Indicators of Education Systems

BENCHMARKING THE QUALITY OF LEARNING OUTCOMES:
SESSION 2 OF THE FOURTH INES GENERAL ASSEMBLY

ISSUES FOR DISCUSSION

Fourth General Assembly to be held in Tokyo
on 11-13 September 2000
TABLE OF CONTENTS

BENCHMARKING THE QUALITY OF LEARNING OUTCOMES SESSION 2 OF THE 4TH INES GENERAL ASSEMBLY  ISSUES FOR DISCUSSION................................................................. 3

Where do we stand today? *Looking back*.......................................................................................... 3
Where do we need to be? *Looking forward* ...................................................................................... 4
How can we get there and what can/should be the contribution of INES? *Challenges and issues*........ 5
  What should be the scope of instruments for monitoring learning outcomes in terms of the knowledge and skills to be assessed?.................................................................................. 5
  At what age-groups/ populations should future instruments be targeted?......................................... 6
  How can the analysis and utilisation of assessment results be enhanced? .......................................... 6
  How can different international survey instruments be more closely co-ordinated and how can their integration with national evaluation programmes be strengthened? ............................... 9
  How can INES find the right balance between substantive objectives and considerations of technical feasibility, resource requirements and cost-effectiveness .......................................................... 10

ANNEX .................................................................................................................................................. 12
  The DESECO programme on the definition, selection and assessment of knowledge and skills......... 12
BENCHMARKING THE QUALITY OF LEARNING OUTCOMES
SESSION 2 OF THE 4TH INES GENERAL ASSEMBLY

ISSUES FOR DISCUSSION

Where do we stand today? Looking back

1. A shift in public and governmental concern, away from the mere control over the resources and content of education toward a focus on outcomes has influenced both policy agendas and the practice of data collection and reporting. Periodic assessments of student performance are now common in many OECD Member countries, and the results are widely reported and used in public debate as well as by those concerned with school improvement.

2. At the 1995 INES General Assembly, Member countries noted the relevance of the international comparative dimension in assessment and evaluation and encouraged the OECD to develop international benchmarks that would allow assessing how far students approaching the end of compulsory education have acquired some of the knowledge and skills that are essential for full participation in society. Subsequently countries began, through the INES Network A, to develop methods for assessing students that would be strong at measuring relevant skills, provide a predictable tool for monitoring educational progress and guiding policy decisions, and ensure that comparisons would be valid across countries and cultures. The result of this work has been the Programme for International Student Assessment (PISA), which Member governments launched in October 1997.

3. The first PISA data collection cycle was successfully implemented in 2000 in 32 countries and included a comprehensive assessment of 15-year-olds in reading, mathematical and scientific literacy as well as selected cross-curriculum competencies. Beyond providing a baseline profile of the knowledge and skills of students, PISA was also designed to provide insights into how knowledge and skills relate to important educational, social, economic and demographic variables and how countries compare in terms of disparities in the access to educational opportunities and outcomes among students from various social and economic backgrounds. In addition to establishing a reliable measurement instrument, a particular strength of PISA has also been the creation of an effective and collaborative management structure which has brought together expertise from all OECD countries, steered jointly by their governments, on the basis of shared, policy-driven interests.

4. General agreement on the scope and design for the second PISA survey cycle was reached in March 2000 and almost all OECD Member countries, as well as a number of non-Member countries, have already signed up to implement this second cycle in 2003.

5. Eighteen OECD countries have also fielded the International Adult Literacy Survey (IALS) under the auspices of a joint OECD/Statistics Canada programme. This survey provided important data on the stock of skills in the adult population and the role that skills play in creating inequity in economic outcomes for both recent graduates and labour forces. Some countries are currently exploring to extend this work in the Adult Literacy and Lifeskills Survey (ALLS), a project jointly administered by the governments of Canada and the United States and the OECD.
Where do we need to be? *Looking forward*

6. While the consolidation of the PISA instruments and operations are generally considered the most important objective for the next years, some Member countries have expressed an interest to advance internationally comparative assessment in several aspects:

- International assessments provide a unique basis for policy dialogue and collaboration in defining and operationalising educational goals, in ways that reflect judgements about the skills that are relevant to today’s societies and economies. While the first PISA cycle has offered a useful starting point, countries seek to widen the scope of assessments (including a better coverage of cross-curricular competencies) in order to increase the relevance and cross-cultural validity of outcome measures. Countries are also seeking benchmarks of educational quality that are better interpretable in policy terms (including the establishment of proficiency scales).

- Measuring the quality of educational outcomes leads to the question of which educational, social and economic factors contribute to improved education and learning and enhanced returns to investment in education. Analyses that help to explain important relationships between inputs and educational outcomes and promote an understanding of how such relationships are mediated by factors such as instructional and learning processes or educational and social contexts are frequently stated as important objectives for international comparative assessments. They will also be a prerequisite for bringing assessments closer to teachers and schools, which many countries consider another important objective for the years to come (see also Session 4 of the General Assembly).

- Monitoring trends in both absolute performance standards of students, school systems and countries as well as how effectively and equitably education systems operate is a function which all countries expect PISA to serve. This requires stability and predictability in the survey cycle.

- Increasing attention is paid to linking comparative assessment with more targeted qualitative studies. The widespread interest in and utility of the TIMSS video study is an indication that the more in-depth, qualitative information that can be gained through video studies and case studies may become an important complement of assessment in the future. Such efforts can provide detailed information about teaching and learning, the quality of life in schools or parental and student satisfaction that are beyond the reach of current assessment instruments.

- While the development of school-based assessments is on a promising track, there is also need for better information on key transitions within education systems and between education and employment.

- Finally, many countries seek better information on the profile of skills in the adult population and how skills compare for various subgroups within countries; on the relationships between skills and industry, occupation and technology of production; on the supply of skills and mechanisms of skill acquisition and loss over the lifecycle; on how markets for skills function with respect to productivity, employment and wages; and on the social distribution of skills and their impact on economic and social outcomes.

7. The INES General Assembly provides an important opportunity to review the demand and supply of comparative benchmarks of educational performance and to determine priorities for OECD’s future developmental work in this area.
8. In considering priorities for INES, the General Assembly should take into account activities pursued by other organisations, such as the European Union or APEC, as well as research associations such as the International Association for the Evaluation of Educational Achievement (IEA). Whatever substantive directions future development takes, a central objective must be to strengthen the conceptual and managerial co-ordination of the different international assessment programmes as well as their interrelationships with national evaluation and assessment. Effective management and governance structures will be needed that contain risks for participants, ensure broad input, and provide a mechanism for sharing research and development costs.

How can we get there and what can/should be the contribution of INES? Challenges and issues

9. The General Assembly should seek to establish what the most important policy issues are that Member countries seek to address with internationally comparative outcome measures. Considering that many of the goals and objectives set out in the preceding section involve high costs of development and implementation, clear direction on what the priorities are and how they might best be achieved is needed. An unbridled expansion of international assessment will make it difficult to finance and manage these activities and may actually detract from their effectiveness for policy.

10. In establishing priorities for INES, the General Assembly should also review the relationship between national and international assessment activities. It is important that he focus of INES remains on areas where international comparative evaluation adds substantial value to related national work.

11. Some questions which the General Assembly might consider in addressing these questions are set out below.

What should be the scope of instruments for monitoring learning outcomes in terms of the knowledge and skills to be assessed?

What are the most important knowledge and skills that Member countries consider relevant for today's societies and economies, and for which they seek internationally comparative measures and analysis? How should these knowledge and skills be selected, defined and operationalised so that they can be measured in a cross-culturally appropriate and statistically valid way?

12. While recognising the relevance of knowledge and skills in a broad range of domains, Member countries agreed in 1997 to focus the first PISA survey cycle on the assessment of important literacy skills, based on a new and innovative concept of literacy that encompasses reading literacy, mathematical literacy, scientific literacy and technological literacy, and that emphasises the ability of students to apply knowledge in a range of situations, backed by a broad understanding of key concepts, rather than the possession of specific subject matter knowledge. Member countries considered an assessment of these literacy skills as particularly relevant for a cross-national comparison of schooling outcomes, sufficiently founded on scientific theory and sufficiently feasible in the field.

13. At the same time, Member countries are conscious of the relevance of a much broader range of subject areas, including foreign languages and the social sciences for students’ success in life and the importance of cross-curricular competencies including problem solving and the ability to work in teams. A survey of educational goals that the INES Network A undertook in 1996 revealed the importance that education systems place on ensuring that individuals not only acquire knowledge in varied fields, but that they also develop abilities with which to evaluate, compare, and critique; abilities with which to imagine,
hypothesise, discover and invent; and abilities with which to put into practice knowledge and skills. An important concern of education systems is also to ensure that individuals develop the capacity and the motivation to continue learning throughout life. To achieve this, they need to be able to manage their own learning - to set goals, to persevere, to monitor their progress and to adjust their learning strategies as necessary.

14. Whether intentionally or not, evaluation and assessment can have an important normative impact on curriculum development and instruction. When measuring the outcomes of education, it is therefore important that measures carefully reflect desired learning outcomes. A focus on isolated subject matter areas bears the risk that attention - and resources - in education systems are detracted from outcomes that are important but perhaps difficult to measure.

15. Member countries are therefore striving to extend the coverage of PISA subjects and to progressively also include the assessment of cross-curricular competencies. Similarly, the international Adult Literacy and Lifeskills Survey (ALLS) seeks to capture a wider range of knowledge and skills than was the case in the first International Adult Literacy Survey (IALS).

16. It will be important that the widening of the scope of future assessments is guided by an understanding of - and agreement on - the knowledge, skills and attitudes that are most relevant. The establishment of a broad framework for the definition, selection and assessment of knowledge and skills, within which existing measures can be situated and their validity assessed, could therefore be an important objective for INES. The Swiss-led DESECO initiative has been a first step into this direction (see Annex). At the same time, the desired understanding and agreement cannot be achieved without some collection of empirical data. Such data, collected in domains where theory and empirical evidence suggest that comparative data can be provided, will serve to validate key aspects of competing theories.

**At what age-groups/ populations should future instruments be targeted?**

Should PISA be extended to younger and/or older age groups? Should the assessment of school systems be complemented with an international assessment of adult skills as part of the OECD data strategy? Should the development of a longitudinal PISA component be a future objective?

17. These are important questions that will critically influence the analytic power of the OECD assessment strategy in the medium and long term. For example, school-based assessments can contribute to assessing how effectively and equitably education systems function, what the key determinants of educational performance are, and how the delivery of education can be improved. Studies of adult skills can extend this picture and examine how knowledge and skills play out over the lifecycle on various economic and social dimensions. Longitudinal studies provide a means of assessing the extent to which the knowledge and competence acquired through school do, in fact, help to prepare young people for adulthood and the transition to work. While adding to the analytic potential, each additional target population will bring with it significant increases in costs at both national and international levels and thus needs to be carefully weighted to ensure that policy priorities and resources are well balanced.

18. The General Assembly should provide directions concerning the target groups for OECD’s future assessment work, considering both the work currently undertaken within the INES framework as well as that undertaken by other organisations and research associations, most notably the IEA.

19. There is general agreement among countries, that the current focus of PISA on students towards the end of compulsory schooling should remain. At the same time, some countries have indicated an
interest in adding an assessment of children during the early years of primary education, both in order to
gain better insight into skill acquisition in the early school years, and to obtain measures of growth
between the primary-age cohort and the PISA population at the end of compulsory schooling.

20. Also the assessment of more advanced levels of education, including general and vocational
programmes at the upper secondary level of education and, indeed, tertiary education programmes have
been identified as areas of policy interest - although assessments at these levels of education are largely
unexplored territory.

21. A particularly important question that Member countries need to answer is whether a survey
approach to the assessment of adult skills should be undertaken within the framework of INES or whether
it would better be left to those concerned with the labour market. There is keen interest among a number of
countries to explore the feasibility of adapting techniques used to assess student competencies to assess
adult competencies. However, this has proved to be difficult: Survey-based assessments of student
outcomes can usually rely on learning objectives to be relatively homogeneous. This makes it possible to
define objectives and measures of competencies. In contrast, the task of assessing meaningfully the
competencies of adults is far more complex conceptually and methodologically. For example, competence
levels for adults may depend on individual preferences as well as requirements that are specific to
occupations and firm-specific features related to work organisation and human resource management
practices. Moreover, there do not exist yet the “institutional arrangements” - analogous to formal education
systems - for evaluating the competencies of adults (in terms other than formal educational attainment),
and applying the results.

22. If agreement is reached that an adult skill assessment should become a goal for INES, new
questions arise: Should INES search for a comprehensive assessment of a broad range of adult skills or is
the strive for a genuine assessment of “Life Skills” an illusionary goal, for example, because of differences
in the relevance of skills in different societal, cultural or occupational contexts or simply for practical
reasons and constraints? Should INES choose, instead, a more targeted approach to assessing adult skills
and if yes, what would be an appropriate policy context and reference frame? In addition, should the focus
be on all adults (at the risk that the instruments loose in resolution) or should the focus be on specific age
groups, e.g. the youth population and young adults?

23. Beyond the exploration of additional target populations, some countries have shown an interest in
extending student assessment with a longitudinal follow-up. The essence of longitudinal data is that the
same people are surveyed on a regular basis over time. By surveying the same young people over time it is
possible to build an understanding of the changes taking place in their lives - and the ways that previous
achievements and experiences influence what is happening to them now.

24. The combination of a longitudinal school-to-work transition survey with PISA provides an
attractive potential for policy analysis: PISA aims to assess the degree to which 15-year-old students,
approaching the end of compulsory education have acquired the knowledge and skills that are essential for
full participation in society. A longitudinal survey would allow the examination of: how the skills
measured through PISA play out in the transition from education to work and further education; the
attitudes, aspirations and behaviour of youth entering the labour market; the economic and social benefits
from participation in different forms of education and training; the patterns of transitions from initial
education to subsequent education and training and the labour market; as well as the exploration of how
access to post-secondary education and initial labour market success are conditioned by achievement,
formal qualifications, and social origins.
25. On the other hand, while longitudinal surveys offer considerably more information value per unit investment than repeated cross-sectional surveys, this needs to be balanced against the fact that longitudinal surveys have a very long payoff period in policy terms and are complex to analyse.

**How can the analysis and utilisation of assessment results be enhanced?**

*What can/should be done to improve the dissemination of assessment results at national and international levels? How can international comparative assessments be made more useful for supporting national development and analysis? And how can they be more closely related to teaching and learning processes in the classroom?*

26. The value of international assessments for policy development depends, in part at least, on the analysis of the results and how effectively these reach the actors in various settings, including policymakers and analysts as well as school authorities, teachers and researchers. This is not only an issue of dissemination and public relations but also, and perhaps more importantly, one of developing the right instruments that provide insight into important individual, instructional and institutional factors that influence learning and learning outcomes. Traditionally, the design and instrumentation of international assessments have been weak in this respect.

27. As an instrument that focuses more closely on public and government policy concerns, PISA seeks to progress in this respect and to yield clearer signals for the evaluation of different investment strategies in education, for informing education policy on how to improve the effectiveness of education and learning, and for ensuring that relevant criteria for the successful evaluation of education systems are selected. To achieve this, PISA examines:

- differences in education systems and the instructional context that account for differences in student performance among countries;
- differences in achievement among countries in terms of content and processes;
- the extent to which schools and teachers can moderate the relationship between students’ performance and the economic, social and cultural capital of their families.
- the proportion of variation in proficiency between (rather than within) schools and differences in this value across countries;
- relations between school factors such as size and resources, and student proficiency levels - and the differences between countries in these relations;

28. However, the development of methods for establishing linkages between performance measures and learning processes is, in many areas, still at an early stage of development. In particular, the process of teaching, which is at the heart of education and the expertise, pedagogical know-how, organisational and technical competence of teachers are all skill beyond the reach of PISA and other internationally comparative assessment programmes, such as TIMSS.

29. The development of a survey on teachers, teaching and instructional technology has been considered by the INES Network C as one way to overcome these limitations and a group of countries will pursue the development of an optional teacher survey in the second PISA cycle as a first, and limited step,
into this direction. The implementation of a full survey on the delivery of education, however, would be an undertaking in its own right that would have significant resource implications.

30. Advice from the General Assembly on priorities for analysis and the need for additional instruments will be important for the future development of international comparative benchmarks.

**How can different international survey instruments be more closely co-ordinated and how can their integration with national evaluation programmes be strengthened?**

31. A framework for the definition, selection and assessment of knowledge and skills can provide the basis for the conceptual development and integration of international surveys. Clear analytic priorities will allow Member countries to decide on target populations and the relative weight of future school and adult skills surveys. A third, but no less important step will be to ensure the articulation, complementarity and co-ordination of the different and varied international surveys (such as PISA, TIMSS, ALLS, IALS and PIRLS) through the different providers (such as the OECD, the IEA and Statistics Canada).

32. Existing instruments differ in many aspects, including the target populations, the breadth and depth of the measurement instruments, the analytic objectives and the assessment methods. It is neither feasible nor cost-effective to develop a single instrument that will address all of these purposes at all times. This would be like building cars that can also swim and fly – and trying to do so might yield instruments that will do none of this effectively and become prohibitively expensive. Instead, the focus must be on ensuring close co-ordination among the different instruments and complementarity in their analytic potential.

33. What steps could be undertaken to strengthen the co-ordination between different assessment and evaluation efforts? The General Assembly might consider several aspects of this question, including:

- **How can different instruments be integrated conceptually?** It will be important to ensure coherence in the assessment frameworks and assessment methods. While instruments such as TIMSS and PISA have a very different conceptual orientation - with TIMSS focusing on the common denominator of national curricula and PISA emphasising a broader and cross-curricular approach to knowledge and skills - it is important to articulate the commonalities and differences between the two approaches so that given the costs, the benefits of participating in each survey become apparent to both the participants and users of these surveys. The conceptual co-ordination between PISA and TIMSS remains unsatisfactory. PISA and IALS/ALL, on the other hand, have achieved some degree of conceptual integration in both the reading and problem solving domains. In the case of reading, integration has extended to the operational level so that it will, for a subset of the items, be possible to link PISA results with the adult skill distribution obtained from IALS. Nevertheless, much further progress is needed to fully capitalise on the synergies of the different assessment activities.

- **How can international and national assessment be more closely interrelated?** It is of utmost importance to ensure that national and international assessments are closely integrated, both to ensure the best possible use of international results for national policy development and to minimise costs and response burden on schools. Otherwise, there is the risk that both time and resource constraints will limit participation by countries and schools in these activities by forcing choice, to the detriment of this important work. While OECD countries have decided that PISA should not be driven by national curricula or their common denominator, but for the interpretation of international benchmarks it is important for
countries to understand where national curricula are situated within the international assessments. In the longer term, explicit linkages and comparisons of international assessment results with national estimates could perhaps be sought through the inclusion of intact blocks of international assessment items in national assessments or vice versa.

− How can the design and timing of international surveys be better co-ordinated?
Insufficient co-ordination among different surveys has resulted in increased costs and duplication of operations in participating countries and placed great burden on participating schools in terms of access to student response time and dealing with differences in assessment procedures. The co-ordination of the survey and sampling operations between the second PISA survey cycle and the second TIMSS repeat in 2003, that is currently being negotiated between the OECD and the IEA, is a first step towards improvement. Further progress will not only depend on the international organisations but also on Member countries who, ultimately, decide on the priorities and timing of each of the surveys.

− How can the operation and management of different surveys be more closely co-ordinated at both national and international levels to maximise synergies and minimise response burden and costs on countries? Issues of institutional structures and management need to be addressed. PISA has set an example for international collaboration in the development and implementation of comparative assessments. Participating countries are shaping the programme in various ways: as members of the Board of Participating Countries they determine the policy objectives and broad design parameters for the assessment project and guide analysis and reporting; as participants in functional expert groups, they contribute to the development of the conceptual framework and the assessment instruments together with the contracted international consortium; and through National Project Managers, they implement the project at the national level subject to the agreed-upon administration procedures. In addition, through the INES Network A, Member countries continue to engage in long-term development work and to ensure the coherence of the survey programme over time. All countries contribute to the financing of international activities based on an agreed formula. While leaving room for further improvement, might this serve as a model for the management of future international surveys, including those for the assessment of adult skills?

How can INES find the right balance between substantive objectives and considerations of technical feasibility, resource requirements and cost-effectiveness

How can INES find the right balance between the different, and sometimes competing constraints?

34. In considering the substantive objectives for future assessments, the General Assembly needs to consider questions of technical feasibility, resource requirements and cost-effectiveness when establishing priorities. Tradeoffs between various factors will be needed. For example:

− Increasing the scope of assessment instruments in terms of the knowledge and skills to be tested may lower the depth of information obtained for any single assessment domain. In addition, for some of the assessment domains for which benchmarks are desirable from a policy perspective, reliable and comparable measures do not yet exist (e.g. for the assessment of inter-personal skills) or are prohibitively expensive for large scale assessments.
− Placing a premium on the accuracy of skill measures or the authenticity of measures may detract response time and resources from obtaining important contextual information (e.g., on students, teachers and schools) that ultimately determines the analytic power of the instruments.

− There is also a tension between ensuring that surveys remain responsive to advances in theory and measurement and the need to ensure coherence of assessments over time that is required to establish reliable measures of trends.

− While the number of target populations in a survey greatly enhances its analytic power, each target population adds significantly to the costs and complexity of the survey instruments.

− There may also be a tension between international comparability and national relevance of assessment measures.

− The frequency with which assessments can be feasibly undertaken needs to be weighed against the use that is made of the results for policy development.
ANNEX

The DESECO programme on the definition, selection and assessment of knowledge and skills

35. The establishment of a broad framework for the definition, selection and assessment of knowledge and skills, within which existing measures can be situated and their validity assessed, has been the goal of the Swiss-led INES DeSeCo activity. Such a framework is sought to provide a basis for:

- working towards a better, and more operational, understanding of the competencies that are important for today’s societies and economies and the societal mechanisms through which they are defined;
- establishing coherence in how assessment domains are defined and operationalised by different instruments and, in particular, providing a bridge between school assessments and assessments of adult skills;
- further advancing the theoretical underpinning of the measurement of skills and competencies;
- seeking reference points for the validation of assessment indicators and for more accurate and appropriate interpretation of empirical results (in that existing measures can be interpreted within a wider framework);
- providing feedback for education policy by helping to ensure that relevant criteria for the successful evaluation of education systems are selected on the basis of a theoretically grounded framework;
- providing the theoretical foundation for the integration of instruments for the assessment of student competencies and future measures of adult skills.

36. DeSeCo has considered various approaches to establishing such a framework and has identified a number of aspects which a future framework for the measurement of learning outcomes might cover [see also document GA(2000)12]:

- These include basic literacy skills, including reading/writing literacy, mathematical literacy, scientific literacy and the use of modern information and communication technologies.
- They also extend to the skills that allow individuals to cope with complexity and to solve problems of various kinds. Such skills reflect the need for individuals to be able to analyse and recognise patterns and establish analogies between experienced situations and new ones in order to cope with change and to deal with the ambiguity and diversity that is characteristic of modern life.
The capacity to perceive situations appropriately and comprehensively, to discriminate between relevant and irrelevant factors, to manage emotion and motivation are other aspects highlighted by DESECO. This involves various intra-personal skills, including motivation and meta-cognition.

DESECO also considers interpersonal skills and highlights that successful participation in modern life assumes the capacity to join, form and function effectively and democratically in multiple, complex and heterogeneous groups.

Finally, DESECO highlights the importance for individuals to be able to assess, to choose proportionate means in order to reach given ends, to appreciate various possibilities offered, to make moral judgements and to apply them.

Obviously, many competencies are interdependent with the economic, social or cultural contexts of individuals. Furthermore, the relevance of competencies may change across these contexts or across different parts of the lifecycle. The implication is that a future framework needs to be sufficiently broad, with different aspects relevant to different types of assessments, particularly as assessments relate to different target and age groups.

It is also clear that limitations in assessment methodology and considerations of cost-effectiveness will continue to constrain the scope of comparative assessments. However, despite the fact that it is yet to be demonstrated that a satisfactory theoretical framework for the measurement of knowledge and skills can, in fact, be constructed, and despite the fact that many of the facets of such a framework will remain beyond the reach of large-scale assessments for some time, work on this will allow to evaluate and situate current outcome measures and assess their relevance for measuring the quality of educational outcomes. This, in itself, could be seen as an important goal for INES.