

## EDUCATIONAL EVALUATION AND ASSESSMENT IN THE NETHERLANDS

Country background report for the OECD study on *Evaluation and Assessment Frameworks for Improving School Outcomes*<sup>1</sup>; *executive summary*

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

### *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 sum 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.

---

<sup>1</sup> This study was supported by a grant from the Ministry of Education and the Inspectorate of Education in the Netherlands

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 of 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 decision 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.