

Benchmarking the Quality of Education

JAAP SCHEERENS & MARIA HENDRIKS
(project co-ordinators/editors)

Contributors:

Belgium: Peter Van Petegem & Jan Vanhoof

Denmark: Niels Egelund & Morin Thygesen

France: Gérard Bonnet, Jean-Claude Emin & Julie Sedel

Italy: Giovanna Barzanò, Emanuela Brumana & Claudia Cremonesi

Netherlands: Maria Hendriks & Jaap Scheerens

Spain: Noelia Alvarez, Alejandro Tiana & Jezabel Vico

United Kingdom: Wen-Jung Peng & Sally Thomas

Contents

Executive Summary	105
Chapter 1: Perspectives on Education Quality, Education Indicators and Benchmarking	115
1.1 Introduction	
1.2 Facets of Education Quality	
1.3 Illustration of Indicator Sets at System Level	
1.4 The Micro Perspective: School Quality	
1.5 About Criteria, Standards and Benchmarks	
1.6 Some Theoretical Ideas on Feedback and use of Evaluative Information	
1.7 Research Questions	
1.8 References	
Chapter 2: Three Sets of Indicators on Education: <i>Education at a Glance</i> (OECD), <i>Key Data on Education</i> (European Union), <i>The State of Education</i> (French Ministry of Education). Elements of Comparison and Analysis	139
2.1 General Introduction	
2.2 Presentation of the Documents	
2.2.1 General presentation	
2.2.2 Detailed description of cover and content	
2.2.3 Aims and tone adopted	
2.2.4 Visions of problems linked to education	
2.2.5 Methodological tools: sources, countries, nomenclatures	

- 2.2.6 Conclusion
 - 2.3 “Manuals of Reference”
 - 2.3.1 Manuals of reference in the field of international evaluation indicators
 - 2.3.2 Reference documents for Member States
 - 2.3.3 A model to copy for other ministers
 - 2.3.4 Conclusion
 - 2.4 Changes over time in the different editions: resemblance and contrasts
 - 2.4.1 Education at a Glance
 - 2.4.2 Key Data on Education
 - 2.4.3 The State of Education
 - 2.4.4 Conclusion
 - 2.5 Perspectives
 - 2.5.1 The productive perspective
 - 2.5.2 The disjointed perspective
 - 2.5.3 The adaptation perspective
 - 2.5.4 The instrumental perspective
 - 2.5.5 The equity perspective
 - 2.5.6 Conclusion
 - 2.6 General Conclusion
 - 2.7 References
- Annex: The use of international indicators by the European Union and French Ministry of Education documents

Chapter 3: The Use of Educational Standards and Benchmarks in Indicator Publications

177

- 3.1 Introduction
 - 3.2 Issues in Defining Standards and Benchmarks
 - 3.3 Use of Standards and Benchmarks in *Education at a Glance* (OECD, 2001b)
 - 3.4 Use of Standards and Benchmarks in *Key Data on Education in Europe* (EU, 2000)
 - 3.5 Use of Standards, Benchmarks and Targets in the UK
 - 3.5.1 Examples of standards
 - 3.5.2 Examples of benchmarking (England, Scotland, Wales, NI)
 - 3.5.3 Examples of targets
 - 3.6 Conclusions
 - 3.6.1 Use of benchmarks and standards in indicator publications
 - 3.6.2 Usefulness and feasibility of absolute as compared to relative international benchmarking of the quality of education
 - 3.7 References
- Appendix A: Examples of Benchmark information provided to schools in England 2001
- Appendix B: Examples of Benchmarking data in Scotland

Chapter 4: National Interpretations of the International Indicators

213

- 4.1 Introduction
- 4.2 Documents Produced by EU Countries Based on International Indicators
 - 4.2.1 Belgium (Flemish community)
 - 4.2.2 Belgium (French community)

4.2.3	Denmark	
4.2.4	France	
4.2.5	Italy	
4.2.6	The Netherlands	
4.2.7	Spain	
4.2.8	United Kingdom	
4.3	Some Conclusions from the Analysis of the Reception and Interpretation of International Indicators in EU Countries	
4.3.1	A diversity of national initiatives	
4.3.2	A variety of sources	
4.3.3	A disjointed perspective of quality	
4.3.4	An interest for thematic approaches	
4.4	Reflections on the Interest of Synthetic, Composite Indicators for doing Thematic Analyses	
 Chapter 5: Feedback of Indicators to Schools		246
5.1	General Introduction	
5.2	Is there a Need for Feedback?	
5.2.1	From the government's point of view	
5.2.2	From the schools' point of view	
5.3	Two Flemish Case Studies	
5.3.1	School reports based on TIMSS-R	
5.3.2	School reports based on PISA	
5.4	Creating Models for Feedback	
5.4.1	General aspects of a feedback model	
5.4.2	Technical aspects of the model	
5.5	General Conclusion	
5.6	References	
 Chapter 6: Attention that Indicators Receive in the Press		278
6.1	Introduction	
6.2	OECD's Press Clippings	
6.3	Impact of OECD Indicators on the Italian Press	
6.4	Summary and Conclusions	
Appendix		
 Chapter 7: Reports on the Interviews about Reception and Use of International Indicators		305
7.1	Belgium (Flanders)	
7.2	Denmark	
7.3	France	
7.4	Italy	
7.5	The Netherlands	
7.6	Spain	
7.7	United Kingdom	
 Chapter 8: Reports on the Interviews with School Principals		377
8.1	Belgium (Flanders)	

- 8.1.2 Results
- 8.1.3 Summary
- 8.2 France
 - 8.2.1 Introduction
 - 8.2.2 Conclusion

**Chapter 9: Benchmarking the Quality of Education: Discrepancies and
Correspondence between Theory and Practice in Seven European
Countries**

392

- 9.1 Introduction
- 9.2 Recommendations

Executive Summary

Background and design of the study

Presently, the idea of benchmarking educational quality is one of the corner stones of the EU's education agenda (EU, 2002).

Interest in education indicators has taken a high flight in recent years. Publications like Education at a Glance by the OECD and EURYDICE's "Key Data" stimulate a lot of attention in the public media, and at first sight are also frequently used in policy documents.

An interesting phenomenon is that many countries publish national interpretations of the international publications. In these national publications there is of course room for highlighting those elements that are of particular interest to the country, but also for providing background information that "contextualises" the information and may give explanation for certain figures.

The ultimate purpose of education indicators can be described as providing a basis for monitoring the quality of education. Even though the current indicator sets have become quite comprehensive, and there are clear signs of their use in policy-debates, there are still some important aspects in which their impact on educational policy and practice could be strengthened.

Addressing this issue appeared to be extremely timely in terms of informing developments of evidence based policy and practice (e.g. in the DfES, UK) and was the main motive to carry out this study. Our objectives were concerned with several conceptual issues, with more theme-based ways of reporting indicators and gearing them to policy issues and with feedback to lower level units of education systems, notably schools.

The overall purpose of this study can be seen as learning from national use and application of educational indicators to provide suggestions for enhancing the supra national effort. In addition the aim was also to provide suggestions for the improvement of national applications through the mechanism of international exchange and countries learning from one another.

Perspectives on quality of education

The available sets of education indicators provide data for different interpretations of quality and outputs in education. Examples are:

- quality in terms of participation and graduation rates

- quality in terms of levels of attainment in key subject-matter areas and competencies
- quality in terms of a distribution of educational “goods” that is equitable over sub-groups of the population
- quality in terms of levels of financing education
- quality in terms of human and physical resources and instructional technology that meet more or less explicit standards of educational “good practice”, or which may be seen as having the potential to foster educational success.

Despite the level of comprehensiveness of the available indicator sets there is an unequal coverage of different aspects of quality. Also users might have preference for other priorities. Investigating the perceived seriousness of possible discrepancies between supply and demand on this issue was seen as a basis for further improvement of the indicator sets.

Criteria and standards

Criteria are the dimensions or constructs that identify what is measured by an indicator. Standards define a norm, threshold or “cutting score” on an indicator. Precise standards, preferably specified in advance, allow for optimal clarity about the structure of the evaluation to which an indicator should ultimately lead. In their turn standards in international indicator studies can be used as benchmarks. The most common type of standards are relative comparative standards, that depend on the distribution of levels of “scores” across countries, i.e. countries that are above or below the overall average. However, the use and applicability of absolute “criterion referenced” standards was also considered of interest. Possible trade-offs between accuracy and political viability could be made explicit in such an exercise.

Theme-based presentation of indicators and explanatory background information

As stated before, many countries have developed background reports or national interpretations of the international indicators. From the way these presentations are organised, in terms of which issues are put together, the degree of connectivity with broader policy issues may be determined. For example, in a Dutch study, (cf. Luyten, 2000) various indicators on expenditure and participation in secondary education were assembled, in order to provide a further analyses of the alleged efficiency of the current secondary education system, in the Netherlands. A further analysis of these reports was expected to provide interesting suggestions for a more integrated and even more policy relevant presentation of the indicators. The analyses of these country reports or internal notes within Ministry’s of Education was also expected to yield the kind of explanatory background information provided to contextualise the presentation of the indicators.

Feedback to schools

Although most international indicators are defined at the national level, there is also a growing body of information that is collected from schools. Examples are the INES Network C primary school survey (1996) and the current school survey that was part of the PISA project. It was considered that the overall relevance of indicator systems would be enhanced if information feedback to schools could also be achieved. Methods that have been developed within various frameworks of school self-evaluation could be useful in exploring this possibility (cf. Tiana et al, 1999)

Thus the basic rationale and general aim of this study includes providing suggestions for optimising the impact of international indicators on educational policy and practice, assessing the current state of affairs on each of them and suggesting ways to improvement .

In more specific terms the objectives of the current study were:

To assess possible discrepancies in the way current education indicators address quality in education and priorities of policy planners, particularly in relation to European comparisons.

To analyse current practice and feasible future options for benchmarking and application of (quality) standards.

To assess needs for a more synthetic and “theme –based” presentation of indicators and explore a possible convergence of needs for country specific background information.

To explore the feasibility of methods of (survey) data feedback to individual schools.

The following methodological approach was used to attain the study’s objectives:

Desk research was carried out to compare and contrast specific sub-sets of indicators to relevant research concerning different perceptions of quality in education. This resulted in a conceptual chapter (chapter 1) that was used as a basis for a checklist which, in its turn, was used to question policy-planners, politicians and other stake-holders.

A similar approach (desk research and questioning of a panel) was used with respect to the analysis concerning the use of standards, and the exploration of the desirability of more precise, absolute criterion referenced standards.

Systematic content analyses of national interpretations of the INES and Key-Data indicator sets, was used to address the issue of different formats of presenting indicators.. A similar approach was followed in summarising the use of background or contextual information in the national reports. These results were also used for the development of the interview schedule for interviewing high ranking officers in the respective ministries of education.

Analytic study on the possibilities of feeding back information to lower levels in the educational system was considered and empirically studied on the basis of interviews

and a descriptive case study in a sub-set of countries (Belgium-Flanders, Italy and France).

Summary of the contents of the report

The report consists of five thematic chapters and country-based summary chapters of the interviews that were conducted with system level officials and politicians on the one hand, and school directors on the other.

Chapter 1, titled; “*Perspectives on education quality, education indicators and benchmarking*” provides six perspectives on educational quality:

a) the productivity view

According to this view, the success an educational system is seen as depending on the attainment of the aspired outputs/outcomes, for example in the sense of a satisfactory quantity of graduates that have attained a specific level (which may be formalised as a diploma), or in terms of an acceptable level of employment of students with a certain diploma. According to this view output/outcome/impact indicators are predominant or even the only type of quality indicators that need to be monitored.

b) the instrumental effectiveness view

According to the instrumental effectiveness view there is a clear perspective for the selection of context, input and process indicators, namely their expected effect on outcomes. To the extent that effectiveness or production functions can be completely specified, in other words outcomes can be totally predicted, context, input and process indicators could replace outcome indicators. The value of certain levels and forms of inputs and processes is determined by there instrumental potential. Clearly the instrumental perspective offers more dynamic handles for policy, as it considers not only given constraints but also factors that are policy malleable.

c) the adaptation perspective

This view “transcends” the instrumental effectiveness perspective by not only looking at the question how to do things right, but first of all considering the question on how to do the right things. In other words the adaptation perspective would lead to a critical analysis of educational *goals*. Conditions that allow for a continuous sounding of changing contextual conditions for the education province would receive emphasis as *means*, while labour market outcomes or “cultural capital” could be considered as *ends*, according to this view.

d) the equity perspective

When inputs, processes and outcomes are analysed for their equal or “fair” distribution among participants in education with different characteristics, equity is the primary facet of judging educational quality.

e) the efficiency perspective

This perspective can be seen as a further demand on the productivity and instrumental effectiveness view, by considering the highest possible outcomes at the lowest possible costs.

f) the disjointed view

Combinations or relations seen within a basic systems model, comprising inputs, processes, context and outcomes (see figure 1 in Chapter 1) were central in the previous views that represent a particular perspective on education quality. An alternative view is to consider each element “on its own” and judge whether it is manifested in an acceptable way, or at an acceptable level. In this way one could, for example, consider levels of teacher training, as a (minimum) requirement for being allowed to function as a teacher, class sizes could be judged in terms of being acceptable for being “manageable” units for teachers and students, and teaching strategies could be rated according to norms of good practice.

The disjointed view is descriptively the simplest one, although in an evaluative sense it is perhaps the most arbitrary one.

These perspectives were used in the further analysis of documents and also as underlying dimensions in the interviews-schedules.

In addition the first chapter placed indicators among other types of educational assessment, monitoring and evaluation and distinguished indicators systems defined at macro-level and multilevel indicator systems.

In the second chapter, titled “*Indicators on education published by the OECD, the European Union and France*” three indicator systems (those published in *Education at a Glance*, *Key Data* and “*The state of education*”) are compared to one another and to the various perspectives on educational quality. With respect to the latter it is concluded that the perspective in all three publications is predominantly the “disjointed” one, where each indicator is more or less presented “on its own”, without being related to other indicators. Nevertheless there are some differences between the three publications. *Education at a Glance* has a design that is built on the input, process, output and context model, whereas *Key Data* is more focused on description of the systems per se. Some detectable background orientations in the French “*The state of education*”, are

considered, namely democratisation and equality in education. The authors also express some views on the degree to which the various publications make the step from description to prescription and conclude that Education at A Glance goes furthest in making or implying policy recommendations, Key Data remains more descriptive. As far as the French indicator system is concerned the authors speak of “timid suggestions” that are being made.

The third chapter discusses “*The use of standards and benchmarks in indicator publications*”. An important part of the chapter is based on practices in the United Kingdom. The chapter highlights the fact that a lot of variation exists in the conceptual interpretations and definitions of these terms. The loosest sense in which the term standards is used is the one where standards are seen as identical to *educational objectives*. This pro-active interpretation of standards is also contained in the term *targets*, which perhaps has an added underlining of precision in the statement of the objective. Examples from the UK are provided in which targets have been formulated as absolute standards, for example: 80% of 11 year olds reaching the expected standard for their age in literacy. The chapter further states examples of specific statistics, for example country means, that are used as comparative standards in the various indicator publications. The term benchmarking is loosely used in the sense of “comparing to the best”. An important conclusion from the chapter is that no evidence was found of absolute standards in the reports on international comparisons, whereas at the country level, notably for the UK evidence was found of “expected” standards, similar to absolute standards. It is interesting to note that the current interest within the EU for European benchmarks in education is quite similar to this kind of “proactive” interpretation of absolute standards as noted in the UK.

Chapter 4 addresses “*National interpretations of the international indicators*”. In fact three kinds of national initiatives were found in the countries that participated in the study. The first approach consists of briefing notes after the release of new editions of EAG or Key data; applications of which were found in Denmark, Italy (only once), The Netherlands, Spain and the United Kingdom. The second approach consists of national version or reports on the international indicators, following a similar structure. These documents intend to adapt OECD or EU indicators to national or regional situations. Examples are publications from the Flemish speaking and the French speaking parts of Belgium and from Catalonia. The third approach is characterised as using international indicators for completing national indicators. Examples of these were found in Denmark, France, Belgium and the Netherlands.

As it comes to the frequency in which international indicator publication are used in these national publications there is an overwhelming predominance of EAG over Key Data. This latter publication has only been used in a few cases.

When the national publications are compared to the six quality perspectives there is again a strong prevalence of the disjointed perspective, because quality elements are considered on their own and not combined in specific interrelationships.

As stated in the above in many cases the national reports were more or less structured as summaries of the international publications. Only in a few cases the publications concentrated on a few core issues, generally reflecting the most pressing domains of educational policy at a particular time.

The OECD volumes titled “Education policy analysis” provide a more synthetic use of indicator based information. Re-occurring themes are expenditure, teachers and outcomes. The national reports do not show advances in the development of more integrated, composite indicators.

The fifth chapter of the report has looked specifically at the issue of “*Feedback of indicators to schools*”. It contains a case-study from Flanders. Feedback from the TIMSS and PISA study to schools was studied. The chapter addresses important methodological issues concerning the interpretation of school effects. Particularly when value-added outcome indicators are presented the within-country differences between schools differ only in a statistically significant way at the extremes of the distribution. The chapter also has graphic examples of the way achievement results can be presented to schools. At this stage what could be shown to schools was their relative position as compared to the country average, to averages of other, comparable countries, or to international averages. More ambitious types of performance feedback might differentiate results for different sub-groups of students (the idea of “differential effectiveness”) and try to include information on process indicators (school and classroom characteristics) to allow for a more instrumental interpretation. If this could be realised this would be a clear example of the educational effectiveness perspective in using indicators.

The information from interviews with school directors, collected in Belgium, France and Italy, further illustrates that actual use of international indicator information by schools is still an idea that has not been realised by far. Apart from the already mentioned technical complexities, the results from Belgium and France indicate that schools are not used to this kind of information, and treat it sometimes with a certain degree of suspicion. The Italian respondents have a more open and interested attitude, but it should be noted that they are all part of an innovatory project on school self-evaluation that has provided them with more background and experience in using data-feedback.

It was also concluded that general information provision about international indicators does not seem to be particularly “cut” to the needs of schools and therefore remains largely unnoticed for probably the bulk of schools in the countries concerned.

The sixth thematic chapter has looked at the *“Impact of indicators”* by means of an analysis of press-coverage on the international indicator publication in Italy. First of all it was noted that the most recurrent topics referred to indicators on financial and human resources and access to and participation in education, and in the Italian case on the transition from school to work and on teacher pay. For the rest it was noted that in the Italian context, where a national report on international indicators was published only once, the press is of dominant importance in the way the information from international indicators is selected and interpreted. The authors state that “Indicators are often used for conveying opinions that have nothing to do with the indicator itself, or even for influencing public opinion and choices regarding school policies”. This picture of an “eclectic” use of indicators was supported by the interview data. Choice seems to be determined by the educational issue that is in the focus of attention at a particular time for other reasons than the indicator publications. When it comes to selecting areas from the total of indicators, there appears to be a tendency to focus at those areas in which a country does poor. See for example the chapter on the Danish interviews.

The country-based chapters on the interviews with state level officials and politicians have looked at the degree to which the international indicator publications are known, and which, possibly selective, use is made of the information. In addition they have addressed the extent of more synthetic publications and composite indicators and the need for more absolute standards.

The following main results were obtained:

In general the interviewees across countries are more familiar with Education at a Glance than with Key data. Education at a Glance is considered to be a difficult and complex publication, top level officials usually depend on scientific advisors and analysts to advise them on the contents. These specialists have a more complete knowledge of the indicator publications.

Use of indicators is eclectic, depending on the policy issue that “happens to be important” at the time. Results that are “bad” in an internationally comparative sense may focus the attention.

From several countries (UK, Denmark, Italy France and Belgium) the respondents expressed an interest in more “qualitative” information. In some cases at least, this was interpreted in the sense of explanatory background information about the “why” of certain performance patterns.

Critical reflections about “cultural biases” were noted from France and Italy. These comments seem to challenge rationalist, technocratic and instrumentalist approaches, as well as quantitative measurement ideas, that are associated with an Anglo-Saxon background.

Respondents overwhelmingly reject the usefulness of absolute standards. They associate these with a kind of international uniformity in educational standards, which they

appear to reject. This result seems to stand in sharp contrast to the current prospect within the European Union to formulate European benchmarks.

A general comment, noted from some country level officials, was to enhance the role and participation of teachers in the development of international indicators.

More details about the results from the interviews are stated in the final summary chapter of the report, which ends with the following recommendations:

Concerning some important issues the results of this study would reinforce current practice rather than point at specific required modifications. A case in point are the national reports that use and incorporate the information from the international indicators publications in combination with national information. At this stage we see no reason to suggest changing this by including more country specific information in the international publications. This would only make them even more voluminous and “heavy” than they perhaps are already. Another important point of this nature is the rejection of absolute international standards and benchmarks; apparently continuation of the current use of relative comparative benchmarks is supported. As far as the wide spread comparative use of country averages is concerned it would be recommendable to provide additional information on the importance of these differences, for example by means of indicating confidence intervals.

Inclusion of micro-level information, a certain emphasis on process indicators and to relating these to outcome data, comes out as a fairly general recommendation. In this sense current practice, see the current PISA study and the long term further agenda of this project, comes close to meeting these wishes.

The strong demand for more “qualitative” information by several countries requires further study. Different interpretations are given to the concept of “qualitative indicators”. The overriding motives behind this requirement seem to be both more attention to micro level policies and a concern for doing more justice to the intricacies of national contexts. Possible answers to this should also address the, at present perhaps fuzzy, border between international indicators and research studies. For international organisations like the EU it could be considered a challenge to think of supporting research studies that follow up on the more general information that is attainable from international indicator projects.

The study confirms that there is a generally felt concern for making international education indicator-work more accessible to schools and teachers. In this area there are interesting challenges for projects that stimulate feedback to schools that have yielded data for the international studies, but perhaps also for projects that seek to fruitfully gear existing types of school based review and school self-evaluation to international data sets. In addition dissemination of the highlights of international indicators to schools is also considered as recommendable. Finally, it would be generally desirable to involve

teachers in the process of indicator development; e.g. by creating response-groups of teachers to review prospective indicators on schooling.

Perhaps not surprisingly, as it is usually one of the conclusions in international comparative research, also this study has incidentally referred to the importance of national cultural contexts in the meaning and interpretation of international data. Again this could be an interesting area for EU stimulated further study. Could we, for example, find internationally acceptable operationalisations of the concept of national education culture?

References

European Union (2002). Communication from the commission. European benchmarks in education and training: follow-up to the Lisbon European Council. Brussels: Commission of the European Communities.

Luyten, J.W. (2000). Sign of the times. Monitoring the position of Dutch education: the O8 project. Stage two: digging deeper. *Secondary education*. Enschede: University of Twente, CHEPS.

Scheerens, J., & Brummelhuis, A. ten (1996). *Indicators on the functioning of primary schools in twelve European countries*. Report on the Network C primary school survey. Paris: OECD.

Tiana, A. (ed.) (1999). *Innovative Approaches in School Evaluation*. Madrid: UNED.