CIDREE

Consortium of Institutions for Development and Research in Education in Europe

Schools for Quality - What Data-based Approaches Can Contribute

Schools for Quality - What Data-based Approaches Can Contribute

CIDREE

Schools for Quality - What Data-based Approaches Can Contribute

Authors: Maria da Conceição Castro Ramos
Peter Dobbelstein
Helder Guerreiro
Paula Hammond
Dougal Hutchison
Bart Maes
Thomas Neidhardt
Rainer Peek
Zora-Rutar Ilc
Ian Schagen
Heinz Schirp
Alejandro Tiana Ferrer
Els Ver Eecke
Veerle Verhaegen
Hana Zufanová

Editor: Peter Dobbelstein
Thomas Neidhardt
Production: Thomas Neidhardt
Renate Prahl
Titelbild: CORBIS
Cover design: Renate Prahl
First published: 2006

ISBN 3-00-020494-6

Order address: CIDREE Administration
Landgoed - Abdijsite Roosendael
B - 2860 Sint-Katelijne-Waver
Belgium
cidree@roosendael.be
www.cidree.org

Foreword

The sixth yearbook of CIDREE addresses one of the hottest current themes of education policy in most European countries: quality evaluation and, particularly, the assessment of school level quality through measured outputs. This theme has been discussed within CIDREE for many years and it has figured long as major item on the research and development agenda of its member institutions. Some of them have directly been involved in developing and using standardised tests, interpreting their results and using them to give professional orientation to schools. This topic has also appeared in various CIDREE collaborative projects.

This is not surprising. CIDREE member institutions are providing research and development services to national governments which, in all European countries, are making serious efforts to improve the quality of school education and to make schools more effective and more accountable. They also share the concerns of these governments about the possible, and not yet sufficiently known, impacts of assessment through standardised tests on teaching and learning and on the behaviour of teachers and schools. Schools and teachers need reliable feedback about the results of their work: this is a condition of continuous improvement of their way to organise teaching and learning. Without such a feedback no real improvement can take place. Making the quality of this feedback better is today one of the highest priorities of applied policy oriented educational research. This is why we talk more and more about the reliability of tests, the necessity to measure added value and the importance of using appropriate techniques to calculate it.

There seem to be an increasing consensus among both politicians and educational researchers that the best way to improve the quality and effectiveness of education is to combine school autonomy with strong feedback mechanisms and a regulatory framework that gives strong incentives to schools to use the feedback they get for self-improvement. Schools can do this only if they behave as intelligent and motivated agents who are capable to interpret the evidence they receive about the outcome of their work, and are willing and capable to translate this evidence into developmental actions. The creation of this type of organisational intelligence requires external support to be provided by governments and development experts.

Schools, in most European countries, are increasingly responsible for their own quality policies and they have to exercise this responsibility in a creative and intelligent way. This is in accordance with the recommendation adopted in 2001 by the European Union on quality evaluation in school education which encourages a good balance between external and internal evaluation, and suggests that evaluation should be conceived as a learning process. This is also in harmony with the notion of intelligent accountability used more and more frequently by politicians and experts committed to educational development.

I am convinced that this yearbook of CIDREE has a great potential to contribute to the improvement of this type of organisational and developmental intelligence. Its reading can be recommended to all of those who are involved in shaping and implementing policies of quality and evaluation in education. I hope this publication, similarly to the previous yearbooks, will be read and used not only by the closer CIDREE community but also by many practitioners, researchers, developers and decision-makers engaged in improving quality in school education.

Gábor Halász
National Institute for Public Education, Hungary
President of CIDREE

Contents

Introduction

Peter Dobbelstein, Thomas Neidhardt

The Quality of Schools between Self-Evaluation and Accountability

Supported by national and international school achievement studies initiatives and approaches to improving the quality of schools in large parts of Europe have been given new orientation and accentuation. Overall, this development has on the one hand led to a greater independence of the educational agent "school" and on the other hand to an empirical turn-about in educational policies and politics: the accountability and quality development of individual schools and the directing and controlling of the whole system are being aligned to data gathered locally and to the effectiveness and output quality of schools that can be measured and observed (cf. e.g. Ackeren, I. van, 2003). In addition this development is characterized by a wide debate about factors of "good" schools among educational experts and educational policy makers.

Starting from results of empirical school achievement research that has had a long tradition in Anglo-American countries and has been based on extensive historical experiences in comparing school achievements, the so-called School-Effectiveness-Research has developed in Europe in the 1980ies - particularly in Great Britain, in the Scandinavian countries and in the Netherlands (cf. among others Bosker/Scheerens, 1997 and before Creemers/Scheerens, 1989). From a variety of studies on benchmarking which over time not only compared pupil achievements with each other but also evaluations of their respective schools by teachers, parents and pupils more and more demanding and differentiated characteristics of "good" schools have been worked out which in addition to the criterion of achievement take into account possible predictor variables: e.g. achievement orientation of the school, pedagogical engagement of the staff, leadership qualities of administration and teaching staff of a school, climate of trust, cooperation between teachers or parent involvement (cf. as an overview of the state of the discussion at the beginning of the 1990ies Steffens/Bargel, 1993 and later Peek/Neumann, 2003).

Quality is multi-dimensional

Evaluation and school innovations linked to it only make sense if they lead to an improvement of the quality of schools. The basic premise is: the controlling of education and the development of schools which take their characteristic features from pedagogy by means of the definition of standards, accountability and evaluation-centred controlling are closely connected and must be related to a concept of quality which can be widely agreed upon (cf. e.g. OECD, 2006).

A concept of quality of schools and "good" school respectively that is sound and can be used for development processes in schools is, however, complex and has many facets (cf. Bildungsdirektion of Kanton Zurich, 2001, 2004). In spite of all differences of possible definitions of "quality", agreement occurs with regard to the complexity and the multi-dimensional nature of the concept, as Alejandro Tiana Ferrer describes it in his article "*Assessing Quality in Education: Concepts, Models and Instruments"* from the perspective of the Spanish system. As a basic factor of the discussion he singles out "*effectiveness*" which is the level at which educational aims were achieved and could be demonstrated e.g. by standardized tests. Further aspects of this multi-dimensional nature are "*efficiency*" meaning time and effort in relation to resources and results and "*the fulfilment of needs and expectations*". The last item is being discussed in a great number of countries with the idea of a stronger orientation of curriculum and school development along educational standards and levels of expectation corresponding to different pressures on schools (e.g. schools with especially severe social problems). "*Appropriateness*" is another building block of an adequate analysis of quality.
Hana Zufanová discusses in her article "*On the Meaning of Internal and External Evaluation for Quality of Education during Educational Reform*" how, since 2005, the Czech school system has tried to take into consideration the complexity of the quality of schools with their internal school evaluation and external evaluation. Accompanying the growing independence of schools the Czech School Inspectorate has made internal school evaluation and external evaluation compulsory. Achievement data are collected at the level of classes and individual pupils; in addition school portfolios on the course of the decision processes have been included. Questionnaires for individual groups or to particular questions as well as observations on school climate

complete the picture. As a novelty the evaluation of teaching staff by pupils was introduced in Czech schools and the teacher's portfolio gathers all documents which show the professional development of the teacher over time.

School as Educational Agent - between Independence and Control

The central tenet of argumentation is that an "optimization of results" or respectively "reaching the aims" can best take place in largely independent schools which are allowed to develop under the specific local conditions. The discussion is determined by the relationship between educational freedom and public responsibility (Liket, 1993) which is defined by stating the degree of autonomy for the individual school. A public school system - and there seems to be a high degree of agreement on this point - can only bestow a partial autonomy on individual schools which has to be clearly described legally in the interest of clear-cut requirements (cf. Kotthoff, 2003).

With the reception of results pertaining school effectiveness but also through studies comparing classes and schools, the importance of schools as educational agents is indisputable. Assuming that every school, within the legal framework and in spite of the general basic structure induced by that, determines its daily life and its specific culture through the way it manages and shapes the tasks of teaching, evaluating, training and educating, the main question is - see Mortimore (1997) in a summary of school-effectiveness-research - how individual schools deal with their "input", assign and use their resources. The different starting points of European educational systems have the assumption in common that the effectiveness and quality of the educational agent school could be enhanced by internal evaluation and the external gathering and evaluation of data with corresponding feedback.
Ian Schagen, Dougal Hutchison and Paula Hammond point out in their article "*League Tables and Health Checks: The Use of Statistical Data for School Accountability and Self-Evaluation*" that the connection between "feedback" and "enhanced performance" is by no means so certain and automatic as has been considered for a long time. Different studies come to the conclusion that a feedback of performance data is helpful but is being used by schools in very

differing ways; school staffs and administrations must be trained to read and use the data. The dates as such do not say anything and only give answers to questions put to them.
The authors place the current points of view between two models that have been discussed for some time which claim to lead to an improvement of schools: the model of responsibility (accountability) where on the basis of "League tables" results are published in the form of rankings and will supposedly lead to an improvement of achievements through public pressure on the basis of comparisons of results, and the model of self-evaluation which starts from the idea that schools will push ahead with their educational activities on the basis of the data made available to them. Other models like e.g. official school inspections are to be placed somewhere in between.
On the scale between on the one hand complete self-evaluation in the sense of exclusive internal evaluation and on the other hand exclusive external control through external evaluation, it becomes obvious that no country favours internal evaluation exclusively. This is shown in Veerle Verhaegen's article "*Equilibrium - On the Balance between Internal and External Evaluation in a Number of European Educational Systems*" which compares different approaches of European countries and regions towards analysing school quality. Whereas some countries and regions are more in favour of external evaluation, the majority aims at a desirable balance between the ends, in the view of the author.
On the second scale which Verhaegen sets up between "pressure" and "support" almost all countries have reached a balanced position. Schools get support through e.g. financial means for self-evaluation, through supportive networks and through opportunities for in-service education and further education. Pressure on the other hand will be exerted by external checks on self-evaluation and through the publishing of test and observation results (ranking). From the perspective of the author also on the scale between "pressure" and "support" a dynamic equilibrium between intrinsic motivation and extrinsic acknowledgement seems desirable.
The third dimension shows the direction of willingness for change (implementation of change): either by hierarchical structures from top to bottom (top down) or brought about by the "customers" of the system (teachers, pupils, parents; bottom up). Again, following Verhaegen's line of argument, the ideal solution is for the top to support the developments at the bottom.

Central Elements of School Development Based on Empirical Data

In spite of all the differences between approaches and processes some central elements in the discussion about data based school development and control can be identified:

- Of crucial importance is *the identification of standards which can bring about orientation*: in the sense of a definition of goals which are highly practicable they accentuate the perspectives for the development of individual schools. Not only the reception of national educational standards and curricula and their school and curricular realization is referred to here. In addition the identification and normative definition of standards refers to all areas of the explicit and implicit set of aims of the individual school.

  At this point definitions of aims by schools and consequently lines of school development will also shape exercise types and the setting of priorities in tests or central examinations as is demonstrated by Zora-Rutar Ilc using Slovenia as an example in her article "*Monitoring the Impact of the Matura on Teaching and Learning*". The contribution from Slovenia points out that pupils regard this examination as determining factor. Consequently pupils and teachers, according to the results of an empirical study, consider the consolidation of knowledge as the appropriate preparation for the matura whereas problem-solving thinking and the development of ideas are less helpful in their opinion. In view of the great weight of "knowledge", teachers mention lack of time as the main argument for the small part played by independent pupil activity. The same pattern of argumentation is true for questions of method. Overall there is a significant increase of the share of methods as routine and a significant decrease of the share of problem-solving and the independent developing of ideas.

  In his review of American research using the example of High Stake Tests Heinz Schirp also critically examines the question which effect central achievement tests may have on teaching and learning *"Like the Fishermen in the Maelstrom!?"Central Quantitative Performance Tests and Qualitative School Development in the USA"*. As a reaction to a critical analysis of

the American educational system High Stake Tests were developed in the USA in 1983. According to the results of such achievement tests published in ranking tables and accompanied by state-wide tests financial awards and sanctions are handed out. The effect is that teachers in the USA, under the pressure of centralized tests, tend to narrow the school curriculum with regard to topics, contents and methods. This "Narrowing the Curriculum" brings about that schools partially fashion their teaching and learning in subjects, timetables and complexity according to the tests. Teachers use a disproportionately high amount of time teaching for "Test-Cleverness". "Teaching to the test" means that what is to be tested will be taught. The final consequence of all of this is "teaching the test itself".

- Closely connected with the identification of standards is "negotiating the meaning of quality" and directly related "identifying quality criteria and factors" and in addition "developing quality indicators" which is extensively dealt with by Alejandro Tiana Ferrer in his article "*Assessing Quality in Education: Concepts, Models and Instruments"*.
  Quality tableaus which form the basis for school inspection and quality analysis respectively e.g. in NRW and many states of Germany (cf. Bartz & Müller, 2005) clearly delineate for the schools which aspects and indicators are used for their evaluation. Considering the wide scope of the discussion about quality such a tableau cannot offer schools the only possible criterion but a possibility of "good teaching and learning" and "good school" which can be widely accepted. Even if these criteria will already get the character of norms by simply being implemented and thus becoming the basis for estimating quality and of feedback, the discussion about what should be achieved with regard to quality in individual instances has not finished, of course, but on the contrary if we talk about the consequences and development perspectives of individual schools (Helmke, 2003) has just been given a foundation to start from.

- In all countries more autonomy and *accountability* are closely connected with growing demands for accountability: the more independent schools are the higher is the expectation that they show and prove how much they live up to standards (orientation following criteria) and how they are placed in relation to other

schools of comparable school types (orientation following norms). Evaluation-based control must be seen in direct connection to that: the debate centres on the question of macro- and micro-control of measures like resources of individual schools to safeguard reaching the defined goals through e.g. a model or a school programme respectively, through standards or value expectations (cf. Peek, 2006). Schools, however, operate under unequal conditions so that one single criterion is considered to be unfair. Ian Schagen, Dougal Hutchison and Paula Hammond take up this discussion in their article "*League Tables and Health Checks: The Use of Statistical Data for School Accountability and Self-Evaluation"* using England as the example. The criticism of using data as criterion for school quality in an uncritical way led to the introduction of the concept of value added in the field of education. When the measuring of achievements brought about e.g. better measurements than could be expected according to the school district, it was seen as having gained value added. After the change of school inspections from HMI to Ofsted (Office for Standards in Education) added value measurements of schools were developed which only considered the comparison of *Keystages* and left out possible social factors of influence. Later models publish *contextualized value added* which take into account entry and exit achievements and background factors.

- In Germany a comparable discussion has brought about e.g. in Northrhine-Westfalia that in the case of centralized achievement tests so-called types of location have been singled out and that with the development of social indexes at the level of the individual school, disadvantaged schools in social problem areas can be differentiated from less disadvantaged schools with parents keen on education and that correspondingly the tools of comparison can be adjusted. Peter Dobbelstein's and Rainer Peek's article "*Potentials of Centralized Proficiency Tests for the Improvement of the Quality of Schools - an Example from North-Rhine-Westphalia"* focusses on two questions: firstly, how external data material can be edited, and secondly, how a feedback for schools can be shaped in order to reach the highest possible impact on the improvement of the teaching and learning process. New horizons can open up for schools: They do not only receive an orientation towards norms ("Our results are better or worse than those of other schools."), but even more a criterion-rated

comparative view ("Which compentences do our pupils reach, which should they reach?").

- Not only schools but institutions for evaluation and instruments of evaluation need quality control and development. They, too, have to prove their effectiveness and the quality of their results which means that they also have to be evaluated and measured on a regular basis. This aspect is dealt with in the article by Conceicao Castro Ramos and Helder Guerreiro "*Perceiving the Impact of Inspections on the Improvement of School Quality*" and it presents a newly developed programme by the Portuguese Inspectorate. It is a programme that allows the team of inspectors to evaluate their own actions and to react in a flexible way. The starting point was the wish of the Portuguese Inspectorate to learn more about the strengths, weaknesses, effectiveness and room for improvement of their own work in schools. Checks within SICI (The Standing International Conference of Inspectorates) showed that most European institutions dealing with inspections successfully use those measures of self-evaluation, in most cases with the help of questionnaires and that they develop midterm and long -term aims from them.

……. with all this variety

The contributions to this book in spite of a great variety of approaches and accentuations show wide areas of an all encompassing consensus. There is agreement e.g. on the notion that the benchmarks mentioned for school development allow no one-sided measures but require a mixed system of self-control and control from outside at the centre of which is the individual school capable of acting on its own.
The concept is as follows: educational authorities set a clear framework. That means for schools that they are directed schools with clear leadership competencies for the administration within the framework of a defined partial autonomy so that they know their degree of autonomy with regard to financial means, organisation and teaching and learning.

Greater autonomy calls for greater responsibility for individual schools so that every school must put into practice its internally conceived quality management system including self-evaluation in order to be able to recognize whether the measures for school

development really lead to qualitative improvements. This internal self-evaluation in conjunction with measures like standardized tests or external quality analysis and school inspection respectively make it possible to name strengths, weaknesses and strategies for improvement; in addition these results offer knowledge on controlling to the system level which can use it for quality improvement of the educational system as a whole.

References

Ackeren, I. van (2003). *Evaluation, Rückmeldung und Schulentwicklung. Erfahrungen mit zentralen Tests, Prüfungen und Inspektionen in England, Frankreich und den Niederlanden.* Münster.

Bartz, A. & Müller, S. (2005). Schulinspektion in NRW - Pilotphase 2005/06, in *Riecke-Baulecke, T. (Hrsg.): Externe Evaluation. Fakten, Gründe. (Schulmanagement-Handbuch 116)*. München, S. 7-26.

Bildungsdirektion des Kantons Zürich (2001). *Handbuch 1 - Schulqualität*. Zürich.

Bildungsdirektion des Kantons Zürich (2004). *Handbuch 2 - Schulqualität*. Zürich.

Bosker, R. & Scheerens, J. (1997). *The Foundations of Educational Effectiveness*. Oxford, New York, Tokyo.

Creemers, B. & Scheerens, J. (1989). Developments in school effectiveness research, in *International Journal of Educational Research 13 (1989)*, pp 685-825.

Helmke, A. (2003). *Unterrichtsqualität erfassen - bewerten - verbessern*. Seelze.

Kotthoff, H.-G. (2003). *Bessere Schulen durch Evaluation? Internationale Erfahrungen*. Münster.

Liket, T. (1993). *Freiheit und Verantwortung*. Gütersloh.

Mortimore, P. (1997). Auf der Suche nach neuen Ressourcen. Die Forschung zur Wirksamkeit von Schule, in *Böttcher, W., Weishaupt, H. & Weiß, M. (Hrsg.): Wege zu einer neuen Bildungsökonomie*. Weinheim, München, S. 61-71.

OECD (Hrsg.) (2006). *Bildung auf einen Blick - OECD-Indikatoren 2006*. Paris.

Peek, R. & Neumann, A. (2003). Schulische und unterrichtliche Prozessvariablen in internationalen Schulleistungsstudien, in *Auernheimer, G. (Hrsg.): PISA - Schieflagen im Bildungssystem. Die Benachteiligung der Migrantenkinder*. Opladen, S. 139-159.

Peek, R. (2006). Dateninduzierte Schulentwicklung, in *Buchen, H. & Rolff, H.-G. (Hrsg.): Professionswissen Schulleitung*. Weinheim, S. 1343-1366.
Steffens, U. & Bargel, T. (1993). *Erkundungen zur Qualität von Schule*. Neuwied.

Assessing Quality in Education: Concepts, Models and Instruments

Alejandro Tiana Ferrer[1]

Quality in education: a complex and polysemous term

Many of the problems that arise when the meaning of the term **quality** is analysed in relation to education are given by its own polysemy and complexity, even though it is so widely used that its ambiguity remains concealed. In fact, the term has being extensively used since the early 80s, and it has become a key element in the educational discourse. The fact that the Education Ministers of the OECD member countries met in 1990 for the conference *High-quality Education and Training for All* (OECD, 1992) clearly shows the importance attached by then to a phenomenon that had been anticipated ten years before (OECD, 1983). As it was stated at the opening of the *IX Congreso español de Pedagogía* (9th Spanish Congress for Education), which was devoted to that same topic, "quality and excellence are desirable goals in education and in every human activity: no one is against excellence and it is difficult to think of a school or organisation satisfied with its mediocrity" (*Sociedad Española de Pedagogía*, 1988:7).

There are external and internal reasons to the education system that explain the increasing interest in the improvement of quality in education. The external reasons include the pressure of the economy on education, due to the globalization process, as well as the new emphasis given to education as a development factor, or the reconsideration of the role of the State as a service provider. The internal reasons refer to factors such as the uneasiness generated in the last decades by the growth and the excessive student numbers in the education systems, the obvious tension between the search for excellence and equity, or the disappointment caused by the educational reform processes (Tiana, 1996 b).

Despite this concern, the term *quality* is far from having a meaning unanimously accepted. In the mid 80s, an OECD report included a

[1] Secretary-General for Education. Spanish Ministry of Education and Science.

statement which is still shared: “In fact, *quality* has different meanings for the different observers and interest groups, since the perceptions of the priorities for a change are not shared by all”, that is why “it is not strange for the statements on quality in education to be usually full of controversy” (OECD, 1991: 21, 39). From a more critical point of view, some authors think that the content different social agents associate to the term and the orientation towards their particular aims is made easier thanks to the high level of ambiguity found in the term (Álvarez-Tostado, 1989).

Agreement as regards the general points and disagreement in relation to the details is the contradiction found when dealing with quality applied to education, a contradiction that is caused by that polysemy, lack of definition and ambiguity.

At least, two conditions have to be met to discuss this question coherently.

The first condition is to accept the *complexity* of the concept, which implies that it has to be analysed from different perspectives. Thus, on the one hand, if quality is analysed from the point of view of the target groups, it is possible to distinguish between an individual and a social perspective, which do not necessarily coincide. For instance, the differentiation or equity established by an education system may be relevant for a social group but not necessarily as relevant for each of its individuals.

On the other hand, either a macroscopic perspective (covering wide areas or the whole education system) or a microscopic one (focusing on specific schools or particular education programmes) may also be taken to approach the question of quality.

Therefore, quality should not be referred to in abstract terms and the point of view taken for its analysis should always be made clear.

The second condition is to acknowledge the *multidimensional* nature of the concept. In fact, it is a construct where several dimensions can be identified.

A first approach to quality in education should be understood as *effectiveness*, that is, as the extent to which the educational aims have

been effectively achieved. These aims have been associated to students' educational performance or achievement and have been assessed through standard tests or the grades awarded by the teachers. This approach is similar to those first adopted by the industrial and productive world, where the perfection of the product was the fundamental quality criterion.

The main outstanding contribution of this approach is that it focusses on relevant and objective elements of the education process: achievement, models and instruments for calibration, and performance assessment. The weak points of this approach include the excessive reduction as regards the identification of the outcomes to be taken into consideration, the lack of reflection on the appropriateness of the achievement, the possible lack of sufficiently clear aims and, above all, the abstraction made of school processes, which finally determine the educational outcomes.

A second approach to quality in education should be seen as *efficiency*, that is, as the level of suitability between what has been achieved and the resources used for that achievement. Although linked to economic factors, the term *resources* refers not only to material or economic components, but also to personal and organisational elements, and the relation between that and the outcomes does not necessarily need to be established in cost-benefit terms.

The development of indicators related to human and economic resources, such as the pupil-teacher *ratio*, the rates of pupils who finish their studies and the drop-out rates, the expenditure on education per pupil and other similar indicators, are one of the main contributions of this concept of quality. This approach and the one mentioned before have a weak point in common: the abstraction made of the school processes, which reinforces the *input-output* model applied to education.

A third approach to quality in education underlines the *fulfilment of needs and expectations*. Its appearance and development is due to the fact that it provides a prominent role for the context as well as for the school processes. The new trends in quality management and overall quality related to the field of production that have influenced the world of education should be included here (López-Rupérez, 1994; Gento, 1996).

This concept of quality in education has been developed along two different lines. On the one hand, it has drawn attention to the *appropriateness* of the educational aims and achievements. It questions the meaning, importance and suitability of the established aims and achievements, raising questions such as the *added value* on the part of the educational institutions. On the other hand, this approach has stressed the need of including *fulfilment* among the dimensions to be taken into account.

A proper discussion on *quality in education* should acknowledge its multidimensional nature, incorporating in the analysis the dimensions of *effectiveness*, *efficiency*, *appropriateness* and *fulfilment*. The recognition of these dimensions opens the door to a relational conception of quality and facilitates a positive answer to the question about its possible assessment.

Assessing quality in education: a possible task under certain conditions

At present, the assessment practice has been generalized to such an extent, that everything seems to be subject to evaluation in the field of education.

There are several reasons that explain this phenomenon, such as those referring to the processes of change inherent to education systems (House, 1993; Tiana, 1996a). A broad consensus seems to exist as regards the increasingly importance and significance given to assessment in education and training.

One of the main reasons to carry out an assessment is the improvement of educational activity. The evaluation of a reality involves the analysis of its innermost features in order to assess: invading the *privacy* of a reality and of the people under assessment implies, as compensation, that attention should be paid to their possibilities for improvement. As a last resort, the emphasis on improvement is a moral demand that falls on the evaluator, as compensation for the power he/she holds in the assessment process.

Among the realities that constitute the possible object of assessment, there is one whose analysis is especially difficult and this is quality in

education, an expression normally used as the determining factor of the term *assessment*. This difficulty cannot be denied. Although it is easy to understand what the assessment of a pupil's learning involves or the running of a school, there are greater difficulties when an education system or the quality in education is assessed. In many cases, the expression is used with a reductionist meaning and/or mainly for propaganda purposes.

Therefore, the expression *quality assessment* should be used with caution and restriction, avoiding its indiscriminate use; and should only be applied to the real quality assessment understood in the complex, multidimensional and relational sense referred to before. According to this criterion a correct use of this expression should meet three additional conditions.

First, the *integrity* of its approach should be guaranteed. In other words, it should only be used for those initiatives whose aim is to cover the main aspects of the reality assessed and not only some of them. Second, the *global nature* of the assessment should be given priority as a whole, and not its individual components. Integrity and global nature are two additional requirements that favour that the object evaluated may be seen as a whole, including its *relevant and significant* aspects. Third, the demand for the *opening to the qualitative dimension*, focusing on aspects more directly related to the concept of the quality adopted.

If these requirements are not met, the activity faced has been probably conceived with the aim of mystifying or without sufficient thought. Therefore, it is correct to talk about assessing the quality of a school or education system, provided that an overall assessment is to be carried out in accordance with certain criteria already established and related to the most relevant and significant dimensions for their qualitative assessment. Another issue is the difficulties that inevitably will arise when explaining the meaning of quality in such contexts and when that quality is assessed.

Therefore, judging from these pages a restrictive use of the expression *assessing quality in education* is favoured. Its use should be restricted to those cases in which the meaning of *quality* applied to education has already been established and/or negotiated, and in which certain methodological and instrumental conditions, which allow for overall

and not partial judgements have been met. In all these cases quality improvement is clearly aimed at. In the event that these conditions are not met, it would be better not to use that expression but to substitute it for one which is less ambitious and more realistic. Consequently, that expression should be restricted to articulated assessment programmes rather than applied to specific assessment actions or initiatives.

Elements for a quality assessment strategy

Negotiating the meaning of quality

In the absence of a unanimously accepted definition of quality education, the meaning of the term has to be negotiated in order to reach, at least, a minimum agreement in relation to its characteristics and the criteria used for its identification.

This demand for negotiation is increasingly present in today's education systems, in which democratic participation is a shared value appreciated at its different levels. Thus, the current trend in the education policy area is an increasing negotiation over the meaning of quality in the case that public debate is used as a strategy prior to the approval of legislative or administrative measures. The *Etats Généraux sur l'éducation*, held in Quebec in the mid 90s, is a significant example of this way of acting in which the main social agents were set to work within a process of a fully open public debate on the reforms to be introduced in the education system. And the same could be said of the process followed in very different countries to define a new curriculum for the different stages of education, or of the various attempts made in several places to define *performance standards* (OECD, 1995). As a last resort, policy negotiation of this kind, which may and must be applied to the agreement on the meaning of quality in education, aims at providing an answer to the increasing demands in favour of a new educational pact; a pact similar to the one encouraged by Juan Carlos Tedesco (Tedesco, 1995), which falls within a tendency that authors like Hans Peter Schneider call "new type decision-making" or decision-making of irreversible consequences.

Negotiation processes on the meaning of *quality in education* are not only conducted in the field of education policy. Similar processes are

also carried out at the level of schools when school projects or documents with similar names are drawn up. The purpose of these plans or documents is to clarify in writing those aims and criteria that an educational community deems important in order to achieve its educational goals. Thus, the design of this type of documents implies a process of negotiation about what is worth preserving and how to go about it. Whether with these or with different words, the main debate is focused on what must be understood by quality education in that particular institutional context.

These modest but undoubtedly promising initiatives, which are not free from difficulties, clearly trace out the route to be followed. For those that understand the agreement as the absence of conflict, it is worth mentioning that the above-mentioned process of negotiation of meaning is a slow and complex process. Since the different legitimate interests found in a school or education system do not always coincide, it is necessary to accept the existence of the conflict along the negotiation process. It is not a question of denying its presence but rather of finding institutional and relational framework that facilitates the integration and handling of the conflict when it arises.

Identifying quality criteria and factors

The main aim of the negotiation processes of the meaning of quality is its definition, either through a specific statement or, more often, through a series of identity features. That is, negotiation does not always lead to a precise definition of what should be understood by quality in education. There are many times in which it only points out several criteria that show its attainment. As a general rule, such criteria specifically include the main educational outcomes to be achieved, although they do not exclude elements such as a favourable school climate, the answer of the school to the demands of its environment, the stimulation of the activity of the pupil, or the democratic participation at the heart of the school community, which are some of the most frequently mentioned elements.

However, a logical analysis of this type of criteria shows that not all of them belong to the same category. Some of them are certainly *identification criteria* of quality, like those referred to the results attained, and do not necessarily need to be formulated only in educational terms. Some others usually point out the predictable factors, which predict the achievement of those objectives. There are

even criteria that belong to both categories at the same time. For instance, including the existence of a good *school climate* among the quality criteria, as it usually happens in many educational school projects, is due to the conviction that it works in two directions: on the one hand, it favours the achievement of a high-quality education, through the provision of the appropriate environment for the members of the educational community and, on the other hand, it is in itself a symbol of the quality of the human relations developed, which is a quality element of the school. The distinction between the criteria of both categories that are present in a definition or agreement on quality in education is important when it comes to the establishment of the appropriate assessment indicators and instruments, as it will be seen later on.

This reflection leads directly to a new problem: the identification of *quality factors* or, in other words, the conditions that favour or facilitate the prediction of its achievement. It is very difficult to clearly distinguish those factors and their effect on quality achievement. Logically, the inclusion of a certain factor in this category should be based on its predictive character, which strictly speaking, cannot be done without referring to its empirical verification. However, the studies carried out are very seldom conclusive. Subsequently, the identification of quality factors is usually based on a combination of the acquired experience correctly checked, a formulation of intentions, and the formulation of hypotheses only occasionally validated by strictly scientific studies.

Despite these difficulties, the importance of the identification and formulation of these factors should be recognized, since they provide the bases for the development of qualitative improvement programmes and plans. In fact, it would be useless to know only the meaning of quality if there were no indications of how it could be achieved.

When it comes to the discussion about *quality factors*, the previous comment about the different perspectives of analysis that might be assumed should be taken into account and applied to this particular case. To be precise, quality factors could be referred to from two complementary points of view: on the one hand, from a macroscopic perspective, whose aim is to cover the whole education system; and on the other hand, from a microscopic perspective, which is focused on the study of what happens at school level.

From a general or macroscopic point of view, several attempts have been made to identify those factors that have a considerable influence on the qualitative improvement of education. In certain particular cases, this identification has prevailed even over the very definition of quality. A clear example of this approach can be found in the Organic Act on the General Organisation of the Education System (LOGSE), which came into effect in Spain in 1990. Its Fourth Title is entirely devoted to quality in education but instead of looking for a definition, which is certainly difficult, it lists the seven factors that have a greater influence on the improvement of education, and issues a number of measures that apply to each of them. These factors are:

- Teacher qualifications and training.
- Curriculum design and development.
- Educational resources and management.
- Educational innovation and research.
- Educational and vocational counselling.
- Educational inspection.
- Assessment of the education system.

According to the recent Organic Act on Education (LOE), enacted in Spain in May 2006, the first fundamental principle is to provide a high-quality education to all citizens at all levels of the education system. Likewise, the firm compromise with the educational aims established by the European Union for the near future has led to the decision of improving quality and effectiveness of the education and training systems, which means to improve teacher training, to develop the necessary skills required by the knowledge society, to guarantee everybody's access to the information and communication technologies, to increase enrolment in scientific, technical and artistic study programmes, and to make the most of the available resources increasing the investment in human resources.

The LOE includes, among the aims that guide the Spanish education system, the factors that favour the quality in education, in particular:

- Teacher qualifications and training.
- Teachers' team work.
- The provision of educational resources.
- Educational research, experimentation and improvement.
- The promotion of reading and library use.

- Educational, organisational and managerial school autonomy.
- School leadership.
- Educational and vocational counselling.
- Educational inspection.
- Assessment.

Similarly, some years ago the OECD prepared an important report on the quality of schools and educational systems in which several of these factors were examined, although not with the aim of exhausting the topic. These factors were referred to as "a selected number of key areas which are part of any major strategy conceived to improve schools and to increase quality through the systems" (OECD, 1991: 71). These were the five most important areas included in the document:

- Curricular design and development in relation to the learning assessment tools.
- Teacher training and professional development.
- Organisation and running of educational establishments.
- Assessment and monitoring of the education system and educational establishments.
- Availability and appropriate use of the required resources.

It is clear that, although the factors included in those lists (and probably in other similar ones developed in other countries or educational contexts) are not identical, the coincidences among them are quite important. Thus, a shared point is the emphasis placed on the key role played by the teacher in relation to the attainment of a high-quality education, which attaches great importance to the systems used for their recruitment, training, improvement and professional development. The same can be said of the process for curricular design and development, provided it is understood in its broad sense: the materialisation of the educational goals at school level. And the same could be said of the assessment and monitoring tasks (understood as instruments for the guidance of the system), and of the organisation and running of the schools (including in this variable the definition and establishment of their area of competence). It is without doubt in those contexts that all education systems face the quality challenge.
From an institutional or microscopic point of view, an increasing interest in identifying the factors that determine the quality of the

schools has been observed over the last decades. This effort has turned into the development of several theoretical models of quality of the educational institutions. However, most of these models could hardly make a convincing selection of those quality factors. One of the most accurate attempts carried out to identify, to define from an operational point of view, and to assess the influence of those factors on the institutional context, has been the one encouraged by the studies on school effectiveness. Despite the ideological and theoretical controversies brought about by this movement, it cannot be denied that its initial premises have been positively extended thanks to its most recent findings and that the importance of many of the factors brought to the front is widely accepted by the educational community. Subsequently, not many people will disagree on the influence that some of the following characteristics have on the quality in education provided by a school. These characteristics have been established by that movement and reformulated in the abovementioned OECD report:

- An agreement with clear and jointly established norms and goals.
- Cooperative planning, participation in decision-making and collegiate work within a framework of experimentation and assessment.
- Positive management of the implementation and preservation of improvement.
- Staff stability.
- A strategy for the continuing development of the staff related to the educational and organisational needs of each educational establishment.
- The development of a carefully planned and coordinated curriculum that guarantees a place for each pupil to acquire the basic knowledge and skills.
- A high level of parental involvement and support.
- The search and recognition of school values rather than individual ones.
- Making the most of the time devoted to learning.
- Active and substantial support of the competent educational authority (OECD, 1991: 184).

The fact that the movement on *school effectiveness* has to some extent been rejected by researchers of different trends might have been due to elements such as the emphasis the movement placed on the

empirical verification of its hypothesis, its initial oblivion of some dimensions related to quality (i.e. appropriateness of the achievements) and a certain note of positivism in its first formulations. However, from the current perspective and taking as a starting point some of the studies lately developed, it should be acknowledged the opening of its theoretical proposals as well as the explicit recognition of several factors not taken into account before and, at the same time, the analytical power still provided by this movement to the discourse on quality in education (Scheerens, 1992). Its current message is focused on the importance of three fundamental principles to guarantee effectiveness of the teaching and learning processes carried out at school level: the clear organization of the teaching and learning activities, the systematic promotion of institutional learning through instruments such as evaluation and feedback, and the search for consensus and coherence in relation to the goals and basic values of the school (Scheerens, 1996).

Developing quality indicators

The next step after the establishment of agreements on the main factors that should be taken into account to assess the quality of an educational system or school is the identification and development of a coherent set of *quality indicators* that reflect its level of attainment. As it has been previously pointed out, quality in education is a complex, multidimensional and non-directly accessible reality. Therefore, it is necessary to have instruments for its indirect approach. It is precisely these instruments that have been called *indicators*, whose main feature is that they are signs or signals capable of capturing or representing aspects of a reality that are not directly accessible for the observer. Using previous examples, it is possible to approach the attainment assessment of an education system through the academic promotion or qualification rates, the academic grades awarded by teachers, the rates of correct answers in a test specially designed for that purpose, or the rates of access to the labour market, just to mention different possibilities. Similarly, the *school climate* of an educational establishment could be assessed through the level of satisfaction reported by the different sectors of the school community in relation to the others.

All these data, rates and percentages of answer or correct answer constitute a like number of indicators either of the system or of the school. Its main aim is to reduce the complexity of the field to which

they refer, to an easily manipulated set of significant data in order to facilitate its interpretation and diagnosis, support decision-making processes and guide the action. Since complexity is inherent to the reality under analysis, in many cases the development of a single indicator will not be enough and it will be necessary to turn to a combination of several indicators, each of which will focus on a specific part of that reality. That is why the expression *systems of indicators* is preferred to particular *indicators*.

The identification of indicators is a key task because in the end it implies the selection of the most important and significant features from a qualitative point of view. The famous definition by Oakes, which refers to the indicator as "a statistic about the education system that reveals something about its performance or health" (Oakes, 1986) implies a previous selection of some important and significant aspects of the educational reality whose assessment is fundamental to form qualitative judgements about it. The negotiation of the sense and meaning of the above-mentioned quality would also cover, at a later point, the identification of the main indicators that facilitate its control, since they imply the operational materialization of those general criteria.

Thus, it is necessary to make the appropriate distinction between *educational indicators* and *indicators on quality in education*. The fundamental difference between the former and the latter is that the latter require a previous definition of quality and a specification of its main factors. Actually, most of the current initiatives to develop educational indicators manage without a similar definition and, therefore, a reference to quality is not included in their title. It could be argued that initiatives similar to those undertaken by the OECD (with its famous INES Project and subsequent editions of *Education at a Glance*/*Régards sur l'éducation*) or by the French Ministry of Education (with its annual publication on *L'Etat de l'école*) have taken a humble and realistic approach in the development of indicators that does not establish priorities among them with a view to assessing quality in education.

This is a desirable way of acting taking into account the current progression of the educational assessment. The expression *indicators on quality in education* should be restricted to those occasions and circumstances in which a previous definition has been established in

relation to the features that should be covered by the notion of quality, as well as to the establishment of its main factors and a negotiation of the indicators that better adjust to some or the others. These are exceptional circumstances in most of the current initiatives and that is why any reference to *quality indicators* should be cautiously made.

In order to finalise this section it is necessary to point out that just as in the past there were references to the *identification* and *predictable* criteria of quality, it is also possible to develop indicators of both types. The aim of the former is to provide information to assert if the desirable and expected quality levels have been achieved or not, whereas the latter assess the extent to which the criteria that predict qualitative improvement are met. Although a proper distinction between both types of indicators is not usually established in common speech, this difference is important from a conceptual point of view.

Developing coherent mechanisms and programmes for the assessment of quality

The final step in the strategy suggested is the establishment of the appropriate mechanisms and programmes for its assessment. In other words, the tasks related to quality negotiation, identification and operational capacity should precede its assessment.

Despite the coherence of such strategy, its feasibility and appropriateness might be limited in view of the international experience in this area. In fact, there are an increasing number of education systems that have implemented systematic programmes for the assessment of education which include an increasing number of areas, as it has been previously mentioned. However, the fact that such programmes have been implemented does not mean that the problems related to the conceptualisation and identification of a high-quality education have been solved. On the contrary, the paradoxes and dilemmas of quality already pointed out are increasingly obvious in many different national contexts (Tiana, 1996b).

The recognition of such paradox might lead to the conclusion that the analysis conducted so far has a serious contradiction that threatens its validity. However, this suspicion vanishes as soon as the necessary distinction is introduced: the fact that the assessment is the last step of a logical process does not mean that it is also the last element from an operational point of view. In fact, the experience is rather the

opposite: most countries have implemented their assessment programmes without having previously clarified and negotiated the meaning of quality in education or established its criteria and indicators. This is due to several good reasons, among which the following stand out: the experience accumulated by the theory and practice of the assessment over the last decades, the difficulties inherent in the definition of quality and the need of taking specific steps in favour of the qualitative improvement of education, even though its definition is not yet possible.

Therefore, the way of assessing several aspects of the educational reality is very well known, and this task has been firmly undertaken. The problem is that sometimes this undertaking is defined as an *assessment of quality in education*. The perspective repeatedly held in this work calls for greater care when it comes to the use of that expression. It is true that many countries are assessing different aspects of the education, but on the other hand, it is not true that quality is always the object of such assessment. Even though the current development of the educational assessment provides an answer to clearly stated needs and demands, it is not always the result of a systematic and strategic reflection on the problems implied by quality in education and the most appropriate tools for its improvement.

It should be stressed that what is defended here is not the abandonment of the line of action followed in recent years nor the reduction of the amount of attention paid to the task assessment. Such a proposal would be meaningless and unrealistic. The opinion defended in this work is that thanks to the progress observed in this area in recent years, the debate on quality in education and its assessment can be approached from a new angle. Assessing different aspects of the educational reality is not enough. It seems that it is time for linking that task to a systematic reflection on that concept so ambiguous, polysemous, complex and elusive known as quality. If, as argued before, this term for very good reasons has become the centre of the educational discourse, it would be better not to use it any more as a symbol or programme but in a coherent way. And this reflexive and constructive use means to take a strategy similar to the one suggested here. At least, the proposal included in these pages is worthy of assessment and discussion.

Finally, it would be interesting to relate these words to an idea mentioned before: if improvement is an unavoidable demand of the assessment practice, it is even more urgent in the case of quality in education. It would not make much sense to go in depth into such a slippery and unstable field as this has proved to be, unless it is to contribute to the improvement of the institutions under assessment.

Bearing in mind those assessment initiatives whose only aim is to express judgements of one type or another; whose hidden purpose is to justify decisions previously made; initiatives that are seen as instruments at the service of a particular strategy, or that are blinded by the knowledge achieved without realizing that the intervention might have deeply affected the people or institutions involved, it is time to ask again for an assessment that is, above all, a tool for the improvement of institutions and educational realities, no matter how sublime other aims may be.

References

Álvarez-Tostado, C. (1989). *La calidad de la educación: análisis de un discurso*, unpublished doctoral thesis. Universidad Complutense, Madrid.

De la Orden, A. (1988). "La calidad de la educación", *Bordón* vol. 40, nº 2, pp 149-161.

Gento Palacios, S. (1996). *Instituciones educativas para la calidad total*. Madrid: La Muralla.

House, E. (1993). *Professional Evaluation. Social Impact and Political Consequences*. Newbury Park, London & New Delhi: Sage.

López Rupérez, F. (1994). *La gestión de calidad en educación*. Madrid: La Muralla.

Oakes, J. (1986). *Education Indicators. A guide for policy-makers*. New Brunswick: Center for Policy Research in Education.

OECD (1983). *Compulsory Schooling in a Changing World*. Paris: OECD.

OECD (1991). *Escuelas y calidad de la enseñanza. Informe internacional*. Barcelona - Buenos Aires - México: Paidós - Ministerio de Educación y Ciencia.

OECD (1992). *High-Quality Education and Training for All*. Paris: OECD.

OECD (1995). *Performance Standards in Education. In Search of Quality.* Paris: OECD.
Santos, M. A. (1993). "Los (ab)usos de la evaluación", *Cuadernos de Pedagogía*, nº 215, pp 70-73.
Scheerens, J. (1992). *Effective Schooling: Research, Theory and Practice*. London: Cassell.
Scheerens , J. (1996). "Can the School Effectiveness Knowledge Base Guide School Management?", in *II Congreso Internacional sobre Dirección de Centros Docentes (Second lnternational Conference on School Management)*. Bilbao: ICE at the Universidad of Deusto, pp 98-119.
Sociedad Española de Pedagogía (1988). *La calidad de los centros educativos*. Alicante: Instituto de Estudios Juan Gil-Albert and Caja de Ahorros Provincial de Alicante.
Tedesco, J. C. (1995). *El nuevo pacto educativo. Educación, competitividad y ciudadanía en la sociedad moderna.* Madrid: Anaya.
Tiana, A. (1996a). "La evaluación de los sistemas educativos", *Revista Iberoamericana de Educación*, nº 10, pp 37-61.
Tiana, A. (1996b). "Los dilemas de la calidad y el papel de la evaluación", *Temas para el Debate*, nº 20, pp 46-51.

On the Meaning of Internal and External Evaluation for the Quality of Education during Educational Reforms

Hana Zufanová

Brief outline of the educational reform

Educational reform is a change in educational policy that sharply influences the course or the conditions of education and training. The main reform of the educational system is when through teaching plans and syllabuses the current educational course is replaced by a framework that states final outcomes. The final outcomes are stated by the definition of educational aims in form of competencies. The choice of the educational course (how the competencies will be achieved) is left to the decision of the school. This should be connected to the conditions in which the school works. If this change connects with the school's legal responsibility (financial independence of the school and the head teacher's responsibility for personnel resources) each school becomes an independent organisation providing service: education. In this case the state emphasises its educational policy as limits but does not state the methods. Education under these conditions is very challenging for school management, individual teachers and for authorities that "cover" the state educational policy. Recently, in many European countries the main problem is to explain providers (schools) and receivers of education (pupils and their parents) that school is a service and the receiver of education is a customer. Therefore, with reference to education it is difficult to create a supply - demand approach.
For a better understanding of school reform I will introduce the recent educational reform in the Czech Republic.
In the Czech Republic, the first fundamental educational reform since the time of the government of Maria Theresa (an Austro-Hungarian ruler) took place with effect from 1st January 2005, when Educational Act no. 561/2004 on Pre-school, Basic, Secondary, Tertiary Professional and Other Education[1] came in force. By this Act schools and educational institutions in the Czech Republic became autonomous legal subjects that provide or will provide education according to School Educational Programmes (section 3 § 3 of the

[1] www.msmt.cz

School Act), designed for each school individually. For establishing these programmes there are published Framework Educational Programmes[2] that stress the final state for given education, compulsory contents, extent and educational conditions. The School Educational Programme is approved neither by the Ministry of Education nor by the school founder. The correspondence between the School Educational Programme and legal regulations and the Framework Educational Programme and its fulfilment is supported and evaluated by the Czech School Inspectorate.
A further and very important change brought about by the new School Act is the schools' duty to conduct self-evaluation. The schools' self-evaluation is discussed in § 12 Section 1 and 2 of the School Act. To be more precise, as it is stated in the School Act under Section 1 § 12 - school evaluation is composed of self-evaluation and evaluation performed by the Czech School Inspectorate. Furthermore, under Section 2 § 12 it is stated that the schools' self-evaluation is the basis for producing an annual report about the schools' activities, and is also part of the main basis for evaluation performed by the Czech School Inspectorate. The Ministry states by legal directive structure, rules and terms of school self-evaluation.
Details about self-evaluation (or precisely: about the report on self-evaluation) are stated in § 8 and § 9 of regulation no. 15/2005 that states the purpose of long-term aims of the annual reports and self-evaluation. The report shall always highlight information about the main objectives of self-evaluation:

- the school's educational conditions;
- the course of education;
- the school's support for pupils and students, cooperation with parents, influence of relations between school, pupils, parents and other persons participating in the field of education;
- attainments and achievements of pupils and students;
- school management and leadership, the quality of the staff's work, the quality of the pedagogical staff's further training;
- the school's level of outcome, particularly in regard of educational conditions and economical resources.

The above emphasised fields are stated in § 8 Section 2 under regulation no. 15/2005.

[2] www.vuppraha.cz and www.nuov.cz

The report on the school's self-evaluation is according to law a compulsory school documentation (§ 28 Section 1, letter e), stated in the School Act.).

To complete the information, it has to be mentioned that parents in the Czech Republic have from their child's pre-school education on absolute freedom in choosing the school which their child will attend.

Outline of the term "evaluation"

Evaluation in education means systematic "data" collection, "data" classification and "data" evaluation. To make further decisions, this happens under criteria about documents, treatments and persons. According to this definition, evaluation shall:

- be systematic in an explicitly stated field of interest and its structure;
- be conducted methodologically correct;
- be conducted regularly;
- obey confirmed criteria, and
- prepare decisions for further planning and action.

According to the character of the evaluation processes there is a distinction between **external** and **internal (self) evaluation**. The two types of evaluation are not opposing but shall be connected and shall serve to fulfil common aims: to improve the quality of education. To conduct a reliable inquiry is only possible when they are interconnected.

Internal evaluation

Internal and external evaluations are two complementary views on school that have different objectives and tasks. Both types of evaluation are used for possible improvements of the educational system and their individual parts are neither exchangeable nor expendable. However, that does not mean similarity of all factors responsible for development of evaluation processes, but different factors may be responsible for them.
Internal evaluation evaluates the quality and effectiveness of internal processes (effectiveness of individual activities, styles, methods and forms of work, individual teachers, etc.) as well as external conditions

(from the viewpoint of the school) that lead to the fulfilment of aims stated by the school on the basis of its educational policy assignment. Self-evaluation is a systematic tool on the level of school management. It provides information about decision making processes related to the profile and aims of the school, judgment and confirmation of educational processes, the development of the school image and the school presentation to the public and preparation for external evaluation.
During self-evaluation the school states its objectives, related appropriate measures and criteria. The school chooses such qualities that satisfy its conception of education and training. It provides information about processes that the school focusses on (for example, it emphasises memorising of information, or conversely, the school emphasises orientation on findings and information search), however, it does not tell anything about processes that the school does not consider to be important.

External Evaluation

External evaluation evaluates the school as an institution and evaluates also outcomes that the school has achieved from the viewpoint of external observers. External evaluation is often considered as a tool for directive leadership and sometimes even repression. A misunderstanding of the external evaluator's role or the wrong choice of evaluation criteria might lead to confusion with the internal evaluation that is going to be conducted by him and his external point of view. This concept of external evaluation is wrong and cannot provide objective information about educational processes realised by the school.
External evaluation has in the first place a political character and serves to control the fulfilment of the educational policy of the state. During this evaluation the aims are always stated from the external side. The choice of methods and forms, measures and evaluation criteria are given by external evaluators or by the authority that assigns the evaluation.
External evaluation in any case cannot replace internal evaluation. Only internal evaluation gives the possibility to express the core of a problem, internal relations and follow long-term development. However, during internal evaluation the evaluators might face a problem mainly of not being objective. Another barrier to objective evaluation might be tight personal relations which may lead to intentional covering of mistakes. Therefore, an ideal combination of different evaluation approaches, methods and activities is required.

Information sources for evaluation

The choice of information sources follows the aim of the evaluation. The more complex the aim of evaluation, the more information must be collected.
Among the most important sources of information for self-evaluation is the complex evaluation of activities of the whole class in one or more subjects, and evaluation of individual pupils. That is a very interesting diagnostic entry in the effectiveness of the pedagogical workers' activities, and concerns only characteristics of individual pupils (talent, interest, volitional ability). Diagnostics of the activity of the whole class provides information about the teachers' work in class, and it also provides information on barriers in interaction and whether tasks are not too complicated for the pupils.
An important source of information is the comparison of the pupils' performance in standardised tests with evaluation of pupils' performance made by the teacher. Such findings should revise the meaning of subjects and the meaning of the term "successful pupil".
Very valuable entries for the evaluator provide also findings based on psychological tests which give a very good picture about pedagogical staff and individual classes but also about relations in school.
Another source of information are questionnaires that aim at individual groups (for example pupils, parents, school stakeholders) or at the revision of particular phenomena (for example difficulties during learning and teaching, school climate). Furthermore, very important sources of information are data expressed in school documentations, stated in legal regulations and directives, but also in documentations on which the management of individual school decides.

Evaluation Tools

In regard to the above expressed problems, it is possible to group evaluation tools into these categories:

Political tools - used mainly for external school evaluation. State and state authorities claim their influence on educational quality through these criteria. This group includes:

- formulation of state educational aims with the help of legislation,
- legal demands towards document planning and evaluation report planning,

- financial system in the educational sphere,
- promotion of global aims by learning aims,
- final state exams and its conditions,
- quality control system of teachers' preparation,
- support of supportive and joint systems (science, consulting, further education),
- system of supervision,
- statement of quality of education.

Tools on school management level - used mainly for school self-evaluation, though some of them are used also for external evaluation. Tools used on school management level include:

- leading of discussion and management of decision-making processes connected to the school's profile and aims,
- judgment and confirmation of teaching programmes,
- determination of curricula in learning and teaching,
- management of learning and teaching processes in the school's framework,
- prompting of evaluation processes,
- brochure publication (public relations),
- judgment and working interviews,
- preparation of external evaluation,
- organization and updating of an informational system,
- organization of test-evaluation performed by internal staff (e.g. school-leaving exams) or by external professionals (e.g. PISA).

Further tools that are possible to use on the level of management are evaluation tools and experiences from other fields - enterprise, business or commercial sphere. From the enterprise field, there might be applied already broadly accessible and used models of self-evaluation such as CAF, EFQM or norm ISO 9001, modified for the educational area. These models and the ISO norm include mainly tools for feedback and evaluation of "management work". The models and ISO norm are better used for vocational education training, though they might be used also for elementary schools and gymnasia, because they include many useful elements and might be used as a good inspiration for the "start" of school self-evaluation.
The use of these tools might be difficult at the beginning, because the tools understand school as a service and pupils as customers. This philosophy is common in many European countries already for a longer time.

All mentioned tools have in common:

- the field of school management is described in detail;
- the aim that the school leads to (continuous improvement);
- for the evaluation, there are stated criteria that are divided into sub-criteria that are further elaborated into individual statements;
- they work by scoring points;
- during the preparation for certification the organization co-operates with a professional adviser;
- they are built on PDCA (Plan - Do - Check - Act) methodology.

Brief outline of the above mentioned evaluation models

CAF
Common assessment frameworks are built on the CAF model. It is a tool designed for the public sector, thus schools are included. The evaluation is made along with a set of criteria developed and accepted by European institutions. To the management and staff, CAF offers the opportunity to access their work from an unusual point to view. This model is a high quality tool for the analysis of school conditions and for the development of a school educational programme. It includes nine criteria for evaluation which can be used for individually measurable questions.

EFQM
EFQM is a managerial model arising from an approach where all organization employees are included in the process of continual development of quality products, services and the whole system of the organisation. It is built on a model of thinking about the organization's objectives, processes and people, ethics and "company" culture included. The organization uses self-evaluation in every field of its activity that is divided into nine criteria and is compared with the best ("Best in class"). Furthermore, benchmarking, a systematic process focused on comparison of self-effectiveness of the organization, is used from the view of performance and quality, among the best organizations.

ISO/IWA 2:2003
Methodology ISO/IWA 2:2003 is the directive for application of ISO 9001:2000 in education. This methodology is based on process management. During such a procedural approach a system of

processes is created in an educational organization. First, processes have to be identified, described, and their common effect must be analyzed and after that it is possible to divide them. This progress emphasizes the added value of processes and thus these processes have to be objectively measured (for example evaluation of pupils, staff, satisfaction of pupils, parents, staff, etc.). With the help of this approach, when a school continually improves processes, the school reaches outcomes in the field of process quality and effectiveness which suggests whether the school is successful.

Tools on class and group level - used for self-evaluation performed by teachers. Into this group it is possible to include these tools:

- Curriculum evaluation
  Continual curriculum evaluation is part of the teachers' basic responsibilities in a good school. A teacher should:
  - monitor academic literature,
  - monitor new possibilities of didactical and methodological character,
  - prepare the pupils for tests,
  - support pupils' individual activities,
  - participate on further education and shall report about its outcomes.

  Every teacher has the tendency to change the contents of teaching after some time and work with other methods. Very important is the teacher's will to change on the basis of what the pupils expect and how they are prepared for the change. The teacher can encourage several organizational and content experiments, however, everything has to stay within the legislative framework. Suggestions for the teachers' self-evaluation might arise from team cooperation on individual subjects but also from external stimulants (employers' pressure).

- Diagnostic and selective tests
  The assessment of the pupils' outcome is one of the most important information sources about the effectiveness of learning and teaching processes. In regard to the aims, there is a distinction between diagnostic and selective tests.
  When monitoring how pupils master learning or what attention they pay to understand learning, it is important not to look only at

the selective use of findings, but also at groups of individual interviews with pupils about possible ways for improvement. Selective tests with its objectives must be announced in advance, so the pupils know what it is for. Selective testing might come in form of entrance exams and interviews, initial exams to dived pupils into groups, etc. The teacher should also distinguish among several forms in order to find more objective information on pupils. Both tests may be performed individually, by groups, by the whole class, year or the whole school. In the case of whole school testing also higher school management is involved.

- Evaluation of databases and databanks
  An important precondition is the compilation of database or databanks with topics and tasks that shall cover the whole curriculum. Further, an approach is discussed of how this database might be used and mastered by individual pupils. It deals with the use of possibilities of group work or individual work that is very effective only if pupils are motivated by self-effort to fulfil fictive of real agreements with the teacher (for example by form of a project). Individual pupils are evaluated periodically and must defend their own outcome in front of the whole group which leads to social and other professional skills.

- Development of questionnaires for gaining relevant data
  This form is known in several variations and with several objectives in many schools and in many classes. The teacher should think very carefully about aims and processing of the questionnaires. That includes especially the formulation of questions, which is a problem about which professionals, psychologists or educationalists have to be consulted, so that the questions really lead to their aims and at the same time leave the pupils some space for their invention. It is important to emphasize why questionnaires are or are not anonymous and how the gained data will be handled. This form is mainly used for gaining information about relations in class, about ethos, about pupils' views, about pupils' needs, pupils' attitude to individual subjects, etc.

- Interviews with teachers and pupils
  Interviews are important and often the only source of information. Particularly guided and focused interviews must be very well

prepared in advance. The guidance of interviews, particularly among teachers, but also among pupils has its own ethnical rules, and the views of the other side must always be respected.
The interview must not lead to convincing the other side about our truth, but it should lead to getting other views about our opinions in a friendly atmosphere. The interview thus might be used as a very good tool for self-reflection.

- Class observation
  Class observation is a very famous form for gaining information. New forms of class observation are, for example, to use audio-visual aids for recording a class on tape, and the discussion of the recording afterwards. Furthermore, after some time changes etc. that have been made might be compared with it. The objectives and the course of the class observation must always be discussed with the observed teacher in advance.

Evaluation methods

It is possible to differentiate evaluation methods according to their use: whether they are used for external and internal evaluation at the same time, or only for self-evaluation.

Common methods

- Analysis of basic school documents assigned by legislation;
- analysis of pupils' work;
- data analysis;
- observation of learning and teaching - however, there is a distinction between external and internal evaluation. During external evaluation, it is focused on how the school management performs class observations, leads interviews and takes decisions. Observation of learning and teaching within internal evaluation is focused on quality or other evaluation aspects.
- Inquiries, questionnaires - written or oral (with the school management, teachers, pupils, parents, public, community, etc.);
- interviews (with the school management, teachers, pupils, parents, public, community, etc.);
- sociometric rating methods for monitoring the school climate.

Internal evaluation methods

- School documentations - records, diaries, chronicles, photo-documentations, video records, intentional documentation carried out according to school management decisions;
- documentation on teaching outcomes (individual teachers; effectiveness = time, energy, stability, applicability);
- analysis of pupils' work, observation of pupils' development;
- didactical tests focused on knowledge, skills, competencies, for example solving of difficult problem, etc.;
- project methods (essays, paintings, activities, games);
- observation of learning and teaching, mutual teachers' observation, visits of classes by parents. Key competencies as proclaimed educational objectives often request fundamental changes in the work of teachers. The choice of methods, didactical progresses and forms of teaching leads consequently to fulfilment of expected outcomes and adequate development of the pupils' key competencies. Even though the aims and principles of observation activity are still valid, the formal side of observations is emphasized in the first phase of curriculum reform. Therefore, it is necessary to focus the observations on fields that the school management regards as essential. There must be more space for observation analysis, and the way of how to deal with observation results should be modified. As a consequence, it is possible to gain a tool of how to monitor teachers' developments in time and tell other teachers main observation findings in a straight forward way. Thus, publicity of exceptional events is meant positively. Observation records have to be known to all teachers, and everybody has to understand what each finding means and how it might be fulfilled. It is appropriate to stress examples of classes and academic publications at the same time. In many cases ideal conditions will be described. Furthermore, it is essential to emphasize that the aim is not to find in every class observation all phenomena stated in the observation sheet. The analysis of outcome of observations conducted that way might indicate to the school, which key competencies are developed in the pedagogical process, which key competencies on the other hand are not developed at all, or which key competencies are developed only to a little extent.
- In times of educational reform questionnaires became more popular within school management. It is a good possibility to gain relatively easy and quickly needed information and feedback from parents, students and teachers. When using questionnaires it must

be kept in mind that, apart from direct information that can be gained through the inquiry, the questions also have formative purpose. Inquiry participants - teachers, pupils, parents, etc. - are finding out which fields are important for the author. If the inquiries take place regularly with the aim to find changes in particular areas, the observed person has the chance to change something in his or her work in the time between the inquiries.

- Sociometric rating methods for monitoring relations and influence of pupils in class and teachers in school show the quality of social relations; however, they have to be used very sensitively.
- Workshops - an activity for one or more days, for common problem solving and mutual help.
- A panel discussion on a known topic by professionals from several academic areas or representatives of several social strata (for example psychologists, paediatricians, economists, head-teachers, teachers, parents). Each participant has a chance to state his or her opinion on the topic discussed. The discussion should lead to clarify different thoughts and finally compromise. It is a good way to solve problems.
- Brainstorming - a method based on the separation of ideas from critical evaluation, thus on overcoming conventional barriers and involving imagination. This method is suitable for searching ideas but also for problem solving. Brainstorming has three phases - problem statement, quick production of ideas (ideas are recorded but not evaluated), evaluation of ideas.
- Mind mapping - in the centre of a big sheet of paper the core of the problem is written, and then we write or draw around it what comes to mind; it’s very good just to follow the course of our hands. After that we might graphically process individual ideas and how they are linked up.
- Situation methods, model situations and their analysis - this method helps to solve certain problematical situations that exist in “real” life or are constructed (in the latter case they have to be credible). The core of this method lies in analysis, the understanding of basic relations that create this situation including their cause and consequence. This method allows us to take a position or decision on how to deal with given situations.
- Presentation methods (role-play, life scenario) - they are used to transfer situations into role-play, i.e. participants take on roles of individual people in the problematical situation. This method helps to solve professional and social problems while mastering conflicts,

but also trains socially effective ways of behaviour and decision making processes. A stress is put on social relations and communication.

The meaning of self-evaluation for schools

In times of curricular reforms self-evaluation of schools is an irreplaceable and very important element for a correct start of changes in the schools' work. To start changes it is very important to conduct evaluations to get findings of the current state of the school. Self-evaluation is then important for determining a strategy for changes to fulfil the stressed aims, the development of the school's educational documentation, i.e. the school's education programme. For the school management it is a very important basis for the determination of further school needs, and thus also a statement of requests towards founders and holders of the state's educational policy.
Self-evaluation is a long-term, never-ending process in the school's work that brings very valuable information on quality and possibilities to change. Self-evaluation is very fruitful only if findings are not purposefully misrepresented.

Good self-evaluation should always have these stages:
1. preparation of evaluation;
2. realization of evaluation;
3. evaluation;
4. adoption of proceedings.

To 1.
During this stage the evaluation team sets up, thinks and processes the school self-evaluation project. It might be stressed that this stage lays the basis for the whole evaluation, because here it is decided what the evaluated subject considers and how the team will proceed, which further strategy and tools will be chosen. The choice of the strategy and especially the evaluation tools is a key element to gain desired evaluation outcomes. At this stage it is important to think and decide how stage no. 3 will be realized and to get prepared for this stage.

To 2.
Preparation of evaluation tools, strategy realization and work with evaluation tools. A very difficult part of evaluation is the preparation

of evaluation tools, so that the team gains the desired information and data. This stage has high demands on knowledge of several evaluation techniques on questionnaires and tests processing. At this stage it is a good idea to ask professional or more experienced colleagues for cooperation and consult with them about their opinion of suitable techniques.

To 3.
Categorization and evaluation of gained data and information and consequential report writing. This stage might be very time-consuming if we do not pay enough attention to its preparation. Furthermore, if the school does not have the necessary hardware and software for data analysis, this stage might be also very time-consuming.

To 4.
This is a very important, however, often neglected part of the evaluation. At this stage findings during evaluation are projected and processed into particular activities. This is the key stage of evaluation that leads to quality improvement.

A well organized system of self-evaluation is progressively becoming part of school organization and school life and does not bother those who are performing evaluation, thus neither the team nor the school. The feedback gained through self-evaluation contributes to improvement or at least to preserving the recent state of the school. Practical experiences of pivotal schools in the Czech Republic suggest that it is very important to include the whole school in the self-evaluation process. All staff has to be aware of the importance of self-evaluation as a tool for quality improvement.
At the first stage it is better to focus on fields of evaluation where it is noticeable to take steps. That is how strong sides are supported because the weak sides are usually marked phenomena which do not threaten schools critically. In other words, self-evaluation is focused on monitoring and evaluating of processes. For results, external tools for verification will be used when the aim is to find “the rate of fulfilling key competencies”.
The manner of evaluation that the school management will use must have formative character and for teachers highly motivating effects.

Team work should be focused on:
- class observation,

- teacher’s personal index,
- questionnaire survey,
- testing knowledge,
- external testing,
- the system of annual exams,
- evaluation of teachers by pupils.

Two new elements appear there - the teacher’s personal index and evaluation of teachers by students. Those two elements might sharply influence school quality.

New elements of evaluation that appeared in some schools in connection with educational reform - examples of practise

Teacher’s portfolio and personal index

The teacher’s portfolio should monitor a teacher’s professional career. In his or her portfolio the teacher gathers all documents essential for his or her subsequent development.

The aim of a teacher’s personal index over the period of a year is to monitor and evaluate his/her long term activities, thus such activities that are not possible to observe during one class observation. The teachers’ activities are part of a school’s fulfillment of aims, the quality of pedagogical processes and also part of aims of curricular reform in individual schools.

The personal index is designed according to a teacher’s commitment to take part in asserting the so-called whole-school strategy stated by the school educational programme. At the beginning of the school year the teacher together with the other colleagues of his/her subject stresses how he/she will help to fulfil the school aims stated by the plan for the year; for which activities and project he/she will be responsible; what steps he/she plans to take in regard of personal development, etc. The stress is put on the commitment to promote a strategy for the whole school on which the pedagogical staff agrees as school priorities. The personal index is discussed at the beginning of the school year with the school management, and is evaluated again at the end of school year together with the school management. An individual approach to teachers is important especially at larger schools where the school management does not have the opportunity to daily observe their teachers’ work. The conditions for a successful

introduction of the personal index are such relations between staff, where the fulfilling of common aims takes first place in the working priorities of most teachers. Teachers should not perceive the introduction of a personal index as a tool to control them, but first of all as tool for continuous quality improvement of their work.

Evaluation of teachers by pupils

Evaluation of teachers by pupils might be beneficial for both sides and the school as a whole. Nevertheless, it is a delicate event and its introduction is important to prepare in advance. It is possible to realize it with pupils of all age groups, however to assure objectivity pupils and teachers must be well prepared for the evaluation. In Czech schools it is a new element, therefore as illustration I will stress an example of successful introduction of this evaluation tool into a school's system of self-evaluation.

Evaluation of teachers by students in our school surpassed the level of vision or intention. This year is the first year when also teachers will receive reports from their students. We do not have particular outcomes yet. I will explain the reason why we made this decision, and the method by which we will analyse the evaluation of teachers. The opinions on the evaluation of teachers by pupils we have asked by questionnaires. The results were quite unambiguous. The majority of students was in favour of the evaluation but also a big part of the teachers. That suggests that our teachers are very self-confident in a positive meaning. Concerns of pupils and teachers against it, which were stated in the questionnaires, are understandable. Teachers might be afraid that in their evaluation "higher demands" will be reflected, and pupils might be afraid of reactions of teachers who were not evaluated positively. However, we successfully explained to both sides that their reasons to be afraid are unfounded: if I teach well, there is not reason to be afraid; and if the evaluation is objective, there is no reason for revenge. If the evaluation of teachers by pupils is mutually understood as a means for improvement of class activities, then the relations between teachers and students should not be in danger. A real possibility to use this element of self-evaluation is influenced in every school by the school climate.

The students' parliament was fully taken by being involved in making "reports on teachers". It was very difficult to find areas in which teachers will be evaluated. The students had at disposal the results of a questionnaire survey of the whole school. The survey by open questions investigated how students see the typical teacher in our

school, and in another group of questions it was asked how the students see the ideal teacher. Owing to truly ‘everybody inclusive’ questionnaires we were successful to find “subjects” that will be emphasized in the teachers’ reports in June.
It is a first attempt, thus rules for evaluation and areas in which teachers are evaluated will be surely adjusted during the following years. Especially in the next years teachers will know in advance in what area (“subject”) they will be evaluated, which this year they did not know.

The meaning of external evaluation for school

External evaluation in times of change has a key meaning for schools, because only the external view provides information for schools whether the aims and goals of the educational reform were well understood, furthermore it has formative character for schools. Cooperation between school management and external evaluators enables the management to fruitful discussions about self-evaluation findings, exchange of experiences and opinions. The discussion is beneficial for both the school and the external evaluator.
External evaluation in schools is not performed only by controllers or school inspectorate. This evaluation might be performed also by founders. External evaluations of the school’s work are results achieved by the school at testing pupils on international or national level.
Evaluation, either external or internal, that is implemented in schools, is not complete and does not fulfil its task if no steps towards improvement or at least remaining of the recent state are taken.

External testing as inseparable part of self-evaluation

The preparation of tests that will verify gained key competencies, possibly tests that will be useful for monitoring the continuous approach to desirable outcomes, is done by state organizations (in the Czech Republic - Centre for quality in education, Centre for state school-leaving exams) or professional organizations that have test preparation competencies (in the Czech Republic: Kalibro, SCIO, etc.). The question is whether tests will be useful for expected educational results, for an evaluation at the end of any separate

educational cycle and also for continuous monitoring of individual years or in the middle of an educational cycle. Outcome testing in nodal educational moments (at the end of an educational cycle), through which the school verifies that pupils have reached the expected outcome, is possible with the use of external tests because expected outcomes at the end of a cycle are according to educational frameworks obligatory for all schools. If external tests for continuous monitoring will not make it possible to choose criteria with regard to a school, then the use of external tests will be problematic or will lead schools to uniformity of their educational documents which would degrade the meaning of educational reform.
External testing offers the opportunity to compare results among schools. Therefore it is required that the tests will take place at the same time. For assurance of the above mentioned conditions it is necessary for schools to cooperate with the founders which can assure fulfilment of these conditions. It is always beneficial for schools when the founder has an apparent interest in the school's quality improvement.
When judging the quality of pedagogical work of a school, it is very important to take into account added value and its comparison with other comparable schools. Such a comparison is impossible without the use of tests that are prepared for schools by external organizations. Identification of added value is very difficult in a system where none of the included input variable social-economic conditions is known.
Work with external testing results is a challenge for school management. Information gained by schools in the Czech Republic is owed to organizations that provide tests for schools so detailed that its appropriate use is a next step to quality improvement of the pedagogical process.

The meaning of self-evaluation for evaluators as managers

The key element in evaluation of school management is that the management proves their clear understanding of who are theirs customers, what are their needs and expectations and how this might be in balance with political assignment. At the same time they shall prove clear obligation towards students/citizens as well as towards other stakeholders.
Regarding the fact that self-evaluation is not a one-man-show but teamwork (if possible the more the better), all who participate in self-

evaluation are at the same time participating in the school's decision making process which supports their identification with the school. For the school management as well as for the members of the team self-evaluation is a very important basis for planning the staff's further training. Not only the management but also the staff gain complex views on school activity, information on the school's weaknesses and strengths.
Every team member gains new views on his/her own work, very important experiences of self-reflection and evaluation. It leads to a development of one's own technique of self-evaluation that is consequently evident in the approach to one's own training in education practise.
If the school decides to prepare its own questionnaires and tests for the evaluation to gain needed information on outcomes, then all members of the team must study the new knowledge, so the intention is possible to realize. However, that leads to an application of gained knowledge also within the educational activity framework. Also the application of professionally proceeded tools is beneficial for every evaluator. It positively shifts his or her qualities.

The meaning of evaluation for customers - pupils and their parents

Self-evaluation has a double meaning for pupils and their parents: First as evaluation actors, second as receivers of educational outcomes. Both meanings are very important. In times of change they might highly influence as evaluation actors the course of the school or provoke corrections in running the school's work.
Pupils enter self-evaluation in two phases. In the first phase, within self-evaluation their knowledge is tested. Knowledge testing is in many Czech schools perceived as a very important component of the outcome of pedagogical work. Schools want their pupils to have a very broad knowledge, even though key competencies enter the main interest of the pedagogical process. Schools would perceive a decrease of knowledge output as a mistake of curricular reform aims. Many schools still put finding successful strategies for the development of key competencies while remaining necessary knowledge among the most important tasks in the near future. The development of evaluation tools that verify gained competencies and knowledge at the same time are not within the schools' power. This is

where the schools expect service from either state or private subjects. The second phase are questionnaires distributed among pupils. That makes a very good and fast verification of needs and reality possible, for example the use of a new didactical tool which was expensive for the school and the use of this tool during class observation is not possible to evaluate.
By using questionnaire surveys to verify outcomes of separate activities - excursions, internships, sport activities, school projects, etc - the school creates space for active participation of pupils and parents on quality improvement of such activities.
Through reports on self-evaluation the school can provide current information to other partners. Recently, as the majority of parents is interested in the schools' work, this is very important. The majority of parents make inquiries before their child enters school to get information about the school. Well performed self-evaluation of the school together with the annual reports about school activities are still the best sources of information.
Relevant information that the school provides is not only good commercially but also supports good relations with the whole community.
External evaluation mainly informs whether school activities do correspond with the legislative framework of the state's educational policy and whether the information that the school provides is truthful.

Relation between internal and external evaluation

The implementation of self-evaluation and the proceeding report of this evaluation is the preparation for external school self-evaluation. During the external evaluation the report is introduced by the school to the external evaluators. During the introduction the school also stresses its qualities and weaknesses and emphasizes works on improvement. If self-evaluation is conducted carefully and does not hide anything, then the school is well prepared for external evaluation and knows what statements will be expressed in the final report. If the school's self-evaluation differs extensively from the external evaluation, topics for discussion arise. During the discussion the school learns why differences appeared and again gets feedback useful for its next self-evaluation and for the improvement of school quality.

The meaning of evaluation for successful educational reform

The success of reform firstly depends on different views on school, mainly from the side of the school itself. The school must notice that providing education is a service that the school offers on the basis of the state's educational policy, and that pupils and their parents are customers. Only schools which accept these principles have a chance to succeed on the market of education providers.
Self-evaluation is a means through which the school learns more about itself, the needs of its customers and the rate of educational policy fulfilment. The external evaluation shows an overall picture of the fulfilment of the objectives of the educational reform. Both evaluations are a feedback for the school and the reform creators. Reform has no chance to succeed if its creators will not have truthful and objective information about the process of the fulfilment of the reform's objectives, about barriers that schools have to face during reform. Not only educational reform, but any reform does not have a chance to be successful without appropriate reaction to its feedback.

References

Common Assessment framework (verze k 12.5.2005). *Národní informacní stredisko pro podporu jakosti*. Praha 2005.

Hanousek, Ch. (1992). *Moderní metody zpracování dat. Matematická statistika pro kazdého*. Praha: Grada.

Havlínová. (1994). *Jak menit a rozvíjet vlastní školu?* Praha: Agentura STROM, Edice NEMES, sv. 2.

Hrabal, Lustigová, Valentová (1994). *Testy a testování ve škole.* Praha: SVU PedF UK.

Janovcová, Prucha, Koudelka (1988). Aktivizující metody v pedagogické praxi stredních škol. Praha: SPN.

Mertin (1996). *Individuální vzdelávací program*. Praha: Portál.

Pelikán, Lukš a kol. (1981). *Metodologie výzkumu úcinnosti pedagogického pusobení ve výchovne vzdeláváacícm procesu*. Praha: VÚOŠ.

Pelikán (1998). *Základy empirického výzkumu pedagogických jevu*. Praha: Karolinum.

Prucha (1996). *Pedagogická evaluace*, MU - CDVU. Brno.

Rýdl (1995). *Model sebeevaluace školy*, grantová studie MŠMT CR.

Praha.
Rýdl, Horská, Dvoráková, Roupec (1998). *Sebehodnocení školy*. Praha: Agentura STROM, Edice ŠKOLA 21, sv. 3.
Smernice pro aplikaci ISO 9001:2000 ve vzdelávání (2004). Národní informacní stredisko pro podporu jakosti. Praha.

League Tables and Health Checks: The Use of Statistical Data for School Accountability and Self-Evaluation

Ian Schagen, Dougal Hutchison, Paula Hammond

Introduction

The use of statistical information on the performance of educational institutions is a growing subject of debate within different educational systems. Very roughly, the uses of such information can be characterised as either "league tables" or "health checks". The term "league tables" represents the use of data within an accountability framework, whereby information on the performance of institutions is published in order to make clear their relative rankings and encourage those ranked more lowly to improve. Often such an accountability agenda is linked to some kind of "virtual market", whereby the "consumers" of education can choose among competing institutions.

On the other hand, we use the term "health checks" to refer to the use of data within a confidential, self-evaluation framework in which information is shared with those within an institution who can affect change, with by and large a "no blame" framework. These two extremes are not quite as rigid as we have made out, as there are examples which fall somewhere between, and systems (such as England's) which are attempting to carry out both simultaneously. However, we shall retain this somewhat simplistic polarity for this paper, and refer to the two extreme concepts as:

- "League tables" = accountability = public information
- "Health checks" = self-evaluation = confidential information.

In order to see how these two concepts can be operationalised and the effects they may have on education systems, we shall consider the UK as a worked example, and in particular England. Many of the possible ways of using data to drive educational improvement have been tried in England, and the system has evolved in different and sometimes contradictory ways due to different impulses and shifts in government policy and educational fashion. This chapter is split into three major sections, which will explore the following themes:

- The use of performance data for accountability, including the debates about “league tables” and “value-added analysis”;
- The evidence for the transformational power of “health checks” for schools, including recent analyses of data carried out by the *National Foundation for Educational Research* (NFER);
- The different ways in which statistical data is used within different education systems in the UK, based on an NFER review, and important lessons for the effective use of this kind of information to drive school improvement.

Finally we shall draw together the different strands to give some general guidance for educational systems on the use of performance data for educational improvement based on the UK experience.

Performance Data and Accountability

The question of accountability in education, as in all public services, is an important and complex one. How are we to decide whether we are getting value for money or not? Such a decision is relatively easy in a straightforward commercial transaction, where those who do not provide good value are liable not to sell their products and consequently to go out of business. The situation is more complex in education. What is the product? How do you assess it? And what happens if it is not considered satisfactory?

How, in short, are schools and teachers accountable? One the one hand, there is the question of competence: are they doing their job well? However, it is not just this: another and arguably equally important question is to ask, regardless of competence, whether what they are doing is what is required. Finally there is the question: “Accountable to whom?” - to officialdom, to parents, to children, to society as a whole?

National assessment in England

In Victorian times in England, teachers were subject to “Payment by Results” a rigid method of accountability associated with English and Welsh elementary education during the second half of the nineteenth century, whereby a school’s governmental grant depended for the most part on how well pupils answered in the annual examination conducted by Her Majesty’s Inspectors (Rapple, 1994). However by

the 1950s, employment was virtually secure except in extreme cases, and accountability was only within the education sector, for example particularly to HM Inspectorate and the Local Education Authority.

The 1970s saw an early attempt to assess the performance of the system as a whole across the UK, as opposed to its constituent elements, with the Assessment of Performance Unit (APU) - situated in the then Department of Education and Science (DES). In the 1970s and 1980s this commissioned a series of studies of pupil performance in a range of subjects (see e.g. Foxman et al. 1991). APU surveys comprised lightly sampled, low stakes, testing exercises with the aim of assessing the nation's overall state of achievement at certain ages. Significantly these were specifically designed not to provide an identifiable performance measure for any entity in the system, school, teacher or pupil.
Despite the important information gained by the APU surveys, they were discontinued at the time of the Education Reform Act of 1988 and the advent of national curriculum testing. Accountability was now to be located at the individual school level. The message that "schools matter" had been taken on board, but national test results were to become all-important, as far as performance was concerned.

The new national assessments were expected to perform a number of functions, but among these was to measure the level of performance of the education system as a whole and its constituent schools. Further, the accountability audience was to change from the more supportive arrangement of reporting to other education professionals to the harsher gaze of consumers of education in the shape of pupils and their parents. After the 1992 Education Act, schools had to provide summary details of the examination performance of their pupils. The first set of school performance tables comprised summaries of results in the relatively new GCSE examinations. Although schools were not presented in rank order, the media swiftly converted the raw data as presented into "league tables".

The "Value-Added" Movement

However, a reaction against the publication of raw results and "league tables" began fairly quickly. Researchers were quick to point out the many fallacies inherent in their uncritical use as measures of school quality (see, e.g. Goldstein and Spiegelhalter, 1996). On the one hand such results were more dependent on the characteristics of the school

intake than on anything else. On the other, results were unstable from year to year, especially with small schools. It is surprising that there was so much emphasis on the provision of such a poor and unsatisfactory measure of school performance, and it seems likely that this was at least partly due to dependence on "market-led" dogma than to objective evaluation of the evidence.

To highlight these points the term **value added** was introduced. Under this approach a school was supposed to give positive value added if on average the pupils' attainment was higher than would be predicted on the basis of pupil characteristics, for example attainment at entry to school and home background. On the introduction of "league tables" based on raw outcomes, it gave opponents of the latter a standard around which they could rally (see Saunders, 1999). During the 1990s there was a steady growth in value-added services to schools, driven partly by the need to look behind the crude characterisation of schools provided by "raw league tables". Most of these services were provided by LEAs, sometimes assisted by university researchers. Another area where the demand for value-added data grew in strength was the provision of quantitative information to individual schools for self-evaluation. A number of organisations began to provide this kind of service to schools. A consortium of metropolitan authorities, mainly in London, commissioned first the Institute of Education and later NFER to deliver value-added analysis of their schools' GCSE results, matching pupils' outcomes to prior attainment on a London verbal reasoning test (see Thomas, Pan and Goldstein, 1994; Kendall, 1995). Researchers based at the University of Newcastle-upon-Tyne (later at Durham University) began to develop value-added feedback for schools, beginning with a GCSE to A-level system (see Fitz-Gibbon, 1992). NFER provided its QUASE service (Quantitative Analysis for Self-Evaluation) for GCSE results, including detailed breakdowns of GCSE subject-level performance (see Schagen, 1996; Schagen & Morrison, 1998, 1999).

At about this time, governmental organisations began to respond to the drawbacks of the existing performance tables and the arguments in favour of a value-added approach. One might have expected that this coincided with the election of a Labour government, but in fact it started under the less doctrinaire Conservative administration elected in 1992. The then Department for Education (DfE) produced a

briefing paper (DfE, 1995) which, although relatively uncomplicated in its approach, did begin to address the issue, and also acknowledged the problems related to data collection and management which would be associated with any kind of national system. In 1994 the School Curriculum and Assessment Authority (SCAA) attempted to review the current situation and produce recommendations for future work (SCAA, 1994) and commissioned of a study, carried out by researchers at the University of Durham, into the possibilities for a national value-added system. This report (Fitz-Gibbon, 1997) covered a fair amount of work on this topic and made a number of recommendations, including advocacy of value-added information mainly for internal school information to aid school improvement, rather than for public consumption. Value-added information would only be incorporated in "league tables" voluntarily, or subject to a number of constraints about the minimum cohort size in a school.

The Drive for More Pupil Data

One consequence of the gradual acknowledgement of the strength of what might be called the "value-added argument" has been the recognition of the need for more and better data at the individual pupil level. As the demand for value-added information grew, it became apparent that one of the obstacles to its use for informing school self-evaluation and self-improvement was the need either for schools to provide their own data, with prior attainment in some form attached, or to provide funds for this to be done for them. Neither was a reasonable option for many schools, and it became apparent that only centralised systems for data collection and maintenance could ensure that such data was available locally and nationally in a consistent fashion.

Towards the end of the 1990s QCA and DfES, realising the need for national value-added data, convened several meetings of "value-added service providers" to discuss the way forward. Combined with a combination of an increasing central demand for individual information, and increasing IT human and machine capacity to match pupil information over time the eventual outcome was a quite extensive data set containing individual level data for all pupils. Since January 2002 schools have been required to complete the Pupil Level Annual School Census (PLASC) about each pupil in their school, instead of the previous school-level information. The process of collecting and cleaning this data, and matching it to national

curriculum outcomes, has been in hand for several years, with the outcome of high-quality *longitudinal* data for every pupil in England.

We are therefore in a situation in England where extensive and exhaustive national datasets are readily available, giving detailed information about individual pupils including their progress between key stages. Surprisingly perhaps other countries on the whole have not adopted value added methodology. This is partly due to the unavailability of the requisite data on a regular basis (England's record in the availability of matched national longitudinal pupil level data is unrivalled), and partly due to philosophical objections. Examples in other countries tend to be research-based, rather than official, one-off or patchy in parts of the system. Examples of such exercises can be found in Schagen and Hutchison (2005), Hutchison, Mifsud, Morrison and Grech (2003) and Lissitz (2005).

No advance in practice is ever entirely uncontroversial, and value added analysis is no exception. The most fundamental objections are to what the adjustment process does, or, indeed, to doing it at all. Kenneth Clarke when Secretary of State for Education dismissed the process as "sociology", and such attitudes have some respectable intellectual arguments underlying them. First, of course, a prospective employer is going to be more interested in one's qualifications and what thay show one can actually do, than in adjusting them for social background. Second, it is argued, this adjustment process can lead to "writing off" children from disadvantaged backgrounds. This argument however really only applies to adjustment for home background: it does not have the same force for adjustments for attainment at intake.
However well the researcher carries out the analysis, the results can have a historical quality about them. If it takes five years for a pupil to go from secondary school entry to the end of compulsory schooling, then the results may not be relevant to looking at the future performance of the current intake of pupils. Another related issue is the question of the stability of school effects over time. As mentioned earlier, if these are highly unstable then they become irrelevant to attempts at school improvement.

Inspection and Performance Data

By the 1980s HM Inspectors of Schools had to a large extent moved away from a role of assessing the accountability of individual schools

and their role has been described as largely advisory and policy related. Under John Major, however, his "big idea" of the Citizen's Charter (HMSO, July 1991) and Parent's Charter (DES, September 1991) aimed to turn inspection of schools "inside out", that is towards parents, school governors, and the general public, and away from central government (Smith, 2000). The aim was to carry out and publish inspections of all schools on a regular basis of a four-yearly cycle. In the event, when Ofsted took over from HMI there was a massive increase in the number of reports. At its peak HMI had published a maximum of 150 full inspection reports a year while by 1994-5, of the order of 3000 inspection reports were produced under Ofsted by 1994-5 (Smith, 2000). While HMI inspections could certainly not be described as cosy, they were on a within-profession basis, but Ofsted inspections put schools in the much more exposed glare of the public arena. Ofsted defines its role as contributing to the provision of better education and care through effective inspection and regulation. However it was widely considered that this was damaging to many schools, and stressful to the staff (Fitz-Gibbon, 1997). Certainly anecdotal evidence abounds of good and dedicated teachers finding the adverse public criticism of an Ofsted inspection the last straw that drove them out of the profession.

Eventually protests by academics and others about the inequity of judging schools based purely on their "raw" examination results led the Department for Education and Skills (DfES, as it had become) to develop value-added measures of school performance, which began to be published early in the 21st century but did not replace the performance figures based on raw results. These initial value-added measures were based on a simple "median line" model and took account only of prior attainment at the start of the "key stage", with no allowance for other factors which are known to affect performance, such as deprivation, special educational needs or ethnic minority status.

More recently DfES, alongside Ofsted, has begun to publish "contextualised value-added" (CVA) measures which are based on a more sophisticated model taking account of a range of background factors in addition to prior attainment. The CVA scores come with confidence intervals to express the reality that they are estimates of an underlying school characteristic based on finite data, and are a definite technical and theoretical advance on anything published

previously. These are presented to schools annually in a document known as a PANDA (Performance And Assessment), which contains other detailed information about the school and is intended to be used by the school to identify areas of weakness and help fulfil the government's school improvement agenda.

Thus in England we have woven a path over 30 years from performance data being collected purely for a national snapshot, to "league tables" for school accountability based entirely on raw examination results, to a mixed system where "league tables" and "health checks" co-exist with national monitoring of performance in a sometimes uneasy relationship.

Does Performance Information Help to Improve Schools?

As described above, in England there has been a shift from the pure accountability agenda to one which recognises the role of performance data to help schools raise attainment. Recently, the government as part of the "New Relationship with Schools" (Miliband, 2004) has been encouraging the use of data to drive school improvement. There is a wide variety of texts available that discuss feedback effects in depth (Kluger & DeNisi, 1996; Neubert, 1998, Locke & Latham, 1990), all of which agree about the benefits of feedback, but are unclear about what exactly "*feedback*" is or what kind of "*performance*" can be improved by it. For the benefit of this chapter we will use Coe's (2002) broad definition of "*feedback*".

> *"…'information relating to some aspect of task performance'."*

Coe (2002) has discussed in length the wide variety of texts that are available on the effects of feedback. Visscher and Coe (2002) discuss all the theoretical ideas of providing feedback to schools to aid with school improvement. Their book specifically looks at "*school performance feedback systems*" and how these systems can provide confidential information on a school's performance.

Recently, Kirkup et al. (2005) undertook a study of primary, secondary and special maintained schools in England. This study used a questionnaire survey and focus group research to gather evidence to

validate the assumption that the provision of data to schools leads to improvements in performance. The finds from this study suggest that "the impact of data on teaching and learning operates at two levels: directly by means of interventions targeted at individual pupils; and indirectly by means of whole-school approaches". The teachers who responded to the questionnaire and participated in the focus groups suggested "that data only becomes effective if it stimulates questions, discussion and action".

Tymms (1995) has argued that providing performance feedback to schools can improve a school's performance, he has conducted a number of studies to investigate the effects of feedback to schools. One of these studies looked at teachers responses to the different types of feedback supplied by The A-Level Information System (ALIS) - this is a performance monitoring system for upper secondary education in England and has been running for the past ten years. The overall findings did not show a significant difference in teachers' attitudes or self-reporting behaviour; however, some differences were noted between different subjects.

In recent years the NFER has undertaken a variety of large-scale longitudinal projects. These have included the Evaluation of the National Literacy and Numeracy Strategies (1999-2003), the national data collection of years 3, 4 and 5 optional tests results for QCA (Qualifications and Curriculum Authority) (1999-2001, 2003-2004) and the Performance Analysis Service and Support (PASS) (1999-2004). For these projects school feedback has been provided in the form of tables and charts illustrating pupil performance and progress in each school compared to national norms.

The Statistics Research and Analysis Group (SRAG) of the NFER has recently conducted a study looking at feedback that has been provided to primary schools by the NFER to see if it helps to enhance a school's performance. The main thrust of this study was a statistical analysis of the 2003 National Pupil Dataset (NPD). This dataset included pupil-level information regarding progress [Key Stage 1 (1999) to Key Stage 2 (2003)] as well as important background data on aspects such as project type, length of participation, and other contextual information. School identifiers allowed us to link this information to the Department for Education and Skills (DfES) school database to provide school contextual data.

As the data is educational there is a hierarchical structure to it, with pupils nested within schools and Local Authorities (LAs). The analysis was therefore carried out using multilevel modelling (Goldstein, 1995), which is widely recognised as one of the most appropriate methods for longitudinal data of this type. Multilevel modelling is a development of "regression analysis", and is an important technique in educational research as well as other social sciences. Multilevel modelling allows for the complexity of the data to be taken into account. This technique takes into consideration the variation between LAs, Schools and pupils and produces predictions and estimates the size of the effect at each level in terms of "random variance".

Within the complex structure of the data there is a single measure of interest; for this study it was one of National Curriculum average point score, English Score or Maths Score at the end of Year 6 for pupils aged 11. Each of these outcomes was considered during the multilevel modelling. The background factors i.e. gender, age, eligibility for Free School Meals (FSM), ethnicity and school type, can be defined at pupil or school level. These variables as a whole provide a suite of measures that can be used to explain the outcome measures. Multilevel modelling identifies which variables have a positive or negative significant effect, and the extent of that effect.

The overall findings for this study show that providing feedback to schools is positively related to Key Stage 2 performance (Hammond & Yeshanew, 2006). This study concentrated solely on primary schools that have received feedback from NFER projects. It would be beneficial to conduct more evaluations on school feedback systems, covering all stages of education, from a wider variety of establishments. This would provide a clearer evidence-base regarding the impact of all school feedback services. It would also be worth doing a follow up study to see whether the way in which these schools have used the feedback has had an impact on their performance. The findings of this study offer much encouragement for school improvement practitioners.

Considering the discussion of Visscher and Coe (2002), Tymms (1995) and the findings of the study of Kirkup et al. (2005) it seems that providing feedback to schools is a positive approach. This approach should be encouraged, but support system for schools,

teachers, policy makers and practitioners should be widely available to enable them to make the most of the richness of the data that is available to them in today's society.

Educational Statistics in Different Systems within the United Kingdom

The UK comprises four quite different education systems within its constituent countries: England, Scotland, Wales and Northern Ireland. In terms of the use of statistical information, they differ significantly, and there have also been clear changes over recent years in the case of at least one of the countries - England.

The Statistics Commission is an independent body set up to help to safeguard the quality of national statistics across the UK. In the summer of 2004, they commissioned the National Foundation for Educational Research (NFER) to carry out a review of the ways in which educational statistics (for those aged five to 18) are collected, presented and used across the four UK countries and the adequacy of the data for these purposes. This review comprised five elements:

- generating a comprehensive database of school-based education statistics, and making a critical review of the situation based on this and on NFER's own experience and expertise;
- collecting the views of key contacts (researchers, policy makers, data producers, the media), via email surveys and interviews;
- a survey of staff in all 213 local authorities (LAs) across the four countries;
- a survey of key school staff in a sample of 1000 schools;
- case-study visits to five LAs and two schools in each.

The full report can be found in Statistics Commission (2005), and a summary is given in Schagen & Ridley (2005).

The situation in **England** has been well described in an earlier section from a historical perspective, showing how it has evolved over a 30-year period. It is now the case that England has more detailed information about the pupils in its maintained school system than any other country, including not only results from the latest round of national testing but also matched results from earlier rounds and a

great deal of background information about each pupil. This National Pupil Database (NPD) forms the basis of a great deal of detailed work intended to put the data at the service of the system, including:

- Information about national attainment in core subjects, to feed into government targets;
- Performance data for school and local authority accountability, both "raw" and value-added;
- Information for schools on performance of previous cohorts, to inform self-evaluation and hence school improvement (mainly via the PANDA information packs);
- Software to help schools predict the performance of individual pupils based on national information - the Pupil Attainment Tracker or PAT system;
- Analysis of national data by a number of agencies (including NFER) to answer particular research questions or evaluate the impact of different educational initiatives.

Despite the availability of high-quality data in England, there are still many issues to be resolved and pitfalls to be avoided in its use, not least the question of the extent to which key educational staff understand the statistics presented and can make use of the results to inform their task of educational improvement.

Wales initially was part of the English system in the early 1990s in terms of national curriculum, testing and the publication of league tables. With the devolution of responsibility for education in Wales to the Welsh Assembly, however, the system began to diverge from that in England while retaining the same national curriculum structure. After a few years of continuing to publish school performance data in league tables, the Welsh Assembly accepted the arguments against these. Instead of moving, like England, to the publication of value-added measures to supplement the raw data, the Welsh decided to stop publishing school data altogether. However, more recently they have set up a system to compute contextualised value-added measures similar to England's. These will be reported to schools for their own self-evaluation, but not published in league tables.

Scotland has for centuries had a completely separate education system from England and Wales, with a different examination structure and to some extent philosophy of education. With an

approach similar to the APU surveys, the Scottish Survey of Achievement (SSA[1]) gives a national picture of performance in different subjects. Otherwise, attainment by all pupils in national tests is analysed but not published centrally. A central data system for Scottish schools (ScotXed[2]) allows schools access to data for their own self-evaluation and school improvement work. This whole system, including good links with local authorities relating to data use, seems to be close to a model of good practice as far as the results of our review could tell us.

Northern Ireland has long had its own separate education system, mainly based on selection at age 11 by tests and allocation to different types of school. This selective system is in the process of being ended following a critical review, but currently there is virtually no performance data available for schools to use for self-evaluation or for central government to use for school accountability.

As can be seen from the above brief overview, the UK is in some ways a microcosm of different approaches to using performance data. On the one hand is the situation (in Northern Ireland) where almost no use of data for school evaluation occurs. On the other hand (in England), there is a large and sophisticated mechanism for collecting test results and using them to provide information on national performance as well as measures of school accountability and statistics which can be used by individual schools for self-evaluation and (hopefully) improvement. It is probably fair to say that England has more data on the pupils in its maintained school system than any other country in the world - the challenge is to ensure that this data is used effectively to improve the nation's education.

On top of the collection of national data, most countries of the UK take part in international studies of performance run by the IEA and OECD - PIRLS, TIMSS and PISA. Results from these studies do provide some information, mainly on national performance against other countries, but also some evidence on changes over time. However, they cannot provide the detailed school-level information which is needed for self-evaluation.

[1] see http://www.ltscotland.org.uk/assess/of/ssa/index.asp
[2] see https://www.scotxed.net/jahia/Jahia/cache/offonce/lang/en/pid/87

At the end of our comprehensive survey of educational data within the UK, we produced some general recommendations, some of which are relevant to the subject of this paper. These are summarised below.

Move towards a coherent, "uses-driven" system for educational data

The uses to which educational statistics are or could be put should become paramount. In most cases these uses could be best met by a comprehensive "data warehouse" system linked to careful and detailed analysis. It should be noted, however, that this cannot be successful without high-quality analysis applied to the data - data does not speak for itself, and has to be carefully interpreted so that it can be used for this variety of purposes.

Monitoring national performance

The use of "high-stakes" national tests is not ideal but may be the best means of measuring educational improvement that we have available at present. It should be noted that national tests are in place to fulfil a number of purposes, of which monitoring national performance is just one. The use of international studies goes some way to addressing the need for further monitoring of this sort. We recommend that government studies of these issues be set up, in order to determine the best and most cost-effective way forward in the area of national performance monitoring.

School performance data

The production of school-level statistics has two main functions, which can be labelled as accountability and school improvement. It is clear that schools have sometimes experienced tensions between these two functions. We recommend that school performance data should be publicly available, but presented in a form that attempts to measure (as far as possible) the impact of the school on a range of outcomes, taking into account factors outside the school's control, together with indications of the uncertainty in these measures. The use of performance data for school improvement is also important, although more research is needed into whether and how it leads to such improvement. Governments should set up a cross-country body or team to study and advise on these developments.

Interpretation of statistics

It seems that there is a need for a more coherent approach to helping

school staff to understand statistics and use them effectively, and that this should be coordinated more centrally. This approach should consider the needs of all school staff (and governors). One possibility would be for a national qualification or training programme. With regard to the interpretation of statistics (and educational research in general), statisticians and researchers should present findings that are valid and accessible, but at the same time there should be high levels of statistical and research understanding amongst those developing policy. It would be beneficial to establish regular forums where the two sides could come together and share ideas.

Final summary and conclusions

The availability of national data on children's examination results leads fairly naturally to the development of school-level aggregate measures which are intended to indicate schools' relative performance. Within the logic of this process there arises a tension, however, between two different purposes for such measures: public accountability versus self-evaluation, or "league tables" versus "health checks". In this chapter we have shown how the United Kingdom in general, and England in particular, can be used as a case study of these tensions and the different ways in which performance data can be put to use within an educational system.
There seem to be several purposes to which individual pupil performance data can be put:

- Assessment of national standards of attainment, and whether these have changed.
- Deriving measures of school performance for public accountability.
- Feedback to schools for self-evaluation leading to improvement.
- Certificating pupils' levels of learning.

Some of these purposes are "high stakes" and inevitably give rise to security issues in testing and the need for new tests on an annual cycle. In such a system, the maintenance of standards can be challenging, and thus the consistent measurement of national performance over time becomes difficult. The current English system attempts to fulfil all the above purposes with one set of test results, which in practice is very hard to do across the board.

We have shown earlier that there is some evidence that feedback to schools of performance data can lead to improvement. However, the link between feedback and enhanced performance is by no means clear or automatic. It will depend critically on how detailed and well-focused the feedback is, how well school staff can interpret it and put into practice changes based on it, and to what extent those changes actually raise attainment. The English educational system is now taking this area more seriously, but it remains to be seen if the pendulum will swing all the way from an emphasis on accountability towards one on school self-evaluation.

Finally, what advice can we give to policy makers in countries which are considering setting up a national performance monitoring system? From our experience of the English system and its development over the years, we may make the following recommendations.

- Consider carefully which of the above purposes is the primary aim of your system. It may be able to address more than one purpose, but multiple aims will tend to lead to a loss of effectiveness.
- If the main aim is monitoring national performance, and changes in this over time, then consider a low-stakes, lightly sampled testing regime with repeated tests or items, similar to the UK's previous APU system. International monitoring studies (TIMSS, PIRLS, PISA) may help with this, but may not give you all the information you need.
- If the main aim is school accountability, consider carefully how you are going to derive fair measures of each school's achievements with the pupils it has and in its own context. Publication of raw performance data can be very misleading and reinforce social segregation and stereotyping. The outcome on which the main focus of measurement is set should also be chosen carefully, because a poor choice can distort educational processes[3]. The tests also become high-stakes, which means that more effort has to go into developing new tests each year and ensuring standards are maintained.
- If the main aim is school self-evaluation, then results can be confidential to the school and there is the possibility of using low-stakes testing. Careful analysis of data, good presentation of

[3] An example is the use of the indicator in English leagues tables of "percentage achieving five or more GCSEs at grade C or above". This can lead to a focus on "borderline" pupils to the exclusion of those well above or below this threshold.

results and training staff to interpret and apply the results of the analysis to improve teaching and learning are all essential elements of such a system.

- If the aim is to certificate individual pupils' levels of learning, then again a secure high-stakes system may be appropriate. If such certification takes place at the end of a phase of education, then it can be combined with school accountability measures, subject to the provisos set out above. Alternatively, testing during a phase of education can be used to provide diagnostic feedback at the pupil level, which may be an even more effective way of improving teaching and learning.

Common to all these recommendations is the fact that data is mute - it does not speak for itself, and needs careful analysis within a coherent framework in order to be used effectively to improve education. When this is the case, great things are possible.

References

Coe, R. (2002). Evidence on the Role and Impact of Performance Feedback in Schools, in *School Improvement Through Performance Feedback (editors Visscher, A. J., & Coe, R.)*. Lisse: Swets & Zeitlinger.

Foxman, D., Ruddock, G., McCallum, I. & Schagen, I. (1991). *APU Mathematics Monitoring (Phase 2)*. Slough: NFER.

Department of education and science (1991). *The Parent's charter: you and your child's education*. London: DES/HMI.

Department of education (1995). *Value Added in Education*. London: DFE.

Fitz-Gibbon, C. T. (1992). School effects at A Level: genesis of an information system? in *School Effectiveness (eds Reynolds, D. & Cuttance, P.)*. Cassell: London.

Fitz-Gibbon, C. T. (1997). *Feasibility Studies for a National System of Value-Added Indicators*. London: SCAA.

Fitz-Gibbon, C. T. (1997). "Ofsted: time to go?", *Managing Schools today*, 7, 6, 22-25.

Goldstein, H. (1995). Multilevel Statistical Models 2nd Edition. London: Edward Arnold.

Goldstein, H. & Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional

performance, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 159(3), pp 385-443.
Hammond, P. & Yeshanew, T. (2006). "The Impact of Feedback to Schools", submitted to Educational Studies.
HMSO (1991). The Citizen's Charter: Raising the Standard, Cm. 1599.
Hutchison, D., Mifsud, C., Morrison, J. & Grech, R. (2003). "The Malta Literacy Value-Added Project - a template for value added in small islands?", *Research Papers in Education 20*, 3, 303-345.
Kendall, L. (1995). *Examinations Results in Context: report on the analysis of 1994 examination results*. London: Association of Metropolitan Authorities.
Kirkup, C., Sizmur, J., Sturman, L. and Lewis, K. (2005). *Schools' Use of Data in Teaching and Learning. Research Report No 671*. Slough: NFER.
Kluger, A. N. & Denisi, A. (1996). The effects of Feedback Interventions on performance: a historical review, a meta-analysis, and a preliminary Feedback Intervention Theory, *Psychological Bulletin*, 119, 2, 254-284.
Lissitz, R. (ed) (2005). *Value Added Models in Education: Theory and Applications*. Maple Grove, Minnesota: JAM Press.
Locke, E. A. & Latham, G. P. (1990). A Theory of Goal Setting and Task Performance. Englewood Cliffs, New Jersey: Prentice Hall.
Miliband, D. (2004). "Personalised learning: building a new relationship with schools." Speech by David Miliband, Minister of State for School Standards. North of England Education Conference, Belfast, 8 January [online]. Available: http://www.dfes.gov.uk/speeches/media/documents/NorthOfEngland-final.doc [7 October 2004].
Neubert, M. J. (1998). The value of feedback and goal setting over goal setting alone and potential moderators of this effect: a meta-analysis. *Human Performance*, 11 (4) 321-335.
Rapple, B. (1994). "Payment by Results: An Example of Assessment in Elementary Education from Nineteenth Century Britain." Education Policy Analysis Archives, 2, 1, 1-23.
Saunders, L. (1999). A brief history of educational "value added": how did we get to where we are?, *School Effectiveness and School Improvement*, 10(2), pp 233-256.
School curriculum and assessment authority (1994). *Value-Added Performance Indicators for Schools*. London: SCAA.
Schagen, I. (1996). *QUASE: Quantitative Analysis for Self-*

Evaluation. Technical Report 1996: Analysis of GCSE Cohorts 1993 to 1995. Slough: NFER.
Schagen, I. & Hutchison, D. (2003). "Adding value in educational research - the marriage of data and analytical power" in *British Educational Research Journal*, Vol. 29, No. 5, pp 749-765.
Schagen, I. & Morrison, J. (1999). A methodology for judging departmental performance within schools, *Educational Research*, 41(1), 3-10.
Statistics commission (2005). *Report No. 26: School Education Statistics: User Perspectives*. London: Statistics Commission.
Thomas, S., Pan, H. & Goldstein, H. (1994). *Report on the Analysis of 1992 Examination Results*. London: Association of Metropolitan Authorities.
Tymms, P. B. (1995). Influencing educational practice through performance indicators, *School Effectiveness and School Improvement*, 6, 2, 123-145.
Visscher, A. J. and COE, R. (2002). *School Improvement Through Performance Feedback*. Lisse: Swets & Zeitlinger.

Equilibrium
On the Balance between Internal and External Evaluation in a Number of European Educational Systems

Bart Maes, Els Ver Eecke, Veerle Verhaegen (DVO) [1]

Introduction

The pursuit of quality is often part of studies and innovations nowadays. Also in education there is this persistent call for quality assurance.
This article evolved from the question how we, in Flanders, can maintain and improve quality in our education system through school self-evaluation. Knowing that we are not alone in this pursuit, it was a natural reflex to find out how other countries are managing. In an environment in which Europe's importance increases steadily, we can only gain by learning from others' experiences.

Schools have a responsibility to guard and improve the quality of learning processes. Therefore we believe that the quality of our education system lies primarily in the schools: we have started our research by listing all the agencies and authorities which contribute to the education system as a whole, and we then have considered how these interact with each other, and particularly with schools. In this, we have concentrated on the internal evaluation in schools.

The instruments for quality assurance and their implementation within an education system can vary enormously. Some will bring about pressure, while others are perceived as supportive. It is clear that sometimes pressure can evolve from an instrument that was meant to support, or the other way round. We realise that it can vary between schools and countries how people perceive an instrument.

[1] The division Curriculum (afdeling Curriculum), formerly known as the Department of Educational Development (Dienst voor Onderwijsontwikkeling), is part of the Flemish Ministry of Education in Belgium. It is a multi-disciplinary team of researchers and advisers: next to their initial subject training, all members of the team are educational specialists.
This article is based on the book Equilibrium, written by several members of our team, along with former advisers: Christine De Coninck, Rita Dunon, Laurent Osaer, Chris Van Woensel, Els Ver Eecke and Roland Voet.
Bart Maes, Els Ver Eecke and Veerle Verhaegen worked on the translation for this article.

Our ultimate aim was policy advice to the Flemish Department of Education and Training in Belgium, so we needed a means to identify similarities and differences in education systems. We found one in the work by MacBeath et al. (1999). It presents a conceptual framework for balance in an education system, rather than just a balance between internal and external evaluation.

This framework made us also look at the dynamics of changes within the various systems: who makes the changes happen, schools or policy makers?

We then tried to scale the instruments for quality assurance in the framework of MacBeath. This resulted in three separate diagrams for the studied systems and a combined one, and we will present those further on. By comparing the different systems, we hope to clarify which measures can tilt the balance one way or another and, more importantly, what policy makers can do to keep or bring their education system in “equilibrium”.

Overview

Concepts
First we will define the terms: we used educational literature, national and international, and previous work from our team.

Conceptual framework: school evaluation and development
Next, we will present the conceptual framework by MacBeath et al. for school evaluation and development.

Methodology
In the methodology part, the instruments used for quality assurance in the education systems are positioned within the three dimensions of MacBeath’s conceptual framework:

internal - external
top-down - bottom-up
pressure - support

It is evident that we have made choices that determined the course of our research. We will elaborate on these choices. Needless to say our

study could have been different had we opted for a different approach. Also, we may have overlooked an agency or instrument of quality assurance in a given system, if it is not reviewed in literature.

We have selected a range of European education systems for their characteristic properties and their mutual variation. Unfortunately, we could not extend the study to questioning teachers or head teachers, and therefore we did not evaluate how they experience the quality assurance in their education system. For our information, we relied on literature, websites and communication with contact people: from the CIDREE[2] and SICI[3] networks, and from the "European Network of policy makers for the evaluation of education systems".

Throughout this article, we will refer to education systems by their countries or regions. We will write "countries" for the studied countries or regions.
The studied countries are: Denmark, Finland, Ireland, the Netherlands, Austria and Spain.
The studied regions are: Scotland, Northern Ireland and England in the United Kingdom, North Rhine-Westphalia in Germany and Flanders in Belgium.

Three one-dimensional models and one three-dimensional model
In the major part of our article, we will account for our positioning of these education systems. Systems in transition (e.g. Austria and North Rhine-Westphalia) are positioned according to their current situation and not according to what they are planning to implement. In the analysis we will expand on some of their plans.

Towards a balanced system
In our concluding part, we will try to find patterns that become apparent in the comparison between education systems, and that can be helpful if policy makers consider changes within a given system.

[2] Consortium of Institutions for Development and Research in Europe: www.cidree.org/
[3] The Standing International Conference of Inspectorates: www.sici.org.uk

Concepts

In this part we will explain some important terms that we use intensively throughout this article. The terms on internal-external evaluation correspond to these terms in the ESSE-project (Maes, Ver Eecke & Zaman, 2004).[4]

Self-evaluation or internal evaluation is a process undertaken by the school, in which staff systematically gather and analyse evidence, including feedback from a range of stakeholders, and use it to assess and evaluate aspects of the school's performance against agreed standards. This process should produce outcomes that help the school to target its planning or initiatives for school improvement effectively.

It should be mentioned that there seems to be a consensus among most authors that self-evaluation is not an end in itself. This is best expressed in the report of the project "Evaluating quality in school education" (1999, 20): "Self-evaluation is a tool, and is judged by schools according to its impact on the effectiveness and improvement of their school. The positive attitude of schools towards evaluation is not because evaluation is an end in itself, but because it points towards aspects of school life which are significant and worthy of attention." According to the project's questionnaire, schools were more appreciative of the process when self-evaluation was followed by action. Therefore, the process of self-evaluation should be seen in action perspective and driven in accordance with expected outcomes.

This is also the opinion of educational experts and head teachers in a research conducted by Devos et al.: "Good self-evaluation is a permanent process leading to action. It must be part of the whole school policy (and not be an isolated activity). It requires an attitude of critical self-questioning. It is a tool for change or improvement and not an end in itself. However, it doesn't make external evaluation redundant: both are complementary".

External evaluation is a process of quality control undertaken by an agency or institute outside the school. This external agency can be either a person or a group (e.g. an inspector), or an organization (e.g. audit commission). The external agency cannot be part of the school

[4] SICI, Report on the Effective School Self Evaluation (ESSE) Project. SICI, 2003.

or is not involved with the school. Often, this agency represents the federal or local authority.

External evaluation may be justified on the grounds of a need at centralised level to control and guide schools. It attempts to ensure that quality education is provided, that schools use resources efficiently and that they provide "value for money". It has the task of ensuring that differences in school standards are not too discrepant and that agreed outcomes are met. External evaluation also raises the public's general awareness of quality issues by publishing reports on the general health of the education system or of specific schools. While external evaluation is driven primarily by a need for (political) accountability, it may combine this with an improvement perspective. External evaluation can offer feedback to schools on their strengths and weaknesses, drawing up action points, offering support or resources to meet their targets. It can give impetus to school improvement by providing comparative data which can then be used as a management tool for focusing on a school's performance in comparison with others (or benchmarks). (MacBeath, 2000, 91)

Internal evaluation refers to a self-evaluation process that schools can use to monitor quality from within. External evaluation means that quality is guarded by one or more agencies outside the school. In both forms of quality control, it is necessary to obtain data from the learning output of the students. These data can be gathered in various ways.

National programmes which monitor student performance (OECD/INES - network A, 2006, 15) can generally be categorised under three headings - assessment, testing and examination -, distinguished from each other in terms of programme goals, whether the programme focuses on a sample of students or a comprehensive group of students, and at what level results are reported.

The overall purpose of **national assessment programmes**, for example, is to assess students' performance (frequently against national curricula and/or goals) and to monitor and subsequently enhance the quality of the educational system at the national level. Assessment programmes generally are administered to a sample of students and focus on monitoring student achievement in the aggregate, rather than producing individual student scores.

Examples of national assessment programmes:
“National Assessment of Educational Achievement” in Ireland
“Periodieke peilingen” in Flanders

In contrast, student **testing programmes** usually seek to provide information at the school or student level, not just in the aggregate, and most programmes test all students in the target grades. The main goal of testing programmes is to assess individual student achievement in order to provide information on students to their schools, teachers and/or parents.

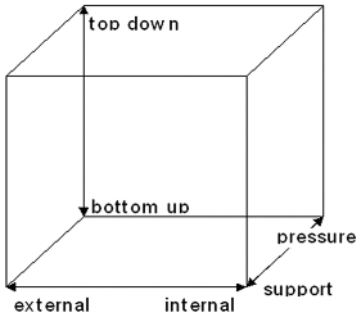
Examples of testing programmes:
“Key Stage Tests” in England
“CITO eindexamentoets” in the Netherlands

Last, **national examination programmes** aim to measure individual student achievement for high-stakes purposes, such as certification of completion or advancement to the next grade or level of education or to a particular occupation. Examination programmes commonly assess students at the end of upper secondary education and usually are not sample-based, though not all students in the target grades take all examinations. Most programmes require either all students or only the select students seeking the next level of education or occupation to sit the exam.

Examples of national examination programmes:
“Ylioppilastutkinto” (Matriculation) in Finland
“General Certificate of Secondary Education” in England

Conceptual framework: school evaluation and development

We can analyse school evaluation and development using the conceptual framework developed by MacBeath et al. (1999).



(MacBeath 1999, 2)

We can scale the following features of an education system in these three dimensions.

External evaluation - internal evaluation

This first dimension represents a continuum between self-evaluation and external control. On the internal side, we find systems that rely only on evaluation carried out by the schools themselves, without external involvement. On the external side, we find systems where the assessment of quality and standards rests exclusively with an external body. It depends on the context, where the balance lies.
In the pursuit of quality it is essential to find an equilibrium that is suitable for the system.

Pressure - support

The second dimension shows the amount of pressure from the system: on one side we find extreme high pressure, while on the other side schools receive a high level of support.
Pressure and support are best understood in terms of what people experience, whether they feel under pressure or supported.
Between the two extremes is a dynamic balance, where people perform best because they are intrinsically motivated and feel extrinsically recognised and rewarded.

Top-down - bottom-up

The third dimension reflects the implementation of change: at the top-down extreme, innovation is imposed from structures above, without participation. Change at the bottom-up extreme comes from teachers, from pupils and parents; from below. Most commentators agree that neither extreme is ideal: the best system seems to be one where bottom-up development is supported and endorsed from the top-down.

We have placed all the countries in this study within MacBeath's conceptual framework and we will present the patterns that we found. We will account for the method that we have used.

Methodology

First we list the instruments that education systems use to evaluate and to improve their quality. We grouped them according to their use, in MacBeath's three dimensions framework.

Dimension: external - internal

External

- monitoring of schools by inspectorate
- annual report by inspectorate
- national examinations or national testing programmes
- national assessment programmes
- monitoring by local authorities
- monitoring by external partners

Internal

- public report by the school on internal functioning
- school improvement plan or school development plan

Dimension: top-down - bottom-up

Top-down

- legislation about quality assurance
- national structures (inspectorate, national, regional and local authorities) with an impact on quality assurance and improvement

Bottom-up

- initiatives from the schools: e.g. applying for a quality assurance audit after an internal evaluation
- consulting advisors on topics related to self-evaluation or school improvement
- initiatives from head teachers, teachers, students or parents to evaluate or improve certain aspects of a school

Dimension: pressure - support

Pressure

- compulsory self-evaluation
- indirect obligation for self-evaluation (self-evaluation is not compulsory, but it is required to meet other legislation, e.g. school improvement plan)
- monitoring or other activities by inspectorate
- monitoring or other activities by local or regional authorities
- monitoring or other activities by other agencies
- inspection of the validity and reliability of self-evaluation
- self-evaluation as the starting-point for control by the inspectorate
- strongly recommended or imposed framework of quality indicators
- publication of school-results, with or without comparison between schools

- publication of external evaluation reports

Support
- examples of frameworks of quality indicators
- material and instruments to support self-evaluation processes: guidelines for target setting, publications, test banks
- guidance and advice
- supporting projects
- financial support
- networks
- databases with output data (e.g. learning outcomes, truancy rates)
- publications without the possibility to rank or identify schools or students
- in-service training or alternating learning projects for head teachers
- support from research

Positioning[5]

Educational experts rated the application for all the above mentioned instruments within every education system, based on the acquired information.
They then scored every education system in all three dimensions. The focal point was less the exact position than the (deviation from) the equilibrium in every dimension.

In the representation we position all the systems relative to the equilibrium in the centre of the diagram.

external	country	|		**internal**
top-down		country |		**bottom-up**
pressure		|	country	**support**

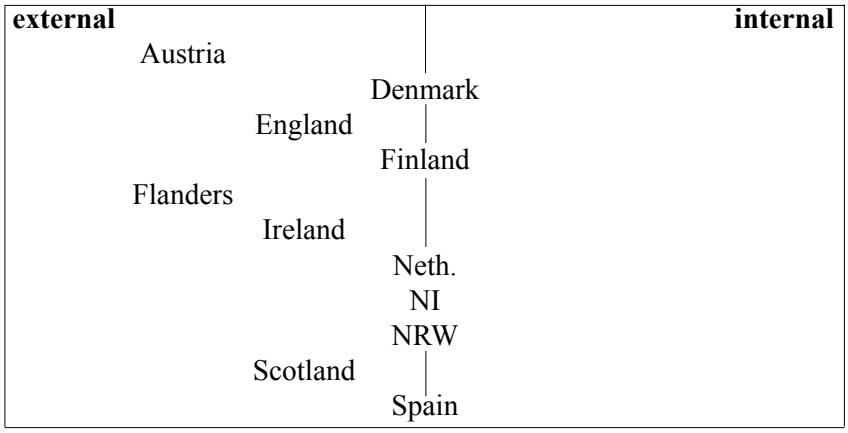
"country" indicates the position of the educational system in the given dimension, relative to the equilibrium (indicated with |)

[5] It is beyond the scope of this article to list all the instruments for every system.
The interested reader can find these lists in our extensive study "Equilibrium" on the balance between internal and external evaluation on the European scene, albeit in Dutch.
http://www.ond.vlaanderen.be/dvo/publicaties/equilibrium/equilibrium.htm
Separate list of references for the studied countries on:
http://www.ond.vlaanderen.be/dvo/publicaties/equilibrium/EVA_deel1_10_10_2005.pdf

We would like to emphasise that a system is positioned relative to a point of equilibrium in that dimension: the point of equilibrium does not reveal the weight on either side of the dimension, only the balance between the two extremes and the relative position of the studied system herein. If the position of a system leans to one side in a given dimension (*e.g. pressure, in the pressure - support dimension*), it only reflects a relative weight from the instruments on that side of the scale, and may well be caused by lack of weight on the opposite side. (*It can be that there is a considerable amount of pressure, or it can be due to poor support*). It is very well possible that the weight on either side differs considerably for all the systems within one diagram.
The benefit of this representation is the ease with which education systems (which can vary considerably) can be compared for three important characteristics. A limitation of this article is the two-dimensional representation of a three-dimensional model.

Analysis per dimension: three one-dimensional models

External / internal quality assessment



Neth. = the Netherlands
NI = Northern Ireland
NRW = North Rhine-Westphalia

None of the researched countries leans to internal evaluation: North Rhine-Westphalia, Northern Ireland, the Netherlands, Denmark, Finland and Spain have reached an equilibrium.

Ireland, England and Scotland incline towards external evaluation, while Austria and Flanders find themselves on the external side. In Flanders, that is not a result of a strong policy of external evaluation, but of poor internal evaluation.

External

In all the studied systems there is some form of external control. Even if there is no inspectorate, there is still control by local authorities.

- Finland, Denmark and Spain have an external evaluation in which various agencies are involved. Finland and Denmark have little or no inspection, but there is a strong monitoring by local authorities.
- Scotland and England have a highly organised approach of initiatives at central level, and of centrally monitored initiatives at the local level. Besides, they have an extended inspectorate for schools and for the local authorities. Furthermore schools are monitored rather closely by local authorities. The central authority evaluates the quality of individual schools, unlike the more northern countries, where schools are not, or less, the focus of external evaluation.
- The inspectorate plays a major role in the external evaluation in the Netherlands, Northern Ireland and Ireland, but there are no intermediate monitoring levels.

Most countries use external information to monitor quality of education: they collect and analyse data on students' performance. Many countries use national examination programmes. Other forms can be national assessment programmes and testing programmes.

- North Rhine-Westphalia will introduce national exams in the near future (after a central decision for the entire Bundesrepublik) and Austria seriously deliberates the option.
- Flanders and Spain have no national examination programmes (and do not intend to implement them) but they organise national assessment programmes.
- Finland organises national assessment programmes next to matriculation exams at the end of secondary education.

Internal
We find self-evaluation to some extent in all the studied systems.

- In Denmark, Spain and Finland self-evaluation is compulsory.
- In England, Ireland, the Netherlands, Northern Ireland, North Rhine-Westphalia and Scotland it is an indirect obligation. Schools have to set targets regularly, make development plans and write reports on their functioning.
- Austria considers the introduction of compulsory self-evaluation.
- In Flanders self-evaluation is only compulsory for schools which receive additional resources to provide equal opportunities.

In almost all the studied systems schools work with school improvement and/or development plans.

- Austria considers the introduction of a compulsory school improvement plan.
- In Flanders a school improvement plan is only compulsory for schools which receive additional resources to provide equal opportunities: the plan only needs to deal with the school policy on equal opportunities.

In almost all the studied systems the report, by the school, about school functioning is public.

- Austria considers the introduction of a public report on school functioning.
- In Ireland and Spain reports about school functioning are not imposed.

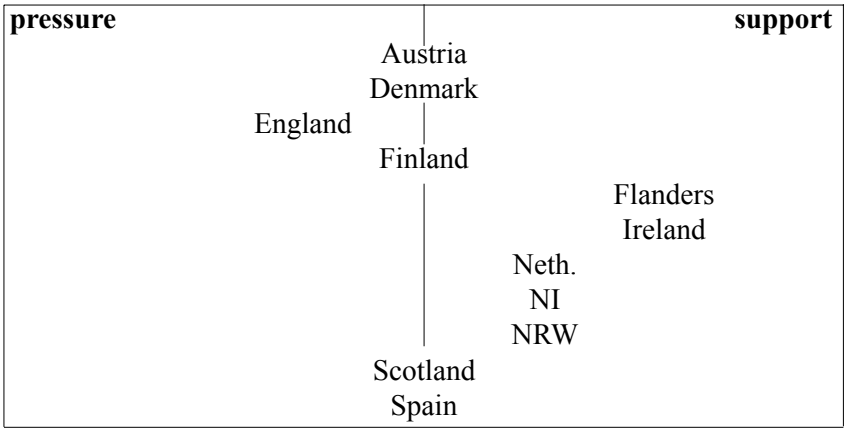
Trends
On the internal and external evaluation dimension, most education systems approach or reach an equilibrium. In Flanders and Austria the emphasis lies with external evaluation and there is little clarity on internal evaluation. We found no country where internal evaluation prevails.

It is striking that in systems with a profound external evaluation scheme, internal evaluation is also in various ways supported and directed towards the external quality indicators or criteria. The former does not seem to bar the latter, or vice versa. To maintain a balance

between internal and external evaluation, the system needs a clear framework for quality assurance, in which all the participants have a place.

The research of these systems pictures the role of the central authority regarding quality assurance. The more northern countries (e.g. Finland and Denmark) have developed a system in which the central authority mainly monitors the educational quality at system level. The monitoring of schools is left to local authorities. In that design, there is no central authority or inspectorate which engages in school monitoring.

Pressure / support



Neth. = the Netherlands
NI = Northern Ireland
NRW = North Rhine-Westphalia

Five out of the eleven education systems that we researched have reached an equilibrium between pressure and support: Austria, Scotland, Denmark, Finland and Spain. In England the pressure that schools experience is greater than the support, although they receive considerable support. This support may be more substantial than the support in the systems where a balance was reached.
North Rhine-Westphalia, Northern Ireland and the Netherlands lean towards the support side of the scale, whereas Ireland and Flanders have definitely more support than pressure. Here also, the inclination towards support does not necessarily imply a lot of support.

Pressure

In most education systems, there is a fair amount of pressure due to the required self-evaluation, either direct or indirect.

- Only in Austria and Flanders, self-evaluation is not compulsory.
- Austria has the intention to impose it.

In most education systems the inspectorate, the local authority or another public agency reviews whether the self-evaluation or the school improvement plan functions.

- In Flanders, the quality of self-evaluation is only inspected in schools that receive additional resources to create equal opportunities. Within these schools, only the self-evaluation of the policy on equal opportunities is inspected.

Most authorities verify the validity and the reliability of self-evaluation.

- This is not the case in Finland and North Rhine-Westphalia. In Flanders this verification is not yet fully effective, but it is in development.

Self-evaluation can be the onset for external review.

- This is the case in Denmark, Austria, Scotland and the Netherlands.
- In most other countries this is only partially the case e.g. in premeditated controls or when aimed at particular schools.
- In Ireland and Flanders, self-evaluation is only the starting point for external review to a limited extend.
- In North Rhine-Westphalia and in Spain, self-evaluation is not the starting point for external review.

One or more authorities can put pressure on schools to implement self-evaluation through various activities (e.g. providing benchmark data). When this is the case, the inspectorate often plays an important role.

- This is not the case for Finland and Denmark, without a real inspectorate. Yet in Finland another governmental body pressurises schools through her activities.

In most countries there is a certain pressure (or a strong instigation without formal enforcement) to use the government's framework for school self-evaluation purposes.

- So far this is only compulsory in Finland.
- There is no governmental framework in North Rhine-Westphalia.

Some countries opt for publication of test results in which schools can be identified.

- This happens in Denmark, England, Scotland, the Netherlands and Northern Ireland.
- This is not the case in Finland, Ireland, North Rhine-Westphalia, Spain and Flanders.

External control reports can be accessed by the public in seven of the studied education systems.

- This is not the case in North Rhine-Westphalia, Austria and Spain.
- Ireland's choice to stay away from public reports collides with the public nature of administration. It was subject for debate, and it has recently been decided that inspection reports will be published on the website of the department of Education and Science, starting from June 2006.

Support
Advice and counselling are offered to schools in all education systems. Almost all the countries offer instruments and material to support schools in the improvement of their quality assurance processes.

In most countries a framework of quality indicators is offered to schools.

- This is not the case in North Rhine-Westphalia.

In a number of countries, certain schools receive financial resources for self-evaluation or school improvement.

- Spain allows grants for schools with a good school improvement plan.
- Finland considers it and is carrying out experiments.

In some countries supporting networks were established.

- Those networks are accessible to all the schools in the Netherlands, Austria and Scotland.
- In Finland and North Rhine-Westphalia the networks are limited to certain schools.

Most education systems established supporting projects with various partners.

- Ireland, Scotland and the Netherlands have some examples of good practice.

Most education systems publicise test results without the possibility to identify schools or students.

- Very recently Austria implemented a process concerning national standards for system monitoring and to give feedback to schools, not for publication of results.

In many education systems schools have access to extended data bases with test results. The schools are allowed to compare their test results with similar schools. For this purpose characteristics of schools and students are linked to the data bases.

- This is not the case in Ireland, North Rhine-Westphalia and Austria.
- In 2006, Flanders is starting a research project to develop a school performance feedback system.

Most countries provide training for head teachers, middle management and/or teachers on quality assurance. Universities and researchers provide support in all countries.

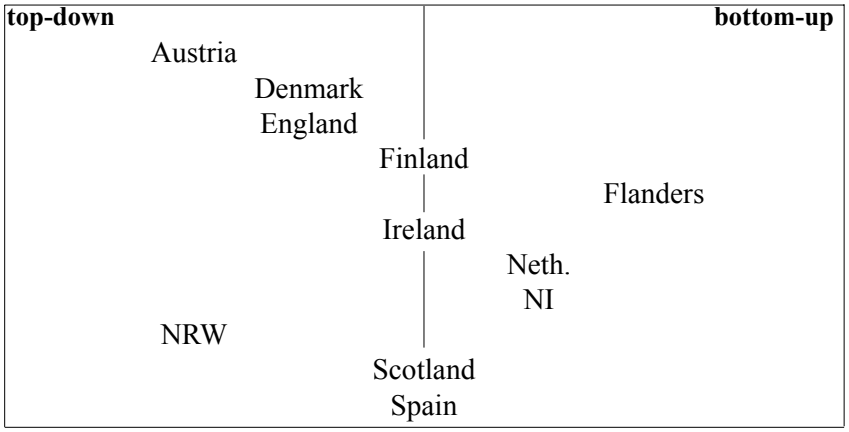
Trends

Some of these education systems reach a balance in this dimension. England emphasises pressure, Flanders and Ireland lean towards support. However, this is due to the absence of pressure. This dimension deals with various realities in different education systems. Initiatives which were meant to support can feel like pressure, and vice versa. It is remarkable that within an education system, pressure and support often go together: England is probably the clearest

example. The league tables and the inspectorate cause considerable pressure, but in the meantime England invests a lot in agencies which provide training, counselling and support for schools. Moreover, they create a feedback system to support the policy of quality assurance, centrally and locally.
Scotland also experiences a lot of support next to a fair amount of pressure, but here it stems from a tradition of consultation.

Many countries invest in data collection and feedback, well aware of the pressure that it can cause. Many authorities provide and stimulate the use of quality frameworks, instruments for self-evaluation and benchmark data. Institutes for quality control often bring about a combination of pressure and support.

Top-down / bottom-up



Neth. = the Netherlands
NI = Northern Ireland
NRW = North Rhine-Westphalia

Austria and North Rhine-Westphalia incline towards top-down, a common feature for systems that change their course drastically for innovations. England and Denmark draw near an equilibrium; Ireland, Finland, Scotland and Spain reach it, Northern Ireland and the Netherlands go beyond and incline to-wards bottom-up.
Once more the Flemish education system is the odd one out: this means that in Flanders the schools have a lot of freedom for initiatives on quality assurance. This does not imply that schools use this freedom effectively. Again, an incline towards bottom-up does

not mean there are no top-down initiatives, but there is room for schools to launch initiatives.

Top-down
Most of the studied education systems have mechanisms for central steering, e.g. legal framework or regulation, national structures and an inspectorate.

- North Rhine-Westphalia is very much regulated from the top, through an extensive regulation of self-evaluation and quality assurance, and through a complex establishment of national, regional and local authorities.
- Education systems in England and Scotland are governed by extensive regulation about target setting and performance/quality management. They have a clear framework of quality indicators for self-evaluation.
- Steering in Denmark, Finland and Spain is powerful. This is because of their statutory frameworks for self-evaluation, and also due to the impact of the national, regional and local authorities.
- In the Netherlands, self evaluation is not compulsory, but schools are by law responsible for the quality of education. The government defines the quality concept.
- Northern Ireland has similarly limited steering provision for target setting, and self-evaluation is promoted through the school development plan.
- Austria and Flanders do not have comparable legal provisions: Austria considers imposing self-evaluation, and Flanders only imposes self-evaluation for schools that receive additional resources to realise equal opportunities.

Bottom-up
Most countries leave enough room for initiatives from schools, head teachers and teachers, parents and students to shape quality internally.

- In North Rhine-Westphalia schools have rather limited autonomy.
- So far, England and Scotland have little room for initiative, but are evolving towards more.
- Denmark, Finland, Ireland and Spain leave substantial room within a given framework.
- The Netherlands, Northern Ireland and Flanders allow freedom for schools.

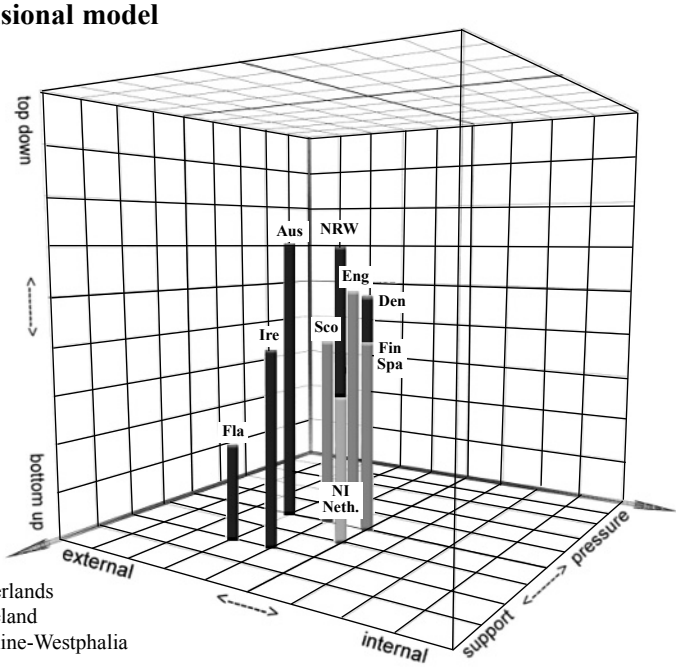
- The Netherlands, Austria and Scotland encourage the creation of networks between schools as a bottom up strategy.

Trends

On the top-down - bottom-up dimension, we find most diversity between education systems. Austria and North Rhine-Westphalia lean towards central steering, as opposed to Flanders, where bottom-up prevails noticeably.

This dimension is the one that reflects how education systems allow autonomy to schools. Again, in some countries, both extremes can exist at the same time. Those countries which opt for regulating the authorities' expectations, (endorsement of self-evaluation, target setting, performance/quality management, school development plan) will often leave much room for initiatives from schools. In Scotland and Finland, regulations do not impede the development of a largely supported evaluation culture. On the contrary, schools, teachers and students are encouraged by expectations which are well formulated in frameworks for quality assurance.

Synthesis in one three-dimensional model



Aus = Austria
Den = Denmark
Eng = England
Fin = Finland
Fla = Flanders
Ire = Ireland
Neth. = the Netherlands
NI = Northern Ireland
NRW = North Rhine-Westphalia
Sco = Scotland
Spa = Spain

Conclusion: towards a balanced system

From the previous we learn that it is not straightforward to reach the equilibrium. Flanders has recently made a number of policy decisions concerning autonomy, accountability and quality assurance in schools. Every decision affects the balance between internal and external evaluation, and the manner in which educational quality will be guarded.
The current policy on quality of education allows steering power to the authorities in Flanders. But the manner of steering is changing: the authorities want schools to gain more insight in their impact on pupils' learning. Therefore the schools will need access to data and instruments to monitor the progress of the learners.

Supported by various examples of good practice from the systems in the study, we have drawn up some recommendations which can improve the balance in quality assurance. These recommendations should be considered within the context of each education system.

A culture of evaluation

The quality of education is subject to permanent discussion. This debate ought to take place at national and local level, using reliable output data. It is important to establish an environment which instigates various forms of evaluation. At every level the partners should participate in evaluation. Ideally there is choice between forms and instruments for evaluation to gather relevant information.

- A coherent approach of external evaluation at national and local level, and self-evaluation by local authorities, schools, teachers and students exists in Finland. All use different instruments adapted to the level, but these instruments harmonise within a common framework of reference.
- In Austria authorities are trying to warm the various partners to an innovative quality concept. Different ideas for initiatives on evaluation were brought together in a white paper, which evolved into a social debate.

Data on performance

In a climate where evaluation is important, a framework for a coherent policy to measure students' performance is necessary. It is expedient that the authorities have the returns of schools available.

When authorities provide these, schools will operate in an evidence based environment: they can improve their own functioning and the learning processes and outcomes of their students. Authorities should consider how to present and use the national performance data: they need to decide which initiatives are required to teach schools to interpret the data correctly and use them accordingly.

- In England the authorities present a yearly "package" with test data to schools and local authorities. These benchmark data can be used to examine aspects of a school's performance against all schools nationally or against a group of similar schools. These data are very important in England. They are considered an essential element of self-evaluation, target setting and school improvement.
- In Denmark, schools can use social background information and benchmark data on pass grades and dropout rates provided by the Ministry of Education.

A framework for quality

If a system wants to establish evaluation which supports the implementation of the educational goals, it needs a coordinated approach between different levels in a broad and unambiguous framework. The framework must indicate the responsibilities of all partners, and the relationship between internal and external evaluation must be clear. A system for quality assurance needs to comply with social and cultural standards in the country or region.

- In Finland, the National Board of Education made a model to evaluate learning outcomes on the national level. It is conceived around efficiency of teaching, profit of learning and economy of finances. The model was developed for use at national level, but it can be used in schools for self-evaluation.
- In Spain, INECSE (the Institute for quality) adapted the European framework for quality (EFQM) to the Spanish system, and offers this to schools.
- Ireland, Scotland, Northern Ireland and England offer the inspection's quality framework to schools. Its use is encouraged.

A law for quality

Some education systems evolve towards a quality law for education. Within this law, the authorities can specify all aspects of quality assurance.

- In 2002 Spain launched a law that regulates the integral quality assurance in the education system. It aims at better learning outcomes and the guarantee of equal opportunities, imposing national assessment programmes and compulsory self-evaluation, including a school improvement plan.
- The Standards in Scotland's Schools Act from 2000 is a legal framework with principles and structures for quality care in education. The applications are reviewed every five years.
- Finland imposed in 1999 a law that schools realise self-evaluation and participate in external evaluation. The law indicates what to do with the evaluation results, and it accounts for the role of NBE[6].
- Following the introduction of compulsory school development plan, the Irish Ministry of Education introduced two initiatives to help schools meet their requirements: the "School Development Planning" for primary schools, and the "School Development Initiative" for secondary schools. With these supporting initiatives, the authorities want to offer help and resources in various ways, together with the inspectorate and the Network of Education Centres.

System monitoring

Many countries have an institute to monitor evaluation at national level, like Denmark, Finland, Austria and Spain.

- The Danish Evaluation Institute (EVA) was established in 1999 as the only organisation (under the auspices of the ministry of Education) with responsibility for evaluation of all funded education institutes.
- INECSE is a national institute for quality and evaluation in the Spanish education system. They outline the evaluation for the entire system in association with the autonomous communities in Spain, and design a national system of indicators to evaluate the efficiency and returns of the system. They also suggest initiatives to improve educational quality.
- The Finnish Education Evaluation Council is related to the Ministry of Education, and is responsible for planning, development, coordination and organization of national evaluation in education.

[6] NBE: National Board of Education.

The inspectorate

Self-evaluation is subject to external control. Many countries seek for a more explicit control by the inspectorate, of the self-evaluation and of the school improvement plan.
Many countries aim at a well-defined link between self-evaluation and external evaluation. One way to achieve this is meta-evaluation: an external scrutiny of the self-evaluation. It is important that policy makers at central or regional level and external evaluators (e.g. the inspectorate) communicate well with schools about the relationship between internal and external evaluation, and about the framework offered by authorities or by external evaluators.

Active networks

Supportive networks appear to have a positive effect on the implementation of self-evaluation. These networks can grow at different levels and can take various shapes.

- Scotland opts for participation before implementation of new ideas. The collaborative approach broadens the social base for changes.
- The Quality Initiative in Scottish Schools (QISS) is a partnership between schools, local authorities and the inspectorate. It helps schools raise standards and launch permanent quality improvement.
- In Austria the “Qualität in Schulen” (Quality in schools, QIS) is a project with a network and an interactive website. It offers a framework of reference for quality care. The network provides material, guidelines, support, scientific information, instruments and policy information, and it updates all these regularly. The network also offers an interactive platform to discuss and exchange experiences.
- In the Netherlands, networks as “Q*primair” for primary schools, and “Q5” for secondary schools, stimulate the quality assurance in schools. They provide information on self-evaluation instruments and support quality care in schools.

The balance that we can gain within an educational system is subtle and dynamic. Because education is ingrained in the culture of the country, it is not feasible to write a unique guide to the very best of education systems. Thanks to different cultural and social circumstances, every country must seek its balance within its education system, rooted in the local cultural society.

References

Blanquaert, A., Maes, B. & Ver Eecke, E. (2001). De relatie tussen de zelfevaluatie en externe evaluatie. Evenwichtsoefening tussen verantwoording en schoolverbetering? *Impuls*, 32 (1), 11-16.

Bosker, R. J. (Ed) (2001). Kwaliteitszorg, in *Onderwijskundig Lexikon,* Editie III. Alphen aan den Rijn: Kluwer.

Breye, V. (2005). *Het inspecteren van de indicator "ethos" in Schotland, Noord-Ierland, Nederland, Engeland en Vlaanderen*. Gent: Universiteit Gent, Faculteit Psychologie en Pedagogische wetenschappen, scriptie.

CIDREE: www.cidree.org

CIDREE (2004). *Becoming the best Educational ambitions for Europe. Brussels: Consortium of Institutes for Development and research in Education in Europe.* CIDREE.

Daniels, K. & Kemps, T. (2001). *Schooldoorlichtingen in het Vlaams Secundair Onderwijs. Een kwalitatief onderzoek naar de bijdrage van doorlichting tot schoolontwikkeling.* KU Leuven, paper.

Departement Onderwijs: www.ond.vlaanderen.be

Dienst voor Onderwijsontwikkeling: www.ond.vlaanderen.be/dvo

Dienst voor Onderwijsontwikkeling (DVO) (1998): *Een geïntegreerd model van prestatiemetingen voor kwaliteitsbewaking en kwaliteitsbevordering vanuit de overheid*. Brussel: DVO, Interne nota.

Dienst voor Onderwijsontwikkeling (DVO) werkgroep evaluatie (2005). *Equilibrium. Evenwicht tussen interne en externe evaluatie op de Europese scène*. Brussel: DVO.

http://www.ond.vlaanderen.be/dvo/publicaties/equilibrium/equilibrium.htm

http://www.ond.vlaanderen.be/dvo/publicaties/equilibrium/EVA_deel1_10_10_2005.pdf

Dienst voor Onderwijsontwikkeling (DVO) (1999). *(Overheids)zorg om kwaliteit. Een geïntegreerd model van prestatiemetingen in het onderwijs van de Vlaamse Gemaanschap. Een gefaseerde aanpak bij het begin van het nieuwe millenium 2000-2010. Een beleidsvoorbereidend rapport*. Brussel: DVO, Interne nota.

Fullan, M. (2001). *Leading in a culture of change*. San Francisco: Jossey-Bars.

Fullan, M. (2001). *The new meaning of educational change*, Third Edition. New York: Colombia University, Teachers Colledge Press.

Harris, A. (2002). *School improvement. What's in it for schools?*

London-New York: Routledge/Palmer.
Inspectie Van Het Onderwijs (2005). *Expert Meeting, Proportional Supervision and School Improvement in an International Perspective (PROPINT)*. Utrecht: Inspectie van het Onderwijs.
Inspectie Onderwijs (2003). *Toezicht gebundeld. Actuele ontwikkelingen in het toezicht op het onderwijs*. Utrecht: Inspectie van het Onderwijs.
MacBeath, J. (1999). *Schools must speak for themselves. The case for school self-evaluation*. London/New York, Routledge/Palmer.
MacBeath, J. et al. (2000). *Self-evaluation in European schools. A story of change*. London/New York: Routledge/Palmer.
MacBeath, J. et al. (1999). *Evaluating quality in schools. A European pilot project. Final report*. Brussels: European Commission.
Maes, B., Ver Eecke, E., & Zaman, M. (2000). *Inspectorates of education in Europe. A descriptive study*. Utrecht/Brussels: SICI-DVO.
Maes, B., Ver Eecke, E. & Zaman, M. (2004). Gezocht: vruchtbare relatie tussen interne en externe evaluatie, *Impuls*, 34 (4), 215-237.
Maes, B., Ver Eecke, E. & Zaman, M. (1999). Schooldoorlichting van Ierland tot Tsjechië, *Impuls*, 30 (1), 13-26.
OECD/INES/network A, Assessment, Testing and Examinations in OECD Countries - A snapshot, *Review of Assessment Activities*, January February 2006.
SICI (2003). *Report on the Effective School Self Evaluation (ESSE) Project*. SICI: www.sici.org.uk
Standaert, R. (1990). *De vlag in de top. Het secundair onderwijsbeleid in Duitsland, England en Frankrijk*. Leuven: Acco.
Standaert, R. (2002). *Inspectorates of education in Europe. A critical analysis*. Leuven: Acco.
Standaert, R. (1998). Loodsen of sturen. Sturingsmechanismen voor inhouden op school, *Impuls*, 29 (2), 123-129.
Standaert, R. (2005). Onderwijsvormen in Europa, *Informatie Vernieuwing Onderwijs*, 26, n 99, 19-47.
Standaert, R. (2004). *Vergelijken van onderwijssystemen*. Leuven: Acco.
Van Petegem, P., Verhoeven, J. C. et al. (2005). *De implementatie van het gelijke onderwijskansendecreet I, Evaluatie-onderzoek naar het zelfevaluerend vermogen van scholen en het ondersteunend aanbod. Eindrapport OBPWO* - project 02.05. Universiteit Antwerpen, KU Leuven.
Van de Perre, L. (2001). Vlaamse onderwijsindicatoren ter

ondersteuning van het onderwijs-beleid op microniveau, *Impuls*, 32 (2), 95-103.
Vanderploeg, P. et al. (1999). *De overheid als bovenmeester*. Baar: Intro.
Wiborg, S. (2004). *Education and social integration. A comparative study of the comprehensive school system in Scandinavia*. Copenhagen: The Danish University of Copenhagen, paper.

Potentials of Centralized Proficiency Tests for the Improvement of the Quality of Schools - an Example from North-Rhine-Westphalia

Rainer Peek, Peter Dobbelstein

Setting and testing of standards as central elements of the discussion on quality

Centralized tests, quality analyses respectively school inspections and regular centralized proficiency tests which cover all pupils of a year (grade) are rather new instruments in the States of the Federal Republic of Germany. Here the picture is rather non- uniform: whereas a fair number of States can look back on quite a few years of experience already, several of these instruments are now just being introduced in some States.

This article will focus on the centralized proficiency tests which are being carried out at the moment for the third time, and which are to be made a project covering the whole Federal Republic respectively more than one State from 2007/2008.

It is the basic idea of this instrument to professionalize the specialist discourse in schools by offering them empirically validated reference values with regard to achieved levels of competence of their pupils in important domains of the subjects German, English and Mathematics. Through concrete comments to professional strengths and weaknesses schools get impulses for the further development of teaching and learning.

The two parameters setting of standards and testing of standards are of central importance, and looking at the keyword setting of standards a dominant role is played by an expertise by Klieme et al. (2003) “On the development of national educational standards” commissioned by the Federal Government in 2002/2003. Starting from the PISA results of 2000 this expertise calls for national educational standards respectively centralized educational aims and for measures how the compliance with them can be appropriately tested. As main assertions of the expertise can be stated that the standards should comprise general educational aims and stipulate which competencies pupils

should have acquired at a certain point in time of their schooling. To allow an empirical testing of the standards achieved the competencies will be phrased in a way that they can be tested and that they are related to subjects. They are definable, i.e. by using suitable test procedures it can be determined whether a pupil has acquired a certain competency or not. To reach this aim the competencies must be described in such a concrete manner that they can be converted into tasks and can be recorded with the help of tests.

Proficiency Tests and Standards of Comparison

State-wide and area-wide proficiency tests are taken in North-Rhine-Westphalia at the elementary level in year 4 (from 2007 in year 3) and at secondary level I in year 9 (from 2007 in year 8). The following statements focus on proficiency tests at the secondary I level (to the concept and procedures at the elementary school cf. in detail Helmke, 2005; Helmke & Hosenfeld, 2005a, 2005b).

The tasks for proficiency tests are centrally developed by teams of teachers who are supported by educational experts in didactics and subject specialists of the Landesinstitut and the School/Quality Agency in Soest. The individual tasks which have proved reliable in pre-piloting at selected schools and the evaluation guidelines will be checked - on the basis of a comprehensive pilot study - for their suitability by an independent scientific taskforce at the University Duisburg-Essen (empirical assessment of levels of difficulty, interrater - reliability, fast scalability etc.). Finally tests are introduced to fields of the subjects German, English and Mathematics which always present a central complex of tasks for all school forms and additions for individual types of schools. In this way a state-wide scaling of data for separate test areas is made possible.

To enable comparisons, uniform evaluation criteria for the test items must be available which take their orientation from didactic and subject content demands on the one hand and competency requirements of central curricula on the other hand. The decisive point is that these criteria can be uniformly understood and used by teachers when marking the tests of pupils beyond individual classes and schools. The evaluation of tests of individual pupils is done - on the basis of extensive evaluation manuals - through teachers of the

schools and a probe of individual pupil tests from 250 schools is then centrally requested through the State Institute for School/Quality Agency and corrected a second time by independent teachers with the same guidelines. Only those items will be part of the following comparative school feedback where the interrater- agreement is sufficiently high.

The data for individual pupils respectively classes are - likewise by teachers of the schools - immediately after the taking of the standardized tests respectively as part of the standardized test evaluation entered into a prepared data mask to which every school has access with its own password.

Teachers get to know through the computer supported feedback from the proficiency tests to which degree important standards have been met by their pupils and in which particular areas of tasks their pupils have particular strengths or weaknesses, too. This database and information base serve the professionalisation of specialist discourse in schools: on this basis the subject conferences are to agree on measures about advancing and developing teaching and learning. A state-wide centrally evaluated probe additionally supplies in the sense of system monitoring more important information to analyse development trends in North-Rhine-Westphalia (cf. the scientific report on proficiency tests 2005 at http: www. learn -line. nrw.de/ angebote/lernstand8/download/ergebn_05/lse-ergebnisse_2005.pdf).

Traditionally teachers base their judgement on pupil achievements on comparisons. Parameters for teacher judgements in regular lessons are considered to be the criterion benchmark meaning the achieved proficiency in relation to subject demands, the social benchmark meaning the achieved proficiency compared to the average level of performance of the class, and the individual benchmark meaning the development of learning of an individual pupil over time. Centralized proficiency tests have as their aim - over and above the above mentioned criterion connections to the basic curricula - the comparability with norm values which are valid for all school forms and state-wide.*

* Proficiency tests are part of a professional culture of evaluation to enhance instruction and learning. In the sense of a snapshot they provide insights in proficiency achieved by groups of learners. As snapshot that can neither record learning development nor learning processes they suffer clear limits and need additions with regard to personal respectively individual standards of comparison: at a professional level through continual observation in regular lessons to see how far proficiency in the subjects has been reached and with regard to individual support through the observation of learning processes and developments of learning of the individual pupils.

In Germany as a whole and within States schools of the same type in some cases show considerable differences with regard to pupil achievements.

The PISA study showed within schools of the same type e.g. spreads of the average learning levels in Reading Comprehension, Mathematics and Sciences which are considerable according to school type and subject - in time up to three years of learning. From PISA and from many other proficiency studies it has become obvious that differences in pupil achievement are largely caused by characteristics outside the schools and outside teaching and learning - especially socio-demographic factors like social background, affinity to education within the home, status of migration or lingua franca of pupils.

To be able to make statements on the effectiveness of work done in schools and in the classrooms it has become customary for achievement studies to include beyond general school specific benchmarks the difference between the actually produced achievement values of pupils of individual schools (respectively individual classes) and an empirically set so-called expectancy value (adjusting). The effectiveness of educational work is measured as a difference: the average achievement of a school (respectively of a class) is compared to the achievement of other schools (respectively classes) with similar learning related factors (the initial learning situation, social status etc.).

To enable schools to get a “fair” position finding and a realistic estimate of their results in the proficiency tests, it is also not enough to make available as benchmarks representative results of different school forms. Schools with unfavourable learning conditions of their pupils would tend to be in danger of underestimating the quality of their efforts at improvement, schools, however, with a favourable set up would tend to overestimate themselves (cf. Arnold, 1999).

In North-Rhine-Westphalia it is not possible to undertake a setting of expectancy values for all schools respectively classes involved as was done with scientific comparative achievement studies e.g. LAU, QuaSUM or PISA, when evaluating centralized proficiency tests: the evaluation of the tests takes place in the schools; parent and pupil interviews as customary with the above mentioned scientific comparative achievement tests or even tests of cognitive basic abilities cannot be evaluated by the schools on their own. To nevertheless

make meaningful comparative opportunities available to schools the following procedure will be chosen: the assigning to a group of schools of the same school form with similar characteristics of the pupil population is done by the school itself - (as a rule by the school administration) prior to the execution of the achievement test. To this purpose the schools have to estimate selected basic conditions of their location. Even when schools have no hard data with regard to individual parameters (e.g. to the affinity to education in parental homes) and they have to rely on estimates, school research has shown: school administrations possess a largely realistic and unerring picture of the socio-demographic composition of their schools.

In the run-up to the proficiency tests of 2004 the Hauptschulen and the Comprehensive Schools assigned themselves to one of the three given types, the Realschulen and Grammar Schools to one of two types. The description of the different types always contains characteristics of the socio-demographic composition of the pupil population and regional characteristics of the school environment. The descriptions of the types concentrate on characteristics that have a high correlation with the achieved subject results of the pupils in their individual school forms.

Feedback Formats and Evaluation Perspectives for Schools

There are several levels for an evaluation of the contents of proficiency tests:

In the centre of result feedback which every school can access in a protected Internet data base is information of the achieved level of competency of the pupils. For reading comprehension, listening comprehension and processing and writing in German, for listening comprehension and writing in English and for problem-solving and modelling in Mathematics schools now have information which percentage of their year and of individual classes respectively courses have reached a certain level of competency and what the distribution is within the corresponding comparison groups (class/course, school form, location type).

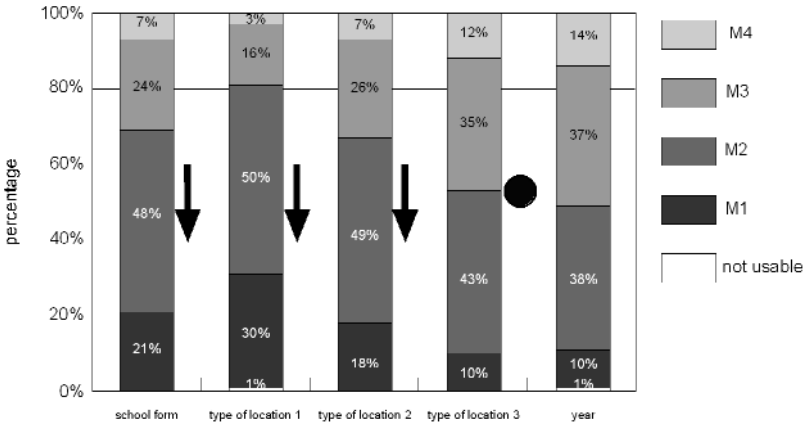


diagram 1: Proficiency test 2005: feedback on competence levels achieved; the distribution in year 9 (example E-courses Mathematics Comprehensive Schools) in comparison to the distribution in the school form and in the three location types

It will be helpful for a thorough debate on the subject if teachers take a closer look at individual tasks with their specific demands beyond the levels of competency. In addition teachers have also access with selected tasks to solution quota and explanations on specifications of these tasks.

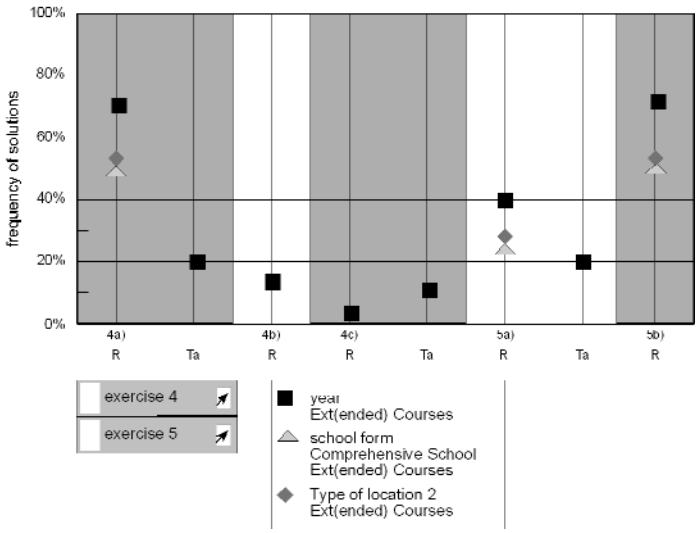


diagram 2: proficiency test 2005: feedback on exercises solved ; the solution frequency of individual exercises in a year 9 (example E-courses Mathematics Comprehensive Schools) compared to school form and location type 2; to increase clearness only those comparison data are shown that deviate in a statistically significant way from their own results

In the given diagrams only those comparative data are indicated with a symbol that deviate in a statistically relevant way from the year (grade) of their own school. If no other symbol appears at the task except for the data of the year it means that the result of the year does not deviate from any comparison group.

Proficiency Tests and the Use of the Results in Schools

The centralized proficiency tests are concentrated more on years towards the upper end of school forms. Here potentials can be seen in interpreting and in discussing the results of classes respectively courses as a consequence of teaching and learning in the school as a whole (cf. Peek & Dobbelstein, 2003; Dobbelstein, Peek & Schmalor, 2004, Dobbelstein & Peek, 2005).

It has been arranged with the proficiency tests that the test items stay in the schools and can be analysed within the school (especially within the corresponding subject conferences) with the idea of setting standards. That allows a debate within the school on standards achieved which goes beyond the comparison of empirically collected norm values (from centralized respectively norm rating probes) and permits an analysis directed by didactics.
The advantage of proficiency tests lies in the fact that they allow a "wider perspective" on the achievement profile of one's own class/ one's own course and the year through curriculum - oriented (criterion parameter) and reference group oriented (social parameter) possibilities of comparison which are normally not available in daily school life. On this basis the following questions arise for the individual subject teacher as well as for the subject group respectively the subject conference of a school which could be the basis for the setting of specific priorities in teaching and learning.

Individual evaluation at the level of the subject teacher

- Which results are striking/unexpected/in need of explanation?
- Do the difficulties concern subjects respectively parts of subjects or a certain area/type of exercise?
- Are there indications of lack of effort, lack of understanding the teaching, lack of test writing expertise of pupils?
- Looking at the concentration of errors, are these errors accidental

or systematic?
- What is the relation between the results to school report marks, parallel exercises, class exercises?
- Have the tested partial areas of subjects and competencies been appropriately dealt with in the classroom?
- Are there special "dimensions" (with the written exercises in German respectively English) with especially striking results? How can they be interpreted?
- Which consequences have to be brought up in the subject group, which in the subject conference?

Sustainability resulting from feedback procedures appears less at the level of individual teachers but more at the level of subject committees. From the experiences of the reception study to QuaSUM (cf. Peek, 2004) and on the background of the published results to conditions for success with school development processes (cf. Rolff, 2001) subject conferences and subject groups at schools have become the primary addressees of results from centralized comparative tests - with the conviction that these committees can first and foremost produce data based development of teaching and learning in the subjects. In this sense, sustainability for the development in schools and in the classroom from feedback-strategies can especially also be expected in schools where forms of working together cooperatively have already been established.

Evaluation at the level of a subject group (teachers of a year(grade) and the subject conference

- exchange of experiences with the preparation, execution and evaluation of proficiency tests
- screening of the results of classes: are there similar patterns in all classes? Are the results within the range of one's own expectations?
- conspicuous factors: are there special "breakaways"? What could be the reasons? How could those conditions (short term, mid term) be changed? Who has to be approached to that end?
- description of possible reasons for deficits with regard to linking contents and method elements of the school curriculum with the stipulations of the basic curriculum
- description of possible reasons for deficits with regard to a possible contents distance of the textbook from the "philosophy" of the basic curriculum

- preliminary work in the subject for the report to the committees/ school inspection; agreements on concrete steps to implement the expected consequences

The process of putting it into practice at schools is followed by reporting to the school inspection in which the school has to present its work within a given report matrix (see www.learnline.de/ angebote/lernstand9/umgang_aufsicht.html). This offers a chance to the school inspection to get an insight into the specific situations of the schools, and possibly ask critical questions and with regard to a more general evaluation also identify schools with special needs.

In addition, within the framework of quality analysis respectively school inspection the results of proficiency tests as well as their use at schools will be integrated in the analysis.

Positive effects, however, do not come about automatically. How far proficiency tests can give positive impulses to schools and can contribute to further developing didactic and educational aims will decisively depend on the quality of the test items, on the debate and acceptance of proficiency tests amongst the teaching staffs and especially how the results are treated in the schools as well as by parties involved in educational policies and politics. It is absolutely necessary that schools can use the exercises (connection to the culture of exercises in schools and the reality of teaching and learning). Exercises must not counteract the philosophy and intentions of a subject (they must not become mere exercise formats in the classroom). Teachers need good evaluation and judgement advice (typical error patterns, judgement criteria, levels of difficulty and demands for sectors etc). There must be success in establishing proficiency tests in such a way that schools stand behind them and that their potential for the development of schools and teaching and learning can unfold itself.

Many questions are still open how schools will deal with externally collected evaluation data. The first explorative reception studies show that the topic of “external evaluation” will naturally have to become part of teacher training and the training of administrations. The embedding of external evaluation in local school development processes can only succeed in the mid and long term if teachers can competently deal with the results of proficiency studies and

centralized comparative exercises, which means if they can use their potential but also recognize their limits. Sound evaluation studies, which examine the effect of proficiency studies and comparative exercises on development processes in schools and in teaching and learning in the participating schools - the pedagogical use of empirically collected proficiency data with regard to reception, reflection and action (cf. Hosenfeld, 2005) - are not yet available at the moment.

References

Arnold, K.-H. (1999). *Fairneß bei Schulsystemvergleichen: Diagnostische Kompetenzen von Schulleistungsstudien für die unterrichtliche Leistungsbewertung und binnenschulische Evaluation*. Münster.

Dobbelstein, P. & Peek, R. (2005). Von der Bestandsaufnahme zur Förderung. Diagnostische Potenziale von Lernstandserhebungen und die Verbindung zur gezielten Förderung von Schülerinnen und Schülern, in *Forum Schule. Magazin für Lehrerinnen und Lehrer,* Heft 1/2005, S. 24-25.

Dobbelstein, P., Peek, R. & Schmalor, H. (2004). An Ergebnissen orientieren. Ein Paradigma für alle Fächer und Lernbereiche? in *Forum Schule. Magazin für Lehrerinnen und Lehrer*, Heft 1/2004, S. 18-26.

Helmke, A. & Hosenfeld, I. (2005a). Fehleranalyse als Gegenstand automatisierter Ergebnisrückmeldungen in Vergleichsarbeiten, in *A. Helmke & R. S. Jäger (Hrsg.): Fehler- und Notenkultur* (Empirische Pädagogik, Themenheft). Landau (in print).

Helmke, A. & Hosenfeld, I. (2005b). *VERA 2004: Erste Ergebnisse des Ländervergleichs.* http://www.uni-landau.de/vera/downloads/Laenderkurzbericht.pdf.

Helmke, A. (2005). *Von der Evaluation zur Innovation: Nutzbarmachung von Vergleichs-, Orientierungs- und Parallelarbeiten und Lernstandserhebungen für die Unterrichtsentwicklung*. Seelze (in print).

Hosenfeld, I. (2005). Rezeption - Reflexion - Aktion. Wie lassen sich Lernstandserhebungen und Vergleichsarbeiten pädagogisch nutzen? in *Becker, G., Bremerich-Vos, A., Demmer, M., Maag-Merki, K., Priebe, B., Schwippert, K., Stäubel, L. & Tillmann, K.-J. (Hrsg.): Standards, Unterrichten zwischen Kompetenzen, zentralen*

Prüfungen und Vergleichsarbeiten (Friedrichs Jahresheft XXIII 2005), S. 112-114.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. (2003). *Zur Entwicklung nationaler Bildungsstandards - Eine Expertise*. Frankfurt/a.M. Deutsches Institut für Internationale Forschung.

Peek, R. & Dobbelstein, P. (2003). Mehr als Wiegen und Messen. Zentrale Lernstandserhebungen in Nordrhein-Westfalen, in *Forum Schule. Magazin für Lehrerinnen und Lehrer*, Heft 2/2003, S. 14-18.

Peek, R. (2004). Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (QuaSUM) - Klassenbezogene Ergebnisrückmeldungen und ihre Rezeption in Brandenburger Schulen, in *Kohler, B. & Schrader, K.-F. (Hrsg.): Ergebnisrückmeldung und Rezeption. Von der externen Evaluation zur Entwicklung von Schule und Unterricht* (Empirische Pädagogik, 18, (Themenheft)). Landau, S. 82-114.

Rolff, H.-G. (2001). Was bringt die vergleichende Messung von Schulleistungen für die pädagogische Arbeit in Schulen? in *Weinert, F.E. (Hrsg.): Leistungsmessungen in Schulen*. Weinheim und Basel, S. 337-352.

Perceiving the Impact of Inspections on the Improvement of School Quality

Maria da Conceição Castro Ramos, Helder Guerreiro

Introduction

Traditionally, Inspectorates are known for their systematic and very often systemic examination of education systems. And that is so! But the organisations that envisage improvement cannot limit their action to this external look. Looking inside is necessary when there is a clear intention of innovating and bettering internal management and external action. And these are the main purposes of the self-evaluation programme that is the core of this paper.

The Portuguese Inspectorate of Education (IGE) is developing a self-evaluation programme that enables the Inspectorate to be aware of the impact inspections have on schools' quality and on schools' self-evaluation devices. This programme also gives the Inspectorate the necessary feedback to support a reflexive attitude about the quality of the service.

Firstly, the paper makes some general considerations about the mission (and the vision) of the Portuguese Inspectorate of Education in order to introduce the Portuguese inspection system and the overall types/models of action.

Then, it provides information about the self-evaluation programme of IGE, much inspired in the Common Assessment Framework (CAF) model: the reasons underlying its development and the preparatory work, as well as some features of this recently launched programme. We will also provide an overview of the management of inspection activities and the evaluations of impact.

Afterwards, the processes related to school surveys will be focused - the application, the processing and analysis of data and the dissemination of results. Then, we will provide information about the results of two inspection activities that were surveyed.

Finally, in the seventh section we will make some general

considerations about the process and consequences of these surveys.

Annex I to this paper is the standard version of the "Questionnaire of Impact" built up by the Inspectorate in order to gather feedback about the impact of fieldwork activities, whereas Annex II includes summarised information about two implemented surveys and a synthesis of their results.

Mission of the Portuguese Inspectorate of Education

Mission of the Inspectorate: Whatever the focus of school inspections - teaching and learning, organisational procedures, financial management, equity policies - their aim is, on the one hand, to ensure that the orientations established by policy-makers are followed and, on the other, to contribute to the improvement of educational provision. The inspectors' core activities (inspections) are based on norms and guidelines established by decision-makers, which correspond to quality standards to be achieved by education providers.

The Portuguese Inspectorate of Education is much in line with this international trend of assuring the quality of the education system. But IGE also has a word to say about equity policies and practices developed through the different layers of education system - services of the Ministry of Education, schools, the classroom, individuals. Indeed, IGE has the role of surveillance, so that equal opportunities are given to students in their academic life, contributing, thus, to a fairer system.

Through diversified inspection programmes, addressing different layers of the education system, and reporting on the findings, the Inspectorate activity contributes to make the education system accountable. For instance, every report on school inspections is made available to the school community, and the national reports are made public on the Inspectorate's website and copies are delivered to institutional stakeholders.

The Portuguese Inspection System, on the whole, has two general goals - to make the education system accountable; and to induce better practices and higher standards among the school players. It consists of five main programmes:

- *Monitoring*, which focusses on the organisation of class activities, such as curricular management and interaction teacher/student;
- *Control* - that aims at checking the compliance with the norms, in order to safeguard equity and to check the accomplishment of quality standards. Examples of these activities are those focused on the organisation of the school year and national examinations;
- *Audits* - these are analysis-oriented, with regard to conformity, effectiveness, efficiency, pertinence and coherence of management activities. Audits focus on school financial and budgetary administration;
- *Appraisal* is an evaluation programme that focusses on diverse areas and has a very strong component of induction by means of the interaction between the inspector and the school players. This programme comprises inspection activities such as "The effectiveness of basic education" or "The effectiveness of school self-evaluation";
- *Ombudsmanship and Disciplinary proceedings* - this aims at safeguarding the interests of the players and of the education system users. It encompasses the analysis of complaints that can lead to disciplinary proceedings, whenever the seriousness of the situation demands so.

Each programme comprises several activities, as it is above exemplified in the "Appraisal" programme.

The inductive strength of inspections: Although their design, purposes and consequences differ, inspections are expected to contribute to school improvement at three moments: before, during and after they take place.

- Impact prior to the inspection:
  - The mere existence of an Inspectorate of Education contributes to schools' permanent efforts to improve their quality, as the awareness of an external look is always a potential factor of pressure. When the consequences of external evaluations are associated with rewards and punishment, this effect is even amplified.
  - The notice of an inspection is another factor of pressure, though not always with the same intensity - it varies depending on the type and scope of the inspection: control, audit appraisal, etc. A visit of inspectors means that the performance of schools is

made visible to school players and education authorities, at a first stage, and to all stakeholders, next.

- Impact during school inspection:
  - A strong interaction between inspectors and school players takes place at this moment.
  - School players know the resources and the environment of their own school better than anyone else. Very often, they are aware of similarities and differences between their own and the neighbour schools. They are committed in finding solutions that better fit their particular school, as well. However, they tend to ignore some aspects of their own schoolwork and to resist to deep changes in the *status quo*.
  - Inspectors have a general and consistent view of the education system, as they have a permanent contact with far different schools, situations and contexts. Whenever inspectors visit schools, they benchmark them, explicitly or implicitly, against other schools they had visited before. This means that inspectors' activity enables them to learn with every single school, to build up a broad view of the education system and to be acquainted with a multiplicity of solutions to the problems schools face. On the other hand, inspectors' tools are built upon norms and standards that reflect the expectations of decision-makers towards education, in general, and schools, in particular.
  - Meetings between inspectors and school players are very often a confrontation between broad views of the system versus focused views of the school, between standards blind to the contexts and contexts blind to the standards. It is a fruitful and dialectic interaction when inspectors find how useful applying their tools to a specific school is and the school players find the applicability of standards and solutions existing beyond the school walls.
- Impact after school inspection:
  - School players can now use the input brought by inspectors - tools and methods, reports, recommendations - and mould them to the school context, taking into account their adequacy and feasibility.

Besides the inspection system: The external look of the Inspectorate over the education system, which is made operational by means of the inspection system, is complemented by looking inside its activities and performance, which has effects on further external

looks. And when we (Inspectorate) look inside, but through the eyes of the others (schools), there can be independent and consistent views about the effectiveness and quality of our provision. That is why the Inspectorate set up a "Quality Assurance System" programme, called the *Inspectorate's self-evaluation programme*, aimed at producing information that will support internal management and control, whose main source of feedback information are questionnaires applied to inspected schools.

A new approach to self-evaluation

Antecedents

The above-mentioned self-evaluation programme was built upon some needs that were felt by the Inspectorate's leaders and inspectors, and relies on a set of assumptions envisaging improvement and quality control of the service provided by IGE.

Identified needs

Setting up systematic self-evaluation processes emerged from needs that were identified by the Senior Managers of the Inspectorate, namely:

- the need to find out how effective the activity developed by the Inspectorate was with regard to the improvement of the education system;
- the need to find out how satisfied with the Inspectorate's activity school players were;
- the need to find out about the Inspectorate's weaknesses, so that improvement strategies could be introduced;
- and, at last, the need to find out how satisfied with their own job inspectors were.

Conceptual principles

The self-evaluation programme was drawn upon 3 assumptions:

- self-evaluation contributes to improve the performance of organisations;
- the organisations that are not aware of what they are and of what they do, are unable to improve;
- schools can provide the Inspectorate very relevant feedback about its action.

The information gathered can be used as an important basis to internal reflection, in order to make the organisation more efficient and effective.

The process of building up a self-evaluation programme

It lasted about one year between launching the initiative and the implementation of the first activity developed in the scope of the *self-evaluation programme.* :

- During that year, it was discussed whether to materialise a self-evaluation programme or not.
- The model that would inspire our own self-evaluation and that would better fit our purposes was also discussed.
- The Inspectorate's self-evaluation matrix was outlined.
- The most adequate methodologies and instruments, according to the foci of the evaluation and the target respondents, were chosen.
- Among the inspectors, the purposes, components and progresses of the self-evaluation programme were disseminated.

The Senior Chief Inspector appointed a working team and assigned its members the mission of designing and launching the self-evaluation programme. Its members were:

- the Deputy Senior Chief Inspector, making thus evident the commitment of the Senior Management in the success of the initiative;
- the five Regional Chief Inspectors, who worked as interlocutors and brought to the working team the sensibility of the regions and of field inspectors;
- a steering team with two members, who were assigned the tasks of
  - preparing discussion papers for the meetings, with proposals,
  - taking part in the discussions,
  - collecting new proposals,
  - incorporating those proposals in the former papers, whenever agreed,
  - designing and making operational the dissemination process,
  - developing some research, in order to make the proposals more consistent and in line with the procedures of other public services.

Learning from others' experiences - the SICI survey

Setting up a self-evaluation device in Public Administration organisations still is an innovative and challenging deed. And if you

intend to use the *customers*' opinions in this process it is even risky. Therefore, it sounded wise to consult similar public organisations about their own self-evaluation devices and to prepare the Inspectorate staff for innovation.

This was why the steering team produced a newsletter and surveyed European Inspectorates about their own self-evaluation processes.

The newsletter was delivered among inspectors and other senior staff. It informed about the progresses made in the self-evaluation programme and about the Common Assessment Framework, presented self-evaluation processes developed by other European public services and informed about events related to self-evaluation.

The questionnaire was applied to the representatives of European Inspectorates in the SICI (Standing International Conference of the Inspectorates) Extraordinary General Assembly, held in Sesimbra, Portugal, in October 2004. The representatives, who usually are senior managers of the Inspectorates, were requested to answer a questionnaire about internal evaluation procedures within their own organisations. The data material collected by means of this questionnaire was intended to give an overview of the main trends on self-evaluation in the European Inspectorates.

Thirteen questionnaires, representing the same number of Inspectorates, were returned and only one reported the inexistence of self-evaluation procedures. The other twelve provided the Portuguese Inspectorate with useful information about European trends, thus shedding light on the path towards a quality assurance system. Below, we sum up some of the findings:

- most Inspectorates develop some kind of self-evaluation;
- the majority mentioned that self-evaluation is part of the organisation's culture, whereas one third referred it is compulsory;
- about half of the Inspectorates include Senior or Middle Managers in the evaluation team;
- the main focus of the self-evaluations undertaken by the Inspectorates are spread in a balanced way among "Performance", "Objectives achieved", "Human Resources management" and "Quality of the Provision";
- surveys are the preferred evaluation methodology, although they

are sometimes complemented by other techniques;
- among the main expected impacts we can find the short- and long-term action plans, as well as the re-definition of objectives.

The results of this surveyed were spread by means of the newsletter, and they strengthened the Senior Managers' purpose of setting up in a short period of time the self-evaluation programme.

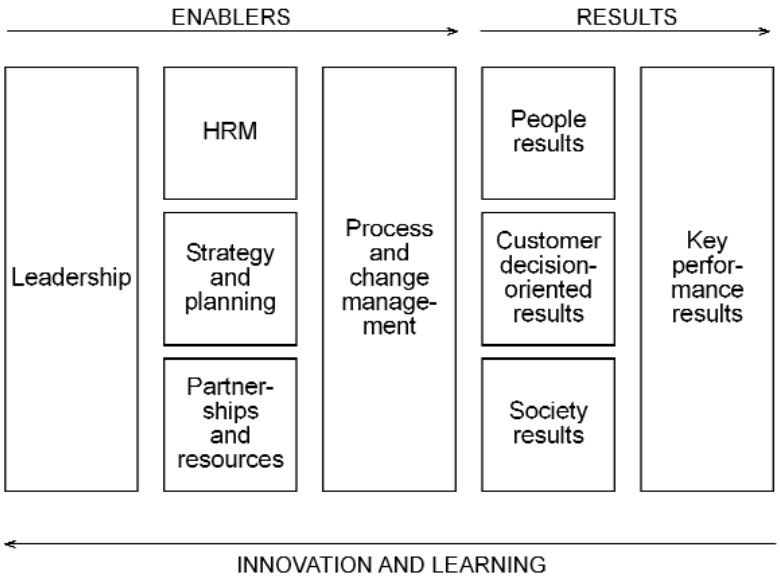
The Inspectorate's self-evaluation programme

The programme:

How does the Portuguese Inspectorate find out about the contribution of inspections to changes and improvement in schools and whether schools profit from inspections to develop their own quality assurance devices?

In the past two years, the self-evaluation programme of the Inspectorate has been introduced gradually. It follows very closely the Common Assessment Framework (CAF model)[1] , whose indicators are organised in two groups of criteria: enablers and results.

Figure 1 - the CAF model



[1] Source: European Institute of Public Administration, The Common Assessment Framework - improving an organisation through self-assessment, 2002

So far, the self-evaluation programme has made use of two evaluation techniques: a staff panel to evaluate some enablers, a questionnaire of impact to evaluate the effects on schools. Until the end of 2006, a questionnaire about staff satisfaction will be applied.

Questionnaire*[2] *of impact

Bearing in mind the sub-criteria underpinning the indicator "Customer decision-oriented results", the Inspectorate built up a questionnaire that is now being used to survey schools. The survey aims at examining the impact of inspections - what is done, how it is done and its effectiveness. By means of this process, the Inspectorate gathers valuable information both concerning the changes brought about by inspections, as well as feedback that will help to revise tools, procedures and even the conceptual framework of the evaluated inspection activity.

The questionnaire has a standard format with standard questions and it is divided into three sections according to the moment the effect is expected: before the inspection (preparatory phase), during (data collection and induction phases) and after (the effects beyond the inspection).

The questions applied are basically the same. The inspector who coordinates the surveyed inspection analyses the applicability of all the questions and if some specific terminology should replace the standard one.
Each part of the questionnaire has three types of questions:

- Agreement scale, regarding certain statements. The scale ranges from level 1 (*I completely disagree*) to level 5 (*I completely agree*);
- Multiple choice (pairs of opposite adjectives);
- Open question (remarks).

Foci of the survey:

- Phase prior to the inspection:
  - Inspection notice
  - Information about proceedings
  - Readiness of the Inspectorate to make clear some doubts

[2] see Annex 1

- Knowledge of inspection instruments.

- During the inspection:
  - Diversity and amount of analysed aspects
  - Feedback provided by inspectors
  - Cordiality of inspectors
  - Report and/or recommendations
  - Correspondence between the report and the remarks made by inspectors made in the course of inspection.
- After the inspection:
  - The recognition of school work by inspectors
  - The contribution of inspection to school improvement
  - The overall relationship between the school and the Inspectorate
  - The usefulness of inspectors' instruments to school work (planning, self-evaluation, guidelines, etc.)

About management and roles - the inspection activities and the evaluations of impact

All activities undertaken by the Inspectorate have a management structure behind them. Below, we describe the different managerial steps within each inspection activity and within each self-evaluation survey.

Organisational structure of the inspection activities
There are four players in the organisational structure of each inspection activity undertaken in the field: the Senior Management (SM), the Coordinator of the Activity (CA), the Regional Interlocutors (RI), and the Field Inspectors (FI).

ACTION	PLAYER
Decision about launching or revising an inspection activity	SM
Deciding on sampling design	SM
Setting guidelines for the inspection	SM
Designing the inspection activity with regard to the guidelines	CA

ACTION	PLAYER
Preparing the inspectors' handbooks and other instruments	CA
Validating the proposed inspection design and instruments	SA
Cooperating with the Coordinator of Activity in the design of the inspection and in the preparation of the handbook	RI
Bringing input to the Regional Services of the Inspectorate	RI
Undertaking inspections	RI[3] , FI
Writing the School Report	RI, FI
Giving feedback about the fieldwork to the Coordinator	RI
Re-designing the inspection activity and revising instruments	CA
Validating changes	SM

Organisational structure of the evaluation of an inspection activity
There are four players involved, who have a complementary role: the Senior Management (SM), the Coordinator of the surveyed Activity (CA), the Steering Team (ST) of the self-evaluation activity, and School Heads (SH).

ACTION	PLAYER
Deciding the scope of the survey (i.e. the activities that will be surveyed)	SM
Proposing a self-evaluation plan	ST
Agreeing on the self-evaluation plan proposed by the Steering Team	SM
Submitting the questionnaire to the appreciation of the Coordinator of Activity to check its applicability	ST
Providing the steering team with feedback about the applicability of questions and terminology of the standard questionnaire	CA

[3] Regional interlocutors are field inspectors, too.

ACTION	PLAYER
Providing the steering team with the list of schools that were inspected in the scope of that activity	CA
Analysing and deciding upon the recommendations of the CA	ST
Preparing paper and electronic versions of the questionnaire and sending them to the schools sampled	ST
Performing the role of "contact person" and providing schools any additional information required by them	ST
Preparing a spreadsheet, where data from school answers will be inserted	ST
Filling out the questionnaires and sending them to the Steering Team	SH
Processing data, analysing the results, reporting and suggesting areas for improvement	ST
Improving the design in the inspectors' handbook, in order to overcome weaknesses identified in the survey report	CA
Planning the dissemination process, taking into account the different target public and their particular needs	ST
Deciding about dissemination procedures and resources to be allocated	SM

The School Surveys - an external look

Each survey requires a number of tasks that we now try to describe shortly.

Procedures for questionnaire application

- Schools are surveyed three months after the inspection has taken place.
- The steering team sends the questionnaire to all the schools inspected in the scope of one activity;
- the surveys are sent to the schools both in paper and electronic versions, attached to a cover letter written by the Senior Chief Inspector;
- the responses are given on a volunteer basis, and schools can make an option between using the paper or the electronic versions;
- schools are given approximately 2 weeks to fill out and to send back the questionnaires.

Respondents
Headteachers. They can ask for the support of school staff that interacted with inspectors in the course of the evaluated inspection.

Processing and analysing data
Data material is processed in order to provide useful information about strengths and weaknesses of the inspection activity and to support the reflexive process that will come afterwards.

After having inserted the collected data in the spreadsheet, the Steering Team will analyse and report on the following issues:

Questionnaire, in general
- response rates;
- most frequent non-responses.

Agreement scale - statements
- mean level in each section - before, during and after;
- distribution of the schools in intervals, according to the mean level they obtained in each section of the questionnaire;
- the mean and the mode in each question;
- the statements that scored below the mean of the section.

Multiple-choice adjectives
- total number of chosen adjectives with a positive connotation;
- more frequently chosen adjectives with positive connotation;
- least chosen adjective with positive connotation;
- most chosen adjectives with negative connotation.

Remarks
- number of remarks made by respondents, per questionnaire section;
- number of remarks per category (for example, criticism, praise, suggestions).

Interpretation of data
Some special attention is given to signals that come from those statements whose score is lower or when some negative remarks and adjectives with negative connotation appear with regularity. They are contrasting spots, when the answers, in general, show that inspection activities that have been much appreciated by schools.

Dissemination procedures

The dissemination of the results is targeted mainly to the Coordinator of the inspection Activity, the Regional Interlocutors, the Field Inspectors, and the inspected schools. Besides informing about the results, the dissemination process is meant to have impacts:

- the coordinators of the activities are inspectors working in the Central Services of the Inspectorate and are entitled to introduce changes in some conceptual and operational aspects of the activity they are in charge of (inspectors' handbook, fieldwork procedures, organisational procedures, etc.). The self-evaluation report provides them with relevant feedback about the weaknesses to be overcome and strengths;
- the Regional Interlocutors ensure the link between the Central Services and the Field Inspectors, who are placed in the Regional Delegations. The feedback report provided to Regional Interlocutor will be transmitted to other inspectors involved in the activity and together they will analyse the results of the report, the recommendations that were made, and they can propose the coordinator of the activity changes that will contribute to improve the activity;
- the Field Inspectors involved in the fieldwork receive a copy of the report, as well. Besides the joint analysis with the interlocutor inspector, there is the individual use of information in self-reflexive processes and the individual inspector has autonomy enough to change some aspects of the interaction he/she establishes with the school players;
- The surveyed schools receive a copy of the report, too. Therefore, they are enabled to compare their own answers (which are kept confidential) with those reported and to evaluate whether the quality and the effects of inspection developed in that school is aligned with the general trend. To a certain extent, the "self-evaluation report" complements the "inspection report" schools receive after each inspection, which points out the weaknesses found and gives recommendations to improve performance, and that is usually discussed in the school management boards.

Survey results

General information about the 2005/2006 surveys
By the end of the school year 2005/2006 four surveys will have been undertaken. One inspection activity per programme, as stated below:

• ***Monitoring***: Preparation of students to lower secondary education examinations and the organisation of the examinations.	***Status***: complete and reported;
• ***Audit***: Social support to students.	***Status:*** complete and reported;
• ***Control***: Organisation of the School year.	***Status:*** complete. A report is now being produced;
• ***Appraisal***: The Effectiveness of Basic Education.	***Status:*** still underway.

Therefore, this paper will focus exclusively on the results of the surveys referring to the already completed and reported monitoring and audit activities.

Surveyed inspection activity: Monitoring the preparation of students to lower secondary education examinations, which has also a smaller component of Control.

- Context: First time that national examinations in regular lower secondary education (ISCED 2) took place.
- Objectives of the inspection activity:
  It is a two-stage inspection - Monitoring and Control - with specific objectives defined to each stage.

- The first stage - Monitoring - is developed in the course of the school year and aims at:
  - monitoring the preparation of students to examinations (Math and Portuguese): curriculum management and measures to prevent dropouts;
  - reinforcing the curricular coordination among the diverse pedagogic structures within schools;
  - verifying the way the curriculum is managed and whether the syllabuses are fully taught;
  - ensuring that schools have launched the necessary strategies to promote students' attainment and to prevent dropouts.
- The second stage - Control - is developed in the course of the examinations and aims at:
  - supervising the examinations process, in order to ensure that confidentiality and equity are guaranteed;
  - evaluating the adequacy of procedures and resources made available.

- Key Findings[4]:
  There was a very positive general reaction towards this inspection at the 3 moments of its development. However, a more fine-grained analysis of each section enables to identify weaker aspects, which need to be improved:

  before inspection
  - the content of the inspection notice, which must be more informative;
  - the clarity of the inspection notice;
  - the short time between the notice and the inspection.

  during inspection
  - the complexity of the inspections' instruments
  - the communication: some inspectors did not make clear the purpose of the observation and document analysis they were undertaking.

  after inspection
  - the time it takes to provide schools with feedback;
  - a formal recognition of the work developed by schools.

[4] Find details about the characteristics of the surveyed inspection activities and a summary a summary of the results in Annex II - A.

Surveyed inspection activity: Audit of the programme of Social support to students

- Context: it is a regular inspection, which runs on a yearly basis
- Objectives of the inspection activity:
  - to verify the use made of school autonomy (management and administration) in compulsory and Upper secondary education;
  - to evaluate the adequacy of human resources and organisational mechanisms;
  - to audit payments, by checking their conformity with legislation
  - to promote effective and efficient resources management
- Key findings[5]:
  There was a very positive reaction from schools, in general, mainly towards the interaction between inspectors and school players. However, a more fine-grained analysis of each section enables to identify weaker aspects, which need to be improved:

  before inspection
  - the clarity of the inspection notice;
  - the short time between the notice and the inspection.

  during inspection
  - the complexity of the inspection instruments;
  - inspectors did not make so clear to schools the pertinence of all observations and recommendations.

  after inspection
  - the recognition of schools' efforts to meet the standards.

Cross-analysis of the two surveys:

A cross-analysis of the evaluation of the two inspection activities, remarks included, led to the following conclusions:

- the results are much alike, although the evaluated activities have a different nature, and the schools surveyed are not the same;

[5] Find details about the characteristics of the surveyed inspection activities and a summary a summary of the results in Annex II - B.

- school players find inspections very positive and, in general, they are welcomed by them;
- the inspections were valued by schools - they underlined mostly their usefulness, relevance and opportunity;
- some weaknesses were pointed out:
- (before) the short notice and the scarce information about the inspection provided to schools, prior to the inspection;
- (during) the complexity of the instruments used by the inspectors, as well as the relevance and pertinence of some of the inspectors procedures that were not understood by school interlocutors;
- (after) the time schools wait before they receive a formal feedback of inspections.

Consequences of the questionnaire on impact

These questionnaires of impact were something innovative in the Portuguese educational administration. Therefore, when it was introduced the first reaction was rather negative:

- schools were surprised with such a questionnaire. Despite the purposes mentioned in the cover letter, school heads regarded it with suspicion and contacted the Steering Team in order to confirm the objectives. Some dared to congratulate the Inspectorate for the initiative itself;
- inspectors, in general, were afraid of public exposure of their own work and of the consequences of schools evaluating inspections, as it could affect their individual evaluation;
- the inspectors' union complained against this self-evaluation process. Trade union complaints had a stronger meaning, as this initiative was coincident in time with a new framework for the evaluation of inspectors;
- coordinators of Activity always asked why the Board of Chief Inspectors had chosen their activity to be evaluated. They also feared the school answers to the questionnaires.

About ten months after the first questionnaire was applied, and at a moment the fourth is underway, there was a clear shift in the reactions of the different players. It is undoubted that the very positive answers contributed to overcome suspicion. But now there is a strong trust on the purposes underlying the surveys:

- it is no longer one of the major concerns of the Union, although they are attentive to any negative consequences that may arise from the schools' answers;
- the Coordinators of Activity show interest in the schools' answers and intend to consider them to revise the activities they are leading;
- schools do not suspect the actual purposes of the survey any longer and their phone calls mainly focus on technical aspects of the questionnaire. They also show interest in having a copy of the report on the results. More and more, the steering team, that is processing and analysing data, realises that schools are writing more remarks in a more open manner.

After having overcome difficult and diverse obstacles, mainly originating from inside of the organisation, there is a firm belief that the questionnaires of impact succeeded and are now rooted in the culture of the Inspectorate. Inspectors, inspector Coordinators of Activity, the Senior Management and others, are all interested in having a consistent view of the impact of inspections.

On the other hand, schools have lost their initial fear and shyness, and are aware that by means of their answers, they have a word to say about the quality of inspections and about their influence on the quality of schools. They are also aware that they can make a contribution to improve the quality of the service provided by the Portuguese Inspectorate of Education.

ANNEX I
Questionnaire of impact

Evaluation of Inspections	Activity: • School/School Grouping: _____ • Code: _____ • Date of inspection: _____

BRIEF NOTE:

In the scope of a self-evaluation process, which intends to promote internal reflection and to improve its performance, the Portuguese Inspectorate of Education (IGE) intends to evaluate the way its activities are being run and its consequences for schools and school groupings. The answers to this questionnaire report to the inspection above identified, are a valuable contribution to that process.

Questions are organized into 3 sections: “Before the inspection”, “During the inspection” and “After the inspection”.

Answers must be given according to a 5 level-scale:

- level 5: I completely agree
- level 4: I generally agree
- level 3: I agree more than I disagree
- level 2: I disagree more than I agree
- level 1: I completely disagree

Nr	Questions	Level				
		5	4	3	2	1

A. Before the inspection

1	The information received before the inspection was clear about the objectives to achieve.	☐	☐	☐	☐	☐
2	This notice was received in due time.	☐	☐	☐	☐	☐
3	The inspector / team of inspectors in charge of the visit made a previous contact with school staff directly involved.	☐	☐	☐	☐	☐
4	The school was informed about the procedures that would be followed, before the beginning of the inspection.	☐	☐	☐	☐	☐
5	The contacts between the school and IGE, asking for explanations about the inspection, were answered quickly and kindly.	☐	☐	☐	☐	☐
6	Those contacts were useful for the requested explanations.	☐	☐	☐	☐	☐
7	The Headteacher was informed about the content of the inspectors' handbooks.	☐	☐	☐	☐	☐
8	School responsibles knew about the handbooks used to support the inspection.	☐	☐	☐	☐	☐

Choose the three adjectives that better qualify the preparation of the inspection.

● clear	❑	● confused	❑
● necessary	❑	● unnecessary	❑
● easy	❑	● difficult	❑
● relevant	❑	● irrelevant	❑
● opportune	❑	● inopportune	❑
● useful	❑	● useless	❑

If you want to make any comment you can use this box. Please synthesize it as possible to facilitate its processing.

Nr	Questions	Level				
		5	4	3	2	1

B. During the inspection

1	Taking into account the objectives of the inspection, the staff that interacted with the inspector(s) considered pertinent the amount and variety of items observed and analysed.	☐	☐	☐	☐	☐
2	The requests for information and the occasional change of school routines caused by the inspection were considered pertinent.	☐	☐	☐	☐	☐
3	The inspector / team of inspectors gave feedback to each person directly involved, in the analysed items, including comments and suggestions.	☐	☐	☐	☐	☐
4	The inspection was developed with cordiality.	☐	☐	☐	☐	☐
5	The inspector / team of inspectors showed openness with regard to school characteristics and difficulties.	☐	☐	☐	☐	☐
6	There were important items the inspector / team of inspectors did not consider.	☐	☐	☐	☐	☐
7	At the end of the inspection, the school received a report or useful recommendations to improve its organisation and performance.	☐	☐	☐	☐	☐
8	The report / information was in line with the remarks made during and at the end of the inspection.	☐	☐	☐	☐	☐

9 Choose the three adjectives that better qualify the preparation of the inspection.

● complete ❑	● useful ❑	● stressing ❑
● focused ❑	● opportune ❑	● partial ❑
● simple ❑	● transparent ❑	● distant ❑
● calm ❑	● incomplete ❑	● useless ❑

9	● impartial ❑	● diffused ❑	● inopportune ❑
	● engaging ❑	● complex ❑	● confusing ❑
	Other: _____		

Nr	Questions	Level				
		5	4	3	2	1

C. After the inspection

1	The school found that their work was recognised.	☐	☐	☐	☐	☐
2	The inspection contributed to improve school performance.	☐	☐	☐	☐	☐
3	The inspection allowed a better relationship between the school and IGE	☐	☐	☐	☐	☐
4	The inspectors' handbooks have been useful in other organisational contexts (planning, self-evaluation, information procedures, guidance…).	☐	☐	☐	☐	☐
5	The Headteacher's team considered this action useful for their work.	☐	☐	☐	☐	☐
6	On the whole, the advantages of the inspection made irrelevant any occasional negative aspects.	☐	☐	☐	☐	☐

7 Choose the three adjectives that better qualify the preparation of the action.

● pertinent	❑	● inadequate	❑
● gratifying	❑	● disappointing	❑
● stimulating	❑	● not stimulating	❑
● relevant	❑	● irrelevant	❑
● opportune	❑	● inopportune	❑

Other: _____

If you want to make any comment you can use this box. Please synthesize it as possible to facilitate its processing.

ANNEX II
Surveyed inspection activities: characteristics and survey results

Surveyed inspection activity: Monitoring the Preparation of students to lower secondary education examinations.

Characteristics of the inspection:

Scope:

- ISCED 2 schools, where national examinations took place;
- 68 municipalities identified with high rates of school failure and dropouts were included in the sample.

Methodology:

- 1 inspector/3 days (monitoring) + 1 inspector/1 day (control)
- school visits
- document analysis and interviews with teachers and the management staff
- observation - organisational mechanisms
- analyses of the School Development Plan and data provided by the schools
- report on the weaknesses and strengths
- analysis of the school development plan

The inspectors' handbook - comprises the following components:

- form with the characterisation of the school;
- form with the appraisal of the pedagogic school structures performance;
- checklist with organisational procedures
- grids with standard recommendations

Evaluation results of the self-evaluation survey (questionnaire of impact)

- surveyed schools - 151 (both individual schools and school groupings)
- respondent rate - 83.4%
- section means (1 to 5 scale)
  - before - 4.2
  - during - 4.5
  - after - 4.3

- preferred adjectives
  - before: useful (96/126)[6], necessary (77/126), opportune (71/126); confusing [negative connotation] (5/126)
  - during: useful (71/126), participative (70/126), opportune (52/126); disperse [negative connotation] (4/126)
  - after: pertinent (106/126); opportune (105/126); relevant (78/126); inopportune [negative connotation] (7/126).
- total of adjectives with negative connotation: 3.5%.
- remarks: 64 remarks were collected and they were distributed among the following categories:
  - 19 praised the action of the inspectorate, in general, or of the inspectors, in particular;
  - 29 were critical about the inspectors' performance or about the design of the inspection;
  - 7 included recommendations, in order to improve inspections and their impact;
  - 9 mentioned other aspects that had no direct relation with the inspection.

Surveyed inspection activity: Audit of the programme of Social support to students

Characteristics of the inspection

Scope:

- 54 schools
- administrative and financial procedures in areas with more risks

Methodology:

- 2 inspectors/5 days will conduct interviews and analyse documents + 2 days for the audit report + follow up inspection to verify the implementation of the recommendations;
- previous meeting with the headteacher to explain the inspection procedures and the documentation that is going to be used
- document analysis and interviews;
- audit report.

[6] In brackets: the number of choices/the maximal number of possible choices.

Inspectors' handbooks - they are divided into two parts:

- characterisation of the school
- financial management (characterisation of the situation, the efficiency and quality of services, the management of the school budget and the inner control systems and procedures).

Evaluation results of the self-evaluation survey (questionnaire of impact)

- surveyed schools - 54 (both individual schools and school groupings)
- respondent rate - 72.2%
- section means
  - before - 4.2
  - during - 4.6
  - after - 4.4
- preferred adjectives
  - before: useful (33/39), necessary (22/39), clear 12/39;
  - during: useful (22/39), participative (22/31), transparent/opportune (15/39);
  - after: opportune (36/39), pertinent (35/39), relevant (24/39).
- total of adjectives with negative connotation: 0.9%.
- remarks: 10 remarks were collected and distributed among the following categories:
  - 3 praised the action of the Inspectorate;
  - 5 were critical about the inspectors' performance or about the design of the inspection.

Monitoring the Impact of the *Matura* on Teaching and Learning

Zora-Rutar Ilc

Abstract

In the present study, that combines the qualitative and the quantitative approach, the impact of the *matura* on the didactic aspect of teaching, especially its taxonomic structure, has been monitored. It has been found that the present concept of the *matura* encourages a systematic consolidation of knowledge, which in turn contributes to better retention as well as better organisation of knowledge. There is, however, less emphasis on an independent approach to problem solving and to independent development of ideas. It is, however, very difficult to isolate the impacts of the *matura*, as they evolve from the whole context of the secondary education process in the *gimnazija*.

Introduction

Since the introduction of the present concept of the *matura* the experts as well as the non-professional public have been raising questions that can be divided roughly into two groups: in which way does the *matura* affect the quality of knowledge, and what impact does it have on the workload of students.
Therefore, on the initiative of the National Examination Commission the National Education Institute decided to carry out a complex study that would shed light on the impact of the *matura* on teaching from different angles, including monitoring what actually happens in the classroom.

For this purpose the study was designed on:

- identifying the taxonomic structure of teaching and learning in the classroom;
- identifying the retention of knowledge acquired through education in the *gimnazija* assessed with the *matura* as a final exam;
- identifying the psychosocial aspects of the impact of the *matura* (especially the nature of the perceived students' feelings of the

burden caused by the *matura*).

A part of the study, i.e. the one in which the didactic aspect (the impact of the *matura* on the didactic aspects of teaching and learning and its taxonomic structure) was studied, was designed as a combination of qualitative (with quantitative elements) and quantitative research. The qualitative part sheds light on the actual process (the actual situation in the field), while the quantitative part indicates the trends in the estimations of the process by students and teachers.
The identification of the retention of knowledge and the identification of the psychosocial aspects of the *matura* were the subject of the other parts of the study that are not shown in this contribution. Presented in this article are the findings of the study carried out by the experts of the National Education Institute of Slovenia that identify the didactic and the epistemological aspects.

The Purpose of the Study

By the study we desired to find out in which way the *matura* influences the didactic aspect of teaching and learning and the related taxonomic structure. Our intention was, in cooperation with the National Examination Commission and other experts, to propose changes that - based on the findings of the study - would prove necessary.

For this purpose we tried to find out:

- What (according to the estimates of teachers and students) was the taxonomic structure of the preparation for the *matura*, and what (according to the observations in the classroom) was the taxonomic structure of teaching in the 4th, final year. Our interest was in the status quo in general and in individual subjects.
- In addition to this we were interested in getting a broader insight into the processes, the course of events in teaching and learning. For this reason we observed some didactic areas especially emphasised by the starting points of the reform of the curriculum such as: the importance of process goals, establishing links within subjects and disciplines and between them, taking account of the prior learning of students, problem-solving orientation of teaching

and learning, the active role of students, clarifying instances of incomplete understanding, developing independent and critical judgement, taking account of the interests of students … and potential influence of the *matura* on these areas.
- The estimations of teachers and students regarding the aspects that (according to their opinion) would get more/less attention had there not been the present concept of the *matura*.
- Their estimations of to what extent, according to their opinion, creativity and critical thinking, the interests of individual students and differences between individual students in their capacities of apprehension, learning and expressing themselves, are taken into account in the course of the preparation for the *matura*.

From the methodological point of view a combination of the quantitative and the qualitative approach with quantitative elements was chosen. With the quantitative approach the prevailing trends are indicated, while with the qualitative one the trends are explained in context and at the same time a broad selection of possible usages developed under certain conditions (see e.g. Cohen and Manion 1992, Wragg 1994 or Wajnryb 1998) is described. This means even an individual instance of observation or a statement by an individual teacher, different from the majority, but indicating a potential reaction in a real situation, can tell very much. In this sense this kind of signals are also relevant, not only statistically significant trends. However, the findings of the qualitative part of the study cannot be generalised for the population as a whole - they can only be used to shed more light on the quantitative part. In this context, the findings of the reports on different subjects are only valid for the observed lessons carried out by the monitored sample of teachers and students and not for the subject in general.

Theoretical Background

With regard to the purpose of the research theoretical starting points are highlighted according to which the taxonomic structure of examination items was analysed. Also exposed is a theoretical framework for some other, broader areas that represent the context of teaching and learning and in which the possible impact of the *matura* is investigated.

Broader context

The broader areas were defined according to the starting points of the curricular reform (Izhodišca kurikularne prenove, 1996, p 15), formulated by the National Curriculum Council (NKS):

"to increase the retention of the acquired knowledge

- by developing different strategies of thinking …,
- with problem-solving approach to teaching and with other (active) methods of learning and teaching,
- with integration of knowledge and establishing links,
- by linking theory to practice,
- by taking into consideration the relationship between cognitive, motivational and affective processes and factors,
- by learning to learn …,

… to develop the capacities of independent, creative and critical thinking and judgement; educating for sufficiently self-confident coping with the problems of everyday life and for solving those problems."

Starting from the points above the following areas were defined as the broadest context for monitoring:

- setting the goals of learning;
- establishing links within the teaching contents, between subjects, with own experiences;
- highlighting the practical value of the contents taught;
- the problem solving orientation of the discussion;
- the activity of the students;
- clarifying the instances of incomplete understanding;
- assessing and marking the knowledge;
- taking into consideration the interests, wishes, experiencing, emotions of students in teaching and learning.

To avoid possible misunderstanding as for problem solving orientation of teaching and learning, the concept was defined in accordance with some modern concepts of knowledge (such as Voutilanen et al. 1990, Marzano 1989, Gifford and O'Connor 1992, Baron and Sternberg 1987 and Marentic Pozarnik 1999).

For the operational use (in the training of observers) the following description was suggested for the detection of problem solving orientation in concrete situations:
The contents are dealt with on higher taxonomic levels, in-depth and systematically, according to clearly evident steps, presumed by systematic discussion and procedures of problem solving. The decisive criterion in defining the category is the discussion of the contents aimed at involvement of students, the approach that encourages them to follow it and stimulates them to actively process the contents. For proper judgement it is necessary to consider the context, i.e. to what extent are the students taken into consideration in the discussion, to what extent can they make the content meaningful for themselves and participate in the discussion (though perhaps only mentally). It is demonstrated through teachers' comments and procedures as well as through questions addressed to students and through the discussion of problems by the students and their formulation of questions in the problem situation.

In accordance with the theories mentioned above the problem solving orientation of the discussion was divided into three subcategories:

- pointing out dilemmas and opening problems;
- discussing the problems, outlining the phases of solving the problems or/and proceeding through them (designing the pathways for solving the problems, analysing, synthesising, proving);
- presenting the practical value, linking to real life practices, actualising;
- pointing out the essence, relevant ideas, conclusions, resolutions, comprehension;
- critical judgement and evaluation.

Recognised as a discussion of contents not oriented to problem solving is the presentation, transmission of facts or finite knowledge predominantly on lower taxonomic levels such as mere enumerating or summarising of facts, categories and classifications, terms and symbols (Bloom 1956, Bloom et al. 1981, Nitko 1996). Categorised as a discussion not oriented to problem solving are the ways of dealing with facts and information that are prevailingly results of conventions and tradition and not of research work, such as routine procedures and methods (see also Marentic Pozarnik 1995). In this sense deriving formulae routinely without pointing out possible links

and procedures or even such presentations of interpretation (e.g. of literary works, historic events …) belong to this category (Bloom et al. 1981). Classified into this category can also be such explanation that might be logical and correct, that might even include elements of problem solving orientation, but is not presented to students in a way accessible to them[1] .

Taxonomic structure

The taxonomic structure as one of the most relevant aspects of a possible impact of the *matura* was monitored both quantitatively and qualitatively, analysed was also the taxonomic structure of examination sheets. Due to the differences in the nature of the three ways of monitoring each was assigned its own taxonomic classification.
The classification is based on Bloom's taxonomy, which in the analysis of the examination sheets was used in its original version, while for the quantitative and for the qualitative part of the research it was adapted to the specific characteristics of each approach.[2]

For the **qualitative part**, i.e. monitoring the processes of learning and teaching, the taxonomic structure was detected from the relationship between the presence and absence of problem-solving orientation (see above) and from the relationship between the activity of the teacher and that of the students as well as from the structure of the activities of the students.

Because it involves communication with students and teachers, a less expert terminology was chosen for the **qualitative part**. Bloom's taxonomic categories were translated into empirical categories derived from Bloom's interpretation of individual taxonomic levels (Bloom et al. 1981).

Consolidation of formulae, definitions, procedures, data … was classified as recognition.

The consolidation of explanations and interpretations was classified separately, to distinguish it from independent explanation and/or

[1] Certainly the judgement of accessibility to students is the matter of judgement of an individual observer supported by interviewing the students after the lesson, when they are also asked about this aspect.
[2] The interpretation of individual taxonomy level and its relationship with other concepts was derived from the content interpretation of taxonomy levels in Bloom et al. 1981.

solving problems and independent development of ideas. Summing up explanations and interpretations is namely often confused with independent explaining, substantiating and problem solving, although they belong to two different taxonomic levels and involve qualitatively different mental processes and skills. Consolidating explanations and interpretations can be thought of as belonging to the dividing line between recognising and understanding and routine use.

We also wished to define a clear structure of use, as in the existing classifications it consists of both routine use (use of routine procedures, forms, routine solving of applicative tasks …) as well as more intensive forms of applying knowledge (e.g. searching for useful value of knowledge, actualisation). For this reason separate categories of a routine use of knowledge and searching for a useful value of knowledge were introduced.

The highest taxonomic levels are represented by independent problem solving and development of ideas.

Working Methods and Techiques

In the framework of the quantitative part of the research surveys with questionnaires[3] were used, and in the qualitative[4] part

- observations of lessons,
- structured interviews,
- analysis of material for assessment and marking of knowledge,
- in some subjects also the analysis of anecdote records.

The method of content analysis was used for analysing the examination sheets.
The procedures and methods of monitoring were unified, but also specific requirements and characteristics of individual school subjects were considered. This resulted in adding some specific categories into the schemes of observation (such as working with texts in teaching Slovene or history …). The differences between subjects, and not the

[3] In this we were assisted by the head masters and the school councillors of the schools in the sample, who took care of filling in the questionnaires.
[4] The designing of the instruments was among others based on Drever 1997, Wajnryb 1998 and Cohen and Manion 1992.

least different interests associated with this, were the source of quite a number of problems. In observing classes, and even more so in written reports about it, a wide variance could be perceived, which makes generalisation more difficult.

Difficulties also occurred in the interpretation of categories in the observation scheme, which showed as early as during the training. So e.g. in the initial phase the observers needed a great deal of discussing and aligning in their judgement when teaching was oriented into problem-solving and when not. The initial difficulties originated from subjective notions of knowledge, learning and teaching, as well as from the specifics of individual subjects. The training contributed to diminishing the differences, yet in the end the judgement was left to the individual, though trained, but still subjective observer.
With regard to their characteristics a combination of quantitative and qualitative methods was used in analysing the data. In doing this, the data acquired through quantitative methods were used as a starting point, which were at a later stage confronted and explained with qualitative approaches.

The following was done in the framework of the research:

- on average 3 observations of teaching in the 4th year with 6 to 10 teachers in 10 subjects, all together more than 250 lessons were observed;
- on average 3 interviews with each of the teachers observed (2 interviews about the lesson observed, 1 summative interview about the impact of the *matura*);
- in some subjects anecdotal records of teachers on the processes in 5 consecutive lessons;
- a questionnaire on the impact of the *matura* on a sample of 600 students of the schools observed;
- the questionnaire on the impact of the *matura* on a sample of 80 teachers of the schools observed.

Observation was performed by experts in subjects - advisors of the National Education Institute, trained in 5 whole day workshops on observing classes and performing interviews: video tapes of lessons were analysed, terminology was clarified, exercises taken in filling in observation schemes, and conducting interviews was trained with changing roles and modes of conversation. A preliminary observation

carried out in the first year classes was followed by a whole day reflection. Also a whole day workshop was organised on interpreting and writing reports. All together the observers received over 50 hours of training.
The observers were also actively involved in designing the research and the instruments.
The completed observation lists and observation forms for all the teachers observed are kept in the archives of the observers - the subject advisers. In agreement with the participants the interviews were audio taped and are also kept in archives together with the transcripts and a short record thereof.
The methodology was presented to the members of the National Examination Commission and of the National Subject Commissions and on their request also submitted to some of them in written form. There were no written comments or proposals.
The major part of the methodology and the instruments are designed in such a way, they can be adapted for a broader use in observing classes of upper as well as lower secondary schools and used as a starting point for further monitoring and comparison.
The research was carried out on the sample of 9 *gimnazijas* (3 small schools with up to 4 classes in a year), 3 medium size schools (5-6 classes in each year) and three big schools (more than 7 classes a year). Regional distribution was also respected. 12 different subjects were included in the research.

The Findings

The taxonomic structure of lessons

One of the most important findings of the quantitative and of the qualitative part of the present study refers to the estimates of the taxonomic structure of lessons in preparation for the *matura* (in the quantitative part by the students and teachers, in the qualitative part by the observers).

In the *quantitative part* it was found that - according to the estimations of the students and teachers - consolidation of given explanations and interpretations (26 %) and routine application (26 %) represented approximately half of the activities. They were followed by the consolidation of definitions and formulas (further 20 %), less represented was search for applicative value (14 %) and solving

problems independently or developing ideas (14 %).[5] The structure suggested by the responses of the teachers is similar: consolidation of given explanations and interpretations (26 %) takes the lead, followed by routine application (23 %). Next come consolidation of definitions and formulas and search for applicative value (18 % each) and solving problems independently or developing ideas (15 %).[6]

Table 1: Arithmetic average of the percentage of activities, number and deviation for all subjects as estimated by students

Activity	**Number**	**Arithmetic average**	**Standard deviation**
Consolidation of definitions	561	19,53	13,12
Consolidation of explanations	561	26,16	15,68
Routine application	561	25,98	18,47
Searching for applicative value	561	14,37	9,89
Independent problem-solving	561	14,06	12,14

Table 2: Arithmetic average of the percentage of activities, number and standard deviation for all subjects as estimated by teachers

Activity	**Number**	**Arithmetic average**	**Standard deviation**
Consolidation of definitions	79	18,01	10,81
Consolidation of explanations	79	26,06	16,15
Routine application	79	22,58	14,95
Searching for applicative value	79	18,25	8,57
Independent problem-solving	79	15,09	10,96

[5] The estimation was made by choosing between the description of the categories offered without using the words.

[6] It should be noted here that statistically significant differences were found between schools with consolidation of given explanations and interpretations (the level of statistical significance 0,047) with routine application (0,000) and with search of applicative value (0,004). This means that the taxonomic structure differs from school to school and that it is not the nature of the preparation as such that defines it, but it is also important what strategies have been developed within the school. The influence of school success was also checked into, but it has not proved to be statistically significant on its own.

As a matter of fact there are differences between subjects in their mutual relations in taxonomic level, but the estimate for all is that there is more consolidation of explanation and/or routine problem solving than searching for applicative value and independent problem-solving or developing ideas.

These findings have also been confirmed by the *qualitative part* of the study. Observing learning and teaching in the 4th year shows that knowledge on the highest taxonomic level (e.g. critical judgement) is the least represented and that in a number of subjects lower taxonomic levels take nearly half of the time. In observing the lessons lower taxonomic levels were captured together as *non-problem-solving approach* and the higher levels as *problem-solving approach*. In half of the subjects the relationship between the problem teaching and non-problem teaching is balanced slightly in favour of the problem approach. In two subjects the relationship is considerably in favour of the non-problem approach, in three subjects it is weighed considerably in favour of the problem approach.

Also important is the structure of the problem-solving approach, determined in this way. The observation has shown, that the problem-solving approach concerns the students mainly through the questions set - they are asked questions containing problems they are supposed to think about and answer. The teachers, however, expose problems, provide examples, derive conclusion and judge critically to a much larger extent than the students. Because of the nature of teaching this can be understood in a way, but for developing these skills it is not of minor importance whether they are demonstrated mainly by the teacher himself and the student is only present as a witness, or the students can test their own skills regularly and go through all the phases of solving the problems themselves (certainly with adequate support by the teacher).[7]

The above findings are additionally highlighted and confirmed also by open responses of the students about their preparation for the *matura* in school and at home which in the words of the students indicate that consolidation, repeating, recollection, preparing along the examination sheets etc. prevails.

*

[7] In the concepts of teaching, focused on the teacher it is expected that the majority of activity is on the side of the teacher and the students are mainly expected to react to it. In the concepts of knowledge, however, that foresee a more active role for the students, their activity is explicitly emphasised. They must actively experience all the phases of problem solving and train them regularly. It is by no means enough for the students to witness these activities and that their activities only come as a sample.

The problem of both levels (the quantitative and qualitative) is: to compare them with what? There is no study that would show the situation before the introduction of the present concept of the *matura*. Neither is there a comparable educational programme that would have the present concept of the *matura* as the final exam (control group). Likewise there are no directly comparable international results or standards concerning the recommended taxonomic structure of lessons. Such recommendations can namely only be a result of particular agreement or systemic solutions and the underlying educational philosophies. There can be no such unified recommendation or standard as to the ratio of taxonomic levels in teaching, simply because there is no unified systemic solution or clear theoretical recommendation.
So the best support for the interpretation of our results is provided by the suggestions of the National Examination Commission on the recommended taxonomic levels in the examination sheets. On average each of the three taxonomic levels: 1. knowledge, 2. understanding and application, and 3. higher taxonomic levels (derived from the original six levels), should cover about a third, though the recommended ratio varies to a certain extent from subject to subject. An additional problem concerning this is that the division used in the present study cannot be directly compared either to the generally recommended three-level division nor to any of the divisions recommended by individual subject commissions.
The following parallels could be drawn between the taxonomic structure used in the quantitative part of the study and the three-level structure recommended by the National Examination Commission:

The consolidation of definitions, formulas, data, procedures … corresponds to knowledge.

The consolidation of given explanations and interpretations also presupposes the level of knowledge, but it can also contain understanding or even elements of application. So it is a more complex category, which is not quite comparable to the categories of the three-level taxonomy, yet knowledge is its prevailing component.

Routine application captures knowledge (knowledge of procedures, formulae …), understanding (when to use what) and application. This application, however, is automatic, routine, so it is different from the application in Bloom's original sense (application in new situations, not merely familiar paradigms).

So a new component was introduced into the insufficiently differentiated notion of application labelled *search for applicative value* representing convincing examples of application of knowledge in new situations.
Thus the use of the three-level scale contains both the routine application and the search for applicative value.
The highest taxonomic levels (the third level in the three-level taxonomy) correspond *to independent problem-solving and independent development of ideas*.

Thus comparing the suggested taxonomic structure for test sheets with that of the preparation for the *matura*, obtained in the quantitative part of the research, we find out that knowledge on the first two levels (1. knowledge and 2. understanding and application) is present in more than two thirds, or that knowledge of the third level (in our terms independent problem-solving and development of ideas) is least represented. Even within application the lower aspects prevail, i.e. routine application on the expense of more convincing application of knowledge in new situations called search for applicative value. During the preparation for the *matura*, according to the opinion of our students and teachers, the knowledge on the highest taxonomic levels is least and also too little represented as compared to the recommended taxonomic structure. In application the routine application prevails to the expense of search for applicative value of knowledge.

In observing lessons the desired taxonomic difference was drawn between problem-solving or non problem-solving discussion. The non-problem discussion corresponds to the lowest taxonomic levels of the three-level scale, i.e. knowledge. The problem solving-discussion on the other hand corresponds to the remaining taxonomic levels, i.e. understanding and analysis, synthesis and evaluation.[8]
As has been said, we have found out that in the activities of the teacher in half of the subjects the problem-solving approach slightly prevailed in the observed classes, which in terms of three-level

[8] In observing classes the taxonomic level is attributed both to the teacher's activities and to the activities of the students, so the classical taxonomic categories cannot be used here. E.g.: when the teacher is developing a problem it is difficult to label his/her activities as “understanding” and “application”. This can only be attributed to questions and suggestions addressed to students. So in observing classes we started from developing problems. Individual phases of problem development were assigned a taxonomic level at a later stage and even then only conditionally.

taxonomy means the presence of lower taxonomy levels in little less than half the time observed.[9] Moreover, the problem-solving nature is prevailingly on the side of the teacher. In some subjects, however, the problem-solving approach has been noticed on average in two thirds of the lessons, so it can be concluded that in these subjects in the classes observed the lowest taxonomic levels are represented only in about a third of the time, the same as suggested.[10]

Thus both from the quantitative as well as from the qualitative findings it can be concluded that during the preparation for the *matura* knowledge on the highest taxonomic levels is least represented. Knowledge belonging to the area of consolidating existent explanation and routine application represents the major part. Also search for applicative value is less common than routine application.[11]
We might arrive at the conclusion that during the preparation for the *matura* the aspect of consolidation of routine explanations and routine application prevails because it is assumed by both teachers and students that in this way they would get prepared best for the *matura*. This is also proved by their statements, upon which we comment later. Whether this perception is realistic can only be examined through an in-depth analysis of the examination sheets themselves (see later), but it is of the same importance that such assumptions exist and that they provoke such effects.

[9] It must be pointed out that we talk here about the results as a whole, derived from the results of individuals, which, however, in some cases are quite different. On average a similar structure prevails with all teachers, but there are exceptions that stand out because one or the other approach prevails. These ratios are described in more detail in the Institute's publication already mentioned.

[10] As for the conclusions of the qualitative part of the study we wish to underline again the small size of the sample and the subjectivity of the observers (it can be assumed that in spite of ample training in observing and in the interpretation their ideas about the problem-solving nature still differ to some extent) do not permit us to generalise the observed situation to those subjects in general. We consider the findings of the qualitative part of the study only as an additional explanation of the quantitative part. So in connection with the findings of the qualitative study we speak of signals rather than tendencies.

[11] For all the findings concerning the process of teaching the fact must be taken into account that they only refer to regular lessons and not to special periods of laboratory work, field work or work on research assignments. In these we can assume that higher taxonomic levels are more represented and better care is taken of the creativity of students, their interests and their differences. The question is whether such sharp division between lessons with prevailing lower taxonomic levels and special periods and research assignments is very productive. A number of modern theories of knowledge, teaching and learning (see references) recommend that higher taxonomic levels, more complex mental processes and skills should be practiced regularly as part of regular teaching, otherwise they remain isolated and take more effort to get consolidated and become part of the mental repertoire.

The differences in the estimations of the taxonomic structures between the first and the fourth year

This result is also supported by the results of comparing the taxonomic structures of the first and the fourth year; those of the fourth year show an **increase in the proportion of consolidating routine explanation, interpretation and routine application and a decrease in searching for applicative value and independent problem solving and independent development of ideas**. This decrease can be noticed for all subjects as a whole as well as for each individual subject separately. In all of them statistically significant differences in the sense of decrease in higher and increase of lower taxonomic levels of knowledge can be noticed - except in chemistry, where the sample was too small to show statistical significance, yet the indicated tendencies remain the same. Of lower taxonomic levels a decrease can be noticed only in consolidating definitions and formulae, but this decrease is compensated for by an increase in consolidating given explanations and interpretations. This decrease is noticed both by students and by teachers, the latter even to a greater extent than the former (decrease and increase for about 7 to 9 %), but in both groups it is statistically significant.

Table 3: Significance and the size of differences between individual kinds of activities/taxonomic structures between the 4th and the 1st year according to the estimates of students (t-test):

Activity	year	Number	Arithmetic average	Standard deviation	Leventest	t-test	Level of significance	Medium difference
1	4	488	18,78	12,20	0,006	-3,767	0,000	
	1	495	21,99	14,40		-3,771	**0,000**	**-3,21**
2	4	488	25,67	15,58	0,017	1,994	0,000	
	1	495	23,80	13,89		1,992	**0,047**	**1,88**
3	4	488	26,85	18,53	0,000	7,272	0,000	
	1	495	19,20	14,20		7,259	**0,000**	**7,65**
4	4	488	14,35	10,06	0,005	-5,978	0,000	
	1	495	18,68	12,52		-5,987	**0,000**	**-4,33**
5	4	488	14,47	12,45	0,066	-4,275	**0,000**	**-3,63**
	1	495	18,10	14,13		-4,279	0,000	

Legend

Activities:

1 - consolidation of definitions and formulae

2 - consolidation of given definitions and interpretations

3 - routine application
4 - searching for applicative value
5 - independent problem solving and development of ideas

As a matter of fact in different subjects the direction of the change is different, but in all of them there is a tendency of decrease of higher taxonomic levels and of increase of lower taxonomic levels or even both at the same time.

The taxonomic analysis of examination sheets

As for the taxonomic structure of examination sheets it can be roughly concluded that different taxonomic structures can be determined: from those in which the lowest level prevails (up to 50 %) to those in which the three levels (1 - knowledge, 2 - understanding and simple applications and 3 - analysis, synthesis and evaluation) are approximately balanced. At least part of the differences, however, results from different taxonomic divisions. Common to all of them is a lack of the highest level on the 6-level scale (3 in the three-level taxonomy) i.e. evaluation.

As for the correspondence between the recommended relations the general conclusions are the following: the recommended ratios vary from subject to subject; the taxonomic structures of the examination sheets approach the recommended ratios to various degrees.
With regard to the taxonomic structure of examination sheets and the taxonomic structure of learning and teaching only a conditional and very loose[12] conclusion can be made, that the taxonomic structure of teaching and learning, as determined through observation and through

[12] The first condition for the inference being loosely is the taxonomic division, employed in determining the taxonomic structure of learning and teaching is not identical to the six-level division used for analysing the examination sheets, which makes immediate comparison more difficult, as shown elsewhere.
Even more relevant is the fact that the taxonomic structure in different subjects consists of a number of parts (because the examination sheet is divided into parts) and sometimes it is not reasonable to combine them into a unified result.
It must also be taken into consideration that the taxonomic analysis is an endeavour with only one meaning: context plays a very important part in it (e.g. when determining the synthesis it is important to consider whether the content has already been explained to the students and they have learned it or it is really a case of a convincingly new problem situation and a creative answer), neither are the dividing lines between individual levels very clear and only have one meaning. In our study we tried to manage these problems by carrying out the taxonomic analysis in moderation groups of 3 to 5 subject experts, who tried to align their estimates.
In addition to this the processes of learning and teaching are influenced by a very complex set of factors that are not always in cause and consequence relationship.

the quantitative approach, is similar to the taxonomic structure of the examination sheets.
So e.g. the structure of skills planned in the test sheets is reflected in the classroom. The biology examination sheets only contain 20 % of the lowest taxonomic levels, which is also the case when observing classes, but in both subjects the highest levels are only represented to a very low extent, understanding and application prevail. The percentage of knowledge on higher levels is, as compared to other subjects, quite high in psychology (40 % - including application) and so does learning and teaching pay very much attention to the applicative value of knowledge. The second part of the physics examination paper puts emphasis on application, which can also be noticed when observing classes - especially the routine aspect of application, while the first part of the paper is more balanced as regards taxonomic levels. The situation in mathematics is similar, where both in examination sheets and in the classroom the stress is on (routine) application. In chemistry, history and geography, where the emphasis lies on knowledge with understanding, also observing classes confirms that nearly half of the time is dedicated to lower taxonomic levels. Sociology is an exception to the rule: in the classroom higher taxonomic levels could be observed than in the examination sheets.
In all subjects a very feeble presence or even total absence of critical judgement and evaluation was noted - level 6 knowledge according to Bloom. In a number of subjects 0 % was noted, with the exception of the essay type of tasks, which makes this element possible. Although in some examination sheets questions appear beginning with the phrase “judge critically” or the verb “evaluate”, the content analysis reveals, also taking account of the context of the lessons observed, that the questions require merely a reproduction of memorised critical thoughts of others rather than a really independent, profound and creative critical judgement.

*

The prevailing of routine explanations and interpretations and of routine application cannot be interpreted quite unanimously. First the conclusion has to be drawn that it cannot be attributed just to *matura*, but to the whole concept of education in the *gimnazija* programmes and its interweaving with the *matura*. This means we cannot claim we have managed to isolate the influence of the *matura* on teaching and learning. The *matura* only reinforces the impact of the concept of the

gimnazija-type of education, and the findings are also valid for the blend of the *matura* with the concept of the *gimnazija* education programme - especially the function of the *matura* for selection of students in the higher education enrolment procedures.
Otherwise the interviews show that both the students and the teachers experience consolidation of knowledge as an adequate preparation for the *matura* or such types of tests as preferred by the *matura*. Complementary they believe that independent problem solving or development of ideas is less appropriate for the *matura*. Their opinions differ, however, in the answer to the question in what way such preparation for the *matura* contributes to their knowledge for life: certainly the consolidation of knowledge contributes to better retention and organisation of knowledge, but on the other hand they are constantly running the risk it could lead into routine i.e. lapse to lower taxonomic levels. They also point out that the positive aspects of consolidating knowledge seem to gradually get lost when the size of the content is extremely large. This aspect was also studied within the associated study on the retention of knowledge carried out by Barica Marentic Pozarnik. A considerable majority of teachers of more than half of the subjects and individual teachers of the rest of the subjects draw attention to the "factographic"[13] orientation of the preparation for the *matura*.
Teachers and students point out the large extent of the contents on various occasions. In the interviews the teachers suggest lack of time as one of the main reasons for the existing relation between the problem-solving and non problem-solving orientation and for the relation between their activity and the activity of the students - in other words for insufficient activity of students. Too little time or too large extent of contents for the time at their disposal is also the reason they give when asked the question what other methods of teaching they would use or why they do not use them (More about this in the reports of individual subjects in the study quoted above).
Both the teachers and the students are aware that routine knowledge prevails over actual application of knowledge. To the question, whether they would change anything if the present concept of the *matura* was not there, more (60 %) teachers give the affirmative answer than a negative one (40 %). Nearly all those who would change something (i.e. more than half of all the teachers included in the sample) would put more emphasis on applicative value and

[13] This is the way they formulate the description themselves, the study tries to avoid it and uses concrete description of activities instead.

independent problem solving and development of ideas. They would put less emphasis on consolidation of definitions and formulae, consolidation of given explanations and interpretations and to routine problem solving.
There is even a lower (17 %) number of those who do not wish any change among students. Among those who wish changes the majority (more than a quarter) would put more emphasis on applicative value. It is interesting, however, that the students wish to search for applicative value much less (even less than consolidation of definitions) than independent problem solving and development of ideas. A possible explanation is that they are not used to this kind of work and that it requires too large a mental effort, a more thorough and in-depth approach to work than the otherwise prevailing consolidation of given solutions. In accordance with the findings quite a number of students (9 %) wish less independent problem solving, although the answers prevail (nearly 55 %) that they wish less lower taxonomic levels. It needs to be stressed that the students were asked (without exposing the *matura*) what they would wish more or wish less in a specific subject, and it can be concluded they estimated what they would like more or like less with regard to what would be more useful for the *matura*.

The estimation of students of some other aspects

The findings above are also supported by the estimates of students that in the course of the preparation for the *matura* they get an opportunity to express their creativity and their constructively critical attitude, their interests in various areas within subjects only rarely or occasionally; different ways of comprehension, learning and expressing themselves are not taken account of. Similarly also the teachers, who estimated these areas a little more favourably, rate none of these elements as occurring frequently or regularly, on average they estimate they are implemented occasionally.

Table 4: Arithmetic means, number and standard deviation for individual items

Item	Number	Arithmetic mean	Standard deviation
1. Encouraging to learn actively and independently	608	3,43	0,96
2. Learning in a different way	606	2,12	0,70
3. How creativity and critical attitude find expression	605	2,65	0,89
4. How interest for individual areas finds expression	606	2,86	0,98
5. How different ways of comprehension and expressing oneself find expression	603	2,65	1,02
6. What is the importance of the knowledge for the *matura* for life	604	2,52	1,14

In items 1 to 5 the students were asked to choose on a 1 to 5 scale, with 1 = never,
2 = seldom,
3 = occasionally,
4 = frequently,
5 = regularly;
and in item 6: 1 = yes, they are very important;
2 = yes, they are important;
3 = they are important, but not much;
4 = I don't know;
5 = they are not important

The above findings of the quantitative part of the research about taking different aspects (creativity, differences, interest etc.) into account during the preparation for the *matura* are also confirmed by the statements of teachers and students in interviews.

The analysis of anecdotal records on experiencing the *matura*

We wished to illustrate the findings of the quantitative part with an analysis of the anecdotal records of the students about relevant impressions, events, thoughts and feelings in connection with the *matura* and events related to it.

We tried to find out mainly:

- what proportion of all students has answered the question at all. They were namely free to decide whether they wished to put down something or not;
- the ratio of answers with a neutral, negative or positive connotation and
- the categories of responses that appear.

Record according to school was also kept of the ratios of answers for connotation.

Out of 338 students 164 put down their impressions on the *matura*. 139 of these had a negative connotation and 25 a neutral or positive connotation. This means 40 % of the students in the survey spontaneously pointed out the negative aspects of their experiencing the *matura* and 7 % the positive aspects. The answers with a negative connotation were classified into the following categories:

Table 5: categories of answers with a negative connotation and their frequency

Category	Frequency
Pointing out the lack of time	50
Pointing out the stress caused by the *matura*	42
Pointing out the excessive number of subjects	22
Pointing out the excessive amount of contents	15
Mentioning knowledge is not useful for life	15

Category	Frequency
Mentioning the *matura* as not a good indicator of real knowledge	10
Pointing out low quality of the preparation	10
Pointing out it suppresses creativity and critical attitudes	7
Mentioning individual aspects of non-objectivity	7
Mentioning teachers use the *matura* to threaten or discipline them	9
Mentioning they consider it as too formal or "bureaucratic"	5
Mentioning it is damaging to life (impact on the enrolment to further studies)	4
Mentioning it is rather a test of psychic endurance than a test of knowledge	3
Pointing out too much emphasis is put on it	3

Remark: the total number of statements exceeds the number of students, because some of them provided more than one statement.

The answers with a neutral or positive connotation were classified into the following categories.

Table 6: categories of answers with a positive or neutral connotation and their frequency

Category	Frequency
The thought the *matura* is not something extraordinary	7
Warning it is an important test in life	5
Warning it is a test of all the knowledge	3
Mentioning positive experience during the preparation	3

Category	Frequency
Mentioning it stimulates teachers to prepare them better and to be nicer to students	3
Mentioning the importance of choice	2
Mentioning they get prepared for the *matura* well	2

It must be pointed out that the variance is considerable from school to school as for the proportion of frequency of individual categories and the proportion of answers with a positive, neutral or negative connotation. As for the reasons of data protection we cannot operate with concrete schools (they could be recognised from the number of responses) we only give a summary description of the differences.

The differences can be noticed in the reaction to keeping records on the *matura*. So in three schools only a third of the students answered and in further two schools only half. In four schools more than half of the students responded. In three of them answers with a distinctly negative connotation prevail and in the fourth the answers with a negative connotation refer to the category “lack of time”. In two of the three schools where the negative connotation distinctly prevails the categories are distributed evenly (especially stress, large number of subjects and lack of time for preparation), in one school the answers pointing out the stress prevail.

Thus it can be concluded that there is quite a lot of room in schools for improvement of the climate in the preparation for the *matura*. The findings of this part of the research ought to be read mainly as spontaneous reactions of students on different aspects of the *matura* in their own language. They serve as an additional illumination as well as a further support to other results gathered in a more “objective” and structured way.

Conclusions and proposals

From the findings above it is not possible to draw any incontestable conclusions about the influence of the *matura* on teaching and learning in school. With the qualitative and the quantitative analysis

predominantly the didactic aspect of the influence was examined. (The psychosocial influence was examined in Janez Becaj's study.) Even about this aspect we found that it cannot be isolated from other factors entirely as the effects of the mature are interwoven with the whole system of education in *gimnazija*. Considering the findings, however, that according to the estimates of both the students and the teachers consolidation of routine explanations and interpretations and routine application prevails during the preparation for the *matura* and considering that the observation of classes in the fourth year yields very much the same image, especially with regard to the activity of students, the conclusion can be drawn that

- the existing concept of the *matura* encourages systematic consolidation of knowledge thus contributing to better memorisation and retention (provided the extent of the content is not too large). This is true not only about the consolidation of definitions and formulae, but also about given explanations and procedures. In connection with this the *matura* has contributed a more systematic approach to teaching and learning and to introducing some approaches and activities such as laboratory exercise, research assignments etc., not previously encouraged by all teachers.
- There is, however, less emphasis on individual approach to problem-solving and independent development of ideas. Through consolidation of knowledge namely the teachers and the students run the risk to reduce the problem-solving approach to mere routine, thus reducing the taxonomic level of the activity. At the same time, due to such emphasis on consolidation of knowledge, for which according to their statements often there is not enough time, there is less emphasis on independent problem solving and on development of ideas. For this there is neither enough time nor interest, because both the teachers and the students assume the *matura* requires more predictable (although also complex) knowledge rather than convincing independent judgement. In connection with this the *matura* causes the effect of "levelling out".

Thus judging the influence of the *matura* is not only simply about opposition between the so-called factographic knowledge and higher levels of knowledge. These relationships are much more complex and the meaning multilayered, which we also tried to demonstrate in our

interpretation. The improvement of the *matura* can thus not start only from the technical level, but implies profoundly conceptual issues, and the issues of the content regarding not just the *matura*, but the whole education in *gimnazija*.

It might be practical to consider the following changes of the *matura*: Because of the many signals about the excessive amount of contents for the *matura*, sooner or later it will become necessary to contemplate what mechanisms it is sensible to apply to reduce the extent. In connection with this quite a number of starting points are at hand such as increasing the extent of choice of topics, a different concept of the *matura* catalogue or a new definition of the relationship between the catalogue and the national curriculum for *gimnazija* and between the syllabus and the subject catalogue.
One of the most upsetting findings regards the low extent of the presence of the highest taxonomic levels and low representation of independent development of ideas and of critical judgement. In this area the potential solutions are open in the direction of increasing the share of research tasks, laboratory exercises and fieldwork to about 50 %. Because of the improvement the culture of performing and evaluating these practices seems more feasible than a few years ago. Noticeable is the contribution of the *matura* to this development.
In connection with the low level of taking into account interests of individual students and the differences in the ways of comprehension, learning and expressing themselves, it would be wise to consider increasing not only the choice of topics, but also of the ways and types of testing the knowledge.

There is some room for change that can be employed immediately in the schools themselves and in the way they are preparing the students for the *matura*. Considering the signals mentioned above, from the open responses about experiencing the *matura* and from interviewing the students, at least in some schools there is not enough time for the preparation for the *matura*; the levers to increase the amount of time dedicated to the preparation for the *matura* in the fourth year can be searched for in the system as a whole as well as in individual schools.

Last but not least, it will also be necessary to consider whether the determined defects can be removed by simply patching up the existing system with individual measures or whether it is necessary to rethink the whole concept, not only of the *matura*, but of the entire

gimnazija educational programme.[14]

References

Baron, J. B. and Sternberg, R. (1987). *Teaching Thinking Skills: Theory and Practice*. New York: W. H. Freeman and Company.

Bloom, B. S. (1956). Taxonomy of Educational Objectives, *Handbook I: Cognitive Domain*. New York: David Mc Kay.

Bloom, B. S., Madaus G. F. and Hastings J. T. (1981). *Evaluation to Improve Learning*. New York: McGraw-Hill.

Cohen, L. and Manion, L. (1992). *Research Methods In Education*. New York, London: Routledge.

Drever, E. (1997). *Using Semi-Structured Interviews in Small-Scale Research*. Glasgow: SCRE.

Gifford, R. B., O'Connor, M. C. (ur.). (1992). *Changing Assessment*. Boston: Kluwer Academic Publishers.

Izhodišca kurikularne prenove. (1996). Ljubljana: Nacionalni kurikularni svet.

Marentic Pozarnik, B. (1995). Pomen operativnega oblikovanja vzgojno-izobrazevalnih smotrov za uspešnejši pouk v Izbrana poglavja iz didaktike. *Pedagoška obzorja*, Novo mesto, str. 5-81.

Marentic Pozarnik, B. (1999). Kako pomembna so pojmovanja znanja, ucenja in poucevanja za uspeh kurikularne prenove (prvi del). *Sodobna pedagogika*, št. 3, 49 (115), str. 244-262.

Marzano, R. J. (1989). *Dimensions of Thinking*. Alexandria: ASCD.

Nitko, A. J. (1996). *Educational Assessment of Students*. Englewood, New Jersey, Ohio, Merrill.

Voutilainen, T., Mehtalainen, J., Niiniluoto, I. (1990). *The Conception of Knowledge*. Helsinki: The National Board of General Education, The Government Printing Centre Helsinki.

Wajnryb, R. (1998). *Classroom Observation Tasks*. Cambridge: Cambridge University Press.

Wragg, E. C. (1994). *An introduction to classroom observation*. London and New York: Routledge.

[14] At the time of publishing this article in the CIDREE yearbook a fundamental renewal of secondary general education, and in this context also of the matura, has been started in Slovenia. (editors' remark)

“Like the Fishermen in the Maelstrom!?” Central Quantitative Performance Tests and Qualitative School Development in the USA

Heinz Schirp

Do you like odd and even macabre stories? If so the name Edgar Allan Poe will certainly be familiar to you. Perhaps you have even read his story “A descent into the maelstrom.” For those of you who can’t quite remember it, here is a short summary:

Two fishermen, two brothers in fact, were with their boat on the Norwegian coast when they got pulled into a dangerous whirlpool in terrible maelstrom. In their fear and panic they grab onto the planks of the boat and watch how their boat on its frenzied journey into the whirlpool, gradually sinks deeper and deeper, and very quickly approaches the destructive gullet of the maelstrom. After a few minutes of panic the older of the two brothers gradually starts to think clearly. He starts to look more closely and watch what is going on, and suddenly it comes to him. He realises that not all objects disappear into the whirlpool at the same speed. The lighter objects obviously swim for somewhat longer on top, whereas the heavy boats and bits of wreck are pulled under much quicker. He calls out to his brother and says that it would be sensible to leave the boat now, to tie oneself to an empty barrel and leave the boat. But his younger brother is so paralysed with fear and panic that he does not listen to his brother’s good advice and does not want to leave the boat. So on his own, the older brother ties himself to the barrel and jumps over board into the thundering sea. And in fact, while the heavy fishing boat is being pulled under into the whirlpool faster and faster, the barrel with the fisherman attached to it, floats for longer around the top edge of the maelstrom. Shortly after the heavy boat has been totally devoured by the whirlpool and broken into smithereens at the bottom of the sea, the sea gradually calms down and the maelstrom disappears into nothing and the fisherman that was tied to the barrel is saved.

Incidentally, Norbert Elias uses this image of a maelstrom, to clarify the relationship between “commitment” and “distance” in social development and decision-making processes (Elias, 1983, p 74 ff). I would like to take up both of these analytic terms and take this

somewhat macabre tale as a metaphor for the current situation in education policy. The main points are the reception, the debate about the results of the international benchmarking, and the significance of centrally quantitative performance tests.
Firstly there is the metaphor of the maelstrom, which should clarify that we should look more closely where the whirlpool of performance tests could possibly drive us to. Then against the background of the analytic terms “commitment” and “distance” I would like to work out where there are good reasons for a critical discussion about the central tests and the linked consequences as well as whether there are meaningful forms of quantitative performance tests.

Let us just remember the processing of the results for the PISA study. During the first stage school researchers, educationists, and educational policy makers obviously reacted like the two fishermen in the maelstrom: they were panicky. It was a bit like what we had with the PISA shock, complete with all the effects and side effects that this sort of shock brings with it. First there was speechlessness and helplessness, then a sort of last minute panic in educational policy with the motto: “something has to be done.” Now we are like the older fisherman in the boat - at a stage where we can look around in a somewhat calmer and more exact way, at what central school performance tests mean for schools.

Because of their comparative and ranking structures, empirical and quantitative studies have one thing in common which is of special importance for the school development processes. In the researched areas they always produce winners and losers. Even the countries which find themselves in the middle field, are already coming under pressure to give explanations, because their results are just suboptimal - often irrespective of how far apart the results from the different countries are. Actually it is only the ranking place that seems to count. It is the headlines in the newspapers that do the rest to put the test results next to the educational catastrophes.

Incidentally, the next international OECD study will take place in 2008. It will test the social competencies of the pupils. According to the results of the IEA-civic study, you do not even need to have the gift of prophecy to be able to predict that Germany, once again, will not be one of the winning countries. And so we are always a step behind these suboptimal results with our educational efforts - a

laborious and rather unsuccessful strategy. Many teachers, head teachers and education authority officers articulate their uneasiness towards this dynamic of performance tests. Partly, they feel downright run down by the demands, which are a result of these output-orientated performance tests.

The new paradigm of “output-oriented steering” had an effect on national curriculum planning in a number of different ways. In Germany, for instance, this led to the development of core curricula with clearly defined standards and competences, to standardised achievement tests and to the introduction of centralised federal A-Level examinations.
The ways in which the results of the pupils in these tests are established and used differ considerably from those in the USA.There results are made public in the form of league tables. As a consequence, schools are judged by their empirically ascertained results, they compete with these results and they have to take appropriate measures as a result of these findings.
As detailed below, a look over the fence into the USA shall show how those American tests affect teaching and learning in American schools.

This is a question I would like to tackle from three different points of view.

The first view will be a systemic one. Very up to date, well proved and empirical investigations are available which show how central performance tests affect a system as a whole. Even though we have not (yet) introduced the US American test procedures and test concepts, it can still be very helpful to have a critical look at their system.

The second view is rooted in a perspective of learning development theory: I would like to go into the question, whether and, if so, in which ways centrally raised performance results and their feedback are consistent with the results from learning and cognitive science.
And lastly, I would like to add to both of these critical perspectives a third. It should be asked and outlined, whether and how central measuring and feedback systems could affect the perspectives and development of a future culture of education and school development.

How do central school achievement tests affect a school system? A look over the fence to the USA

In 2002 the current President George W. Bush announced the so called “No Child Left Behind” law, in which - at least in the programmatic title - it aims to care for every single child, and to meet the appropriate individual needs of each child. At the heart of this concept is of course the setting of standards, and with the help of central tests, to find out whether the pupils have fulfilled these.

In fact, for over 40 years the American education system has been setting centrally developed standards and centrally standardised tests in order to measure the pupils’ achievements. The production of tests and materials which should prepare for them has become a lucrative billion business for the publishing companies in the USA. In the sense of the already mentioned output-orientation comes the underlying philosophy in educational policy: If one confronts the schools with the results of their work, then a competitive and competition situation arises, in which the schools themselves strive towards their qualitative further development. The use of standardised tests both nationwide and - additionally - on the level of the federal states are to provide the basic foundation for the pedagogic reform efforts.
These achievement tests are called High Stakes Tests (HST), because they incorporate cleverly devised, highly standardised and validated standard-orientated test tasks. High Stakes Testing (HST) is consequently a specific reform approach which ties together both gratifications and sanctions for schools to the results of such comparison tests. High Stake Tests are used in all federal states and should provide comparable data about the academic achievements in schools. Furthermore, some federal states still develop achievement tests of their own, e.g. for single subjects or at the end of certain academic years, and some at the end of Junior High School. The publication of the results of these various tests combined with state-wide school rankings allow the parents to decide to which school they want to send their child.

It is important to point out that we in Germany are still a little way off from this reality of High Stakes Tests, though the structure of our assessment of achievement levels does meet some criteria of the High Stakes Tests. However their results are not (yet) used for the compilation of ranking lists. In this respect, we also have the

opportunity to distance ourselves from methods, which are from our view less suitable, for giving the schools valid and reliable feedback about pupils' achievements.
All the more important is to pay attention to the effects and side-effects which can arise from central achievement tests. Therefore it is well worth having a look at a study which critically debates the test methods and test results in the USA.

Campbell's Law: the significance and reliability of central performance tests (High Stakes Tests)

Sharon Nichols (University of Texas at San Antonio) and David Berliner (Arizona State University) put forward a research study in March 2005, which has led to some excitement in the USA. David Berliner is one of the most renowned educationalists in America. The title speaks for itself: "The Inevitable Corruption of Indicators and Educators through High Stakes Testing" (Nichols/Berliner, 2005).
In their study Nichols and Berliner prove what was formulated by the American social psychologist Donald Campbell and became known as Campbell's Law. It reads: *"The more any quantative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor"* (Campbell, 1975, p 35).
Incidentally, Campbell's Law applies to all social institutions, not just those to do with schools and education. Bluntly spoken, this means for our issue: when quantatitive achievement results are made to be the exclusive benchmark for the definition and evaluation of school quality, then there is the danger that a blurred picture arises from the schools' actual achievements. Evidently schools make every attempt possible to dress up their results, or put more commonly: cheat and wangle the system.
That sounds quite surprising at first. But when you think about it though, that in some federal countries it is down to good or bad cuts of a school, whether or not the parents decide to keep signing their children up there, whether or not schools have to be afraid of sanctions of the school surveillance, whether or not further on financial resources will be available, then you can understand this justification.
Nichols and Berliner prove the negative effects of tests with an abundance of examples from empirical school research studies, school reports, articles in pedagogic magazines and reports and interviews

from teachers and education authority advisors. The authors of the study sum up that in total there are 10 heavy weighing factors which point out how the central test results have a negative effect on learning, school development and achievement development. Here I shall concentrate on three main aspects:

- on the unreliability of the test results,
- on the effects concerning school curricula and subjects,
- on the effects concerning the self-assurance of teachers as educational and pedagogical professionals.

First on the unreliability of test results: Many schools and teachers cheat in the tests, awarding the wrong marks, so that they do not slip under the average of their school district and have to face appropriate sanctions. In the study, almost 50% of the pupils asked, admitted to cheating in the tests. Therefore it is assumed that the hidden figure is actually quite a bit higher, and apparently there are some teachers who tolerate, or even support this, in order to get better test results. Pupils with low achievement results and poor achievement prognosis are systematically disqualified from the tests to improve the overall results and to get a better place in the league tables in comparison with other schools.

Pupils with bad marks and slim prospects of success in certain educational paths are advised to choose other high school equivalent educational paths. That way, they are not classed as school dropouts and they do not worsen the balance of the school.

The results presented from the study make it emphatically clear how great the pressure for schools is to score somewhat well in these central tests. They also show how great the danger is as some schools will even try to improve these results with unacceptable tricks. This means that the significance of the whole test and the results from it is completely falsified and worthless. The schools which are honest and adhere to the guidelines, come off as the fools and feel they are unfairly evaluated because they did not trick and cheat the system. It is recognisable from the teacher interviews that next time some schools intend to do what a lot of other schools have already done: to cheat in the test. Naturally this totally destroys the validity and reliability of the whole HST system. Campbell’s law is classically confirmed.

Centrally developed and centrally deployed tests must be applied in such a way that they work for all federal states, regardless of which specific main focuses, profiles, pedagogic or curricula programmes are used each time. The test exercises must inevitably neglect the specific learning conditions and learning guidelines "on the spot". They can of course be used everywhere, but they have little to do with the learning reality of the pupils.
The study by Nichols and Berliner sums up the following systemic effects:

- High Stakes Tests reflect an unrealistic image of the pupils' achievements.
- If test results are put together with rankings and sanctions, obviously schools are encouraged to commit an institutionalised fraud.
- The higher the pressure to achieve, the less the test results say about the actual achievements of pupils and schools.
- As the stakes associated with a test go up, so does the uncertainty about the meaning of a score on the test.

Central tests and their effects on school curricula, subjects and learning processes

The study proves emphatically that through focusing hard on central tests, and through the pressure to produce the best possible results in a class, the school curriculum becomes narrower and narrower, that is to say that it becomes test compliant.

Hence teachers overemphasise a kind of knowledge which can be acquired quickly and easily by training and also repetition and the multiple choice tasks linked with this. The self conducted narrowing of the curriculum finally leads to teaching and learning processes increasingly orientating themselves to the expected tasks in the tests. Complex tasks are simply avoided. They do not appear in most tests, since their development and standardised evaluation is really quite complicated.
It is increasingly so that test training is linked with the clear narrowing of the curriculum, its topics, contents and methods. The skills will be learnt and practiced, so the pupils can deal with multiple choice test questions as cleverly as possible. Teachers are obviously using far too much time for this.
As a result of this, a process can be observed, which is known as "teaching to the test". It is not tested whether or not the pupils have

learnt or understood what was taught and worked on, rather, they are taught more and more what should be tested.
Through this disproportionate use of time which is dedicated to test preparation methods, other subjects and learning areas are clearly pushed back not only in terms of their didactic meaning, but also in school curriculum. This "narrowing the curriculum" has of course effects on the perception of lessons and schools by the pupils and on the routine in which school days are structured.

- Central achievement tests tend to lead to a narrowing of the school curriculum.
- Learning processes are aimed less and less at the development of skills and complex understanding and learning processes, but more at knowledge that is asked for in the test.
- Lessons develop themselves relating to learning content and methods in the direction of higher test conformity.
- The narrowing of the test demands leads to the under-representation of social, interactive and musically creative learning contents and methods.

The effects of central tests on teachers' educational work

"It is true that if we have a good test, then teaching to that test is appropriate. Instruction and assessment would be aligned, as they should be. That line is often hard to see and it is possible that some educators and test preparation companies deliberately cross that line. This would corrupt the indicator, making it quite difficult to interpret any assessment so compromised" (Nichols/Berliner, 2005, p 89).
The other result of this is that teachers use fewer creative, didactically challenging and pupil motivating learning situations and learning approaches, rather, they feel impelled to practice and to prepare the test. For this they use the past test papers for training purposes. Nichols and Berliner worked out that 80% of teachers in the primary school sector in North Carolina use more than 20% of the lesson time to practice for the High Stakes Tests. Furthermore, a third of these teachers admit that more than 60% of the lesson time is used to prepare for the test, which is used in North Carolina.

When tests and results are of such an oversized importance for teachers, when they hang over their heads like the sword of Damocles, then they are not seen as help which will improve the lesson, but as a rather recurrent threat.

Teachers are finally pushed into the roll of test preparers. They can only do less than what they would actually like to do in their role as educational experts. More and more they feel they have less time to take care of the individual learning processes, interests, capabilities and individual ways of learning, even when they want to and consider it necessary. In the long term there is the risk that the didactic aspects of an education orientated around the students' needs fall increasingly into the background.

Unsatisfactory test results increase the pressure to do well the next time. The tendency to get hold of standardised, preparatory test materials and methods, will certainly become greater. For test associated preparation the increasingly appropriate pre-tested and standardised materials are used. Naturally this has further side effects. Since the test results have no or only very limited diagnostic qualities, the teachers receive no evidence, how they can continually improve their pupils' achievement skills. What is missing is some sort of evidence, why the pupils' achievements are unsatisfactory and how one should start support in order to improve them.

Summing this up, there are the following effects on teachers' educational work:

- Teachers are pushed into the role of successful test preparers.
- Individual concepts of learning seem to disappear in favour of standardised methods.
- "Teaching to the test" entails a narrowing of the school curriculum.
- Both the didactic and pedagogic skills of teachers and the significance of interesting, inspiring, student oriented learning arrangements become less important.

These negative effects are systemically linked with the realisation of central quantitative achievement tests. They are "inevitable", as Campbell states.
Instead of gathering realistic, valid and necessary analytic information, the test results distort and falsify the evaluative view on the actual attainment of pupils and schools.
All in all, this is neither helpful to the schools concerned, nor to the administration and policy makers.
Ronald Wolk, who is the publisher of the renowned magazine Education Week, supported standardized tests for a long time, but

holds a pessimistic view today: “Now driven by the federal No Child Left Behind legislation, standards-based reform is actually reinforcing the least desirable features of the traditional school: obsession with testing and test prep, inflexibility and inefficient use of time, overemphasis of on coverage in curriculum and memorization, and increased standardization that offers a one-size-fits-all education to students regardless of their differences.” (Wolk, 2006, p 52)

Central performance tests and the development of school quality

After we evaluated the standardised test methods and test results in America, we should analyse, if and how central tests can contribute to the development of school quality in Germany. Hereby, we are well advised to consider the differences in America’s and Germany’s educational policy.

The quality of learning depends on individual learning processes and intrinsic motivation

Empirical studies reveal that test-oriented processes of teaching and learning, meaning studying for the next test, prove to be little effective in the long run. This is because pupils quickly forget knowledge and skills after the test is passed.
Thus, it appears that the quality of learning rather depends on the individually experienced sense of learning, the pupils’ ability to connect new facts with already known ones, and intrinsic motivation, for instance. It would be more reasonable then to be concerned about the development of individual learning processes that consider the pupils’ interests and capabilities.
Briefly speaking: “The quality of learning depends on the individual learning development of students.”

Individual performance assessments and central quantitative performance tests are a different kettle of fish

The realisation of central achievement tests should not prevent us from developing innovative concepts of learning.
We could rather follow the example of the Finnish education policy which favours the realisation of anonymous central tests. A small number of pupils is tested regularly, thus the administration of education gathers necessary information and monitoring knowledge

on the schools' development of achievement. Test results are not published in form of rankings and league tables. It is the single school and its representatives which will be informed about its results. By this there is no blaming and shaming neither of schools nor of classes or students.
This is probably the reason, why Finnish schools voluntarily apply for the anonymous testing procedures, and moreover, pay money for it. These empirical studies help to establish reference level, which determine the standard achievements of pupils. Thus, teachers and the administration of education can clearly define the achievements which are expected from pupils with regard to single subjects or at the end of certain academic years.
In addition it has to be considered that social, economic and inter-cultural conditions have a strong influence on the achievements of pupils and schools. As a consequence, expectations are scaled down in the cases of schools that meet difficult conditions.

Test developers need to include didactic-methodical models

The results presented show that we should be concerned with the development of learning concepts, which integrate the pupils' individual capabilities and interests. Therefore, it is of great importance to design test tasks in such a way that we can draw useful conclusions from them regarding the improvement of learning methods and school forms. In order to improve the quality of tests, test developers need to consider didactic-methodical models.

Output-orientation needs input-orientation

Many teachers feel unprepared when being confronted with diagnostic material, data-based learning development, and standard oriented curricula.
Therefore, professional help would be necessary in order to demonstrate how teachers can derive benefit from a combination of both, individual concepts of learning and standard oriented methods. An intensive investment into teachers' training makes it possible that test procedures and results are understood as helpful rather than threatening. This seems all the more important in the cases of those schools that already established innovative concepts of pupil-oriented teaching. The concentration on test results should not prevent them from continuing with the development of learning concepts, which focus on individual capabilities and interests. Hence, we should

closely observe in how far standardised tests affect such pupil-oriented school forms on the long run.
Concerning this, we will be well advised to gain more practical experience and invest into these innovative school forms rather than cutting financial support for In-service Teacher Training.
Briefly speaking, governmental support is mandatory in order to improve teachers' skills and competences.

Appreciation and cooperation vs. control und competition

The paradigm of output-orientation combined with testing is directly linked - intentionally or not - with keywords such as competition and control. We should pose the question whether competition and control are really beneficial instruments in order to improve school and learning quality. Could it not be possible that a model which is based on cooperation, appreciation and individual support works out much better? Taking a look at countries that do without school rankings, we can see that their philosophy of cooperation is most effective.
Finland's excellent PISA placement indicates that its education policy, which aims at avoiding an atmosphere and a pressure of competition between schools, is successful. Furthermore, it is recognizable that the intensive investment into teachers' education is a far-sighted and most advantageous decision. Negative side effects, as, for instance, a test-oriented learning attitude or the narrowing of the school curriculum, could be prevented this way.
Additionally, we should follow Finland's example and involve parents as well as the local community in school development processes.

I began this text with a slightly macabre story. I would like to end now with a rather idyllic metaphor. Hartmut von Hentig presents us the following conclusion: "The unintentional effects of the PISA achievement study are ambivalent. The German educational system could have had a great stroke of luck like the two little rabbits in the popular nursery rhyme. Having been shot down by a hunter they lie in the grass paralysed by fear. After a while they discover that they are still alive. So they struggle to their feet again and quickly they hop off, happy to have survived. Like the rabbits in the song, the PISA shock should have stirred us, should have made us vigilantly observe both the risks and the chances of testing and surveying the quality of our schools. Meanwhile the paralysing shock should have left us and - like the rabbits - we should gradually regain our consciousness and understand that the education of children and the development of

schools are slow processes which demand time. So, let us take and use time to evaluate and to test carefully - also the tests." (von Hentig, 2003, p 233)

References

Archer, J. (Jan. 2006). The Road Less Travelled, in *Education Week: Quality counts at 10: A Decade of Standards-Based Education*, Vol 25, Issue 17, p 34-37.

Böttcher, W. (2006). Bildungsstandards und Evaluation im Paradigma der Outputsteuerung. muti. cop. Manuskript, to be published in *Böttcher, W., Holtappels, H. G., Brohm, M. (Hrsg.): Evaluation im Bildungswesen*. Weinheim und München.

Campbell, D. T. Assessing the Impact of Planned Social Change, in *Lyons, G.*, a.a.O., p 35.

Comer, J. P. (Jan. 2006). Our Mission. It Takes More Than Tests to prepare the Young for Success in Life, in *Education Week: Quality counts at 10: A Decade of Standards-Based Education*, Vol 25, Issue 17, p 59-61.

Elias, N. (1983). Die Fischer im Mahlstrom, in *ders.: Engagement und Distanzierung. Arbeiten zur Wissenssoziologie I.* Frankfurt/Main, S. 74-120.

Flick, U., von Kardorff, E., Stemke, I. (Hrsg.) (2000). *Qualitative Forschung. Ein Handbuch*. Reinbek.

Floden, R. E. (Ed.) (2003). *Review of Research in Education, AERA*. Washington DC.

Hentig, H. von (April/Mai/Juni 2003). Die vermessene Bildung. Die ungewollten Folgen von TIMSS und PISA, in *Neue Sammlung, h. 2,* S. 211-233. (Übers. H. Sch.)

Lersch, R. (2006). Unterricht zwischen Standardisierung und individueller Förderung. Überlegungen zu einer neuen Lernkultur angesichts der bevorstehenden Einführung von Bildungsstandards, in *DDS , H. 1*, S. 28-40.

Lyons, G. (Ed.) (1975). *Social Research and Public Policy*, Hanover: NH. Dartmouth College.

Meyer, H. (2004). *Was ist guter Unterricht?*. Berlin.

Moll, S. (23.Mai 2006). Lernen nach der Bush-Methode, in *FR*, S. 39.

Mortensen, P. O. *Why Danes Detest Test. The Danish Approach to Curriculum Thinking - a Remnant from the Past or a Model for the Future?*. The Danish Educational System.

http://www.hltmag.co.uk/sep05/sart08.htm
http://eng.uvm.dk/education/General/general.htm?menuid=1505
Nichols, Sh. L., Berliner, D. C. (2005). *The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing*. Arizona State University (EPSL). Tempe, AZ.
Oser, F. (1998). *Ethos - die Vermenschlichung des Erfolgs. Zur Psychologie der Berufsmoral von Lehrpersonen*. Opladen.
Peek, R., Dobbelstein, P. (2006). Zielsetzung: Ergebnisorientierte Schul- und Unterrichtsentwicklung. Potenziale und Grenzen der nordrhein-westfälischen Lernstandserhebungen, in *Böttcher, W., Holtappels H. G., Brohm, M. (Hrsg.): Evaluation im Bildungswesen. Eine Einführung in Grundlagen und Praxisbeispiele*. Weinheim und München, S. 177-193.
Paige, R. R. (2001). No Child Left Behind, in *Carnegie Reporter*, Vol. 1, No. 2, The Back Page. www.carnegie.org/reporter/02/backpage
Patton, M. Qu. (1990). *Qualitative Evaluation and Research Method*. London (2nd ed.).
Poensgen, O. H. Zur praktischen Nutzung von Forschungsergebnissen, in *Witte, E.*, a.a.O., S. 1-12.
Roeder, P. M. (2003). TIMSS und PISA - Chancen eines neuen Anfangs in Bildungspolitik, -planung, -verwaltung und Unterricht, in *Zeitschrift für Pädagogik,* 49. Jg. 2003, H. 2, S. 180-197.
Ryan, R. M., Sapp, A. (2005). Zum Einfluss testbasierter Reformen: High Stakes Testing (HST). Motivation und Leistung aus Sicht der Selbstbestimmungstheorie, in *Unterrichtswissenschaft*, 33. Jg., H. 2, S. 143-159.
Strauss, A. L. (1991). *Grundlagen qualitativer Sozialforschung*. München.
Witte, E. (Hrsg.) (1981). *Der praktische Nutzen empirischer Forschung*. Tübingen.
Wößmann, L. (2003). Zentrale Prüfungen als “Währung” des Bildungssystems: Zur Komplementarität von Schulautonomie und Zentralprüfungen, in *Vierteljahrshefte zur Wirtschaftsforschung 72*, H. 2, S. 220-237.

About the authors

Maria da Conceição Castro Ramos

Maria da Conceição Castro Ramos has a PhD in Sciences of Education. She is the Senior Chief Inspector of Education. In the recent past she was a Professor of "School analysis and Management" and "Education and Development Policies" at the University of Lisbon and she also coordinates Masters and post-graduations in Sciences of Education. Also, she had several leading positions in the Educational Administration, such as Director-General of the Department of Educational Resources, Director-General of School Administration and Regional Director of School Administration in the Autonomous Region of Azores. She has published several books about educational issues.
mccr@mail.fct.unl.pt

Peter Dobbelstein

Peter Dobbelstein taught German and Social Sciences at various types of schools. First as a scientific assistant, later as a scientific consultant at Northrhine-Westfalia's State Institute for Schools / Quality Agency (LfS/QA) in Soest, he worked on numerous innovation experiments. From 2002 onwards he was subsequently project manager of the standardized achievement tests for Northrhine-Westfalia (VERA, "Lernstandserhebungen"). He was also concerned with national and international studies on the assessment of teaching and learning. In his publications, these aspects cover his field of educational research. He is contact person for CIDREE, too. Since August 2006 he has been working for school inspections and quality assessment with the regional administration in Detmold, Northrhine-Westfalia.
peter.dobbelstein@brdt.nrw.de

Helder Guerreiro

Helder Guerreiro has a Masters degree in Sciences of Education. He worked in Secondary Schools, as a teacher and lately as a headteacher, and in a Local Education Authority. As inspector of

education he has been involved in different stages of inspection activities: from designing to fieldwork and reporting. At present, he participates in INES-OECD projects and he is also member of the steering groups responsible for the “Inspectorate’s self-evaluation” and a programme for “Schools” meta-evaluation’.
helder.guerreiro@ige.min-edu.pt

Zora-Rutar Ilc

Zora Rutar-Ilc graduated in Psychology at the Faculty of Arts of the University in Ljubljana and got her PhD in Sociology from the same faculty. As a senior adviser for secondary education in the National Education Institute of the Republic of Slovenia she has dedicated her research work mainly to the concepts of knowledge and learning. Her extensive bibliography predominantly focusses on the issues of assessing knowledge both in primary and in secondary education. Currently she is head of a four years project aiming at improving the methods and techniques of teaching in selected general secondary schools (gimnazija).
Mirko.Zorman@zrss.si

Bart Maes, Els Ver Eecke and Veerle Verhaegen

Bart Maes, Els Ver Eecke and Veerle Verhaegen are the authors of the yearbook’s contribution “Equilibrium”. Their article is an abstract of the book “Equilibrium” by Christine De Coninck, Rita Dunon, Laurent Osaer, Chris Van Woensel, Els Ver Eecke and Roland Voet. All of them are members of the division Curriculum (Entiteit Curriculum), formerly known as Department of Educational Development (Dienst voor Onderwijsontwikkeling) in the Flemish Ministry of Education in Belgium. “Curriculum” is a multidisciplinary team of researchers and advisers. Next to their initial subject training, all members of the team are educational specialists. The division is engaged in the following curriculum aspects: contents of the core curriculum, structures and qualifications, and evaluation at system level.
veerle.verhaegen@ond.vlaanderen.be

Thomas Neidhardt

Thomas Neidhardt studied English and Latin language and literature. He has been working at comprehensive and grammar schools in the German Federal States of Lower Saxony, Brandenburg and Northrhine Westfalia. He has been awarded various prizes for encouraging and innovative teaching concepts. He has also been involved in teacher training and curriculum planning. At present, he is pedagogical assistant at Northrhine Westfalia's State Institute for Schools / Quality Agency (LfS / QA) in Soest and predominantly concerned with the development of standardized achievement tests in year 8/9 ("Lernstandserhebungen").
thomas.neidhardt@mail.lfs.nrw.de

Rainer Peek

Rainer Peek has been a scientific consultant for standardized achievement tests (VERA, LSE) at the State Institute for Schools / Quality Agency (LfS / QA) in Soest, Northrhine-Westfalia, Germany, since 2003. He had been an academic assistant at the Institute for Educational Science of the Humboldt University in Berlin before. He has a PhD in Education. His publications include many articles in various periodicals and books on national and international studies about the quality of schools, the adoption of standardized achievement tests including the use of external data material in schools and about school developments related to standards in education.
rainer.peek@mail.lfs.nrw.de

Ian Schagen, Dougal Hutchison, Paula Hammond

Ian Schagen, Dougal Hutchison, Paula Hammond are all members of the Statistics Research and Analysis Group (SRAG). Dr Ian Schagen is Head of Statistics with overall management responsibility, while Dr Dougal Hutchison is Chief Statistician with a lead role in methodology. Paula Hammond is a Senior Statistician within the group. SRAG is headed by Dr Ian Schagen and comprises twelve well qualified statisticians who together have many years' experience in all types of statistical analysis techniques, particularly those which relate

to educational research. The group has extensive experience in analysing large and complex datasets, with specific emphasis on multilevel modelling techniques. It provides statistical support to the Foundation's projects and is responsible for all aspects of quantitative data analysis. It also offers a statistical analysis service to outside users, and conducts a number of its own theoretical and applied research projects.
http://www.nfer.ac.uk/about-nfer/departments/statistics-research-and-analysis-group.cfm

Heinz Schirp

Heinz Schirp is Vice-Director of the State Institute for Schools /Quality Agency (LfS / QA) in Soest, Northrhine-Westfalia, Germany. He has got a PhD in Education and is meanwhile a professor at the Department of Education of the University Bielefeld. Both his current research activities and his publications focus on moral-cognitive development, neurodidactic models of teaching, learning and education, and the development of the quality of schools.
heinz.schirp@mail.lfs.nrw.de

Alejandro Tiana Ferrer

Alejandro Tiana Ferrer is currently the Secretary-General of Education in Spain. He is professor of Theory and History of Education at the Spanish Open University. He has a PhD in Arts, specialised in Pedagogy, from the Universidad Complutense de Madrid.
He has been Director-General for Innovation and Development of the Organisation of Ibero-American States; Vice-Chancellor for Innovation and Evaluation at the Spanish Open University; Chairman of the International Association for the Evaluation of Educational Achievement; Director of the National Institute for Quality and Evaluation and Director of the Centre for Research, Documentation and Evaluation.
sg.educacion@mec.es

Hana Zufanová

Hana Zufanová studied Mathematics and Physics at Charles University, Prague. She taught at primary schools, grammar schools and a school for pupils with special needs. At the grammar school she also dealt with aspects of further education. She has been working in the Czech School Inspectorate since 1998. For the last three years she has been executing the position of Head of the department of external relations, which includes a co-operation with international partners of the Czech School Inspectorate. She is involved in the development of a methodology and further training of inspectors, too. She acts as an external tutor in the Centre of School Management at Charles University, Prague.
zufanova@csicr.cz

CIDREE

Consortium of Institutions for
Development and Research
in Education in Europe

ISBN 3-00-020494-6