentation. Schools are assessed to be failing when student achievement results are below standards for three or more years and when two or more norm indicators in the teaching-learning process or pupil monitoring and support are evaluated as insufficient.

---

[11] De Wolf, I.F., &Verkroost, J.J.H. (2011) Evaluatie en praktijk van het nieuwe inspectie toezicht. *Tijdschrift voor Toezicht*, 2

In case the primary detection shows no risks, a school receives no quality investigation during that year. However, a meeting with the governing board (of short duration) will be arranged to discuss the schools' results.

The aim of risk analyses is to increase the efficiency of the Inspectorate by scheduling visits only in schools that need it the most. However, the House of Representatives of the States General agreed in October 2007 to add a provision of four yearly visits to the risk-based school inspections. All schools in primary and secondary education (also high quality schools) should be visited by the Inspectorate of Education at least once every four years. These visits are instated to provide for a 'reality check' and to prevent schools from having no inspection visit for a long period of time (Inspectorate of Education 2010c). The visit is however explicitly not instated to evaluate schools on all the standards in the inspection framework, but only on a selection of standards that is relevant given the previous results of the early warning analysis of the school or the agenda of the Inspectorate.

Schools at risk are scheduled for additional monitoring (expert analysis) and interventions. Additional monitoring includes desk research of additional results and school documents (for example, test scores in intermediate grades or annual reports), interviews with the school board and potentially also quality investigations where the inspection framework is used to assess educational quality in the school as sufficient, failing or highly underdeveloped (Inspectorate of Education, 2009b). This amounts to a supervision arrangement, which can be an "Adapted Arrangement" or a "Basic Arrangement". Schools with an adapted arrangement are put under some form of intensified supervision. An arrangement for weak schools implies preventive supervision: the school must not become weak. An arrangement for very weak schools leads to the inspectorate monitoring the improvement plans. The school board has to formulate an improvement plan to address insufficiencies that have been identified by the Inspectorate. The inspectorate monitors the implementation of this plan. Additional monitoring may also include specific evaluation of the quality of teaching personnel on standards that have been added to the Supervision Act in 2012, including the quality of school management and personnel policy.

It should be noted that this process of risk-based inspection introduces a new kind of proportionality, in the sense that a major effort is being made to turn around very weak schools. An additional measure to prevent that schools "at risk" decline further, consists of the Inspectorate contacting the competent authority, i.e. the school board, already at the stage when the school's achievement results are below the norm for just one year. Recently, the Minister has sharpened the pressure on school improvement of very weak schools, by reducing the period in which they should implement the Inspectorates suggestions for improvement from two to just one year.

In case the school does not improve, sanctions may be enacted such as official warnings or administrative and/or financial sanctions (Inspectorate of Education, 2009b). Sanctions can be enacted when the school fails to meet minimum student achievement results and fails to meet legal requirements as specified in the Act on Primary Education (WPO and WEC), the Act on

Secondary Education (WPO and WVO) and the Act on 'Good Education, Good Governance'. The Department of Education decides on sanctioning of schools.

Within the given legal frameworks schools independently determine quality targets and norms as well as the way in which these are to be measured and assessed. On the basis of their own judgements, the schools determine to what extent quality improvements are required, as well as the contents of such improvements. Supervision (Dutch: *toezicht*) by the Education Inspectorate is to function as an addition to the self regulatory mechanisms by which schools are expected to carry out their self evaluation and quality management.

The supervision frameworks

Since the late eighties the Inspectorate uses elaborate supervision frameworks for school inspection, based on quality aspects, quality indicators and standards. Over time this framework has been adapted several times, and the way it is being applied has also changed, particularly with the fairly recent move to risk based inspection, that was described in the above. An early description of the supervision framework (in English) is provided by Van Bruggen, (2002).
http://www.see-educoop.net/education_in/pdf/workshop/reconstructing_ml_org/pdf/netherlands_reconstructing_ml_org.pdf

More recent descriptions, 2007 and 2012, are available from the SICI website:
http://www.sici-inspectorates.eu/en/contactUs/Secretariat

The central question of school supervision is: "What is the quality of education in a particular school like?" (Inspectorate, 2002, p. 9) This core question is differentiated according to three sub questions:
1) What is the school's quality care like?
2) What is the school's quality in teaching and learning?
3) What is the quality of the learning results? (ibid, p.9)
These questions indicate three domains of the supervision framework.

In more recent documents this core function is maintained but the particular characteristics of risk based inspection, including tailored inspection in case of weak functioning of schools are added. See: http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2010/Risk-based+Inspection+as+of+2009.pdf;
http://www.onderwijsinspectie.nl/binaries/content/assets/Actueel_publicaties/2011/Brochure+Toezichtkader+2011+po-vo.pdf; http://www.rijksoverheid.nl/ministeries/ocw/documenten-en-publicaties/rapporten/2012/06/11/bijlage-bij-de-brief-van-minister-bijsterveldt-vliegenthart-ocw-aan-de-tweede-kamer-over-het-toezichtkader-po-vo-2012-van-de-inspectie-van-het-onderwijs.html .

The four quality domains are divided in the following quality aspects; as indicated in the table below, based on the Inspection Framework for Primary Schools.

Quality domain	Quality aspects per domain
Outcomes	1. Learning results in basic subjects 2. Progress in student development
Teaching personnel policy	3. (if necessary) 4. requirements with respect to competencies 5. sustainable assurance of the quality of teaching personnel
Teaching and Learning	6. Subject matter coverage 7. Time 8. Stimulating and supportive teaching and learning process 9. Safe, supportive and stimulating school climate 10. Special care for children with learning difficulties 11. the content, level and execution of assessments and exams
Quality care	12. Systematic quality care by the school
Financial compliance	13. financial continuity 14. financial compliance
Other legal requirements	15. Law on parent participation in school decisions (e.g.)

For each of the quality aspects a number of indicators have been specified, for example:

- The school systematically evaluates the quality of learning outcomes and teaching and learning processes
- The school uses a coherent system of standardized tests and procedures for monitoring student achievement and development
- Teachers monitor and analyze systematically student progress
- Subject matter coverage is such that it prepares the pupils for secondary education
- Subject matter coverage is integrated
- The school knows the educational needs of its school population
- Learning time is sufficient for the students to have them master the subject matter
- The school programs sufficient teaching time
- The teaching activities are well structured and effective
- The teachers take care of their teaching being adaptive to the learning needs of the students
- School staff and pupils interact in a positive way
- The school stimulates the involvement of parents

- The school guarantees safety
- The school provides a pleasant and stimulating environment for the students
- The school provides a pleasant and stimulating working environment for its staff

The most recent version of the supervision frameworks for primary and secondary education 2012 is available from http://www.rijksoverheid.nl/ministeries/ocw/documenten-en-publicaties/rapporten/2012/06/11/bijlage-bij-de-brief-van-minister-bijsterveldt-vliegenthart-ocw-aan-de-tweede-kamer-over-het-toezichtkader-po-vo-2012-van-de-inspectie-van-het-onderwijs.html

To give one further example, the indicators for the quality aspect "learning results", in the framework for secondary education, according to the 2012 adaptation of the framework are as follows:

The pupils attain the achievement results relative to their capacities
1.1   The pupils attain the achievement level that is to be expected given national averages.
1.2   The pupils stay close to the minimum amount of time to finish the program for each of the program variants in Dutch secondary education.
1.3   The pupils obtain the marks that may be expected of them on the final examinations, relative to national averages, for each of the program variants of Dutch secondary education.
1.4   The differences between the marks of the school examination and the central examination are of an acceptable level."

Details about how the outcome indicators are evaluated by the Inspectorate, are presented in Inspectorate of Education, (2010).

In summary, the current form of school inspection consists of the Inspectorate conducting school inspections in primary, secondary, vocational and adult education and in special education. This leads to summary conclusions about the quality of a school: so-called inspection arrangements. An arrangement can be: basic (no risks for the quality of education), or adapted (weak or very weak quality).[12]
A review on the scientific basis of the indicators of the Dutch Inspectorate supervision frameworks was carried out by Scheerens et al. (2005). These authors checked whether the set of process indicators on teaching and learning in the Inspection Frameworks, corresponds well with the set of variables that has received empirical support in educational effectiveness research. One of their final conclusions is that "the research syntheses described, ….indicate a good match

---

[12] The total number of inspection arrangements in 2010 were: 7.584 for primary and special primary education, 2.781 for secondary education, 627 for special education.
Very weak schools are a minority of all inspection arrangements.

between the set of indicators of the Inspection framework and the set of variables that is frequently studied in educational effectiveness research". They go on to say that: "In our research synthesis we find little support for expanding the current set on indicators that is primarily concentrated at the classroom level with a lot of school level indicators, for example on support and management functions. However, ongoing developments concerning school management, task enlargement of teachers and human resources management at school may be seen as arguments to give more room for indicators that are aimed at measuring such school level phenomena" (ibid, p. 344). As a matter of fact, additions to the 2012 Supervision Act include the quality of school management and personnel policy.

Internal evaluation studies by the Inspectorate, carried out in 2010 (Inspectie 2010 a and b)[13] pointed out that schools rated the work of the inspectorate positively (on average 80% satisfied or very satisfied). Risk based inspection was also judged positively on a number of criteria. Initial figures show a decline of the number of very weak schools, and efficiency gains were noted for both schools and the Inspectorate (a lower administrative burden for schools, and efficiency gains for the Inspection organization). A declining number of (very) weak schools can be seen as an important benchmark for the quality of education in the country, and can also be seen as a central criterion for evaluating the success of the current version of risk-based inspection.

4.1.2.2 Quality Cards
Since 1998 the Inspectorate annually publishes quality cards, including judgements of achievement results, for all secondary schools in the country. Basic information included in the quality cards were examination results and attainment indicators (transfer figures, with respect to obtaining the diploma without delay, class repetition and drop-out), as well as process indicators from the supervision framework. The school report cards thus contained rubrics on result, atmosphere, quality of teaching and methods and material (cf. Janssens, 2012, in press). In 2003 quality cards on primary schools came into existence as well. From 2007 onwards the publication of these, relatively extended, quality cards was stopped and replaced by an indication of the kind of supervision arrangement the school was receiving. At that time school inspection changed to the earlier described risk-based approach, which implied that far less information was obtained on the whole population of schools in the country.

Currently the term "supervision card", which includes information on the kind of inspection regime schools are assigned to (basic, more intense or extended), has replaced the term quality card. To the extent that school reports have been produced, they are publicly available (on line) from the report archives of the Inspectorate. For secondary schools so called "attainment cards" (*opbrengstkaarten*) are annually available. These cards contain average examination marks on school subjects, participation rates to the different subject matter profiles and an indication of

---

[13] Tevredenheidsonderzoek Periode juni- oktober, 2010. Utrecht: Inspectie van het Onderwijs, 2010 a. Risico gericht toezicht beoordeeld. Utrecht: 2010b/

attainment in lower secondary education. The various formats of reporting school information in the context of the project "Windows for Accountability", to be described in a subsequent paragraph, can be seen as substitutes of the original quality cards. A major difference is, however, that the "Windows" do not include norm-based judgments.

4.1.2.3 School internal quality care and school self-evaluation
Quality care in schools involves all the measures taken at school- and classroom level to maintain and improve the quality of education. It relates to all activities aimed at formulating quality goals and achieving, maintaining and systematically improving the quality of education. It involves determining the goals of the school and the way the goals can be reached, the execution and evaluation of quality policy and the deciding on actions that might be necessary for maintaining and improving the quality of the school (Hendriks, 2001; Hendriks, Doolaard & Bosker, 2002).

School self-evaluation is a crucial aspect of quality care. School self-evaluation is an internal evaluation of the school as a whole or of sub-units of the school. It could be 'completely internal', but might include extensive use of external capacity. The decisive point is that the school is the initiator and the prime audience of the self-evaluation (Scheerens, Glas & Thomas, 2003). School self-evaluation can be defined as a procedure started by the school for gaining information on the design and goals of education in order to take policy decisions for maintaining and improving the quality of the school (according to Voogt, 1995; Van Petegem, 2001).
Self-evaluation could be carried out at the outset of applying the well-known quality care cycle, in order to map the strong and weak aspects of the school. In this case the self-evaluation is aimed at diagnosing needs and setting goals for improvement. But self-evaluation can also be used to determine whether the intended goals of quality improvement have been achieved. The aim is then to determine if the school, given the intended goals, is on the right track

The fact that internal school quality care was seen as one of the pillars of "proportional supervision" in the Law on the Supervision of Education of 2002, (see section 3.1.1) underlines the high expectations that existed at the time about the school's own responsibility for the monitoring of its quality.
In the 2005 version of the Supervision Framework (the version for secondary education) the following quality aspects and indicators on school quality care were used (cf. Hendriks, 2005):

Quality Aspect 1:

The school takes care of the assurance and improvement of the quality of education.

- o *The school knows its entrance situation, including the specific needs of the student body.*
- o *The school systematically evaluates the quality of its performance in terms of learning results*
- o *The school systematically evaluates the quality of learning, teaching and counseling*
- o *The school has formulated measurable improvement targets.*
- o *The school carries out improvement activities in a systematic way*
- o *The school guarantees the quality of learning and teaching*
- o *The school guarantees the quality of the school examination and of other evaluation instruments.*
- o *The school reports about the realized quality of education to interested parties (parents, students, competent authorities, funding agencies and sponsors).*

Quality Aspect 2:

The conditions for quality care are in place

- o *School management initiates and steers the quality care*
- o *Quality care is connected to the school's vision with respect to learning and teaching as stated in the school plan.*
- o *The school management takes care of a professional school culture*
- o *The school takes care of an effective communication about the quality of education.*
- o *Staff, school management, pupils, parents and competent authorities are all of them being involved in the school's quality care.*

In the supervision framework for secondary education, version 2012, the first quality aspect for quality care has remained more or less the same. The second aspect on the conditions for quality care however, has disappeared completely. A motivation for dismissing this quality aspect may be the fact that these conditions particularly included the school management's role in quality assurance. As the policy changes described in 'Good Governance and Good Education' now emphasize the school boards' role (instead of the school management) in providing these conditions, there is no need any more to evaluate the functioning of the school management in implementing conditions for quality care.

From 2002 onwards the application and implementation of school quality care and self evaluation was not just stimulated by the involvement of the Inspectorate of Education but also by two dedicated organizations, one for primary education and one for secondary education.
In order to support schools and their governing boards with their quality assurance and to give a strong impulse to the development and implementation of systematic quality assurance, with subsidy of the Ministry of Education, the organizations of governing bodies of the schools, the

organization of school managers and the teachers' unions (these latter organizations were only involved in primary education) set up the national project groups Q5 for secondary education and Q* Primair for primary education.

Q5 was initiated in 1999, and expired in December 2005. The project attempted to stimulate systematic and integral quality care in schools for secondary education, i.e. it was aimed at stimulating schools to:
- Develop a system of quality management as an integral part of school development;
- Involve all relevant groups inside and outside schools;
- Present results of self-evaluation to third parties (other schools, experts, 'critical friends');
- Participate in networks, in order to exchange information and expertise;
- Publish information about their quality (Beelaerts, Bousché, De Goeij, De Graaff, Horsman & Klifman, 1999).

*Q*Primair*
Q*Primair started in 2001 and expired in December 2006. It set the following list of attainment targets (which were to be realized by 2006):
- 80 per cent of the schools has formulated their quality care policy in such a way that it is clear how the school satisfies on the one hand the legal requirements with regard to quality care as well as on the other hand the quality goals as set and aimed for by the school itself (starting situation: applies to 20 per cent of the schools);
- 80 per cent of the schools systematically monitors the quality of their education and takes action to maintain and improve the quality (starting situation: applies to 31 per cent of the schools);
- In their school plan, 80 per cent of the schools has formulated targets for the quality of teaching and learning as well as its performance in terms of learning results (starting situation: applies to 40 per cent of the schools);
- 80 per cent of the schools systematically and regularly evaluates the quality of its teaching and learning and performance (starting situation: applies to 47 per cent of the schools);
- Based on their evaluation of the quality, 80 per cent of the schools has planned an improvement route for the long term (starting situation: applies to 49 per cent of the schools);
- Based on their evaluation of the quality, 90 per cent of the schools has planned improvement actions for the current school year (starting situation: applies to 77 per cent of the schools);
- 80 per cent of the schools uses a system of instruments and procedures for monitoring progress of the pupils and takes care of guidance an counseling in case of problems (starting situation: applies to 54 per cent of the schools);
- 80 per cent of schools involves parents and pupils in the school's quality care but preferably also other stakeholders (starting situation: applies to 12 per cent of the schools);

- 80 per cent of the schools reports about the realized quality of education to interested parties (in each case parents, school board and personnel (starting situation unknown);
- In order to objectify the self-evaluation, 30 per cent of the schools have put the results of the self-evaluation at least once to independent third parties (starting situation unknown) (Hofman, Dijkstra, Hofman & De Boom, 2004).

It should be noted that the criteria and standards of the Inspectorate and the Q organizations differed on one important aspect. The inspection criteria emphasized quantitative data on student performance while the criteria by the Q organizations mainly involved process dimensions of quality care.

During the first decade of 2000, schools could choose from a large range of commercial school self-evaluation instruments. The most frequently used ones were rated by Q5 and Q*Primary on issues like reliability, validity, utility, quality aspects covered, and suggestions for school development (cf. Hofman, Dijkstra, Hofman & De Boer, 2004).

Results on the actual attainment of these targets and the implementation of proportional supervision will be discussed in the section on implementation and use of school assessments in a subsequent paragraph.

4.1.2.4 "Windows for accountability"

"Windows for accountability" is a project in which all quantitative information, as well as additional qualitative information, about secondary schools is presented in one online system. The system uses (amongst others) information from DUO, a central administrative unit of the Ministry of Education, the Inspectorate and the schools themselves. The information is summarized and visualized. "Windows" provide benchmark information and summary statistics on key indicators of individual schools; schools are expected to use these windows to inform their stakeholders. The project aims to develop quality standards that allow, among others, a comparison of schools (benchmarking) and to align vertical accountability (towards the Inspectorate of Education) and horizontal accountability, (towards parents, other school sectors – e.g. Windows on secondary school are made available to primary schools- and municipalities). The horizontal element, informing the stakeholders of the school, is predominant (VO-Raad, 2011)[14]. The alignment to vertical accountability particularly includes an agreement with the Inspectorate that schools who have published accurate and substantive "Windows" will not be asked by the Inspectorate to supply the same information to the Inspectorate again.

---

[14] VO Raad: Werkboek Horizontale Verantwoording VO. VO Raad, april 2011.

Windows for secondary schools has been operational since 2010 and includes information on the following indicators:

1) Number of students
2) Success rates on examinations
3) Examination marks
4) Transfer and school leavers
5) Premature school leaving (drop out)
6) The school plan
7) Cooperation
8) Student background characteristics, more precisely, the percentage of students needing extra care
9) Choice of school sectors and subject matter profiles
10) The so called "care plan" (for students in need of extra support and care)
11) The satisfaction of students
12) The satisfaction of parents
13) Characteristics of schools from which students enter and schools to which students leave
14) External evaluations
15) School climate and safety provisions
16) Market share in the catchment area
17) Teaching time
18) Personnel (age composition, teacher absenteeism, professionalization budgets)
19) Financial indicators (solvability, financial buffers, per student expenditure, school costs)

The first 5 indicators are described as "central" indicators, because the information is provided by national organisations (DUO, the former Cfi department of the Ministry of Education). Other central indicators are indicators number 8, 9, 16, 18; indicator 19, is partly central and partly local (school costs are local). All other indicators are local, i.e. the school provides the information.

Information from the indicators is presented in different ways for various audiences; i.e. there are windows presenting information for the immediate school environment and stakeholders, the national level, school management and school governors.

The coordinating unit for Windows secondary education resides with the VO Council, the organization that unites the employers in education and has responsibility for educational innovation.

http://www.venstersvoorverantwoording.nl/nl-NL/pages/147/Onderzoek.aspx

Since 2011 school files based on Windows for secondary education are available at http://www.schoolvo.nl/

From 2012 onwards a project "Windows primary education" will start and as result a new foundation has been established (*Stiching SchoolInfo*) which will also incorporate the Windows for Secondary Education project The foundation operates under the auspices of the VO and PO council and will coordinate all activities for both secondary and primary education. See http://www.poraad.nl/content/po-raad-en-vo-raad-werken-samen-stichting-schoolinfo

4.2 Implementation and use of school evaluation

4.2.1 Use of school inspections
Publishing school inspection reports may serve three distinct functions:
  a) public disclosure of school performance, as an example of accountability in an administrative context;
  b) informing the consumers of education, as an example of market based accountability, encouraging parents to "vote with their feet", and influence school policies;
  c) providing feed back to schools and teachers, as a basis for school improvement; this latter approach is sometimes described as "professionally oriented accountability".

Janssens, (2012, in press), analyzed the use that schools and parents make of the information from quality cards and inspection reports. His conclusions are rather negative, as far as the use of this information by parents is concerned. Only a small proportion of parents actively use this information in the context of choosing a school for their child. He also found little evidence that supported the hypotheses that quality card information is used by parents, through parent voice, in influencing school policies. At the same time there is evidence that school principals and teachers do make active use of inspection reports and school quality cards. Janssens (ibid) cites various authors who found that the results from school inspections shape the choice of priorities and methods by schools. Not always is this use by schools interpreted uncritically, as it is sometimes taken as leading to approaches considered too narrow and targeted. On the other hand this "teaching to inspection", could be seen as a stimulant of school improvement, when one assumes that inspection frameworks are successful in targeting good quality education. Reviews indicate that this latter condition would actually apply to the frameworks of the Dutch inspectorate (Scheerens et al., 2005).

The evidence on the use of inspection reports in the Netherlands is such that (a) administrative accountability, consisting of incentives and redress for poor performance has only been introduced very recently, in the 2012 adaptation of the Supervision Act, in the form of the Inspectorate being mandated to provide sanctions to none complying schools (b) consumer oriented accountability is intended but has not really set foot yet, and that (c) professional use by schools seems to have been less explicitly intended, but is strongest in being actually applied.

A recent study by Bekkers, Kool & Straten (2012), about the use and impact of inspection reports confirms the above. These authors conclude that the school information on the Inspectorate's website hardly plays a role in parents' school choice. Only 2% of the responding parents took the inspection reports into account. At the same time they conclude that school inspections have a relatively strong impact on school policies as enforced by school governors and school management.

4.2.2   Implementation of quality assurance and use of school self evaluations

The functioning of school internal quality assurance is, first of all, monitored annually –on a sample base- by the Inspectorate. Results are published in the Inspectorate's annual report. In 2003, for example, 13 per cent of the primary schools had a quality care system that satisfied the full set of indicators on quality care. Twenty-six per cent of schools satisfied the minimum requirements for quality assurance (Source: Hendriks, 2005). In 2010 just over 50% of all primary and secondary schools satisfied all the indicators on quality care. This is a clear improvement but still far removed from the standard of 80%, to be met in 2012, as mentioned in the Quality Action Plans of the Ministry of Education.

http://www.onderwijsinspectie.nl/binaries/content/assets/Onderwijsverslagen/2011/Onderwijsverslag+2009+2010+printversie.pdf

See also: Amelsvoort, van H.C.H.M& Wolf, I., (2005)[15]

Following the strong stimulation projects on school self evaluation, Q5 and Q*Primair in the period between 2000 and 2006, several more in depth evaluation studies about the functioning of school self evaluation were carried out. A summary and overview of the results of these studies is presented in Hendriks (2005), who sums up the conclusions as follows:

The idea of proportional supervision is a creative application of decentralization and "subsidiarity" in education, meaning that all that can possibly and reasonably be carried out at a lower administrative level, should not be carried out at a higher level (Scheerens, 1997). But perhaps this formulation also indicates the Achilles heel of this strategy. Can school self evaluation really meet the demands of external evaluation? What was shown is that a lot of developmental effort has been and is being invested in enabling, facilitating and stimulating school self evaluation and school quality care in the Netherlands. A relatively large set of school self evaluation systems and instruments has been developed, and organizations like Q5 and Q Primair have invested in the development of criteria to asses these instruments and procedures and in experimental good practice applications. It is interesting to note that the quality standards developed and applied by Q5 and Q Primair differ from those applied in studies of the Education Inspectorate in the sense of the importance given to what Stufflebeam et al. (1971) subsume under

---

[15] Amelsvoort, van H.C.H.M. & Wolf, I.(2005) *Riks based inspection in Europe.* Utrecht: Inspectie van het Onderwijs.

'accuracy standards'. The Inspectorate is more consistent in seeing criteria such as reliability and validity as essential for the value of school self evaluation and is generally more severe and critical in assessing the available set of instruments.

What is being practiced under headings such as school self-evaluation and quality care is a mixture of elements that have a somewhat different orientation and strongly different traditions. The main orientations are: a) a scientific evaluation orientation, emphasizing accuracy, reliability and validity of procedures; b) an administrative and client oriented branch of quality management systems, and finally, c) a more qualitative, opinion based school improvement orientation. From the user perspective of schools each have their strong and weak points:

- schools may appreciate the objectivity and rigor of the use of structured instruments that have been tested for reliability and validity, but most of all find themselves at odds with the standardization implications, strictness of procedures, and the targeted scope that is easily condemned as narrow and reductionist;
- quality care systems may have the fashionable appeal of the fast moving business world and the more real asset of being responsive to clients; yet these systems usually fail to address the primary process of teaching and learning, and they may paint caricatures in their efforts to "proceduralize" educational processes;
- most suspect among these three, at least from an evaluation perspective, is the school improvement perspective that forgoes rigor in establishing fact, and may favor superficial consensus about change over sound diagnosis; in fact there is a perpetual dilemma in educational analysis and action that is caused by a discrepancy between the urge to adapt and change, and knowledge that is at best partial and contested" (Hendriks, 2005, p.318).

Blok, Sleegers and Karsten (2008), are also critical in their conclusions about school self evaluation as part of proportional supervision. Their study of 27 Dutch primary schools indicate that school self evaluations are often of very low quality. According to these authors school self-evaluation is a very difficult task for most schools. They recommend external support and capacity building over a longer period of time in order to improve this state of affairs. Moelands (2005) sketches an advanced approach for quantitative school self evaluation, based on data from school administration systems and pupil monitoring systems. He concludes that an Expertise Centre for Quality Assurance would be crucial for schools to work with such a system. Schildkamp (2007) investigated the use that schools made of school self evaluation results, after they had participated in applying a comprehensive instrument for school self-evaluation (ZEBO). She concludes that only a minimal amount of schools were able to use the self-evaluation results to improve the quality of their education.

Improving the capacity building at the school level and involving school governors, school managers and teachers in continuous development seem to be crucial to furthering the implementation and use of internal school evaluation in the Netherlands. Alignment of internal and external school assessment is also an important issue as new forms of evaluation and

assessment (Windows for Accountability) are being implemented. A very interesting third element is currently underway as part of the recent policies on "achievement orientation", which includes a strong data-driven element (see the next section).

4.3   Policy initiatives

A number of policy initiatives which were mentioned in the final section of Chapter 3 (the Quality Agendas for primary and secondary education, the stimulation of research studies that can support evidence based educational policy, and pilot studies to produce value-added school performance indicators), are also relevant for school assessment.

Initiatives to stimulate an achievement oriented approach in schools may also be relevant to school evaluation.

The Inspectorate of Education (2010) describes Achievement Oriented Work as a focus on performance of students through the setting of clear goals and evaluating students' achievement and progress towards these goals (p. 8). In another publication the Inspectorate (2010) describes Achievement Oriented Work, as consisting of the following elements:

- statement of clear goals and objectives
- assuring that teachers know exactly what subject matter they need to teach their students
- adapting teaching to the needs of the students
- analyses of problems that students experience who do not reach the objectives
- resolving problems through an adequate system of pupil care
- monitoring performance of groups of students and the school as a whole, annually
- rapid intervention if performance is falling behind

http://www.owinsp.nl/actueel/nieuwsberichten/Betere+prestaties+door+opbrengstgericht+werken.html

In some government supported projects the element of performance monitoring and performance feedback play a central role, next to rational planning techniques and adequate didactic support to schools and teachers (Visscher and Ehren, 2011). In primary education the Cito pupil monitoring system offers favorable conditions for application and implementation of an achievement orientation that is strongly data-driven. The strong "micro" level orientation, i.e. focusing the learning process of groups and individual students, in the achievement oriented policy could be seen as a compensation of (risk based) school inspections and internal assessments (horizontal supervision according to the Mirrors for Accountability) moving away from the primary process of teaching and learning. In risk based inspections, only failing schools are monitored on the basis of process indicators relating to teaching and learning, whereas the large majority of schools is not. To the extent that Windows for Accountability have replaced earlier comprehensive school

self-evaluation approaches, that included teaching and learning, they imply a shift to a more managerial orientation in information provision.

A final point to be made about school assessment policy in the Netherlands is that the interplay of external accountability and internal self evaluation has been a constant element. As external accountability is gradually becoming more high stakes, this might function as a further stimulus for internal accountability. As shown by Carnoy et al (2003) this can be an effective combination. One very interesting facet of this interplay is the shift that has occurred with respect to what one might indicate as selective inspection. Both proportional inspection, and risk based inspection can be seen as instances of selective inspection. In the current risk based inspection approach, internal school evaluations are again considered important in the arrangement of internal supervision and horizontal and vertical accountability. On the other hand, the external support that schools obtain in their current data provision within the framework of Windows for Accountability might reflect a more prudent (perhaps more realistic) approach, in the face of the somewhat disappointing development of "autonomous" school self-evaluation.

CHAPTER 5: TEACHER APPRAISAL

5.1   Current practice

Teacher appraisal belongs to the discretion of the competent authorities of the school, i.e. school boards and municipalities. External appraisal of teachers does not exist in the Netherlands. Recent national policies to strengthen the teacher profession, among many other measures, stimulate schools to implement HRM procedures that include an evaluation of the functioning of teachers. Data from monitoring studies on the implementation of these policies provide some insight in the degree to which these policies are actually implemented. Next, the Inspectorate evaluates good teaching practices, as part of their school supervision task, and as an object of national studies, such as the Annual Inspection Report. The Inspectorate in the Netherlands, however, does not evaluate the functioning of individual teachers.

Government stimulation of internal teacher appraisal

The Education Professions Act, which was implemented on 1 August 2006, regulates standards of competence for both teachers and other people working in education-related jobs in primary education and secondary education. These standards have been specified by the Teacher Council into seven competencies (interpersonal, pedagogical, subject-specific, didactical, organizational, cooperative, self-reflection and developmental competencies). These competencies were included in the Education Professions Act with an additional condition of scheduled revisions by the association of teachers once every six years. Also see the report *Teaching 2020, a strong profession!* (ECBO, 2011). http://www.ecbo.nl/ECBO/ReferNet/docs/11-0315_Teacher_2020.pdf

School boards are required to maintain competency dossiers of their teaching personnel in which they describe (according to a set of regulated criteria) the level of competencies of each staff member and how these competences will be maintained. The act aims to ensure that all teaching personnel (particularly the ones that start teaching without a regular teaching degree) meet minimum requirements, both at the start and throughout their careers. Schools boards are invited to send the Inspectorate dossiers of teaching staff that are new to the teaching profession and don't have a regular teaching degree, once every six months. School boards are also responsible for making sure teaching staff meet the minimum competency levels and have to implement and describe their personnel policy for doing so; they may set additional requirements on competencies for their teaching staff. School boards are responsible for hiring teaching personnel and for firing personnel that doesn't meet the competencies. They may also delegate this task (and the task of setting and evaluating competency dossiers) to the school principal through a management contract.

A new teachers' association, which was instated in October 2011, has opened a national register of qualified teachers who meet the requirements of the Education Professions act. As of 2018, teachers will be required to submit their records to this register.

New policy initiatives, which might imply assessing teacher performance on the basis of "value added" student achievement outcomes, are in a very early stage of development. Pilot projects have been started in 2011 to develop value-added performance indicators.

The functioning of teachers as monitored by the Inspectorate

The Inspectorate of Education evaluates whether school boards meet the requirements of having competency dossiers for their teachers and of having a personnel policy in place to make sure teachers meet the required set of competencies, particularly in failing schools.

In the changes concerning the Law on School Inspection of 2009, more attention for teacher functioning by the Inspectorate was arranged. As the evaluation or appraisal of individual teachers belongs to the jurisdiction of the Competent Authority of the school, the inspectorate evaluates the quality of the staff as a whole, and focuses on the following three indicators:
- checking whether teachers are fully certified or not;
- obtaining impressions of teacher quality at school on the basis of observation of *teaching* in classrooms;
- assessing the personnel policy of the school.

The current Annual Inspection Report only provides summary impressions of teacher functioning based on classroom observations. They are a facet of school evaluation or system evaluation, and not an appraisal of individual teachers.

By way of illustration a brief overview is given of the statements on teacher quality in the 2011 Inspection Report, based on the observation of 2000 lessons in primary schools and 1100 lessons in secondary schools.
- most teachers can keep order in the classroom, give explanations and engage students;
- one third of the primary school teachers, and half of the teachers in lower secondary education adapt teaching to the individual levels of the students; the inspectorate calls this the Achilles heel of instruction;
- insufficient quality of adaptive education for students with special needs

(Inspectie van het Onderwijs, Onderwijsverslag 2011, p.22)
http://www.onderwijsinspectie.nl/binaries/content/assets/Onderwijsverslagen/2011/Onderwijsvers lag+2009+2010+printversie.pdf

5.2   Implementation and use of teacher appraisal

The law on Professions in Education (BIO) was established in 2006. Since then school boards should register, in so called competency dossiers, which competencies their teachers possess and how they make sure that these competencies are continuously developed and updated. School boards are, according to the plan on "Teaching 2010, a strong profession" required to outline how they implement competency dossiers and how they make information in the dossiers available to the wider public.

Bokdam et al. (2011, p.19) found that, in 2010, about one fifth to one third of teachers was familiar with the new competency demands. In primary education 66% of school principals were familiar with these demands. Only 25 % of teachers reported that they had a competency dossier. In 2010 a large majority of teachers (73%), in all school categories, reported to have had a performance review. Topics that school principals addressed in these reviews were:

- Further schooling
- Career development
- Achievement orientation
- Participation in inter vision and coaching
- Observation visits with other teachers
- Other way to keep competencies up to date
- Extra time to keep competencies up to date
- Multi year schooling plan
- Advancement to a higher salary scale
- A financial bonus for keeping competencies up to date
- Agreements on a professionalization trajectory for subjects in which the teacher is not qualified.

(Bokdam et al., 2011)

http://www.research.nl/files/rvb/reportcenter/Rapporten/B3883/B3883TusenmetingConvenantLeerkracht2011def.PDF

Teacher appraisal is aimed at improving the professional development of teachers in schools, it is considered to be part of the so-called Integrated Personnel Policy (IPB; *Integraal Personeelsbeleid*) of schools. By implementing IPB, school leaders can align the competences of teachers to the goals of their schools, and the professional development of individual teachers to the development of the schools as an organization. Although the Dutch government, school boards, school leaders and Human Resource Management departments of schools have put a lot of effort in implementing IPB in schools (Teurlings & Vermeulen, 2004), research on the implementation of HRM practices in Dutch schools has shown that the impact of these HRM practices is still limited, due to the fact that IPB is implemented in a rather operational, narrow way (Runhaar & Sanders, 2007).

Research on the effects of integral personnel policy instruments on teachers' professional development within schools is scarce (Van Wonderen, 2005; Runhaar & Sanders 2007).

Recent research into age-related personnel policy in primary and secondary schools has shown that schools differ a lot with regard to the implementation of age-related personnel policy (Jacobs & Vrielink, 2009). Only a small minority of schools (15%) consider age-related personnel policy as an integral part of personnel policy. Moreover, a great majority of the schools (75%) in primary education do not have a vision on age-related personnel policy. In almost all schools, teacher appraisal interviews were held with the staff and a lot of staff members did have a personnel development plan. In addition, in most schools the personnel policy is attuned to the goals of education.

5.3   New policy initiatives

Apart from the national monitoring of HRM policies in schools, and the enhanced attention that the Inspectorate gives to teacher quality at the school level, there are no concrete policy initiatives in the domain of teacher appraisal. New projects, for instance those concerning measuring teacher added value, are at a very early phase of development.

A lack of external teacher appraisal in the Netherland can be seen as part of the strong autonomy of schools and teachers. Judging by the most recent advice from the Education Council "Controlled space", this tradition is not about to be changed. The report speaks highly about the professional autonomy of teachers, and is keen to warn for this autonomy being threatened by too much external control (Onderwijsraad, 2012).

CHAPTER 6: STUDENT ASSESSMENT

6.1 Current practices

6.1.1. Overall framework for student assessment

6.1.1.1 Context

Students are expected to meet core learning objectives in Dutch language, English language, arithmetic and mathematics, social studies, science, arts and physical education in primary education, and in a number of subjects within one of four profiles that the student has chosen in secondary education (nature and technique, nature and science, economics and society, culture and society). Core learning objectives are specified for each of these different stages and tracks. These core objectives provide the legal basis for what knowledge, insight and skills pupils should have achieved at the end of primary and secondary education. They are described in very global terms, but have recently been specified in more detailed performance benchmarks or reference levels for literacy and arithmetic. As of 2010, schools in primary and secondary education are obliged to implement these benchmarks. In 2010, a canon of historical and cultural topics was also specified that need to be addressed in both primary and secondary education.

The aim of the benchmarks in literacy and arithmetic is to align exit and entrance levels of the different educational tracks, to set a common framework of expectations and learning trajectories across all educational tracks and to increase performance of students in the core subjects of language and arithmetic.

Benchmarks are provided for the four key stages in the education system: at the end of primary education, and at the end of three educational tracks in secondary education. For each key stage, the benchmarks include a description of a 'fundamental' level that all students are expected to meet and a more 'advanced' level for gifted students.

Benchmarks in literacy describe oral fluency, fluency in reading and writing, and proper use of definitions, terms, grammar and spelling rules. Each benchmark provides a general description, a description of the tasks students should be able to perform and the criteria these tasks have to meet. Benchmarks in mathematics include four domains: number sense and operations, expressions and equations, measurement, ratio and proportional relations. Each of these benchmarks is specified in a description of the mathematical language students need to be able to use, connection between definitions, notations and numbers they need to be able to make and their use of the mathematical operation to solve problems. Students need to know facts, definitions and operations, they need to be able to use them to solve problems and they need to know why they are using certain concepts and methods to solve problems.

At the end of primary education, schools are required to use tests to advise students on the appropriate track in secondary education. A vast majority of about 85% of the schools currently administers a standardized external 'End of Primary Education test', developed by testing company Cito (Van der Lubbe 2007; Resit 2009); other schools apply other tests and assessments. New legislation requires primary schools, as of 2014, to administer a common national external standardized test. This test will be aligned to the performance standards in literacy and arithmetic; schools may administer an additional component to assess study skills of students and their performance in history, geography and science (not required). The test is designed by Cito (under the authority of a central examination committee) and therefore closely matches the Cito-test the majority of schools are familiar with.

The primary function of the external standardized test is to provide advice for the most suitable track of secondary education for a student. Additionally, schools in primary education will also be required by law to implement a pupil monitoring system that enables them to assess learning strengths and needs and remediate students' learning needs throughout their school careers. Schools are autonomous in choosing a specific type of system and how often they measure progress of students, as long as they make sure to use valid and reliable assessments. These additional assessments of students' progress and achievement are expected to improve student achievement as schools improve their use of data to differentiate their teaching.

At the end of secondary education, students are to take a set of final examinations in a number of subjects within a profile that the student has chosen. These final examinations are compulsory and assess whether students meet the core learning objects and performance benchmarks and qualify to graduate. The final examination is divided into two parts: a national examination and a school examination. The final mark for each subject is an average of the mark for the national and the school examination. Dutch language is a compulsory subject in the national examination in all types of secondary education. English language and some form of mathematics are compulsory elements in the national examination in pre-university and senior-general secondary education. Other compulsory subjects depend on the profiles (pre-university and senior-general secondary education) or type of vocational training the student has chosen. As of 2014, all students are required to take an additional arithmetic test.

The elements to be tested in the national examination are specified in the examination syllabus, approved by the Ministry of Education, Culture and Science. The syllabus also specifies the number and length of the tests that make up the national examination. Schools are responsible for setting up the school examination. Every year schools are required to submit their own school examination syllabus to the Inspectorate, showing which elements of the syllabus will be tested, when, and how marks are calculated, including the weight allocated to these tests and re-sit opportunities. Usually, a school examination consists of two or more tests per subject. These may be oral, practical or written. The school examinations are produced by the schools themselves or by test institutes. The school examinations are marked by the pupils' own teacher. There are also

practical assignments for which no marks are given, only an acknowledgement that the examinee has completed them properly. The school examination must be completed and the results submitted to the Inspectorate before the national examinations start.

6.1.1.2 More detailed overview of instruments

In summary, the main strands of student assessment in the Netherlands are: a standardized test taken at the end of primary education to facilitate the choice of one of the tracks of secondary education (the "Cito test", - school leavers test); the use of pupil monitoring system (formative assessment) in primary and secondary schools, and examinations at the end of secondary education. Aggregated results from these main strands of student assessment are also being used for either school assessment or system evaluation, but such applications will not be discussed in this chapter. Test taking as part of national assessment and within the framework of cohort studies, are seen as system evaluation, and have already been discussed in Chapter 3.

The Cito test's prime objective is to support the choice of a specific secondary school track. Such a use is obviously inspired by the strong stratification of the Dutch secondary school system, see Chapter 1. Since the mid-eighties primary schools started to make use of a pupil monitoring system, also developed by Cito, the LVS (*Leerling Volg Systeem*). Later on pupil monitoring systems were also implemented in secondary schools and currently every secondary schools has a pupil monitoring system.

As is the case for the Cito test at the end of primary school, the pupil monitoring system (LVS) is purchased by schools (at their own cost and initiative). The primary objective of the LVS is the formative assessment of student achievement and individual students' mastery of key subject matter areas in relation to their grade level. More recently, both the Cito end of primary education test and the pupil monitoring system have been used for other purposes as well, notably accountability oriented school assessment, and improvement oriented school self evaluation. Currently the LVS is used within the framework of performance feedback, as part of the new government policy on Achievement Oriented Work (Visscher and Ehren, 2011).

Examinations at the end of secondary education have a long tradition in the Netherlands. Thinking in terms of checks and balances in national educational governance, one might say that formalized output control, on the basis of examinations, has functioned as a counterbalance against the traditional school autonomy based on the principle of "freedom of education", see Chapter 1.

Opportunities for the longitudinal use of student assessment data are strongly increasing with the implementation of a unique student number. This number enables a comparison of a student's achievement on the Cito test, to his or her examination results. Item Response Theory is used to vertically equate students' scores in the LVS tests which allow for a calculation of student growth trajectories in primary school.

As stated, primary schools are responsible for the use and application of the Cito test and the LVS. Examinations at the end of the various secondary tracks include internal school examinations (which are developed by schools under strict external regulations), and central examinations.. The Minster of Education is responsible for regulating the examination programs and key qualifications in secondary education, while the '*College Voor Examens*' (CVE) , as an independent administrative body, is responsible for the actual production of national examinations. "It (CVE) provides all the services of a national examination board, including publishing the examination syllabus and specifications in accordance with the current Regulations. The syllabus sets out in detail the topics which will be examined in the national examinations." (Ofqual, 2012). In its turn CVE contracts Cito to develop the assessment materials and provide other services.

6.1.2 Student assessment procedures

6.1.2.1 Formative assessment; the Cito pupil monitoring system LVS

The Cito pupil monitoring system (LVS) for primary education is a consistent set of nationally standardized tests for longitudinal assessment of a pupil's achievement throughout primary education, as well as a system for manual or automated registration of pupil progress. The LVS covers Language, (including decoding and reading comprehension), Arithmetic, World Orientation (Geography, History, Biology), Social-emotional development, English , Science and Technology (source Van der Lubbe, undated). An overview of test administration is provided in the table below.

Table 6.1: *Overview of test taking: Cito LVS*

Grades (4 -12 years of age)	1	2	3	4	5	6	7	8
Ordering	x	x						
Language	x	x						
Orientation in Space and Time	x	x						
Technical Reading			x	x	x	x	x	x
Reading Comprehension			x	x	x	x	x	x
Listening Comprehension			x	x	x	x	x	x
Vocabulary			x	x	x	x	x	x
Spelling			x	x	x	x	x	x
General Language Ability				x	x	x	x	x
Arithmetic/Mathematics			x	x	x	x	x	x
World Orientation						x	x	x
Social-emotional development			x	x	x	x	x	x
English							x	x
Science and Technology						x	x	x

During the primary school period, tests are usually taken once or twice a year.

Apart from a system of tests, the LVS also contains specific feedback formats, guidelines on how to interpret individual students' scores and accompanying suggestions for remedial actions.

The LVS tests are constructed to confirm to the requirements of a particular Item Response Model. This has a number of interpretative advantages, particularly the possibility of "vertical equation", which allows for students' scores being comparable over time. This means, for example, that the result of a particular student on a math test in grade four can be depicted at the same scale as the results obtained by that student six months before on the math test of end grade 3. In addition the students' score can be compared to that of other pupils nationally (norm referenced comparison). Finally, tests confirming to the IRT model in question make it possible to draw conclusions about the degree of mastery of a particular subject matter (domain- or content-referenced interpretation).

The following presentation formats are made available on the basis of the LVS:

- The *pupil report,* which is a graph in which the pupil's progress is visible throughout the years. Data available in the national surveys are used as a frame of reference, based on percentiles, so that the position of an individual pupil with regards to 5 reference groups (25% highest scoring pupils, just above average, just below average, far below average, and the 10% lowest scoring pupils) is immediately visible from the corresponding graph.
- For children with special education needs, and who visit special education schools, an *alternative pupil report* is made available. This report also shows at what level a pupil is functioning and how to interpret the results of the pupil compared to children of the same age who attend mainstream primary education.
- In the so called *group survey* the results of all the pupils from a group over a number of years are presented in a table. For each pupil the scale of ability score at the successive measuring moments is shown along with the level score.

When schools have been using the LVS for several years, the results can be aggregated in various ways, and compared to national averages, for purposes of school self-evaluation. For example a cross section shows histograms based on the 5 scoring levels indicated above, for the different grade levels, compared to the national average. In this way, a school can see at a glance how well it is doing at the various grade levels. Next, scores are made visible in the form of the results of cohorts of pupils over the year, compared again to the national mean in the different grades (Van der Lubbe, *ibid*).

Since 2003 the LVS also contains computer-based tests, some of which are adaptive. This means that the computer selects items based on the answer given (and ability estimated) on the previous question. This makes test taking more efficient as better information is gathered in less testing time.

In addition to the monitoring and evaluation system for primary education, there is also a monitoring and evaluation system for students for the first two years of secondary education. This system makes use of computer-based tests, the results of which are analyzed and reported by Cito. The main purposes are: To help teachers monitor their students' development and to provide tools to help students decide on the type of schooling they should choose after their first two years in secondary education and to monitor the quality of the educational process.

The monitoring and evaluation system for students consists of an entrance test, a test after the first year and a test after the second year of secondary school. All tests are in multiple-choice format. They are available at three levels of difficulty, starting from the lower vocational education level. The range of subjects is the same for each level:
- Dutch reading comprehension
- English reading comprehension
- Mathematics
- Study skills

Every test consists of seven assignments, each requiring 45 minutes. It is up to the teacher to determine which skill is to be tested, in which order, and when the test should be given.

A scoring service and a reporting service is part of the system. A student report shows what level the student has mastered per subject. The report also compares the student with three reference groups. Gaps in a student's development will therefore become clear, showing which areas need remedial work. The report also provides insight into how students should continue their school careers. In addition to student reports, reports on the results of a group as a whole, as well as the school as whole are provided. These reporting formats are provided to support schools in monitoring the quality of their education. See:

http://www.iaea.info/documents/paper_1162d1e9fe.pdf

https://springerlink3.metapress.com/content/l052186563m5n631/resource-secured/?target=fulltext.pdf&sid=lrponc3zhwfc0iwutzagdslf&sh=www.springerlink.com

In addition to the use of the pupil monitoring system, schools often use the so called Cito Entrance test, which may be administered from grade 5 until group 7. The Entrance test covers the same subject matter areas as the school leavers test (see below), namely language, arithmetic/mathematics and study skills. The purpose of the test is to provide a summary review of strong and weak points in the performance of individual pupils. This information is provided to inform teachers in targeting the learning needs of their students; more specifically to identify subject matter areas where pupils need extra support. More detailed information on the presentation of the results of the Entrance test are available at: http://www.Cito.nl/onderwijs/primair%20onderwijs/Cito_volgsysteem_po/entreetoets/entree_voor_ouders.aspx

6.1.2.2 The school leavers test (Cito-test) at the end of primary education
About 85 % of primary schools in the Netherlands administer the Cito school leaver attainment test (the "Cito-test"). Most recently, February 2012, 144.708 pupils took the Cito test. This learning progress test evaluates the knowledge pupils have acquired during their eight years of primary school. As the score indicates the most appropriate level of subsequent education, the test is supposed to serve as a tool for parents, teachers and the prospective secondary schools in selecting the most appropriate form of secondary education.

Content areas that are covered in the Cito test are language, arithmetic/mathematics and study skills. World orientation is an optional subject in the test. Currently the Cito test is administered in February, but in the near future the test will be taken closer to the end of the school-year (May or June). The Cito test only uses multiple choice items. Schools are free to use the Cito test, or another similar test, and have to pay for the test to Cito. According to EURYDICE (2009)[16], the Netherlands is the only country where schools have to pay for national tests. Concrete policy plans exist to implement a a compulsory exit test for primary schools; the foreseen year of implementation is 2014.

The Cito test has a specific core purpose, which is closely tied to the highly tracked system of secondary schooling in the Netherlands. Apart from facilitating school choice in this particular context, the Cito test is also being used for other purposes. Municipalities have used the test to assess schools, and even have used specific average score levels as attainment standards for schools. Currently, within the framework of the Better Performance Action plan, the Minister, together with the PO and VO council have agreed to set higher targets for the national average on the Cito test.

6.1.2.3 Examinations at the end of secondary school
There are three kinds of end-of-secondary examinations: one for VMBO, one for HAVO and one for VWO.
The examinations have civil effects and lead up to formally recognized entrance qualifications on the labor market and give entrance to higher education levels. Students with a VMBO diploma obtain access to HAVO or senior secondary vocational education, MBO. Students who have completed HAVO get access to higher professional education HBO and students with a VWO diploma obtain access to universities.

Performance criteria and standards for the examinations are prepared in the following manner: Examinations are prepared on the basis of a syllabus for each subject. The syllabus details the topics that are to be taught and examined. In the case of the school examinations, schools compose

---

[16] EURYDICE (2009) National testing of pupils in Europe: Objectives, Organization and Use of Results. Brussels: Education, Audiovisual and Cultural Agency.
http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/109EN.pdf
http://eacea.ec.europa.eu/education/eurydice/documents/eurybase/national_summary_sheets/047_NL_EN.pdf

the examinations, which are then reviewed and approved by the Inspectorate. The Central or National examination is prepared by Cito, who arranges the input from subject matter experts and teachers. For each subject the length of the examination scale is determined and a norm term (N-term) is decided upon by the CVE, to do justice to annual differences in difficulty; the N- term, which varies between 0 and 2.0, is high when the examinations is considered as difficult and low when the examination is considered easy. Examination marks are on a scale form 1 to 10, and 6 means a pass.
http://www2.Cito.nl/vo/ce/omzet2011_vmbo/tabellen/11184_omzettingstabel.htm

Worth mentioning is the role that CVE and Cito have in test equating and stabilizing the norms. Procedures are described in: http://toetswijzer.kennisnet.nl/html/normering/default.shtm

Excerpts of the Examination programs for major subject matter areas – Dutch language, chemistry, mathematics and history- (in English) are presented in Ofqual (2012). In the tables below, cited from Ofqual (2012), major characteristics of the Dutch secondary school examinations are summarized.

Table 6.2: *Summary of features, the Netherlands*

Country	Netherlands
Upper secondary graduation rates (general and vocational programs)[17]	No overall rates. 39% of population have graduated from a general upper secondary education, 58% from a vocational program (OECD, 2011a).
Where does responsibility for education lie?	The Ministry of Education, Culture and Science.
Regulatory framework	The Ministry of Education, Culture and Science. College Voor Examen (responsible for national assessments
Administration of education system	The Ministry of Education, Culture and Science. College Voor Examen.
Administration of examinations and qualifications at senior secondary level	College VoorE xamen (CVE). CVE contract Cito (a national institute for educational measurement) to develop test materials.
Administration of university entrance	University admissions and student grants are managed by the Learning Implementation Service or Dienst Uitvoering Onderwijs (DUO).
Main suite of qualifications/assessments available at senior secondary level (bold = included in this study)	Hoger algemeen voortgezet onderwijs (havo) Voorbereidend wetenschappelijk onderwijs (vwo) voorbereidend middelbaar beroepsonderwijs (vmbo)
Notes (include number of universities in top lists for TES rankings)	The Netherlands has ten universities ranked amongst the top 200 in the world.

[17]Data sourced from OECD, 2011

Table 6.3: *Summary of key features qualifications and assessments*

Education system	Netherlands	Netherlands
Assessment or qualification	Hoger algemeen voortgezet onderwijs (havo)	Voor bereidend wetenschappelijk onderwijs (vwo)
Purpose	Completion of senior secondary education	Completion of senior secondary education
Usual length of study	five years	six years
Type	composite	composite
Number of courses required	five common subjects plus one specialized subject combination and an independent project	seven common subjects plus one specialized subject combination and an independent project
Compulsory subjects	culture and the arts Dutch English physical education social studies plus one subject combination from: culture and society economics and society science and health science and technology	classical culture culture and the arts or general science Dutch English other modern language physical education social studies plus one subject combination from: culture and society economics and society science and health science and technology
Notes	The overall achievement is determined by a combination of the results from school and nationally set assessments. Students may choose to take the VWO examination in specific subjects.	The overall achievement is determined by a combination of the results from school and nationally set assessments.

Table 6.4: *Summary of approaches to assessment - chemistry*

Country	Netherlands	Netherlands
Assessment or qualification Number of courses or examinations	Hoger algemeen voortgezet onderwijs (havo) 5 common subjects plus one specialised combination and an independent project	Voorbereidend wetenschappelijk onderwijs (vwo) 7 common subjects plus one specialised combination and an independent project
Compulsory science element	None	None
Chemistry courses	Chemistry	
Course length	320 hours	440 hours
Compulsory topics	• Skills • Analysis of and reflections on science and technology • Substances and materials 1, inorganic • Substances and materials 2, organic • Substances and materials 3, biochemical • Controlling reactions • Chemical industry • Acids and bases • Reactions and electrical current	• Skills • substances, structures and Bonding • Carbon chemistry • Biochemistry • Features of reactions • Chemical Technology • Acids and Bases • Redox
Options available	None	None
Practical work	(40 hours)	40 hours
Nature of assessments	School assessed and external examination	School assessed and external examination
Notes	150 hours are allocated to the study of topics to be assessed in the school assessments. The school assessment contributes 40% of the overall score. 170 hours are allocated to the study of topics assessed in the external examination. This is a written 3 hour written paper contributing 60% to the overall score.	200 hours are allocated to the study of topics to be assessed in the school assessments, including the practical work. The school assessment contributes 40% of the overall score. 240 hours are allocated to the study of topics assessed in the external examination. This is a written 3 hour written paper contributing 60% to the overall score.

6.1.2.4 National Assessments

National assessments with no civic effects are described in Chapter 1 on System Evaluation (the PPON project). Also student assessment, within the framework of the Cohort Studies has been addressed in that chapter.

6.2   Implementation and use of student assessment

The Netherlands have a high level of expertise in assessing students and using use student assessment results, particularly because of the international renowned testing company Cito. All main assessment instruments discussed in this Chapter depend on the input and technical expertise from Cito. More is to be gained in the areas of the use and application of the assessment results by schools and teachers. The formative use of the pupil monitoring system, to improve teaching and learning is the most critical area. Hands on experience with using the Cito LVS data to inform and improve teaching in primary schools shows that in many cases schools and teachers under utilize the potential of the tests, and are in need of training and support to use the results for improvement oriented action (Visscher and Ehren, 2011). To the extent that aggregated student assessment results are used for school self-evaluation and quality control, the annual Inspection reports shows that the situation is far from perfect. From this one might deduce that skills and competencies of a considerable proportion of school leaders and school governors might be in need of improvement.

As indicated in the above the "coverage" of application of all student assessment procedures that were discussed in this Chapter, is very high. This comes as no surprise as far as the compulsory examination at the end of secondary education is concerned. Participation rates to the Cito LVS and the Cito tests are also very high, but this can be seen as a particularly striking feature, as application depends on the school actually purchasing these test procedures.

Use of the examinations is straightforward. They define the civil effect that a particular level of secondary education has for the students. There are no indications that the secondary examination system does not serve this function. Apart from a certification function, examinations can also be seen as a basis for defining attainment targets at school level, and as a way to give direction to content orientation in the school curriculum and teaching. Finally, examination results, marks and success rates are used in aggregated forms for system level monitoring, in publications like the annual report of the Inspectorate and editions like Key Figures and Trends, from the Ministry of Education.

The key function of the Cito primary school leavers' test is informing teachers, parents and students about school choice. Score levels are presented in "bands", which refer to specific types of secondary education. This is illustrated in the table, used for the Cito test administration in 2012. The actual school choice is informed by the advice of a students' teachers, in which he/she incorporates the Cito advice.

TABLE FOR THE USE Of Cito BAND-WIDTHS			
Advice	Additional research required	Consultation with primary school required	automatically accepted
Practical education	No Cito bandwidths applicable		
Learning routes supporting education	No Cito bandwidths applicable		
vmbo-basic	514 and lower	515 t/m 520	521 and higher
vmbo-basic/cadre	517 and lower	518 t/m 522	523 and higher
vmbo-cadre	520 and lower	521 t/m 528	529 and higher
vmbo-mixed	526 and lower	527 t/m 533	534 and higher
vmbo-theoretical	526 and lower	527 t/m 533	534 and higher
vmbo-theoretical / havo	528 and lower	529 t/m 535	536 and higher
havo	531 and lower	532 t/m 537	538 and higher
havo / vwo	535 and lower	536 t/m 540	541 and higher
vwo	539 and lower	540 t/m 544	545 and higher
kopklas	No Cito bandwidths applicable		

Parents receive a pupil report produced by Cito. This report indicates the test results. The report states the absolute score, the standard score and which school type is the most adequate. Next the pupil report shows the relative position of the pupil in comparison with other pupils. The actual choice is not taken, solely on the basis of the Cito advice. As a matter of fact the information from the Cito test is seen as a complementary source to the advice that is given by the school.
http://www.jmouders.nl/Themas/School/De-Citoeindtoets-hulpmiddel-naast-het-schooladvies.htm

Apart from the core function of advising parents on school choice, the Cito test is used for other purposes as well. Municipalities have sometimes used school averages on the Cito test as a basis for high stakes external accountability. The dissertation study of Mart Visser (2003)[18] describes an experiment in the city of Amsterdam, where the Cito test was made compulsory in the school year 1994-1995 and used to adapt school budgets. This policy failed utterly as a result of strong resistance from schools and other stakeholders.
The Inspectorate uses school results on the Cito test as a basis for outcome indicators of primary schools. Finally, as mentioned in the above, overall results on the Cito test are used as an important criterion for judging the success of the recent Action Plans for better performance.

---

[18] Mart Visser (2003); *Keuzeprocedure, kwaliteit en kwaliteitszorg*. Proefschrift Groningen.

The Cito LVS has high potential for being used for purposes of formative student assessment, and for evidence based school self evaluation. As far as the latter application of evidence based school self-evaluation is concerned, efforts can be described as disappointing (Moelands, 2005, Schildkamp, 2007). A key problem is that schools tend to be sloppy and incomplete in keeping records of different waves of test data.

The aspiration that evaluation and assessment functions as a tool for improvement of teaching and learning is most evident in the formative use of student assessment. The experiences in the Netherlands show that useable formative evaluation comprises a lot more than applying high quality instruments and collecting data. Among others this is illustrated in a recent study in which the Cito LVS is used as a corner stone of achievement oriented work in primary schools (Visscher and Ehren, 2011). In this project the use and application of test based formative student assessment is embedded in a structured approached to didactic action planning.

The project should be placed in fairly recent central policies to improve the quality of education, with a specific emphasis on basic subjects, the quality agenda's for primary and secondary education, and the action plans for better performance. Result oriented work is to be stimulated, and monitored by the Inspectorate. "Result oriented work" is defined by the Inspectorate as "systematic, goal oriented work aimed at maximizing performance" (*ibid*, p. 4). Visscher and Ehren analyze this concept and distinguish three main components;

a. Determining the entrance situation of the school and groups (classes) within the school, in terms of achievement levels and the degree to which pupils are mastering the subject matter. This assessment is to be supported by using information from pupil monitoring systems, and is functional to establishing the educational needs of individuals and groups of pupils.

b. Defining targets (i.e. explicit and clear objectives that indicate future attainment levels), which should stimulate goal oriented rather than activity oriented behavior.

c. Choosing a didactic and instructional approach to bridge the gap between the entrance situation and the target.

In their analysis of the degree to which schools have basic conditions in place to work according to this systematic approach, Visscher and Ehren mention several problematic aspects in the use of assessment tools, in the ability of schools to work goal oriented and also with respect to didactic action planning. In this context weaknesses in the way schools apply assessment tools, in this case the Cito LVS are most relevant. They mention the following points under the heading "underutilization of pupil monitoring systems":

1) Imperfections in the storage and retrieval of test data; so that, for example, it is sometimes impossible to retrieve overviews of items on which pupils have failed.

2) Schools make limited use of the available analysis routines; longitudinal analyses are frequently not carried out.

3) Mistakes in interpreting the LVS output, so that it may happen that groups who would need special attention are not being detected.
4) Use of LVS results is better developed as far as external reporting to parents and the Inspectorate is concerned, than with respect to remedial use for *all* students, not just the extremely low performing students.
5) Schools frequently take little time for discussing the LVS results in the staff team, let alone for defining measures to improve performance.

The authors conclude that a lot of ground needs still to be covered before all schools are in a position to effectively use student achievement data to improve student learning. According to the Inspectorate only about 30% of primary school currently realize this approach. They propose that a "pressure and support" strategy is followed, where external accountability demands stimulate schools to perform better and professional training and support is made available to assist schools in further developing result oriented work.

6.3   Policy initiatives

The policies that have been described in the final sections of Chapters 3 and 4, are also relevant for student assessment policies and practices. In this chapter the role of formative assessment as part of the policies with respect to achievement oriented work has been highlighted. The experiences so far underline that adequate and sophisticated monitoring systems are a necessary but insufficient pre-condition for adequate application and use of these instruments.

In a general sense the emphasis on achievement orientation, improvement of achievement in basic subjects, and public accountability have created a climate that is favorable to a further development and optimization of student assessment. At the same time the intention to move to a central obligatory test at the end of primary schools appears to be a sensitive process. In a recent advice the Education Council, which is the most important advisory body to the Minister of Education, states that it prefers a situation where the central level formulates substantive demands to the test to a uniform obligatory test. The Council fears undesirable side effects of a centrally established, obligatory test for all primary schools (Education Council, 2012, p. 13). As a matter of fact the advice in question, titled "Controlled Space", expresses a concern for school autonomy in realizing a broad set of educational and pedagogical values, in a situation in which much emphasis is given to basic subjects, and measurement of outcomes in basic subjects.

REFERENCES

Amelsvoort, H. C. H. M. v., & Wolf, I. (2005). Risk based inspection in Europe. Utrecht: Inspectie van het Onderwijs.

Beelaerts, M., Bousché, C., De Goeij, L., De Graaff, E., Horsman, K., & Klifman, H. (1999). Projectplan Stimulering Kwaliteitszorg VO. (Available from [Q5, Burg. Reigerstraat 74, Utrecht]).

Béguin, A., & Ehren, M. (2011). Aspects of accountability and assessment in the Netherlands. *Z Erziehungswiss Zeitschrift für Erziehungswissenschaft, 14*(S1), 25-36.

Bekkers, V., de Kool, D., & Straten, G. (2012). Beleidsgericht Onderzoek Primair Onderwijs *Ouderbetrokkenheid bij schoolbeleid* (pp. 108).

Blok, H., Sleegers, P., & Karsten, S. (2008). Looking for a balance between internal and external evaluation of school quality: Evaluation of the SVI model. *Journal of Education Policy, 23*(4), 379-395.

Boer, H. de, Bosker, R. J., & Werf, M. P. C. v. d. (2007). De gevolgen van onder- en overadvisering. In I. v. h. Onderwijs (Ed.), *Inspectie van het Onderwijs, Onderadvisering in beeld* (pp. 83-92). Utrecht: Inspectie van het Onderwijs.

Bokdam, J., Berger, J., Van Oploo, M., & Volker., G. (2011) *Tussenmeting Convenant leerkracht 2011*. Zoetermeer: Regio Plan http://www.research.nl/files/rvb/reportcenter/Rapporten/B3883/B3883TusenmetingConvenantLeerkracht2011def.PDF

Branderhorst, E. M. (2005). *Wat is er mis met MIS?: ontwerp en evaluatie van een training in het gebruik van managementinformatie binnen het voortgezet onderwijs*: University of Twente.

Bruggen, J. C. v. (2007). Inspecteren van scholen: internationale ontwikkelingen en trends. *Impuls voor Onderwijsbegeleiding, 38*, 13-37.

Bruggen, J. C. v., & Mertens, F. (2001). Inspecteren voor openbaarheid vanuit bredere kwaliteitsnoties. In A. B. Dijkstra, S. Karsten, R. Veenstra & A. J. Visscher (Eds.), *Het oog der natie: Scholen op rapport [The eye of the nation: School reports]* (pp. 208-218). Assen: Koninklijke Van Gorcum.

Carnoy, M., Elmore, R. F., & Siskin, L. S. (2003). *The new accountability : high schools and high stakes testing*. New York: RoutledgeFalmer.

Cito Volgsysteem. (2012). Entreetoets. Voor groep 5, 6 en 7. Arnhem: Cito.

Cito Volgsysteem. (2012). Omzettingstabellen - Tijdvak 1 2011, from http://www2.Cito.nl/vo/ce/omzet2011_vmbo/tabellen/11184_omzettingstabel.htm

Creemers, B. P. M., & Hoeben, W. T. J. G. (1985). *Educational evaluation and the professional role of the teacher.* . Paper presented at the Contributions on the occasion of SVO's 20th anniversary The Hague.

De Minister van Onderwijs Cultuur en Wetenschap. (2003). Regeling inspectie van het onderwijs. *Uitleg Gele katern, 11*(13), 11-19.

De Vries & Communicatie bv. (2012). Uw Schoolgids conform wettelijke vereisten. *Schoolgids.org*, from http://www.schoolgids.org/schoolgids_wettelijke_eisen.html

De Wolf, I., & Verkroost, J. (2011). Evaluatie bsn de theorie en praktijk van het nieuwe onderwijstoezicht. *Tijdschrift voor Toezicht, 2*(2), 7-24.

Driessen, G., & Smeets, E. (2007). De relatie tussen prestaties en advies.

Dronkers, J., & Veenstra, R. (2001). Schoolprestatie-indicatoren in het voortgezet onderwijs: start, reacties en vervolg. *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties.*

EACEA P9 Eurydice. (2009). National Testing of Pupils in Europe: Objectives, Organisation and Use of Results (pp. 112): Education, Audiovisual and Culture Executive Agency.

Ecbo (2011) *Teaching 2020, a strong profession!* http://www.ecbo.nl/ECBO/ReferNet/docs/11-0315_Teacher_2020.pdf

Ehren, M., Leeuw, F. L., & Scheerens, J. (2005). On the impact of the Dutch educational supervision act. *American Journal of Evaluation, 26*(1), 60-76.

Eurydice. (2010). National system overview on education systems in Europe, the Netherlands (pp. 9).

Graafland, H. (2011). De Cito-eindtoets: hulpmiddel naast het schooladvies. Retrieved from http://www.jmouders.nl/Themas/School/De-Citoeindtoets-hulpmiddel-naast-het-schooladvies.htm

Hendriks, M., & Bosker, R. (2001). Systemen voor kwaliteitszorg in het primair en secundair onderwijs [Systems for quality care in primary and secondary education]. *Kwaliteitszorg. Onderwijskundig lexicon III*, 47-72.

Hendriks, M., Doolaard, S., & Bosker, R. (2002). Using school effectiveness as a knowledge base for self-evaluation in Dutch schools: The ZEBO-project. In A. J. Visscher & R. Coe (Eds.), School improvement through performance feedback (pp. 115-142). Lisse: Swets & Zeitlinger Publishers.

Hofman, R., Dijkstra, N., Hofman, W., & de Boom, J. (2004). Kwaliteitszorg in het primair onderwijs. *Deelstudie 1: peiling 2003/2004*.

Hofman, R. H., Dukstra, N. J., & Hofman, W. H. A. (2005). School self-evaluation instruments: an assessment framework. *International Journal of Leadership in Education, 8*(3), 253-272.

Inspectie van het Onderwijs. (2011). Tevredenheidsonderzoek PO, Periode januari - oktober 2010. Utrecht: Ministerie van Onderwijs, Cultuur en Wetenschap.

Inspectie van het Onderwijs. (2006). Jaarverslag 2005 (pp. 113). Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs. (2007). *Onderadvisering in beeld*. [Utrecht]; [Den Haag]: Inspectie van het Onderwijs ; Postbus 51 Infolijn [distr.].

Inspectie van het Onderwijs. (2007). Onderadvisering in beeld (pp. 98).

Inspectie van het Onderwijs. (2011). De staat van het onderwijs (C. e. W. Ministerie van Onderwijs, Trans.). In I. v. h. Onderwijs (Ed.), *Onderwijsverslag 2009/2010* (pp. 261): Ministerie van Onderwijs, Cultuur en Wetenschap.

Inspectie van het Onderwijs. (2011). Toezichtkader 2011 primair onderwijs en voortgezet onderwijs (pp. 22).

ITS Nijmegen. Het cohortonderzoek PRIMA Retrieved April, 2012, from http://www.ru.nl/its/onderwijs/afgerond_onderzoek/vm/het_cohortonderzoek/

ITS Nijmegen. (2009). Cool 5-18, Cohortonderzoek onderwijsloopbanen. *Basisonderwijs: Doel & Opzet* Retrieved April, 2012, from http://www.cool5-18.nl/doelenopzetbo/#4

Jacobs, J., Vrielnk, S. (2009). Levensfasefericht personeelsbeleid in het primair en voortgezet onderwijs. Nijmegen: ResearchNed bv.

Janssens, F. J. G. (1986). *De evaluatiepraktijken van leerkrachten. Een beschrijvend onderzoek naar het evalueren tijdens het rekenen in het primair onderwijs*: University of Groningen, Groningen.

Janssesns, F.J.G. & Van Amelsvoort, G. (2008). School self-evaluations and school inspections in Europe: An exploratory study. *Studies in Educational Evaluation, vol. 34,* 15-23.

Janssens, F. J. G. (in press). The evaluation of the impact of the publication of school performance indicators in the Netherlands. *Journal of Education Policies, 6*(1), in press.

Janssens, F. J. G., & Noorlander, C. W. (2010). *Goed onderwijsbestuur: opstellen over een nieuwe besturingsfilosofie van het onderwijs*. Den Haag: Den Haag : Boom Lemma.

Karsten, S., & Visscher, A. (2001). Ervaringen met het openbaar maken van schoolprestaties in Engeland en Frankrijk. *Het oog der natie: scholen op rapport. Standaarden voor de publicatie van schoolprestaties*, 36–53.

Kuyper van der, H., & Werf, M. (2007). De resultaten van VOCL'89, VOCL'93 en VOCL'99: vergelijkende analyses van prestaties en rendement.

Leune, J. M. G. (1983). *De reikwijdte van het onderwijsaanbod in het funderend onderwijs : referaat ten behoeve van het RION-congres op 28 en 29 maart 1983 te Groningen over "De kwaliteit van het onderwijs"*. Rotterdam: EUR.

Leune, J. M. G. (2007). *Verstandig onderwijsbeleid*. Antwerpen; Apeldoorn: Garant.

Lubbe, M. v. d. Pupil Monitoring System (PMS) for Primary Education. Retrieved from http://www.iaea.info/documents/paper_1162d1e9fe.pdf

Lubbe, M. v. d., Verhelst, N., Heuvelmans, T., & Staphorsius, G. (2005). *Verslag van een onderzoek naar de toelating van leerlingen in het voortgezet onderwijs.* Amsterdam.

Market, R. C. f. E. a. t. L. (2009). About ROA. Retrieved April, 2012, from http://www.roa.unimaas.nl/

Menting, D. (2012) De kwaliteitsnormerende functie van deugdelijkheideisen; waarborgen voor goed onderwijs. Den Haag: Onderwijsraad

Meijer, C. J. W., & Peschar, J. (1997). *Weer samen naar school onderweg : de evaluatie van het 'Weer Samen Naar School' beleid*. Groningen: Wolters-Noordhoff.

Middleton, P., & Slikkerveer, L. (2011). *Key figures 2006-2010 : education, culture and science*. [S. l.]: Ministry of Education, Culture and Science.

Ministerie van Onderwijs Cultuur en Wetenschap. (2011). Trends In Beeld 2011 (C. e. W. Ministerie van Onderwijs, Trans.). In D. Kennis (Ed.), (pp. 73): Ministerie van Onderwijs, Cultuur en Wetenschap.

Ministry of Education and Science, the Netherlands (1989). *Richness of the uncompleted : challenges facing Dutch education*. Hague, Netherlands: Ministry of Education and Science.

Ministry of Education and Science, the Netherlands. (1993). *Gezamenlijke richtgevende uitspraken: van de vertegenwoordigers van de onderwijskoepelorganisaties en de minister van Onderwijs en Wetenschappen in het Schevenings Beraad Bestuurlijke Vernieuwing, met betrekking tot gewenste ontwikkelingen in de bestuurlijke verhoudingen in primair en voortgezet onderwijs*. [Zoetermeer]; Leiden: Ministerie van Onderwijs en Wetenschappen ; Distributiecentrum DOP.

Moelands, H. A. (2005). *Schoolzelfevaluatie : het evalueren van en door scholen*. S.l.: s.n.].

Mulder, L., Roeleveld, J., Veen, I. v. d., & Vierke, H. (2005). Onderwijsachterstanden tussen 1988 en 2002: ontwikkelingen in basis- en voortgezet onderwijs. Nijmegen/ Amsterdam: ITS/SCO-Kohnstamm Instituut.

Mullis, I. V. S., Erberber, E., & Preuschoff, C. (2007). The TIMSS 2007 International Benchmarks of Student Achievement in Mathematics and Science. *TIMSS*, 339-347.

Mullis, I.V.S., International Association for the Evaluation of Educational, A., Third International, M., & Science, S.-R. (2000). *TIMSS 1999 : international mathematics report: findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: International Study Center, Boston College, Lynch School of Education.

OECD. *Education at a Glance 1998*: OECD Publishing.

OECD. *Education at a Glance 2011*: OECD Publishing.

OECD. *Highlights from Education at a Glance 2008*: OECD Publishing.

Ofqual (2012) The Netherlands. In: International Comparison of Senior Secondary Assessment (ICoSSA). Coventry: Ofqual.

Onderwijs Consumenten Organistatie. (2011). Cito-score binnen? Lees het leerlingrapport! Retrieved from http://www.onderwijsconsument.nl/ouders/artikel.php?aid=349

Onderwijs Raad. (2001). De basisvorming : aanpassing en toekomstbeeld. (pp. 65). Den Haag.

Onderwijs Raad. (2012). Geregelde ruimte (pp. 76). Den Haag.

Onderwijsraad. (2012). *Artikel 23 Grondwet in maatschappelijk perspectief : nieuwe richtingen aan de vrijheid van onderwijs*. Den Haag: Onderwijsraad.

Onderwijsvernieuwingen C.P.O. (2008). Tijd voor onderwijs. *Zoetermeer, Ministerie van Onderwijs, Cultuur en Wetenschap*.

Postlethwaite, T. N., & Ross, K. N. (1992). Effective Schools in Reading: Implications for Educational Planners. An Exploratory Study: International Association for the Evaluation of Educational, Achievement.

Programmabureau Onderwijs Bewijs. (2010). Prijsvraag: Onderwijs Bewijs. In M. v. O. C. e. Wetenschap (Ed.), (Vol. nr. 08BK2009B029, pp. 8). Den Haag.

RCEC (2012). Research Center for Examinations and Certification, from http://www.rcec.nl/en/

Runhaar, P.R. & K. Sanders (2007). P&O als intermediair tussen management en leraren? Tijdschrift voor HRM. Vol. 10, zomer 2007. pp. 54-77.

Scheerens, J. (1983). *Evaluatie-onderzoek en beleid: methodologische en organisatorische aspecten* (Vol. 68): Stichting voor Onderzoek van het Onderwijs.

Scheerens, J. (1993). Basic School Effectiveness Research: Items for a Research Agenda 1. *School Effectiveness and School Improvement, 4*(1), 17-36.

Scheerens, J. (1997). *De bevordering van schooleffectiviteit in het basisonderwijs: Mogelijkheden tot" flankerend beleid" bij klassenverkleining*: Universiteit Twente, faculteit Toegepaste Onderwijskunde.

Scheerens, J., Glas, C. A. W., Thomas, S. M., & Thomas, S. (2003). *Educational evaluation, assessment, and monitoring: a systemic approach* (Vol. 13): Taylor & Francis.

Scheerens, J., & Hendriks, M. (2004). Benchmarking the quality of education. *European Educational Research Journal, 3*(1), 101-114.

Scheerens, J., Seidel, T., Witziers, B., Hendriks, M., & Doornekamp, G. (2005). Positioning and validating the supervision framework (D. o. E. O. a. Management, Trans.) (pp. 430). Enschede/Kiel: University of Twente.

Schevenings Beraad Bestuurlijke Vernieuwing. (1993). *Gezamenlijke richtgevende uitspraken: van de vertegenwoordigers van de onderwijskoepelorganisaties en de minister van Onderwijs en Wetenschappen in het Schevenings Beraad Bestuurlijke Vernieuwing, met*

betrekking tot gewenste ontwikkelingen in de bestuurlijke verhoudingen in primair en voortgezet onderwijs*. Den Haag: Ministerie van Onderwijs en Wetenschappen.

Schildkamp, K. (2007). *The utilisation of a self-evaluation instrument for primary education*. University of Twente.

Sijtstra, J.M., Aarnoutse, C.A.J., Verhoeven, L., & Expertisecentrum, N. (1999). *Taalontwikkeling van nul tot twaalf : raamplan deel 2*. Nijmegen: Expertisecentrum Nederlands.

Sociaal en Cultureel Planbureau. Primair onderwijs en speciaal onderwijs cohortonderzoeken (PRIMA) Retrieved April 16th, 2012, from http://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Primair_onderwijs_en_speciaal_onderwijs_cohortonderzoeken_PRIMA

Stufflebeam, D.L; Foley, W.J.; Gephart, W.J.; Guba, E.G; Hammond, R.L.; Merriman, H.O.; and Provus, M.M. (1971). *Educational Evaluation and Decision Making*. Itasca, Ill.: F.E. Peacock.

Teurlings, C., & Vermeulen, M. (2004). Leren in veranderende schoolorganisaties. Alphen aan de Rijn: Kluwer.

van den Dool, P. (2012, Februari 16th). Onderwijsraad waarschuwt voor meer verplichtingen op scholen, electronic article, *NRC*. Retrieved from http://www.nrc.nl/nieuws/2012/02/16/onderwijsraad-waarschuwt-voor-verplichtingen-op-scholen/

Van der Schoot, F. (2008). Onderwijs op peil? Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.

Van der Steeg, M., Vermeer, N., & Lanser, D. (2011). Niveau onderwijs daalt. Vooral beste leerlingen blijven achter *Nederlandse onderwijsprestaties in perspectief*.

Van Wonderen, R. (2005). Monitor integraal personeelsbeleid in de Bve-sector. Derde meting (eindrapport). Leiden: Research voor Beleid

Van Petegem, P. (2005). *Vormgeven aan schoolbeleid*: Acco.

Vanhoof, J., & Petegem, V. (2001). Scholen vergelijkenderwijs: een internationaal vergelijkende studie over hoe scholen hun functioneren openbaar (kunnen/moeten) verantwoorden.

Vensters voor Verantwoording. (2012), from http://www.venstersvoorverantwoording.nl/nl-NL/pages/96/Home-kopie.aspx

Verantwoording, V. v. (2012). SchoolVensters Online, from http://www.schoolvo.nl/

PO RAAD (2012). PO-Raad en VO-Raad werken samen in stiching SchoolInfo, from http://www.poraad.nl/content/po-raad-en-vo-raad-werken-samen-stichting-schoolinfo

Vernieuwing Toezicht. (2009). Inspection Reform Programme. More effect, less burden (pp. 20).

Visscher, A., & Wild, P. (1997). The potential of information technology in support of teachers and educational managers managing their work environment. *Education and Information Technologies, 2*(4), 263-274.

Visscher, A.J., & Ehren, M. (2011). *De eenvoud en complexiteit van opbrengstgericht werken*. [Enschede]: [Universiteit Twente, Vakgroep Onderwijsorganisatie en -management].

Visser, M. (2003). *Schoolkeuzeprocedure, kwaliteit en kwaliteitszorg.* Rijksuniversiteit te Groningen.

Vlug, K. F. M. (1997). Because every pupil counts: the success of the pupil monitoring system in The Netherlands. *Education and Information Technologies, 2*(4), 287-306.

Voogt, J.C. (1995). Schooldiagnose. In H. P. M. Cremers (Ed.), *Onderwijskundig Lexicon*. Alphen aan den Rijn: Samsom.

VO-Raad (2011). Werkboek Horizontale Verantwoording VO (VO-Raad, Trans.). In C. Curre & S. Visser (Eds.), (pp. 51). Utrecht: VO-Raad.

WEC RAAD (2007). Kwaliteitsagenda Speciaal Onderwijs 2007-2011 (voor de ondersteuning van scholen voor (voortgezet) speciaal onderwijs en Regionale Expertise Centra) (pp. 11).

Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation, 1997*(76), 41-55.

Werf, M., & Kuyper, H. (2007). De resultaten van VOCL'89, VOCL'93 EN VOCL'99: vergelijkende analyses van prestaties en rendement.

Other unpublished documentary sources

Evaluatie van de Wet op het basisonderwijs (1994).Kamerstukken II 1993-1994, 21 630, nr. 9.

Rijksbegroting voor het jaar 1987 (1987).Kamerstukken II 1986-1987, 19 700 VIII, nr. 65.

Scholen voor morgen. Samen op weg naar duurzame kwaliteit in het primair onderwijs (2007). *Kamerbrief over de voortgang lerarenregister* (2012). 374636

Kwaliteitsagenda Voortgezet Onderwijs. Tekenen voor Kwaliteit. Afspraken voor een beter voorgezet onderwijs 2008 - 2011 (2008).

Parlementair onderzoek Onderwijsvernieuwingen (2008).Kamerstukken II 2007-2008, 31 007, nr. 8.

Primair Onderwijs (2008).Kamerstukken II 2007-2008, 31 293, nr. 1.

Vaststelling van de begrotingsstaten van het Ministerie van Onderwijs, Cultuur en Wetenschap (VIII) voor het jaar 2011 (2011).Kamerstukken II 2010-2011, 32 500 VIII, nr. 211.

Besluit met het tijdstip van inwerkingtreding van de Wet van 2 februari 2012 tot wijziging van de Wet op het onderwijstoezicht (2012).Kamerstukken 2012, 119,

Wet Primair Onderwijs (2012).Kamerstukken II 2012-2013, 33 157, nr. 7.

Wet op het voortgezet onderwijs (1963).

Wet op het primair onderwijs (1981).
Wijzigingswet op de Wet op het onderwijstoezicht (2012).Kamerstukken 2012, 118,

ANNEX

References with Table 2.1, Chapter 2: *The Netherlands' results in TIMSS, PIRLS and PISA surveys*

Survey	Mathematics	Science	Mathematics and Science Literacy	Reading	Problem-solving
	Score	Score	Score	Score	Score
TIMSS 95 - final year			559`		
TIMSS 95 - 3	493#	499##			
TIMSS 95 - 4	549*	530^^			
TIMSS 95 - 7	516~	517~~			
TIMSS 95 - 8	529**	536^			
TIMSS 99 - 8	540**	545^			
TIMSS 03 - 4	540*	525^^			
TIMSS 03 - 8	536**	536^			
TIMSS 07 - 4	535*	523^^			
TIMSS – Advanced 08 - 12	552***				
PIRLS 01				554''''	
PIRLS 06				547''	
PISA 00	564	529		532	
PISA 03	538	521		513	520
PISA 06	531	525		507	
PISA 09	526	522		508	

*** Mullis, I. V. S., Martin, M. O., Robitaille, D. F., & Foy, P. (2009). TIMSS Advanced 2008 International Report: Findings from IEA's Study of Achievement in Advanced Mathematics and Physics in the Final Year of Secondary School. Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

~~ Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1997). The Science Achievement in the Middle School Years Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

~ Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1997). The Mathematics Achievement in the Middle School Years Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1998). The Science Achievement in the Primary School Years Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

* Martin, M. O., Mullis, I. V. S., Foy, P., Olson, J. F., Erberber, E., Preuschoff, C., & Galia, J. (2008). TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

** Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1998). The Mathematics Achievement in the Primary School Years. Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

` Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1998). The Mathematics and Science Achievement in the Final Years of Secondary School. Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

^^ Mullis, I. V. S., Martin, M. O., Foy, P., Olson, J. F., Preuschoff, C., Erberber, E., Galia, J. (2008). TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

^ Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

"" Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools. Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.

" Mullis, I. V. S., Martin, M. O., Kennedy, A. M., & Foy , P. (2007). IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries. Chestnut Hill: MA: TIMSS & PIRLS International Study Center, Boston College.