EU-AEI

EUROPEAN NETWORK FOR EDUCATIONAL RESEARCH
ON ASSESSMENT, EFFECTIVENESS AND INNOVATION

http://www.to.utwente.nl/prj/euaei/index.htm

University of Twente

FINAL REPORT

EUROPEAN NETWORK FOR EDUCATIONAL RESEARCH ON ASSESSMENT, EFFECTIVENESS AND INNOVATION (EU-AEI)

By

W.J. Pelgrum

May 1998

ACRONYMS OF PARTICIPATING MEMBER INSTITUTES

Country	Acronym
Austria	USALZ.IES
Belgium	RUG.VO
Belgium	ULGG.DEPF
Germany	UBER.AP.AEB
Denmark	DPI
Spain	NCE
Finland	UJYV.IER
France	MEN.DEP
United Kingdom	NFER
United Kingdom	ULON.IE
Greece	UAT.DPED
Ireland	SPC.ERC
Italy	CEEDU.PSF
Luxembourg	MENFP.IPT
Netherlands	UAM.SCOKI
Netherlands	RUGR.IER
Netherlands	THT.OCTO
Portugal	IIEPT.ASE
Sweden	NAE.DE

CONTENTS

1 BACKGROUND AND OBJECTIVES OF THE NETWORK

This chapter contains a description of the background of the network, the reasons for founding it, the objectives set forth for the first two-year work programme, and the associated member institutes.

Background of the Network

In April 1995 the European network for educational research on Assessment, Effectiveness, and Innovation (EU-AEI) was founded. The members of the network are educational research institutes from all EU Member States, while Iceland, Cyprus, and Switzerland are represented as associated members. The mission of the network was at the network's founding meeting defined and described as follows:

> *To carry out international comparative research in education in the domain of achievement, assessment, effectiveness and innovation within Europe.*
> *This orientation of the network implies that there is a key interest in the collection and analysis of both outcome oriented data at the level of individual students, and educational background conditions at the level of students, teachers, classroom arrangements, curricula, school management and organisation, and system parameters at the regional and national level (including specific educational policies).*
> *Substantive questions that are at the basis of the network's research interests are questions about micro-, meso- and macro-level conditions that enhance educational performance, and questions about the potential for educational change and innovation.*

In 1996, the network acquired for a period of two years, a grant from the European Commission, which allowed for strengthening its infrastructure and dissemination of its knowledge base.

Objectives 1996-1998

At the outset of the past two years (April 1, 1996-March 31, 1998), the network aimed at providing improved access to the knowledge base and empirical data on educational assessment and effectiveness, which were collected in order to contribute to a better insight into the state of the art of different aspects of and possibilities for improvement of education in Europe.

The more specific aim of the network for the period 1996-1998 was to realise the following set of products:

- Electronic data base (accessible via World Wide Web) containing information about educational systems characteristics, descriptions of the participating research institutes, their research programs and their research projects.
- Exchange and integration of conceptual frameworks regarding educational indicators, and the use of expertise from the partners for producing national and European overviews of statistical information on Mathematics-, Science-, and Technology-education.
- Exploration and identification of possibilities for cooperative pilot projects.

The rational underlying this first workplan was as follows:
Given the available funding resources, the network could not engage itself in extensive new substantive activities. Rather, the network would in its first two years of existence especially focus on creating an infrastructure for research on assessment, effectiveness and innovation. With regard to the thematic embedding, the network intended to concentrate its activities on the issues which have a high priority in the European research agenda.

For the first two years of the network it was seen as crucial that a well-organised infrastructure will be established and that the relevance and added value of the network would be demonstrated. The network should not only be perceived externally as relevant, that is the products from the network should convincingly demonstrate the added value of European cooperation and integration. In addition, and equally important, the internal relevance of the network should be clearly demonstrated, that is by the end of the first two years the participants should perceive their input to the network as paying off in terms of what they receive in return, for example in terms of sharing of expertise, and tools. This would constitute the basis for consolidating the network and for broadening and intensifying the scope of its activities in the future.
Moreover, the network should be able to demonstrate that it can contribute to the research issues as defined by the European Commission.

It was expected that the added value of the data base would consist of an improvement of the current situation with regard to the dissemination and accessibility of up-to-date information about ongoing and previous research efforts in the area of assessment, effectiveness and innovation, with a special emphasis on the research themes which are high on the agenda of the European Commission. As such, this information base could contribute to establish a basis for a system of European educational indicators. It was expected that especially this last aim would require careful tuning to projects and institutes inside and outside the EU, which are especially dedicated to developments in the area of educational indicators. Hence it was expected that the coordination unit would need to communicate intensively with these third parties.

Statistical Overviews

Most network partners have been and are cooperating in international comparative research projects. This cooperation involves several areas, such as education in mathematics, science, new technologies, civics, and foreign languages. The resulting databases, which are standardised across countries provide, in principle, ample opportunities to examine several research themes which are high on the European research agenda. Therefore, the network intended to stimulate that the partners start to cooperate to produce comparative statistics for these themes. At the start of the first two-year period it was seen as somewhat problematic to concretely plan this activity, as there was no systematic overview available of the content and priorities of the national reporting plans of the partners with regard to the ongoing international assessments in which they were involved. Therefore, it was difficult to estimate to which extent such network activities could be performed almost budgetary neutral, which would be the case if the network partners did already plan to address these themes, and the collection of European statistical results would be a matter of exchange of information. In case the network activities for producing European statistics would put too much pressure on the national project budgets, additional funding for assisting partners in conducting these activities would be necessary.

Exploration of Promising Pilot Projects

This activity was intended to follow from the foregoing activities. Originally it was foreseen that the following projects would be undertaken:

- comparison of multi-level school effectiveness models based on existing data from a few countries;
- trend analyses regarding math, science and technology education;
- pilot studies regarding the validity and reliability of new indicators.

The Associated Member Institutes

From the outset the network has capitalised on involving partners which have a considerable experience in conducting international comparative assessments of educational progress. Such experience might be gained in the context of studies conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) and/or the Organisation for Economic Cooperation and Development (OECD).

Table 1

The network's member institutes and representatives

Country	**Institute**	**Member**
Austria	University of Salzburg	Günter Haider
Belgium	University of Gent	Christiane Brusselmans-Dehairs
Belgium	University of Liège	George Henry
Denmark	The Danish National Institute for Education Research	Poul Skov
Finland	University of Jyväskylä	Jouni Välijärvi
France	Ministry of National Education	Gérard Bonnet
Germany	Humboldt University of Berlin	Rainer Lehmann
Greece	University of Athens	Georgia Kontogiannopoulou-Polydorides
Ireland	St. Patrick's College	Thomas Kellaghan
Italy	Centro Europeo Dell 'Educazione	Roberto Melchiori
Luxembourg	Grand-Duché de Luxembourg	Jean-Paul Reeff
Netherlands	University of Groningen	Bert Creemers
Netherlands	S.C.O.-Kohnstamm Institute for Educational Research	Kees de Glopper
Netherlands	Center for Applied Educational Research (OCTO)	Hans Pelgrum / Jaap Scheerens
Portugal	Instituto de Inovacao Educacional	Borges Palma
Spain	Instituto Nacional de Calidad y Evaluación	Guillermo Gil
Sweden	Skolverket	Sten Pettersson
United Kingdom	University of London	Caroline Gipps
United Kingdom	National Foundation for Educational Research	Seamus Hegarty

Country	**Institute**	**Associate Member**
Cyprus	University of Cyprus	Constantinos Papanastasiou
Iceland	Institute for Educational Research	Einar Gudmundsson
Switzerland	Swiss Coordination Centre for Research in Education	Armin Gretler

The board of the network consisted of:

- Jaap Scheerens - The Netherlands (chair);
- Seamus Hegarty - United Kingdom;
- Georgia Kontogiannopoulou-Polydorides - Greece;
- Hans Pelgrum - The Netherlands (ex officio).

2 ACTIVITIES AND OUTCOMES

This chapter describes the activities which were undertaken in the network and the outcomes of the work in terms of the main products which were planned for the period 1996-1998.

Activities

This section reviews the meetings which took place in the past two years and the decisions about cooperative activities which resulted from these meetings.

The following meetings guided the elaboration and review of the work plans of the network:

- Board meeting April 1996
- Plenary meeting October 1996
- Board meeting April 1997
- Plenary meeting September 1997
- Plenary meeting March 1998

The content and results of each of these meetings will be shortly reviewed below.

Board Meeting April 1996

The board met on April 26 in Amsterdam in order to make decisions on the steps to be undertaken in order to execute the workplan which had been negotiated with the EU and which contained the global substantive specifications of the work to be carried out by the network between April 1, 1996 and April 1, 1998.
The workplan of the network as specified in the EU-contract distinguished 3 main products which the network would deliver, namely:

1. an electronic database;
2. an integrated conceptual framework regarding the areas the network is focusing on, and
3. European statistical overviews.

The board discussed the steps to be undertaken for initiating work with regard to the first 2 products and decided that work related to the last product would be reviewed in a later stage.

Ad 1 Electronic database

- The information demand on partners should be minimised. The proposition was to present a model for the partners. The Network Coordination Center (NCC) would present a model for the partners, showing what kind of information we would be looking for. Also a list with 8-10 keywords would be produced, in order to make concrete the meaning of words like 'assessment'.
- In order to minimise the demand on partners for delivery of information, it was decided that the NCC would first scan available information-sources (such as Eurydice) in which some of the information already might be included. Then, when a short description of the countries would be available, it would be send to the partners to check how up to date the information was and to indicate if additional specification would be needed.
- The partner institutes would be asked to describe the subset of research projects relevant for our network in their institutes on the basis of a set of keywords.
- Hans Pelgrum would elaborate (together with Georgia Kontogiannopoulou-Polydorides) the procedure and seek approval of the board before approaching the members of the network (end of June).

Ad 2 An integrated conceptual framework

- A paper describing a proposal for a common conceptual framework and definitions of core concepts would be elaborated by Jaap Scheerens and Seamus Hegarty (early September). Seamus would share the outline of a project which NFER is conducting and which might be useful.
- The countries would be asked to identify indicators which were crucial (at plenary meeting early October 1996).
- The results from the previous steps would be used to plan the production of comparative European statistical overviews (year 2 of the network).

The board also discussed possibilities for acquisition of new projects from TSERII and/or Socrates.

With regard to the links with other international organisations, it was decided that:

- *IEA*: The coordination center would informally discuss with the chair of IEA in which ways the network might be linked to IEA.
- *European network policy makers*, coordinated by the French ministry of education: The NCC would send information on our network to this group and would also refer in the network's newsletter to the work of this group. Moreover the NCC should design a brochure/newsletter about the network.
- *EU-projects*: Earlier a suggestion was made by Brussels that the network should function as a kind of background network for area II of TSER II. The network's contact person in Brussels should be consulted on how this idea might be further implemented.

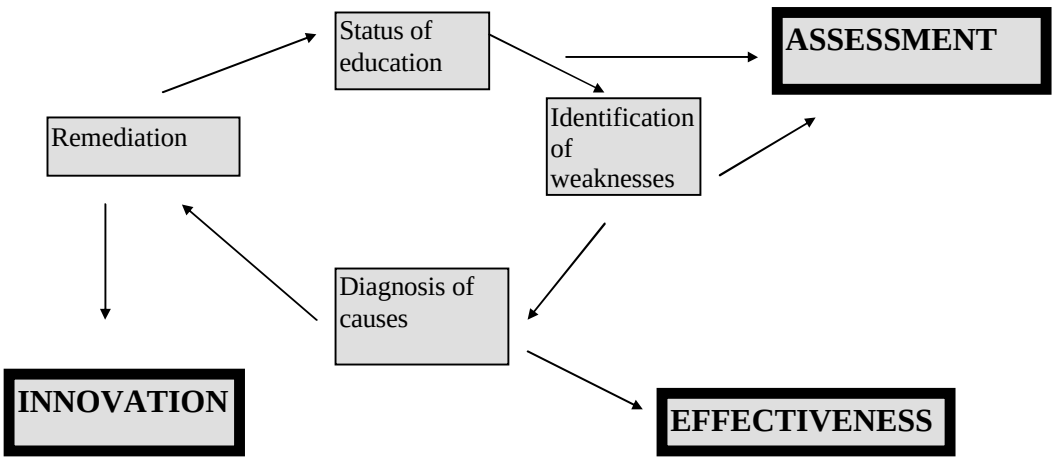
Furthermore the board discussed requests from Iceland, Norway, Switzerland and Cyprus to join the network. It was noted that these countries could not become formal members, because the contract with the

Commission was already signed. Therefore, it was decided that a summary would be written, in cooperation with Dr. Van den Brande, stating how they might take part in the network.

Plenary Meeting October 1996

This meeting was attended by: Gerard Bonnet, France, MEN.DEP; Christiane Brusselmans-Dehairs, Belgium, RUG.VO; Carmo Climaco, Portugal, IIEPT.ASE; Bert Creemers, Netherlands, RUGR.IER; Ruben Fukkink, Netherlands, UAM.SCOKI; Caroline Gipps, United Kingdom, ULON.IE; Armin Gretler, Switzerland; Günter Haider, Austria, USALZ.IES; Seamus Hegarty, United Kingdom, NFER; George Henry, Belgium, ULGG.DEPF; Monique Kole (notes), Georgia Kontogiannopoulou-Polydorides, Greece, UAT.DPED; Rainer Lehmann, Germany, UBER.AP.AEB; Mark Morgan, Ireland, SPC.ERC; Hans Pelgrum, Netherlands, THT.OCTO; Sten Pettersson, Sweden, NAE.DE; Renata Picco, Italy, CEEDU.PSF; Jean-Paul Reeff, Luxembourg, MENFP.IPT; Pam Sammons, United Kingdom, ULON.IE; Jaap Scheerens (chairman), Netherlands, THT.OCTO; Poul Skov, Denmark, DPI; and Jouni Välijärvi, Finland, UJYV.IER.

The **progress report**, which was distributed to the participants at the meeting, was reviewed. It was especially emphasised that in the work so far and also in future work the cohesion between the main areas of the network (Assessment, Effectiveness, and Innovation) should be maintained. In fact the cohesion between the network's study areas followed from the concept of educational monitoring for school improvement:



The meeting underscored the importance which ministries attach to the further analysis of existing databases.

It was decided that the network would focus primarily in this stage (**of creating statistical overviews**) on actions which would result in statistical comparative overviews which could be produced on the short term, on the basis of an inventory of existing international data bases and consultation of the European Commission representatives. For determining which actions could be undertaken on the medium and long term, input from Brussels would be sought.

The meeting agreed with the plan to use WorldWideWeb as basic infrastructure for information storage and exchange.

The draft paper on the **conceptual framework** was presented and reviewed. The paper explained definitions and interrelationships between concepts underlying the main goals of the network with regard to research on assessment, innovation and effectiveness. It was agreed that the paper constituted a good starting point, and that some further elaborations were needed.

The background of the document on data base construction in the network was explained, and it was pointed out that the network needed to show what information could be made available (in terms of indicators), what was known about the quality of these indicators, which indicators were already available and which needed to be developed. Therefore, it would be necessary to create an inventory of available data sets and a description of the contents for those data sets to which the members had direct access and which were created in the last 5 years. For this purpose each institute would need to fill out 'database information sheets'. The coordination center would take care of describing international comparative data sets.
During the discussion it was pointed out that the conceptual framework and the inventory needed to be linked. It was agreed that the coordination center would provide more detail on criteria for inclusion of data sets (representativeness, which educational levels, which categories of indicators had to be represented) Furthermore the meeting agreed that during the first year a minimum number of data sets would be entered in the data base, and that after an evaluation a decision on elaboration would be taken.

The major part of the meeting was devoted to generating ideas on **common research projects**, with a main focus on plans which could be submitted under the second call for proposals of the TSER programme.
Pelgrum pointed (after having talked to Mrs. Van den Brande, the network contract officer in Brussels) to the possibility to resubmit proposals which gained an A qualification in the first round of TSER. There were 3 proposals with an A qualification:

1. Monitoring.
2. The capacity for change and adaptation of schools in the case of effective school improvement.
3. Language study.

Following this introduction a short brainstorm in subgroups on possible themes for elaborating research proposals was organised, resulting in the following proposals:

Group 1 (Chair: Günter Haider, Rapporteur: Hans Pelgrum).

- Capacity change from TSER 1 (Creemers and Hoeben).
  Develop adapted proposals, process the comments of the judges.
- Reanalysis Timss Basic Competencies
- Definition of Basic Competencies: identify the EU Core

- Technology
  - ⇒ video project
  - ⇒ integrated monitor from TSER-1, elaborate telematics component
  - ⇒ Prototype evaluation of new education and training products
- Labour market
  - ⇒ methods development
  - ⇒ school (career - specially unemployment, are there data sets available?)
- Impact diversity and integration on educational outcomes.

Group 2 (Chair: Seamus Hegarty, Rapporteur: Rainer Lehmann).

1. TSER Topic List

- Transition from school to work: mechanisms and patterns of individual and group "trajectories", II.3.1.4
- The capacity for change and adaptation of educational systems, TSER-2 area II.1.2.1
  - ⇒ Suggestion: Bert Creemers should dig up the old proposal and resubmit for TSER-2
  - ⇒ Resistance to change, evaluation (Jaap Scheerens) of impact of programs. Some support, but very long term (4-5 years), very general.
- Assessment of European basic competencies, TSER-2 area II.2.1.3 (students at risk, effectiveness of intervention).
- Inventory and evaluation of opportunities for life-long learning, TSER-2 area II.1.2.2

2. Existing proposals

1. Pam Sammons' proposal on "cultural diversity: impact on school effectiveness"
2. Amsterdam Study Skills proposal
3. Amsterdam Second Language proposal difficult because of work with adult population.
4. IEA-LES should maybe be resubmitted if it is reasonable to assume that it will be successful. Consider the option that the network steps in if IEA fails.

On the basis of the exploration described above, thematic working groups were formed.

There were 4 themes and subgroups.

Theme 1: Basic Competences

Theme 2: Transition

Theme 3: Information Technology

Theme 4: Impact Diversity

Theme 1: Basic Competences; Leader: Günter Haider

Steps:

1. Reanalysis of existing data (Timss Reading Literacy);
2. Add value in collecting data from experts in the EU on minimal competencies using existing item banks;
3. Identify possible indicators to find student at risk groups, and potentially causal indicators;
4. Generate possibilities for action to tackle the problems of students at risk.

Günter Haider would produce a draft outline proposal within a month. He would be supported by Georgia Kontogiannopoulou-Polydorides and Rainer Lehmann, within a month.

Theme 2: Transition; Leader: Poul Skov

It was proposed to perform case studies in order to determine how some institutions managed to prevent drop-out. This should be done in the form of a pilot-project which would be aimed at (1) exploring the possibilities for running a large-scale project and (2) developing the methodology for a l large-scale study. The data collection would be qualitative and based on interviews.

It was suggested to connect to a project in Switzerland studying the discrepancies between qualifications produced by schools and qualifications needed by employers.

Theme 3: Information Technology; Leader: Hans Pelgrum

1. Pro/con of a video-project: the high costs was seen as a problem.
   Hans Pelgrum would ask a colleague (Alfons ten Brummelhuis) who developed a proposal to run a Dutch project aimed at creating videos of good examples. Jouni Välijärvi was very interested in co-authoring such a proposal. The link with the IEA-SITES study should be optimised.
2. Monitor: Revision of Proposal TSER I (A rate) Pelgrum would be responsible.
3. Jean-Paul Reeff would take the lead in developing a prototype-evaluation proposal. The outline would go to the members early November.

Theme 4: Impact Diversity; Leader: Jaap Scheerens

This study was aimed at providing a concrete and empirically founded definition of national educational cultures and cultural diversity. The study would use data from the IEA Reading Literacy study and the OECD school survey. The main steps in this study would be:

1. Concise presentation of the empirical material.
2. National panel studies on the basis of this material: reflections and possible explanations in terms of cultural identity.
3. Smaller set of countries (5 or 6). Come up with a more precise definition of educational culture and conclusions regarding cultural diversity.
4. Conference to present results.

Georgia Kontogiannopoulou-Polydorides offered help in elaborating the proposal.

Capacity for Change from TSER-1; Leader: Bert Creemers

Hans Pelgrum would check in Brussels for the comments.

A new version would be made and sent to the members.

Board Meeting April 1997

The meeting took place on April 4, 1997 in the meeting room of the IEA-headquarters in Amsterdam.

The agenda of the meeting covered the following points:

1. Announcements/News
2. Progress
3. Acquisition new projects
4. Planning of future work
5. Cost statement Year 1 and Budget year 2
6. Plenary meeting Frankfurt: dates

The meeting was attended by Seamus Hegarty, Georgia Kontogiannopoulou-Polydorides, Hans Pelgrum, and Jaap Scheerens (chair).

Pelgrum reported about a recent meeting with Mrs. Ham from DGXII (who is temporarily replacing Mrs. Van den Brande, who is on pregnancy leave) and Mr. Schulte. During that meeting the progress of the network was reviewed, the audit procedures were described, and the prototype of the WWW-statistical overview was discussed. Mr. Schulte suggested that the network activities should be reviewed in due time. In relation to this point Scheerens mentioned that the current TSER had been reviewed by an Austrian evaluator. The coordination center of the network would try to find out more about this review.

Scheerens mentioned that in the draft texts of the 5th framework programme there was hardly any mention of social science research. The board suggested that an exploration of national views on the 5th framework programme should take place, and that Pelgrum would contact Mrs. Ham and Mr. Bures to find out which implications the current plans might have for the future of the network.

The board reviewed the status of proposals submitted by the network. Only the proposal submitted by the GION had any chance of being funded. It was given an A2 rating, which meant that it would not automatically be selected for funding (this only happens with A1 ratings).

The **progress** on the network's main activities (conceptual framework, inventory of assessment data bases, WWW-statistical-overview-site, and overall EU-AEI Web site) was reviewed.

A revised proposal for a **common conceptual framework** was reviewed.
This framework (see Figure 1) provided the ordering principles for the construction of the prototype of the Web site containing statistical overviews of international comparative educational indicators collected by the partners. Moreover, it constituted the basis for the inventory of national assessment data bases which were in the possession of the partners.

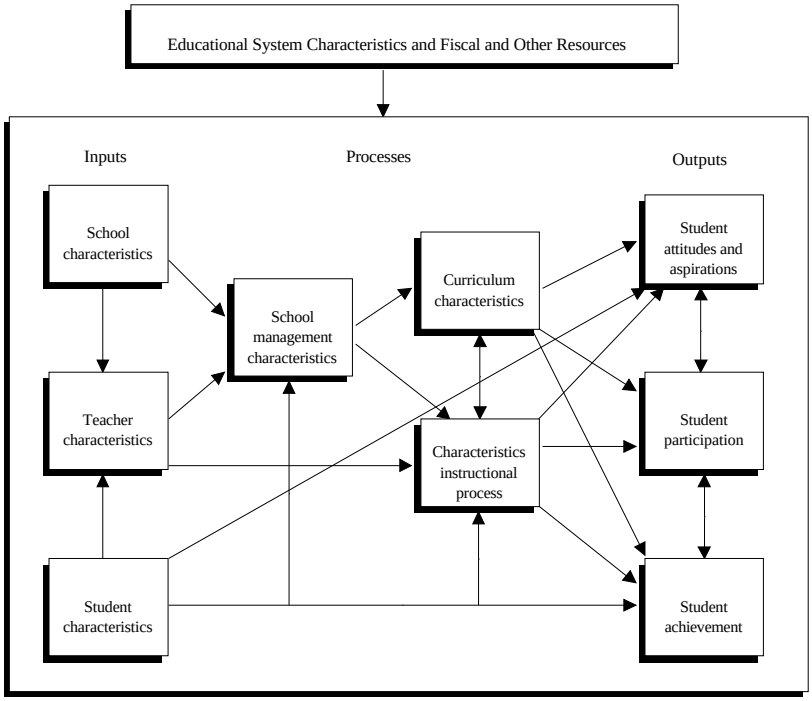


Figure 1. Elements of a conceptual framework for EU-AEI.

It was noted that the arrows in Figure 1 have the character of potential relationships. As indicated in the conceptual framework document (see below), the current empirical evidence did not support the existence of many of these linkages. It was not sure to what extent the non-existence of certain relationships was caused by a lack of quality of instruments which were used to measure these indicators.
The board reviewed the revised conceptual framework document and discussed the problem of potential implications of the degree of centralisation of educational systems for the design of international comparative assessments. It was decided not to include this issue in the current version of the document, but to invite Georgia Kontogiannopoulou-Polydorides to prepare a discussion note for exploring the dimensions of this problem area. Finally, the revised document was adopted.

The progress with regard to the inventory of **national educational assessment data** bases was reviewed.
Pelgrum pointed out that, whereas the partners in the network have cooperated to collect international comparative data on educational achievement, they also are involved in national assessment projects. He

pointed out that national assessments were very costly operations, due to the fact that usually quite large sample sizes were needed and also because instrument construction was a very labour intensive component. Researchers were constantly searching on the basis of theoretical insights and empirical evidence for improvement of the measures collected in national assessment. In this respect the EU could be seen as a large laboratory in which innovations with regard to the design and instrumentation of national assessments were constantly taking place. In order to facilitate the search for data bases which might be relevant during the design period of national assessments, the network partners decided to make an inventory of data bases to which each of the partners had direct access. This inventory was conducted on the basis of a standardised questionnaire, which enabled the partners to describe their data bases in terms of design characteristics and content.

So far, the partners had produces 14 inventories (containing 48 data bases).

Pelgrum proposed to summarise the results in a table of which an example was shown at the meeting. This overview showed which indicators were quite common across the countries. Probably more interesting was the fact that developmental work with regard to some indicators (for instance, school climate indicators) was concentrated in a few countries. Especially this last category could be important for researchers involved in national assessments in other countries.

This table would be made available on the WWW.

The board agreed with this procedure, but suggested in addition that possibilities would be explored to increase the user-friendliness by offering options to search on keywords.

Pelgrum explained the work content and progress with regard to the construction of a WWW-site, which would give access to **European comparative statistics**, which were available for the different categories covered in the conceptual framework.

The AEI statistical overview is a facility for providing access to educational indicators. It existed as a prototype of a WWW-site. The goal of this activity was to increase the accessibility of statistical data on comparative educational indicators which did not yet exist in the current EU-services (Eurostat and Eurydice).

Users searching for particular statistical information would be guided by a visualised conceptual framework, which allowed them to navigate through the available information.

Pelgrum showed the prototype to the board members, who agreed that after a review of DGXII would have taken place, the production work for filling this site could start. The board suggested to keep close contacts with Eurydice, Eurostat, and OECD in order to avoid overlap with already existing comparative statistics from these sources.

The construction of the **AEI-Web site**, which started in May 1996, was reviewed. The site provided information about the mission of the network, the work plan for the period funded by TSER, current members and addresses, and access to project documents. From the list of members, links were established to Web sites of the member institutes. For those member institutes which were not able to construct a site

themselves, the coordination center of the network had entered the information on the members' institute on the Web. At the time of the board meeting, 15 sites of members were linked to the general EU-AEI site.
Seamus Hegarty performed a preliminary review of the site and the links to other institutes. It was suggested that each institute should at least cover the main topics which were given in the example which was created by the AEI-coordination center. Moreover, in order to increase the visibility of the network, it was suggested that members are invited to establish links from their sites to the EU-AEI site.

The board also discussed possibilities for **the acquisition of new projects**. It was suggested that for the submission of network proposals the next rounds of Socrates and TSER-3 might be important. The board suggested that members be invited to generate ideas on possibilities for cooperative projects. Between plenary meetings, the board would decide (on the basis of overlap with the network's mission) which projects might be considered as network-projects.

It was decided that the next plenary meeting would take place in Frankfurt on September 24, in conjunction with the ECER97 conference.
In the meantime, the work on the database inventory and statistical overview would be continued and finalised, while feedback on the realised products would be actively sought from Brussels.

In order to increase the networks visibility, it was suggested that Pelgrum would explore the possibility to present the network during a symposium at the ECER97 conference. Moreover, it was advised that the address of the network's overall Web site would be given a prominent place in the letter heads of the coordination center.

The board considered the question how to evaluate the networks functioning and products, and expressed the intention to assist DGXII in finding ways to define the scope and appropriate time for such a review.

Plenary Meeting September 1997

This meeting took place on September 24 in Frankfurt. The meeting was attended by: Gerard Bonnet, France, MEN.DEP; Caroline Gipps, United Kingdom, ULON.IE; Armin Gretler, Switzerland; Marriette Ham (project officer, European Commission), Seamus Hegarty, United Kingdom, NFER; Wijnand Hoeben, Netherlands, RUGR.IER; Georges Henry, Belgium, ULGG.DEPF; Monique Kole (notes), Rainer Lehmann, Germany, UBER.AP.AEB; Mark Morgan, Ireland, SPC.ERC; Hans Pelgrum, Netherlands, THT.OCTO; Borges Palma, Portugal, IIEPT.ASE; Sten Pettersson, Sweden, NAE.DE; Renata Picco, Italy, CEEDU.PSF; Jean-Paul Reeff, Luxembourg, MENFP.IPT; Jaap Scheerens (chairman), Netherlands, THT.OCTO; Poul Skov, Denmark, DPI; and Jouni Välijärvi, Finland, UJYV.IER.

Pelgrum summarised the progress on main activities of the network with regard to:

- Creating a conceptual framework;
- Creating a proforma for an inventory of national assessment datasets from all countries, and the collection of data from all members on the basis of the proforma;
- Exploring possibilities for cooperative work with regard to different research themes
- Construction of a Web site containing links to the participating institutes and their research programs, as well as an overview of comparative education statistics resulting from international comparative assessments in which the network members are involved.

The AEI Web site contained information about the mission of the network, reports of meetings, and links to the participating institutes, their organisation, staff, projects, and publications.
The latest development on this site was the pilot version of AEI Statistical Overview, which had been organised according to the conceptual framework which was developed by Scheerens and Hegarty in a document, which was finalised early 1997.
The statistics which were included in the AEI statistical overview Web site (address: http://uttou2.to.utwente.nl/AEI/index.htm) were extracted from the TIMSS databases covering EU Member States, Japan, and the USA.
It was concluded that the current version of this statistical overview might be a first step towards the creation of a full-site, containing:

- Trend statistics;
- Effectiveness estimates;
- Help desk;
- Remote data processing;
- Links with training projects for educational researchers.

As a first step towards a potential project it was proposed to undertake the following actions for determining the need for relevance and practicality of the current pilot version:

- Review by members of the network;
- Evaluation of the current site by 2 educational journalists from each country;
- Consultation of policy-makers network;
- Consultation of Eurydice and Eurostat.

After the meeting in Slough, several subgroups in the network had been active in preparing proposals for projects to be submitted for several programs initiated by the Commission.
These were:

- ‘Capacity for change and adaptation of schools: the case of effective school improvement’ – submitted to TSER-2
- Basic competencies – submitted to TSER-2

- A virtual market place for the European learning community – Submitted to the Task Force Multimedia

It was observed that other initiatives generated at the last plenary meeting in Slough had –for various reasons- not led to submissions of proposals.

The general reaction to the work performed on creating the **Web sites** which would give access to the network's knowledge base was positive, although it was noted that a full elaboration of the current pilot versions would require a substantial amount of work, for which additional support would need to be sought. Especially it was noted that the conceptual framework offered an important organising principle for the statistical overview.
The participants decided that a plan should be developed for further integration of the conceptual framework and the statistical overviews. This plan should contain a stepwise procedure, showing how international and national assessment statistics and data would be made available on the Web and how these could be used by different audiences.

The general experience with regard to our cooperative efforts for increasing the functionality of our European involvement in international comparative assessments for addressing issues covered by EU programs had been disappointing. It was observed that it looked like it was difficult to find from the EU commitment for strengthening the European dimension in these international comparative assessments, in which all network members are involved and to which EU national governments altogether spend millions of ECU, sometimes up to a level which could hardly be justified at the national level.
A discussion evolved about the network's position in the EU and suggestion were made as to how to define the network's future role.
It was agreed to elaborate a proposal for a 2nd phase of the project within the framework of TSER-3. The deadline for submission of proposals was January 15, 1998. The board and Caroline Gipps would be involved in the development of the proposal. Before it would be submitted the proposal would be send to the other members for comments and adjustments. Mrs. Ham mentioned that the proposals should be very clear. Everything would need to be spelled out. What would be obvious for us would not be always obvious for others.
Mrs. Ham summarised the content of TSER-3, which was published by the European Commission just before the meeting. The research proposals should address one of the following 3 themes:

1. Competition, change and dialogue.
2. Work, welfare and employment.
3. Innovation and institutional change.

Moreover, within each theme the proposals should reflect how the research would contribute to each of the following areas:

1. Life-long learning and educational goals.

2. Implication of societal development for the educational technology system.
3. Educational implications of the European integration.
4. Labour market and economic growth.
5. Educational technology's contribution to fostering innovation.

Wijnand Hoeben presented the new GION project, 'Capacity for change and adaptation of schools: the case of effective school improvement'. It was submitted to DGXII in January 1997. It was at the time of the meeting in the contract negotiation phase.
Originally there were 3 subprojects: Theoretical analysis, inventory and evaluation, intervention project. The last subproject was excluded because there was an overlap with DGXXII.
The 8 participating countries in the project were: Belgium, Finland, Greece, Italy, the Netherlands, Portugal, Spain and the United Kingdom. The duration of the project would be 2 years.

With regard to the network's future activities, it was decided that:

- EU component: after a lively discussion, it was decided that the network would not capitalise on actions towards defining European components in international comparative educational assessments which were conducted under the auspices of other organisations than the European Commission.
- Seamus Hegarty, Hans Pelgrum, Jaap Scheerens and Caroline Gipps would prepare a proposal for the continuation of 'European network for educational research on assessment, effectiveness and innovation' project. This proposal would contain a description of potential future activities and possibilities for acquiring support from different sources.
- The board would take care of a review of the web site containing the statistical overview.
- Sten Pettersson, George Henry and Hans Pelgrum would form a subgroup for conducting a review of the web site.
- A diskette containing an inventory of databases would be send to the members.
- Members agreed to explore the possibilities of submitting a proposal to the EU. If there were ideas for a proposal a brief outline should be send to Hans Pelgrum by October 17th. He would do some integrating work and send it out to the other members, who could have a quick look at them. After that we could start to prepare the proposals.
- A collection of ideas for new proposals would be send to Mrs. Ham for feedback.

Plenary Meeting March 1998

This meeting was attended by: Gerard Bonnet, France, MEN.DEP; Birgit Eschmann, Luxembourg, MENFP.IPT; Seamus Hegarty, United Kingdom, NFER; Hans Pelgrum, The Netherlands, THT.OCTO; Sten Pettersson, Sweden, NAE.DE; Jaap Scheerens (chair), The Netherlands, THT.OCTO; Sally Thomas, United Kingdom, ULON.IE; and Felicity Wikeley, United Kingdom, ULON.IE.

At the meeting the following issues were addresses:

- Information exchange on EU research programs;
- Review of progress;
- Workplan 1998-2000.

EU Research Programs

The meeting discussed possibilities for submitting proposals in existing research programs of the EU. It was concluded that a main obstacle still was the explicit statement in many documents that there were no possibilities for supporting new data collection. A promising development was the possibility offered by the Socrates programme to support secondary analyses of existing comparative data. It was further suggested that the OECD study (PISA-project) offers an important context for undertaken comparative investigations within the EU. These studies are focussed at Reading, Mathematics, and Science. This means that quite a number of important subject areas (such as Foreign Language, Culture) are not assessed. It seems important that the network will devote its energy to promote comparative assessments in Europe which have a broader coverage than the subject areas covered by the OECD. An important question is if the Commission wants the network to continue its activities. Moreover, it was suggested that a core group of members should take a much more pro-active role in increasing the visibility of the network and establishing good working relationships with the network of European policy-makers. With regard to the last point it was noted that the different missions of EU-AEI and the policy-makers network should be continuously accentuated.

Review of Progress

The discussion focussed on the necessary steps for improving the functionality of the Web site containing the statistical overview. It was felt that user-friendliness and presentation need further improvement before this site will be made widely known. As such these improvements are discussed in the workplan for the period 1998-2000.

Workplan 1998-2000

A distinction was made between: (1) the facilitating and (2) the thinktank function of the network. Some of the planned activities fall under the first category (such as improvement and elaboration of the Web site, Helpdesk, etc.). All current members could be involved in these activities. The second function is covered by activities towards methodological improvement (such as value added and fair testing). It was felt that it would be efficient to create small core groups which can deal with this function.

The following conclusions resulted from the discussions.

1. Instrument bank: this was seen as a very useful activity which is not yet undertaken by other groups. The content selection should rely heavily on the know-how resulting from school effectiveness studies.
2. Value added: this activity should start with an EU workshop in which the technical and transnational issues are spelled out. The core group for this activity consists of: NFER, Univ. London, OCTO, and expert institutes to be appointed by Sweden, and France.
3. Fair testing: this was seen as a very important area. A more precise definition of this issue in an international comparative context needs to be given. Also for this activity a workshop might be a good starting point. The core group for this activity consists of: Uni. London, OCTO, NFER, a Swedish university, and possibly a French university.
4. Web site: the ultimate goal must be that our network finalises a version which can be evaluated by different user-groups. The ultimate goal would be to transfer the activity to specialised institutes like Eurydice.

Next Meeting

It was decided that the next meeting of the network would take place on October 2, 1998 in Amsterdam.

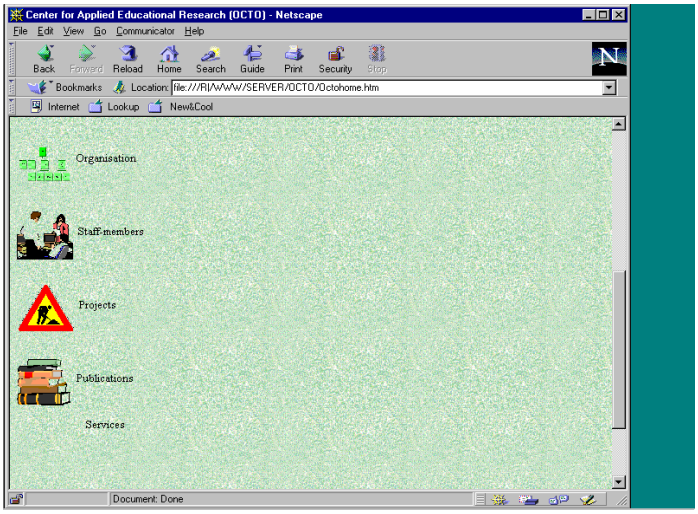
Outcomes

In this section the outcomes of activities undertaken by the network in the past two years are described with reference to the targets which were stated at the outset.

Objective 1

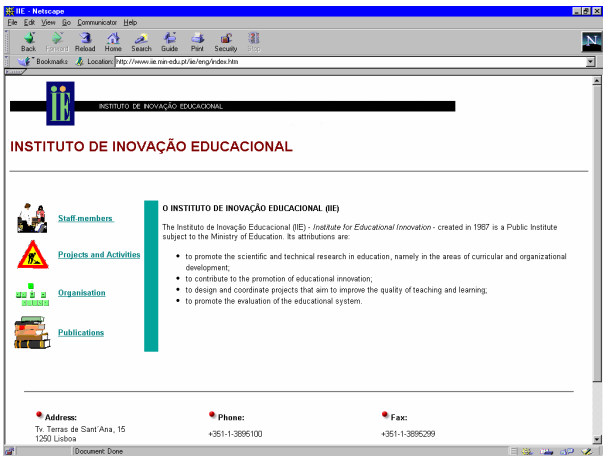
Electronic data base (accessible via World Wide Web) containing information about educational systems characteristics, descriptions of the participating research institutes, their research programs and their research projects.

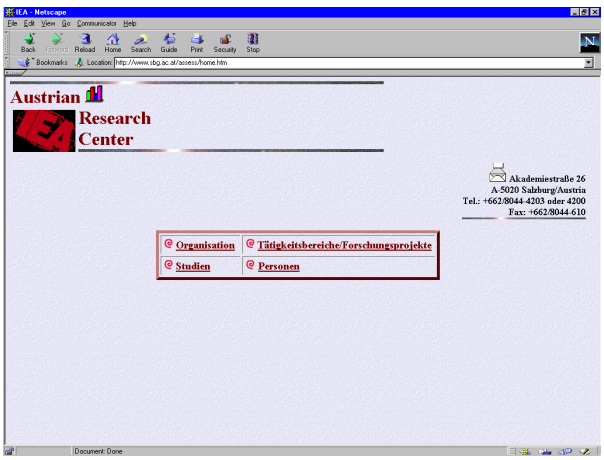
Each participation member institute has undertaken developmental activities for making available information about their institute, research programs and their research projects. As an example of how a Web site could look like, the coordination center developed some exemplary materials which could be copied and amended by the partners. The home page of this example is shown below:



Some partners followed the examples given by the network coordination center, while others (actually most) created their own design.

For example the Portuguese page followed the example given by the network coordination center, while for instance the Austrian page has an original design (see below).





A review of the network's web site from a journalist's point of view (Judit Lukács, 1998) led to the following observations:

Opening Page

It appears that on the opening page not all users can read the lines below the headline. Perhaps another font should be used for the name of the network.

Moreover, the logo of the network (that appears on several of the following pages) takes a lot of data-space, which is shown by the fact that it takes *very* long time to print the pages on a printer.

The word ”Information” is underlined, but there is no link to it. It is not underlined as if it would be a link, but on paper it can easily be mixed. Perhaps this headline should be put in bold.

On this page more information about the network should be made available. Something that gives information of what sort of network it is and what it offers to the reader. Such information should stimulate the reader to go further to the other pages. The current table of contents is good, but not enough.

Also it would be good to show on every page *when* the page was updated the last time, or (which is not necessary, but interesting) a sign that shows the number of visitors to the site.

Summary of the Network's Mission

The information is correct, but could have been made more interesting. This is a formal description of the network, but not the description that makes a reader who is not already interested to read more and be curious about the network and how it works.

A potential improvement could be to start with the things that have been done recently or with the three products that are to be realised.

On this page, the TSER-work programme is mentioned the first time. There is no explanation of what it means. The word appear several times later in the documents. Words like these should be explained every time they appear the first time on a page.

Content of the EU-AEI Project Programme

This is a long and detailed document where some of the information from the former page is found again, but this time more detailed. There are a lot of interesting things to read, but for superficially interested persons, it is not easy to understand what is said. The section about intended milestones should also show how far the network has come in this work today.

Also it seems worthwhile to allow users to get to the top of the page at the end of each of the four pages in this document.

Finally, the word ”Workplan” should appear in the headline, as it is this word which is referred to in the index page.

Organisation

In this document there are two words ”newsletter” and ”EU” that are linked to the same document. It is not clear why, in a very short document, this is necessary.

Members

The table is interesting and contains a lot of information. But users would be helped if links to each person’s e-mail address would be available, which would make it easier to get in contact with him/her.

Moreover it would be valuable to have additional information about the ”roles” of participants in the network, if it is the network coordinator or if he/she has a specific mission.

The reason for the current ordering is not clear. This page really need the date of last update.

Minutes Board Meeting

This is a formal document that the members need to know of to be updated. But it contains nothing that is interesting (the way it is written now) for a broader audience. The document is more than two years old. It makes the reader wonder if the network did not meet since then.

At the end of this document it is impossible to get to the homepage.

Objective 2

Exchange and integration of conceptual frameworks regarding educational indicators, and the use of expertise from the partners for producing national and European overviews of statistical information on Mathematics-, Science-, and Technology-education.

Overall Framework

A systems model of inputs, process or throughputs, outputs and a context or environment provides a convenient starting point for conceptualising the network's domain of interest. (see Figure 2).

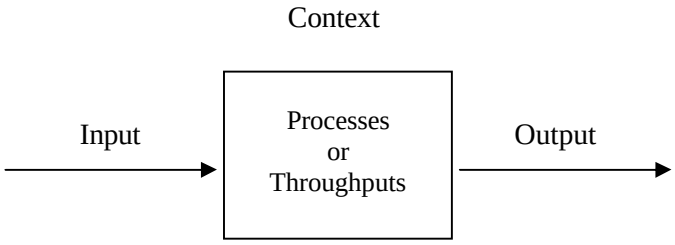


Figure 2. Input- process- output- context model.

The internal part of the model (the inner box in Figure 2) can be defined at various levels of aggregation: the pupil, the classroom, the school, the school district or even the national school system. It is also possible to define the inner box as a multi-level structure. Thus, if the school is chosen as the core of the system, material and financial resources are examples of inputs, school managerial and instructional activities are

examples of processes and student achievement measures are examples of outputs. District level or national regulations could be seen as examples of relevant characteristics of the context. Pupils' place in this framework depends on the focus of interest: they can be seen as the inputs of schooling, when their aptitudes and dispositions at the beginning of a period of schooling are considered; in their achievements they are very commonly seen as the output of schooling. Considering the international comparative scope of the network, national contexts of educational subsystems are of particular importance. Structural (e.g. the patterns of centralisation and decentralisation) as well as cultural conditions (e.g. national philosophies with respect to assessment) of national contexts should be taken into account.
This general framework serves as a basis to delineate the three sub-domains of the network's agenda: assessment, effectiveness and innovation. This is done in the following sections.

Distinctions and Connections between Assessment, Effectiveness and Innovation
Assessment refers to estimating the value of certain things. Educational assessment usually refers to evaluating educational outcomes. For example, the term 'national assessments' in education refers to measuring the achievement levels of pupils in certain subject areas. Strictly speaking, one could also speak of the assessment of the context, the inputs and the processes of schooling. Sometimes, in evaluative schemes such as the use of education indicators, assessments of inputs or processes are seen as substitutes for educational outcomes (cf. Oakes, 1987). For present purposes we propose to stick to the more common use of the term and refer to assessment of educational outcomes whenever the term assessment is used.

Effectiveness is a causal concept in which instrumental or means-to-end relationships are the conceptual core. In terms of the systems model, inputs and processes are seen as means to achieve maximisation of outcomes within a particular set of contextual conditions. The key questions for educational effectiveness research are to discover sets of antecedent conditions that predict certain desired outcomes.

To the degree that different types of inputs or process characteristics are predominant in answering this question, various strands of effectiveness (or educational productivity) research can be distinguished: economically oriented research on education production functions, instruction oriented research on classroom practices, and organisation oriented research on exceptionally effective or ineffective schools. In modern educational effectiveness studies a comprehensive approach is chosen encompassing these three research traditions (cf. Creemers & Scheerens, 1994).

When the focus is on the relationship between output and resource utilisation, the term efficiency rather than effectiveness is used. An important technical issue in effectiveness research is how to estimate the value added by specific educational interventions, over and above the impact of pupils' initial aptitudes and attainment levels.

With respect to the domain of *innovation* in education it is proposed that the network adopts an approach that stays close to the domains of effectiveness and assessment. In the field of educational innovation a lot of attention is usually given to innovation strategies: how can particular innovations be initiated, adapted by schools and implemented. In our case the orientation proposed refers more to the substance of innovation than to the strategy.

If one knows what works in education, which is the ambition in effectiveness research, one has some points of impact for educational improvement. If the more ambitious question of why it works is answered there are even better handles or levers for school improvement. The supposed mechanisms behind the factors that have received some empirical support in educational effectiveness research are therefore quite relevant to the more substantive approach to educational innovation that is proposed. In the final section of this paper, some of the available ideas on levers for school improvement as generated by school effectiveness research will be discussed.

Assessment

For network purposes, assessment studies can be categorised in diverse ways. Table 2 offers a scheme as a starting point for discussion. It assumes that studies can usefully be grouped in terms of their purpose, the content of the assessment, the means by which it is done and the target group. The examples given are illustrative rather than exhaustive.

Table 2

Outline scheme for categorising assessment studies

Purpose, e.g.

- monitoring
- accountability
- raising standards
- diagnostic uses
- resource allocation
- comparisons
- trends.

How, e.g.

- standardised tests
- checklists/normative instruments
- diagnostic measures
- expert judgement.

What, e.g.

- knowledge
- skills
- subject mastery
- IQ/general ability
- attitudes.

Who, e.g.

- age group
- institutional base
- special populations.

Note: Distinguish also between latent and manifest purposes.

This approach to grouping assessment studies is of course analytical and risks obscuring key features of studies. It may well be that network purposes would be better served by a broader, topic-based categorisation. Topics to underpin categorisation in this approach might include

- trends over time;
- international comparisons;
- value-added studies;
- developments in use of indicators;
- manageability of assessment;
- authenticity.

The Selection and Measurement of Context, Input and Process Indicators

Given the centrality of outcome assessment in an international context for the network, the most likely research opportunities in the areas of effectiveness and innovation will be background studies associated with national or international assessment projects. Two kinds of background studies are:

1. in-depth studies, in which questions of effectiveness and efficiency are the predominant purpose.

2. more applied studies, in which input, process and maybe also context "indicators" are seen as additional parts of monitoring systems, next to outcome indicators.

For both types of study the most likely rationale for selecting variables reflecting input, process and context conditions will be based on the state of the art of school effectiveness research.

In a recent review Scheerens and Bosker (1996) use several sources of evidence to review the knowledge base on educational effectiveness research. Their summary table (adopted) is presented on the next page.

Table 3

Review of the evidence from qualitative reviews, international studies and meta analyses

	Qualitative reviews	International studies	Meta analyses
Resource input variables			
Teacher/pupil ratio		0	0
Teacher training			0
Teacher experience			±
Teachers' salaries			0
Expenditure per pupil			±
School organisational factors			
Productive climate & culture	+	0	
Achievement press for basic subjects	+	0	
Educational leadership	+	0	0 (r=.01)
Monitoring/evaluation	+	+	0 (r=.05)
Cooperation/consensus	+		0 (r=.03)
Parental involvement	+		± (r=0.14)
Staff development	+		
High expectations	+	+	
Orderly climate	+		0 (r=0.06)
Instructional conditions			
Opportunity to learn	+	+	+
Time on task/homework	+		++
Structured teaching	+		++
Aspects of structured teaching			
• cooperative learning			++
• feedback			++
• reinforcement			++
Differentiation/adoptive instruction			+

Notes: ++ = relatively strong; + = positive influence; ± = weak or doubtful evidence of positive impact; 0 = considered, but no impact.

Obviously, the results presented in this table evoke quite a bit of cognitive dissonance, particularly the apparently low or negligible impact of resource input and organisational process conditions and the discrepancy between the results of qualitative reviews and quantitative meta analyses in this latter area. This discussion is beyond the scope of this paper, but in future studies to be initiated by the network further reflection is needed to select the most relevant set of process and input variables.

In the modelling of educational effectiveness the multi-level structure of the data on context, process and input variables is now commonly recognised. The predominant conceptual interpretation of such multi-level structures of schools nested within contexts (e.g. districts), classes and teachers nested within schools and pupils nested in classrooms is that higher level conditions somehow facilitate lower level conditions. (For several more precise alternative interpretations of higher level facilitation, see Bosker & Scheerens, 1994.) In more elaborated versions additional complexities like non-linearity, non-recursiveness and dynamic aspects are included (ibid.). The phenomenon of higher level facilitation has not been addressed in many empirical studies, and the results are still inconclusive. The most relevant point at the current state of development of the network is probably that hypothetical model structures as well as the nature of hypothetical effectiveness enhancing conditions should be taken into account when decisions are taken about the design of studies (sampling design, data collection schemes and instruments).

In the light of the still inconclusive evidence on higher level facilitation, the generally small impact of organisational conditions on student outcomes and certain theoretical notions, a further qualification might be required in which the limitations of this principle for educational organisations is stressed. Scheerens and Bosker (1996) borrow the concept of subsidiarity, which originates from political science, for this purpose. The subsidiarity principle can be paraphrased as "do not seek explanation at higher levels if it can be found at lower levels". The principle recognises that in educational organisations the lower levels, i.e. students and teachers, have a lot of autonomy and discretion. Images of self-directed learning in the current constructivist fashion in educational psychology and the traditional idea of the school as a loosely coupled system or a professional bureaucracy emphasise these features. Obviously the subsidiarity-view on relationships between levels within educational systems is dependent on the centralised versus decentralised nature of the educational system. In actual fact, systems may be centralised in some domains (e.g. outcome assessment) and decentralised in other domains (e.g. teaching strategies).

The principle of subsidiarity accounts for the fact that the impact of higher levels on lower levels will usually be small, but does not exclude that there will be a certain degree of impact. If the interpretation according to the subsidiarity principle is valid, it would enforce the need for further study about which type of mechanism has most explanatory power in making the (small) differences that remain. The direction in which answers to this latter question should be looked for is available in social scientific theory, from which effectiveness enhancing mechanisms could be deduced. Since the discovery of such mechanisms is of the utmost

importance to the approach that is proposed with respect to educational innovation too, we will deal with it under the heading of levers for school improvement.

Levers for School Improvement and Theory-Embedded Principles of Effective Schooling

Scheerens (1995) mentions five theory embedded principles, that might account for school effectiveness. They are summarised in Table 4.

Table 4

Five theory embedded principles, that might account for school effectiveness

Principle	Theory	Organisation Type
Proactive planning & structuring	rational decision-making	bureaucracy
Retroactive, evaluation centered adaptation	cybernetics	learning organisation
Market mechanisms	public choice theory, micro-economic theory	autonomous school, "choice"
"Fit"	contingency theory	undetermined
Self-organisation	chaos theory	adhocracy

It is beyond the scope of this paper to discuss and compare these various theories and principles. Scheerens concludes that the cybernetic principle of evaluation/feedback/corrective action is the most promising one in that:

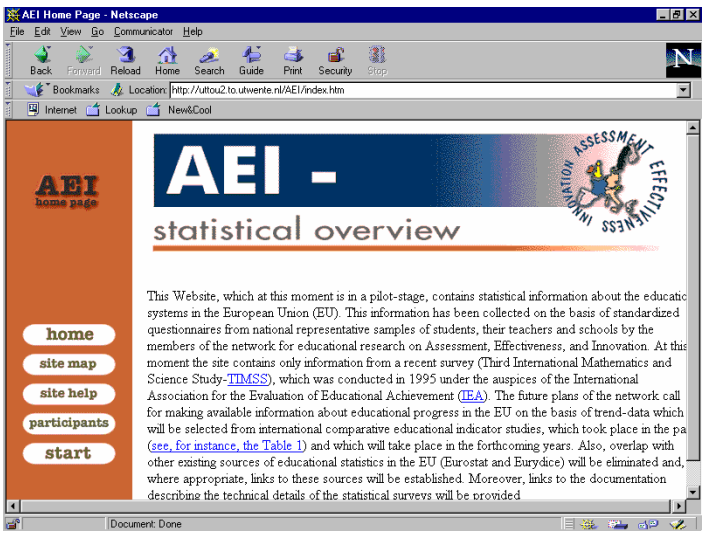
- It has a place in several of the theories considered;

- It accounts for the evidence from school effectiveness research not only in the narrow sense of one key factor, but also because of its implications for other factors such as:
  - ⇒ outcome variation
  - ⇒ coordination
  - ⇒ cognitive and motivational aspects of learning and educational leadership.

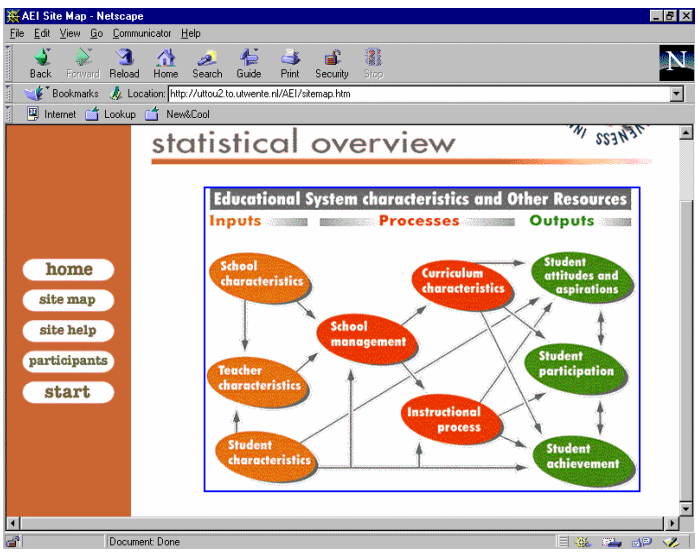
This point of view would support a key interest of the network in quality control and school self-evaluation. Whether this area will indeed become a real area of interest will be determined at a later stage.

The EU-AEI Statistical Overview

On the home page of the EU-AEI Statistical overview (see below) it is explained that the current site is seen as a pilot version which needs to be reviewed and further elaborated in terms of functionality and content. Amongst others, it is pointed out that also potential overlap with WWW statistical information from Eurostat and Eurydice needs to be explored and where appropriate, links to the sites of these institutes need to be established.



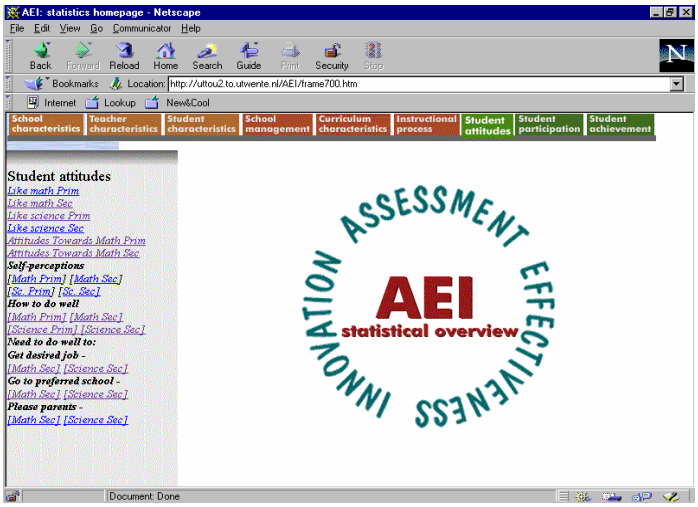
The statistical information at the site is accessible via the conceptual framework of EU-AEI, which was developed by Scheerens and Hegarty (1997). By clicking the category “Site Map” at the home page, the user is linked to a clickable version of this conceptual framework (see next page).



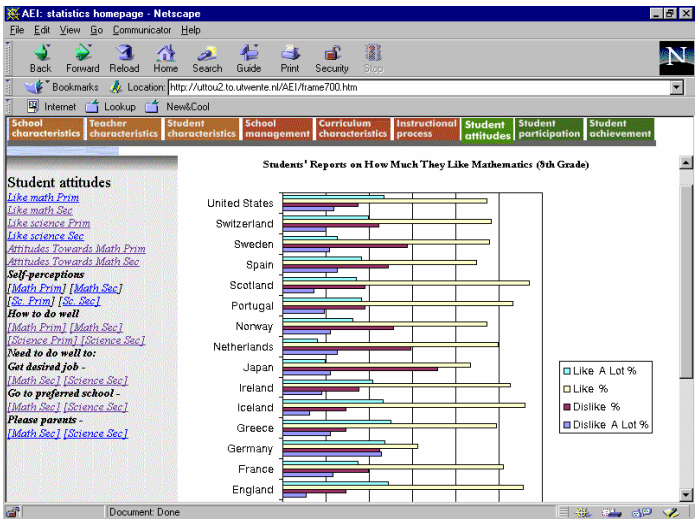
The pilot version of the site contains at this moment a total of 136 comparative statistics for the EU-countries (and Japan and the USA), from the categories school characteristics (4), teacher characteristics (18), student

characteristics (17), Instructional processes (39), Student attitudes (22), and. student achievement (36). These statistics are available for primary and lower secondary schools.

By clicking a particular category on the Site map, like for instance student attitudes the user receives a list of more concrete underlying categories (see below), which shows for which attitudinal aspects comparative statistics are available from representative samples of students from EU-countries, for which educational level (Primary or Secondary education), and for which subject areas.



Finally by clicking the most concrete category (for instance “Like math Sec.”) the statistics can be accessed (see below).



The pilot version of this WWW-site was demonstrated and discussed at the plenary meeting of EU-AEI on September 24 in Frankfurt.

At that meeting it was decided that the current version of this statistical overview might be a first step towards the creation of a full-site, containing:

- Trend statistics;
- Effectiveness estimates;

- Help desk;
- Remote data processing;
- Links with training projects for educational researchers.

Internal Review of the Statistical Overview

A journalist's review of the statistical overview led to the following observations and suggestions for improvement (Judit Lukács, 1998).

Index

A good explanation of the site and what it contains. I would not mind having more links here. The page is too wide (see below).

Site Info

Easy to read and understand, but the page is too wide (the right frame) and when you print it, you only get a part of the text. If you copy and paste the text into a word processor, you get all the words but loose the lay-out.

There is no dates at all at this site either.

Site Map

An interesting map, but it can't be printed (see above). And as I go through the different buttons I find a lot of statistical information. Almost more than I can take at one time. But that doesn't matter, because it is easy to pick out what I need (if I know that).

In the button for School characteristics (frame100), I am supposed to get information about math in primary schools, but the headline is for "lower secondary". The picture though is not the same as the one I get when I go to …ch5/f58.htm. This should be checked.

There are no links at all from some of the buttons, for example "school management", "curriculum characteristics" and "student participation". Perhaps you should explain why you don't have more information.

Involvement of EU Member States in International Assessment

This is an interesting table, but can't neither be printed, nor copied and pasted, which makes it useless. And how come that most of the projects are old? What does the year stand for? The date of starting or ending a project?

This document needs more explanatory information.

Finally, About the Statistical Overview

These documents are all written in an easier language than the documents on the overall EU-AEI sites. They are shorter and contain much more links to make it possible to get a lot of statistical information. I like that.

The statistics (columns and staples) are easy to understand, but I have not tried to print them. I guess that they are too wide as well as the earlier documents.
I miss the dating of each document. It would also be interesting to get short information about each table, but that is quite a lot of job I guess.

Objective 3

Exploration and identification of possibilities for cooperative pilot projects.

From the experience of the network members in conducting international comparative assessments, it appeared that in many societies an increased need is felt for objective empirical data on educational progress. However, given the relatively short history of educational assessments, researchers are still continuously searching for improvements of measures, which can serve as stable indicators of educational processes and outcomes at the macro-level. Moreover, it was also felt that a better insight is needed in the extent to which particular indicators can explain the malfunctioning of (parts of) educational systems. Therefore, it was decided to start –via pilot projects- activities to explore how to share the knowledge base on the quality and explanatory power of indicators collected by the member institutes. In order to realise this goal, the coordinating center devised a task list from which members could choose which contributions they might be able to deliver on the basis of their assessment experiences.
The results of this work can be summarised in four categories: indicators, explanatory analyses, value added approaches, and trend studies.

Originally the intention was to integrate the results of the pilot projects into the Web site. However, the experiences of analysing the materials showed that it would be premature to do so, because of the lack of standardisation, existence of gaps and the need to redesign parts of the existent sites in order to incorporate these materials in a functional way.
Therefore it was decided to summarise the materials and to describe how these might be integrated in a redesigned Web site. The description, which follows, indicates in a final section at which places in the site the materials can be incorporated.

Indicators

MEN.DEP explains how **Civics** knowledge was measured and describes trends by comparing results from samples taken in 1994 and 1995.
RUGR.IER measured **School Climate** with a set of 11 Likert-type items. This measure proved to be fairly reliable.
SPC.ERC showed how students **Home Background Characteristics** were measured (SES, Literacy Press, Pedagogical beliefs of parents, socio-emotional adjustment, and involvement in school) and which reliabilities were obtained.

NFER was focusing on indicator construction on the basis of the OFSTED database, which contains detailed numerical judgements derived from the inspection of the majority of secondary schools in England. Factor analysis techniques were used to derive relevant factors at the school and pupil levels, which were analysed in relation to both lesson observation grades and examination outcomes. The results are discussed in relation to the usefulness of the OFSTED database for school effectiveness research.

ULON.IE describes work on the construction of indicators of **Student Characteristics**, such as self-efficacy, engagement with school, pupil culture, behaviour, teacher perceptions, homework.

ULGG.DEPF created a Web site containing results of the monitoring project of the quality of science teaching in French speaking Belgium. It includes an extensive description of science achievement test, with item-descriptions and test results.

IIEPT.ASE describes the results of constructing indicators of Student expectations, Time on task, Student Homework, Quality of achievement, Student achievement in Geometry, and Level of satisfaction. Reliabilities and relations with several input, process and output variables at achievement and background level are summarised.

NCE was involved in the construction of indicators on the basis of a national assessment project. This work did not yet result in a product, due to unexpected delays.

UAM.SCOKI constructed an indicator of the frequency of leisure-time reading on the basis of data collected with a diary. The diary is filled out during the first five minutes of the school morning, starting on a Tuesday and ending on the next Monday. This indicator proved to have a very high reliability and validity (compared with data collected from parents). SCO concluded that the diary method has proven to be a reliable, valid and relatively efficient instrument for the measurement of the frequency of leisure time reading of books.

THT.OCTO constructed a large number of indicators from TIMSS, which showed that many indicators did not have good measurement quality.

NAE.DE describes a study in which indicators were constructed for investigating the importance of the size of management and teaching resources in school, in relation to other factors which have proved important for explaining pupils' achievement. The questions, which were addressed, were as follows:

- Does the magnitude of school resources make any difference to pupils' achievement?
- How important are school resources for pupils' achievement compared with other factors?
- How can one describe the relation between the pedagogical and social climate (the process) and the economic resources of schools?
- Given a certain socio-cultural background, can variations in pupil achievement be attributed to the pedagogical and social climate in school or to the magnitude of resource allocation to schools.
- What is characteristic of schools with higher and lower achievements than their socio-cultural profile would lead one to expect?

Explanatory Analyses

UBER.AP.AEB showed correlations between student achievement and student characteristics. The beta coefficients are listed in the following summary:

Independent Variable	*Beta*	*Significance t*
Nonverbal intelligence (CFT 20)	.43	.0000
Student age [months]	-.12	.0000
Attitude scale: Own contribution	-.07	.0000
Attitude scale: Abilities relevant for learning	-.04	.0000
Attitude scale: Self-esteem	.13	.0000
Parents' education	.12	.0000
Migrant status [German = high]	-.16	.0000
Amount of television [hrs. Per day]	-.02	.0165
Frequency of reading newspapers	.09	.0000
Frequency of borrowing books from the library	.05	.0000
Number of books in the home	.14	.0000
Availability of own room for student	.04	.0000
Frequency of speaking German at home	.01	.1088

Next, these results were used (see Figure 3) to plot the observed class means against the expected ones for 485 classrooms (13.5 percent of the total of 561 represented had too few participating students to be included in the analysis). As can be seen, there are considerable differences between the observed and the expected class means. Taking error of measurement and the within-class error of sampling into account, there are approximately 12 classrooms with means significantly higher than the expected means and an equal number of classrooms with significantly lower ones. It is planned that the schools where this has been found be investigated further by way of special case studies in order to generate hypotheses as to the reasons behind.

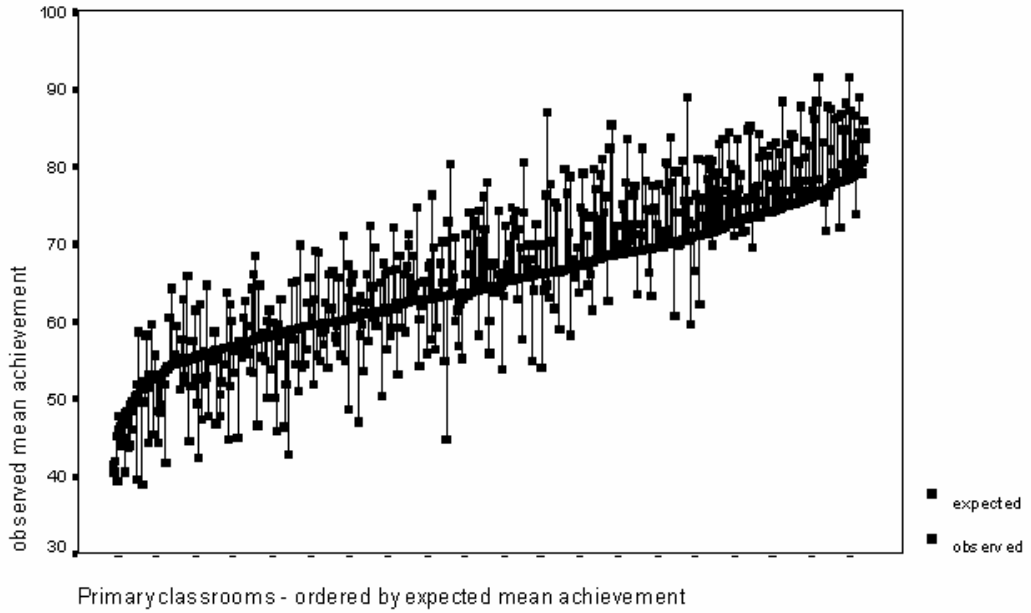


Figure 3 German example of expected (solid curved line) versus observed scores (95% confidence intervals per school).

RUG.VO analysed the correlation between student characteristics (gender, socio-economic status, academic aspiration of parents, language status) and attitudes (liking of learning subject matter, academic aspirations) and scores on mathematics and science tests administered to samples of students in 1995. It was concluded that:

- The older the pupil, the weaker his achievement score for maths,
- The less the student speaks the language of instruction, the weaker his performances in maths,
- The higher the academic aspirations of the student, the higher his maths achievement scores,
- The higher the academic aspiration of the mother for her son/daughter (as perceived and reported by her son or daughter), the better the maths performances,
- The more the student thinks to perform well in maths, the higher his/her maths scores,
- The more the student likes maths, the higher his/her performances for maths,

NFER analyses were focussed on investigating the relationship between student's characteristics, school characteristics, and teachers' characteristics on the one hand and student outcomes on the other hand. The results showed that:

- Boys outperformed girls on average in the Mathematics test, and this effect remained even when pupils' TA levels were taken into account. Girls outperformed boys on average in the Spelling test, although this effect became insignificant when TA level was controlled for.
- Pupils who had experienced some form of nursery education scored higher on the Mathematics test and the Level 3 Reading test than those who had not.

- Pupils with special educational needs scored significantly lower on the Mathematics and Spelling tests than others.
- Older pupils scored higher marks, on average, than younger pupils on all three tests, even when allowance was made for their TA levels.
- Pupils whose ethnic background was recorded as "black" scored lower marks on the Mathematics test, on average, than those in other groups.
- Pupils for whom English was an additional language underperformed in both the Mathematics and Level 3 Reading tests, even relative to their TA levels.
- Independent school pupils had significantly higher scores on all three tests, on average, than those from other types of school. This effect remained when allowance was made for their TA level.
- Scores on the Mathematics test were significantly related not only to the pupils' Mathematics TA levels but also to their awarded level in En 2 (Reading). This seems to imply that performance in such tests at Key Stage 1 is not purely determined by mathematical ability, but is also influenced by children's fluency in reading.

Other NFER results show that:

- Multilevel analysis of lesson grades using general derived factors at the school and subject levels has shown that between one-half and three-quarters of the variance between schools can be explained by these factors.

- The achievement against national norms was most strongly related to the ability level of the class and the overall attainment level at entry of the schools' pupils. Other lesson grades were also related to ability, attainment at entry, and socio-economic circumstances.

- Taking other factors into account, mathematics and science lessons tended to get lower grades than English, and grades in 94/95 tended to be lower than in 93/94.

- Groups comprising Years 8, 9 or 10 tended to get lower grades than those with Year 7 or Year 11. Mixed sex groups got slightly lower grades than single-sex groups.

- At the subject level, the factors relating to equipment, resources and accommodation were positively associated with higher lesson grades.

- The main school-level factors positively associated with higher grades were: ethos and behaviour, SEN provision, leadership and management, equal opportunities, and links beyond school. At the school level, there was a slight tendency for schools with lower ratings on quality of accommodation to receive higher lesson grades.

- Relationships between lesson grades and predefined factors were as might have been expected, with purposeful teaching and high expectations having the highest associations.

- Models with random slopes have shown that the relationship between the ability level of the class and lesson grades is very variable from school to school, and that in some schools it may even be negative (i.e. with lower ability classes getting higher grades).

- Similarly, the relative positions of mathematics, science and English seem to vary considerably from school to school. Mathematics and science grades seem to be strongly correlated, however.

One of the most striking aspects of this analysis was the extent to which the same composite school-level factors were strongly associated with both lesson grades and examination outcomes. Two of these factors (ethos and behaviour, and SEN provision) could be regarded as related to 'school climate' in the broad sense, and another (equal opportunities) was also strongly related to lesson grades. More 'classical' correlates of effective schools (leadership and management, resources) also featured, but in many cases less strongly so.
However, it was also noted that, considering the OFSTED database in terms of a resource for educational research, it must be said that if it were to be re-invented purely for that purpose it would look quite different. It would contain far fewer scales, but many if not all schools would be judged by independent teams to estimate the reliability of the judgements. However, the database exists in its present form, and can be used to give interesting and worthwhile insights into issues of school effectiveness. It would seem short-sighted to reject it as a resource for research because, given its main purpose, it does not wholly satisfy research criteria.

Other NFER analyses focusing on relationships between students' characteristics, school characteristics, parent characteristics and outcomes showed that:

- The responses to 42 of the questions could be reduced to four major factors which explained 30 per cent of the variance in the questionnaire data and summarised pupils' attitudes to school.

- Girls had significantly more positive personal reaction to school than boys, but were more likely to mention bullying. These sex differences should be qualified by the finding that for most factors there were significant variations between schools in male/female differences.

- Black pupils tended to have more negative personal reactions to school than whites, but were significantly more positive on parental attitudes and bullying. Asian pupils were significantly positive on saying it was a 'good school' and on their personal reaction to school. Other non-white ethnic minorities appeared to be positive in identifying their school as 'good' but negative in terms of personal reaction to the school and bullying.

- Of the school-level variables, one of those most consistently identified as related to pupil attitudes was the school's overall average GCSE grade per subject entry. This was positively related to saying it was a 'good school' and to bullying (i.e. lower incidence).

- No school-level factors were quite significant in their relationship to pupils' personal reactions to school, although the size of the year group was almost significant in a negative sense (i.e. smaller schools getting better reactions).

- Pupil-reported parental attitudes appeared to be positively related to having a sixth form, higher staff turnover, and parental attendance at meetings.

- Bullying (lower incidence thereof) appeared to be positively related to overall average GCSE grade and percentage of pupils eligible for free school meals, but negatively to boys' schools, having a sixth form, and higher staff turnover.

Some of the above findings may be intuitively reasonable and others less so. It should be noted that they refer to the best models which fit the data and do not imply any form of direct causality.

Considering the results from the parents' questionnaires, it was concluded that:

- Factor analysis of the 34 questions analysed resulted in three major factors being extracted, which together explained 34 per cent of the variance.

- Parents of girls tended to be more positive on the child-specific factor, and more negative on their reaction to school in general. No significant variation between schools in male–female differences were found.

- The only significant ethnic relationship found was a negative one for parents of Asian children on the child-specific factor.

- School-level GCSE performance (either mean total score or average score) was positively related to all three parental factors.

- Parents in schools with a higher percentage of free school meals had overall more positive responses on the child-specific and Year 11 education factors.

- Parents in smaller schools tended to be more positive on the child-specific factor, whereas those in schools with sixth forms tended to be more negative on the same factor.

- Parents in boys' schools tended to be more positive on the Year 11 education factor.

Again, those relationships highlighted above should not be taken to imply any form of causality.

From the SPC.ERC analyses, it appeared that:
The three relevant areas of proximal influence had a substantial impact on each domain of literacy above the effect of SES. In the case of socio-emotional ratings, the amount of variance accounted for was more than doubled as a result of the addition of these factors, i.e. the increment is greater than the amount of variance accounted for by SES factors alone. Home-school interaction influences are only slightly less powerful in their effects, and in the case of literacy press, while the increment of variance is less than the other areas of influence, the increment is still a major one. As in the case of the multiple correlations without controlling for SES, the pattern of influence was slightly different for the expository domain, with social-emotional ratings yielding a higher increment in prediction than was the case with narrative or document literacy.

From the RUGR.IER analyses on the school climate indicator it appeared that this indicator did not (contrary to expectation on the basis of school effectiveness theories) explain variance in the outcome measures.

The Web site created by ULGG-DEPF also includes the results of explanatory analyses covering student-, and school characteristics such as:

- Student: gender, age, home language, professional expectations, duration of study, attitudes towards science, socio-economic status, homework, auto-appraisal;
- School: home background of students, absenteeism, infrastructure, presence of parental association.

THT.OCTO presented results of explorative path analyses on data collected with grade 8 students and classrooms in eight Western and two Central European education systems. For the ten education systems the resulting general path model explains only 19% or less of the variance in achievement in mathematics. In many systems home educational background and students' attitude towards mathematics showed positive relations with achievement in mathematics, out-of-school activities a negative. Due to the psychometric quality of scales and non-availability of measures of important factors at classroom level (e.g. time on task and teacher's expectation), no significant results were found of factors that can be manipulated by policy-makers.

Value Added
The graph presented by UBER.AP.AEB shows which schools are performing above and below expectation (see Figure 3, above).

Many of the projects from the University of London were focussed on value-added measures.

ULON.IE investigated, amongst others, the relationship between view of parents and outcome measures of students. There was evidence of general stability in parents' views for the majority of schools, but for a minority there were marked changes from 1995 to 1997. There was no straightforward relationship between parents' views and schools' effectiveness in terms of value added measures. Some schools were rated more favourably and were classified as more academically effective. However, some other schools which were classified as ineffective in value added terms were also positively rated by parents.
The results from contextualised multilevel analyses conducted by ULON show evidence of significant relationships between pupils' background characteristics and KS1 (value added) outcomes. These have important equity implications for practitioners and policy-makers.

The following factors were consistently associated with better performance at KS1:

◊ age (older members of the year group doing relatively better);
◊ all forms of pre-school experience had a positive impact in comparison with no experience. However, entry at rising five (i.e. before the term of their fifth birthday - the RISE category) showed the strongest relationship, followed by the category private nursery. (In the case of private nursery, socio-economic factors may also be linked as parents have to pay for the latter type of provision.)

whereas two factors were consistently associated with poorer performance at KS1:

◊ eligibility for free school meals (FSM); and
◊ English as an additional language (EAL).

Gender (girls) showed a strong and significant positive relationship with both reading and writing at KS1. However, whilst weaker, there was a significant negative relationship between gender (girls) and performance in mathematics and science at KS1.

Other analyses conducted by ULON.IE showed that:
The comparison between school residuals for specific academic subjects and an overall academic measure (for secondary pupils) confirm previous findings and show that some schools can vary considerably in their effectiveness across two curriculum areas (English and mathematics). The correlation between primary school residuals for reading and mathematics ranged from 0.39 to 0.47 (note that a perfect positive relationship is represented by 1.00).

The results showed that some schools obtained statistically significant differences in value added results for different groups of pupils categorised by prior attainment measures.

The results from this project indicate that overall there are greater differences between schools in terms of cognitive value added scores than non-cognitive ones, particularly at the secondary level. Therefore the results suggest that schools have a greater impact in terms of pupils' cognitive outcomes in comparison to their reported attitudes. This finding may reflect the possibility of greater focus by Scottish schools on academic goals rather that on the fostering of positive pupil attitudes.

Trends

NFER research on trends showed that the key finding of a survey conducted in 1995 was that the average score rose between 1991 and 1995, and in 1995 had returned to the 1987 level.

The decline between 1987 and 1991 took place in the context of the introduction of the National Curriculum and of increased teacher turnover, and the improvement between 1991 and 1995 when the education system had settled down again. Those changes in the system might be enough to account both for the decline and for the recovery in average reading levels of pupils in Year 3.

In order to examine how the situation in Greek schools changes over time, UAT.DPED compared achievement over time by using data from a representative sub-sample of TIMSS (1995) and data collected in 1994 and 1996 at different grade levels. For these analyses data were used from achievement tests in math and science (per cent correct) as well as a few social and educational (background) variables.

NAE.DE describes the Swedish National Assessment Studies (SNA) which have covered different content areas or schoolsubjects. Every third year students knowledge skills and attitudes related to Swedish, English, Mathematics, Natural and Social science have been tested. These "core subjects" have been followed by other and over the years different school subjects and other areas related to Swedish compulsory school. Examples of these other areas are: Arts, Handicraft, Home economics, French, German, Problem solving in environmental education, Sports, Study and vocational guidance.

The data bases are available as SPSS- files. Codebooks and variable index are available. Interested researchers may have access to the material, which is available on the Internet (although a password is required).

Sundries

In addition to the work on indicator measurement and explanatory power of indicators it is worthwhile to point to one of the NFER analyses which contains approaches for estimating the likelihood that national education targets will be reached in the year 2000.

Incorporating the Products of Network Members in the EU-AEI Web Site

The products from network members, which were described in the sections above, should in the forthcoming period be incorporated in the network’s Web sites. The work on indicators can be made accessible by including an additional icon in the indicators pages, which lead users to the underlying instruments and

documentation on measurement properties of the indicators (reliability, validity and explanatory power). The selection of indicators could be managed by an indicator selection committee, which judges on the basis of standards if certain indicators have sufficient quality.
The results of explanatory analyses might be made available by making clickable the arrows between indicator boxes in the conceptual framework at the Web site containing the statistical overview.
Trends could also be made accessible by including a separate icon at pages containing statistical results collected at one point in time.

Reflections on the Results of the Pilot Projects

The results from the pilot projects showed that across the network there was a rich variety of experiences with many different educational process and outcome indicators. However, it should also be noted that in quite some cases, essential information about the quality of indicators was lacking. Especially information about validity (is the indicator reflecting what was intended to be measured?) was not available. This observation should not be interpreted as criticism, but rather be understood as reflecting the difficulty (and sometimes impossibility) of running (time-consuming) validity studies, given the constraints of budget and timelines which usually are imposed on national educational assessment projects. Such constraints seem to be occurring in many of the projects. Although such observations may lead to pessimistic thoughts about the possibility to systematically improve the quality of national assessments, it may also be taken as a signal that cooperative sharing of work throughout Europe may be beneficial for all parties involved.
One may argue that such efforts would not be profitable for measures, which are culturally sensitive. On the other hand, it may also be said that systematic investigations of the quality of indicators (reliability, validity, as well as explanatory power) would considerably improve the knowledge base of what are promising generic indicators and which are (culturally) specific.
Another observation from the explanatory analyses is that in many cases all variance in the outcome measure is subject of explanation. As argued earlier in this chapter, assessments are part of a cycle of steps aimed at improving the education system or subsystems. Assessment results in displaying variance in the measures used. Variance will always exist and is in itself no reason for worry. What counts is if students are achieving below expectation. For instance, people who do not speak/understand the language of instruction hardly can be expected (ceteris paribus) to function at the same level as native speakers. Therefore, the use of value added measures and especially the improvement of methodologies in this area should be a topic for more intensive European cooperation.

The overall conclusion from the pilot projects is that, throughout Europe, researchers are trying to find appropriate instruments for measuring indicators of factors such as the ones shown in the network's conceptual framework. An important component of this search is the identification of indicators which can explain why unexpected low achievement is occurring in (parts of) the population of students under study.
Realising such task on the basis of single projects is (in the absence of huge investments) hardly possible. Therefore, it may be argued that an increased European effort of sharing the developmental and research

tasks may be beneficial for all parties involved. Such Utopia may only be reachable if a number of conditions are satisfied, such as:

- Rigorous standardisation of reporting indices of indicators (and underlying operationalisations), which need to be easily accessible via instrument banks (including indices of reliability, validity, and explanatory power).
- Systematic replication in many contexts of potential promising indicators. Results should be fed back to the bank.
- Relaxation on the side of researchers (involved in educational assessments) with regard to the goal of trying to be original. Such attitude can be rather detrimental (because of its lack of orientation on replication and/or improvement of existing measures) for the future of educational assessments.
- Middle and long term plan for European cooperation on assessing educational progress and investigating educational effectiveness and its underlying explanatory factors.

3 CONCLUSIONS FOR FUTURE WORK AND REFLECTIONS ON POLICY IMPLICATIONS

This chapter presents the conclusion regarding the network's activities in terms of a work programme for the period 1998-2000. The final section contains some reflections on policy implications.

Introduction

The project reported in this document was a network rather than a substantive research project. It focussed on strengthening the infrastructure for cooperation and explored (via small scale pilot projects) viable paths for future work. The main conclusions for continuing work in the forthcoming period have been laid down in a work programme for the period 1998-2000.
This work programme is described below.

Methodology

The methodological work of the network in the period 1998-2000 will be focussed on the following projects:

1. European instrument bank for educational indicators.
2. Development and sharing of expertise with regard to measures of added value.
3. Development of approaches for improving fair testing of student achievement.

European Instrument Bank for Educational Indicators

The Achilles heal of any educational assessment is the quality of the instruments which are used for measuring aspects of the educational system (see framework in the previous chapter). The construction of educational assessment instruments is a very time consuming and costly activity which in most countries (let alone across countries) is not well coordinated. Such lack of coordination prevented a systematic accumulation of knowledge about which measures work effectively and which do not, despite a large number of publications which have appeared in this field. These publications are usually mostly focussed on descriptions and explanatory analyses, but tend not to disclose details about the quality of instruments which were used (such as reliability and validity). A striking example of the lack of attention for quality indices was given by Pelgrum, Voogt, and Plomp (1996) who showed that for a list of 180 different curriculum indicators which were published, almost no quality indices were reported. For individual researchers and national assessment teams it is usually a too time-consuming and often impossible operation to collect instruments which were used in previous assessments in their own country or other countries.

The EU-AEI database inventory (which was described above) showed that there is quite some overlap between countries with regard to indicators which assessments teams try to measure. Such teams could greatly profit from easy access to instruments, which were already developed and used by other teams (in their own country or other countries). Therefore, the EU-AEI network will start to create an instrument-bank for educational indicators. This bank will be organised according to the conceptual framework, which also constituted the basis for disclosing the assessment statistics.

The bank will have the following content:

1. Actual instrument (in language of origin and in English).
2. Description of samples in which instruments were used.
3. Descriptive statistics (frequencies, means, p-values, standard deviations, standard errors).
4. Reliability and validity estimates.
5. Relationships with educational outcomes and educational process measures.

The first filling of the bank will be done by the members of the EU-AEI network. This will be followed by a dissemination stage and a stage of transferring the bank to an appropriate institutional setting which can maintain and continuously update the bank.

The spin-off of this activity will be that the bank can play a catalysing role in promoting the use of standardised instruments in national assessments throughout Europe. This will create a win-win situation: use from instruments from the bank may result in getting comparative information in return. Such information is crucial for interpreting the results of national assessments.

Parallel to the filling of the bank with existing high-quality indicators, a research plan for indicator and methodology development for the long-term will be developed.

*Development and Sharing of Expertise with Regard to Measures of **Value Added** Approaches*

The value added concept is based on the idea of measuring relative pupil progress, usually in cognitive outcomes such as language attainment during a given period of time. However, the concept can also be applied to non-cognitive outcomes such as pupil attitude scales. Thus, in an education context the term value added is used to refer to the residual value that is added by schools to pupil attainment over and above the progress or improvement that might be expected in a normative sense. In recent years this approach has been increasingly accepted as the most appropriate quantitative method of estimating school and teacher effectiveness and thus an important tool in evaluating educational quality. However, in practise EU national and local educational administrators and policy makers, academic researchers, and teachers have employed a wide variety of different procedures for measuring educational effects using either pupil background factors (such as socio-economic status) or pupil prior attainment data, or both, as well as different levels of sophistication in the analysis (such as employing individual pupil level data and multilevel modelling or analyses of cruder aggregated school level data). In addition, new research has emphasised the importance of examining the internal variations in school effectiveness such as subject or departmental effectiveness or the separate effects for different groups of pupils (such as boys and girls) (Thomas et al, 1997). Therefore, the

first aim of the EU-AEI network is to share and develop expertise across different member countries on the key issues concerning the methodology and limitations of measuring value added. The second aim is to examine issues arising from the intended purpose and subsequent impact on educational quality of utilising value added measures in different EU member countries. The purposes of measuring value added can vary from that of stimulating and informing school self-evaluation on a confidential basis to that of public accountability and the centralised monitoring of educational quality over time. For example, value added data have been provided confidentially to schools by several English local educational authorities to assist school self-evaluation and raise questions about changes and/or consistency in results over time and allowing schools to compare themselves with other schools within limits of statistical uncertainty (Thomas et al, 1996). An alternative purpose of value added measures can be to inform the guidance and monitoring of individual pupils' progress. Therefore, it would be of considerable benefit to EU member countries to share experiences of both the theory and the reality of how using value added information for different purposes impacts on the standards of educational quality over time. The value of schools' educational quality is, however, broader than what can be measured by attainment in a few specific areas of pupil activity. Therefore a third aim is to examine EU member countries approaches to developing a comprehensive value added evaluation framework that encompasses measures related to numerous other aspects of a school's mission, processes and outcomes. Of specific interest are the definition and measures of broader learning processes and pupil outcomes such as pupil attendance, capacity for independent learning, attitudes to school and to learning, behaviour, and self-concept which reflect the overall aims and needs of schooling in the 21st century in EU nations.

The comparative findings of this exchange will assist the EU member countries to work towards a common understanding the multi-faceted nature of school effectiveness and improvement across a range of outcomes and national contexts. It will also assist educational policy makers in understanding how different socio-economic and national educational policy contexts may influence the size, extent and stability over time of value added measures. The EU-AEI network will also address the issue of how value added data collection, analysis and feedback systems may be planned and operationalised in EU nations with the aim of stimulating improved use of evaluation procedures in education.

*Development of Approaches for **Fair Testing** of Student Achievement*

In any use of student achievement measures outside the school it is important that there is confidence in the tests with regard to fairness and equity. Various technical approaches can be used for this: e.g. differential item functioning which is used to determine whether any questions in a test or exam are particularly difficult for a certain group once that group's overall test performance is taken into account; scrutiny of items for 'facial bias'; sensitivity reviews for items which are more relevant to one particular cultural group or gender.

The range of item/task types needs to be examined, set against what we know of how these interact with gender and curriculum experience. It is important, however, also to examine the constructs which are being

assessed; examiners' definitions of the constructs being assessed need to be available to allow examination of construct validity.

This is a complex area: Linn (1992) argues that tests may be differentially valid for members of different groups and may thus have different meanings for these groups.

In order to interpret any findings we also need to know about equal access issues: have all groups had exposure to the constructs being assessed, in similar amounts? IEA data contains some of this information in terms of curriculum coverage in and hours devoted to particular subjects each week.

It is proposed that, for each of the assessment data sets held by the network, an audit be made of the approach to fair testing: statistical analyses; item review; possible examination of construct validity; range of assessment approaches; information about equal access.

Further Development, Dissemination of the AEI-Knowledge Base-Web Site

Following the decisions taken at the plenary meeting of EU-AEI (Frankfurt, 1997) the further development of the EU-AEI Statistical Overview will be concentrated on the following aspects:

- Trend statistics;
- Effectiveness estimates;
- Help desk;
- Remote data processing
- Links with training projects for educational researchers.

The intended work on each of these aspects will be shortly summarised below.

Trend Statistics

The current version of the Web site contains statistics from an international assessment on mathematics and science, which was conducted in 1995 in almost all EU countries. These statistics reflect part of educational reality at a particular point in time. It shows potential strengths and weaknesses of educational systems. However it does not reveal to what extent improvements are taking place. Such trends may come from data collected in earlier international assessments in which EU-countries participated, for instance the ones as listed in Table 5.

Table 5

Involvement of EU Member States in international assessments of student achievement in mathematics, science, and technology

	Austria	Belgium-Flemish	Belgium-French	Denmark	Finland	France	Germany	Greece	Ireland	Italy	Luxembourg	Netherlands	Portugal	Spain	Sweden	England
Mathematics-IEA																
1964		x	x		x	x	x			x		x			x	x1
1981		x	x		x	x			x4	x	x	x			x	x2
1995	x	x	x	x		x	x	x	x	x		x	x	x	x	x1
Science-IEA																
1970		x	x		x	x	x			x		x			x	x1
1984					x					x		x			x	x1
1995	x	x	x	x		x	x	x	x	x		x	x	x	x	x1
CompEd-IEA																
1992	x						x	x				x				
Mathematics-IAEP																
1988									x					x		x3
1991						x			x				x	x		x2
Science-IAEP																
1988									x					x		x3
1991						x			x				x	x		x2

Notes: CompEd: Computers in Education survey. IEA = International Association for the Evaluation of Educational Achievement. IAEP = International Assessment of Educational Progress. 1 = Samples from England; 2= Samples from England and Wales; 3 = Samples from England, Wales, and Scotland; 4 = Curriculum analysis only.

EU-AEI will, between 1998 and 2000 undertake the following activities:

- Map variables from different assessments.
- Locate original databases.
- Extract statistics from published sources or calculate from the original databases.

Effectiveness Estimates

The Site Map (see the conceptual framework shown in Chapter 2) in the current version of the EU-AEI Statistical Overview shows potential links between the blocks of indicators. These links are not yet clickable. Work in the forthcoming 2 years will focus on:

- review of meta-analyses in this field (e.g. Scheerens & Bosker, 1997);
- review and synthesis of analytic studies of international (IEA) data bases;
- additional meta- analyses on the causal relationships between input, context and process variables on the one hand and (adjusted) outcome measures on the other.

Help Desk

A new category, which will be introduced on the home page of the network, is a help desk. This will be the e-mail address of the coordination center to which users may send questions about the statistics. Such questions may include:

- Questions for clarification.
- Assistance with interpretation.
- Questions regarding further analyses.

The coordination center will seek assistance of the experts in the membership for answering these questions. With regard to questions for which additional data processing and analyses are needed, the network will seek ways (via national governments and the European Commission) to support such activities.

Remote Data Processing

For those researchers in Europe, who would like access to the original raw data, the network will set up a facility, which allows for remote data processing

Training Programs

In order to promote that new generations of European researchers get acquainted with the methodology of conducting assessments, EU-AEI will organise two training workshops of one week each, which will be open for researchers working in the field of educational assessments. The topics to be covered at these workshops are:

- Assessment: paradigms and methodologies.
- Monitoring process.
- Sampling.
- Instrumentation.
- Data collection procedures.
- Data processing and analysis.
- Interpretation.
- Dissemination.

Reflections on Policy Implications

The network's main mission in the period 1996-1998 was primarily focussed on improving the infrastructure for comparative assessments in a European context. As such this is not a substantive activity addressing policy issues. Nevertheless, from the discussions and exchanges of information, which took place in the past two years, a few observations can be made which have implications for education policies at the EU-level. From the meeting documents that were summarised in Chapter 2, it appears that the network members agree that at the EU policy level, there is not yet a strong movement towards strengthening the European research activities in order to collect on a regular basis standardised comparative data on student outcomes (and instructional processes) that allow countries to evaluate the development of their educational systems over time. On the other hand, it can be observed that there is an increased interest in comparative data (for instance in the Socrates call for proposals from Spring 1998) and the support for pilot projects for school self evaluation. As such, these activities suggest that an increased need and awareness of the necessity to develop tools that allow for rational judgements of strengths and weaknesses of educational systems (at different levels: micro, meso and macro) is developing. However, from the point of view of the EU-AEI network, the most important basis for further stimulation of the development of a strong comparative assessment research work force would be the political intention and willingness to start an intensive cooperation between countries in order to share the burden of creating high quality systems for comparative assessment and evaluation of educational progress within a European context. In the words of the European Round Table of Industrialists (1995): *It has become a matter of urgency to set up a European Education Information and Monitoring System in order to record experience, evaluate results and improve the quality of education on a European level*. Until such system is realised one will need to rely too much on speculation for making decisions on the way ahead, or in the words of Black (1995), it will be difficult "*to avoid ideology and to identify and tackle real problems*".

Maybe it is a matter of time that such statement ultimately will also be expressed at the EU political level. Once this willingness exists, European institutes (such as Eurostat, Eurydice) can start to develop an infrastructure for systematic evaluation of Europe's educational progress, with the help of a workforce of researchers specialised in international comparative assessment and with the guidance of the European network for policy-makers for the evaluation of educational systems.

ACKNOWLEDGEMENT

This report was written by the coordination center, but it is acknowledged with great appreciation that the network members provided the underlying materials. The coordination center would like to express its gratitude for the cooperation with the participating (associate) member institutes, and is looking forward to the continuation of activities in the future. It is felt that strengthening of the network is very important at this moment that the interest for comparative assessments is increasing. The coordination center would like to express the intention to strengthen the cooperation with OECD and IEA, especially with regard to the interpretation of European data in relation to European policy options.

REFERENCES

Black, P. (1995). *Ideology, evidence & the raising of standards*. London: Kings College.

Bosker, R.J. & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. *International Journal of Educational Research*, vol. 21 (pp. 159-180).

Creemers, B.P.M., & Scheerens, J. (1994). Developments in the educational effectiveness research programme. *International Journal of Educational Research*, 21 (2), 125-140.

European Round Table of Industrialists (1995) *Education for European: Towards the Learning Society*. Brussels:ERT.

Gipps, C. and Murphy, P. (1994). *A Fair Test? Assessment, Achievement and Equity*, Open University Press.

Linn, M. (1992). Gender Differences in Educational Achievement in *Sex Equity in Educational Opportunity, Achievement and Testing*, Princeton, ETS.

Oakes, J. (1987). *Conceptual and measurement issues in the construction of school quality indicators.* Washington D.C., AERA-paper.

Pelgrum, W.J., Voogt, J., & Plomp, Tj. (1995). Curriculum indicators in international comparative research. In: OECD. Measuring the quality of schools. Paris: OECD.

Scheerens, J. (1990). Process Indicators of school functioning. *School Effectiveness and School Improvement*, vol. 1, (pp. 61-80).

Scheerens, J. (1995). Theory embedded principles of effective schooling, *Paper ECER*. Bath.

Scheerens, J., & Bosker, R.J. (1996). The Foundations of School Effectiveness. Oxford: Elsevier.

Scheerens, J., Hegarty, S. (1997) *Educational assessment, effectiveness, and innovation: a conceptual framework*. Enschede: University of Twente, OCTO.

Thomas S. and Mortimore P. (1996). *Comparison of Value Added Models for Secondary School effectiveness*. Research Papers in Education 11 (1) pp 5-33.

Thomas, S. Sammons, P. & Mortimore P. (1997). *Stability and Consistency in Secondary Schools' Effects on Students' GCSE outcomes over 3 years.* School Effectiveness and School Improvement 8, (2):169-197.

ABSTRACT

In April 1995 the European network for educational research on Assessment, Effectiveness, and Innovation (EU-AEI) was founded. The members of the network are educational research institutes from all EU Member States, while Iceland, Cyprus, and Switzerland are represented as associated members. The mission of the network was at the network's founding meeting defined and described as follows:

> *To carry out international comparative research in education in the domain of achievement, assessment, effectiveness and innovation within Europe.*
>
> *This orientation of the network implies that there is a key interest in the collection and analysis of both outcome oriented data at the level of individual students, and educational background conditions at the level of students, teachers, classroom arrangements, curricula, school management and organisation, and system parameters at the regional and national level (including specific educational policies).*
>
> *Substantive questions that are at the basis of the network's research interests are questions about micro-, meso- and macro-level conditions that enhance educational performance, and questions about the potential for educational change and innovation.*

In the past two years the members of the network have cooperated in order to execute cooperative projects, which resulted in:

- Web site with links to the member institutes;
- Inventory of available national assessment data bases;
- A prototype of a Web site containing statistical indicators on educational processes and outcomes;
- Pilot projects focussing on sharing of expertise on indicator construction and explanatory analyses;
- A work programme for the period 1998-2000.

Also several initiatives were undertaken for further cooperation, mainly within the framework of TSER-2. This resulted in one project called *Capacity for Change*, which consists of theoretical analyses, an inventory and evaluation. The 8 participating countries in the project are: Belgium, Finland, Greece, Italy, the Netherlands, Portugal, Spain and the United Kingdom. The duration of the project is 2 years.

One of the main conclusions from the past two year experiences is that the network has a good basis for further cooperation for conducting and sharing expertise with regard to comparative studies of European educational systems. It was concluded that further work on standardisation of indicators on educational progress within a European framework is needed and that efforts towards solving some of the fundamental problems in measuring educational effectiveness should be strengthened.

EU-AEI

EUROPEAN NETWORK FOR EDUCATIONAL RESEARCH ON ASSESSMENT, EFFECTIVENESS AND INNOVATION

http://www.to.utwente.nl/prj/euaei/index.htm

University of Twente

EXECUTIVE SUMMARY

EUROPEAN NETWORK FOR EDUCATIONAL RESEARCH ON ASSESSMENT, EFFECTIVENESS AND INNOVATION (EU-AEI)

By

W.J. Pelgrum

May 1998

Introduction

In April 1995 the European network for educational research on Assessment, Effectiveness, and Innovation (EU-AEI) was founded. The members of the network are educational research institutes from all EU Member States, while Iceland, Cyprus, and Switzerland are represented as associated members. The mission of the network was at the network's founding meeting defined and described as follows:

> *To carry out international comparative research in education in the domain of achievement, assessment, effectiveness and innovation within Europe.*
> *This orientation of the network implies that there is a key interest in the collection and analysis of both outcome oriented data at the level of individual students, and educational background conditions at the level of students, teachers, classroom arrangements, curricula, school management and organisation, and system parameters at the regional and national level (including specific educational policies).*
> *Substantive questions that are at the basis of the network's research interests are questions about micro-, meso- and macro-level conditions that enhance educational performance, and questions about the potential for educational change and innovation.*

In 1996, the network acquired for a period of two years, a grant from the European Commission, which allowed for strengthening its infrastructure and dissemination of its knowledge base.

In this document, a short description of the rational for the network's existence, and a review of work conducted in the past two years is presented.

Background and Rational

The history of standardised and empirically based assessments of educational achievement in Europe is relatively short. While in most European countries educational assessments of student achievement are currently common practice, it is not more than a decade ago that the political and societal awareness of the necessity of using methods of objective measurement of educational processes and outcomes reached a level which justified undertaking large-scale costly operations for getting benchmark data on educational progress. The developments in Europe not only started much later than in the USA, but still have not reached a level of political attention and commitment which is usual in that part of the world. Nevertheless, in recent years it has become clear that in many EU Member States policy makers sympathise increasingly with the idea to replace ideology with statistical facts as a basis for sound communication between educational partners and for educational policy-making. A recent striking example of this increased societal awareness comes from a non-educational source, namely the European Roundtable of Industrialists, which wrote (1995): *It has*

become a matter of urgency to set up a European Education Information and Monitoring System in order to record experience, evaluate results and improve the quality of education on a European level. This citation reflects an ideal of creating a system which can show actors who are directly or indirectly involved in educational matters, what is the condition of education at a particular point in time, to what extent changes are taking place and from the perspective of particular stakeholders how segments of the systems compare in terms of their rate of change.

The current situation with regard to educational statistics in Europe is yet quite distant from this ideal. A typical citizen in Europe (be it a politician, school leader, or parent) will not yet be able to find objective information for finding answers to questions like "How good is this school/school system/etc."

The EU-AEI network is trying to contribute to improve this situation, by creating an information infrastructure which gives access to knowledge about the functioning of educational systems, while at the same time creating means which are aimed at sharing methodologies for measuring and analysing educational progress in a scientifically acceptable way.

Review of Work Conducted Between 1996 and 1998

The main activities of the EU-AEI network in the past one and a half year were the production of Web-pages containing the overall description of the network and the links to the member institutes, the statistical overview and the database containing the description of the content of national assessment data sets, which are available in the member institutes.

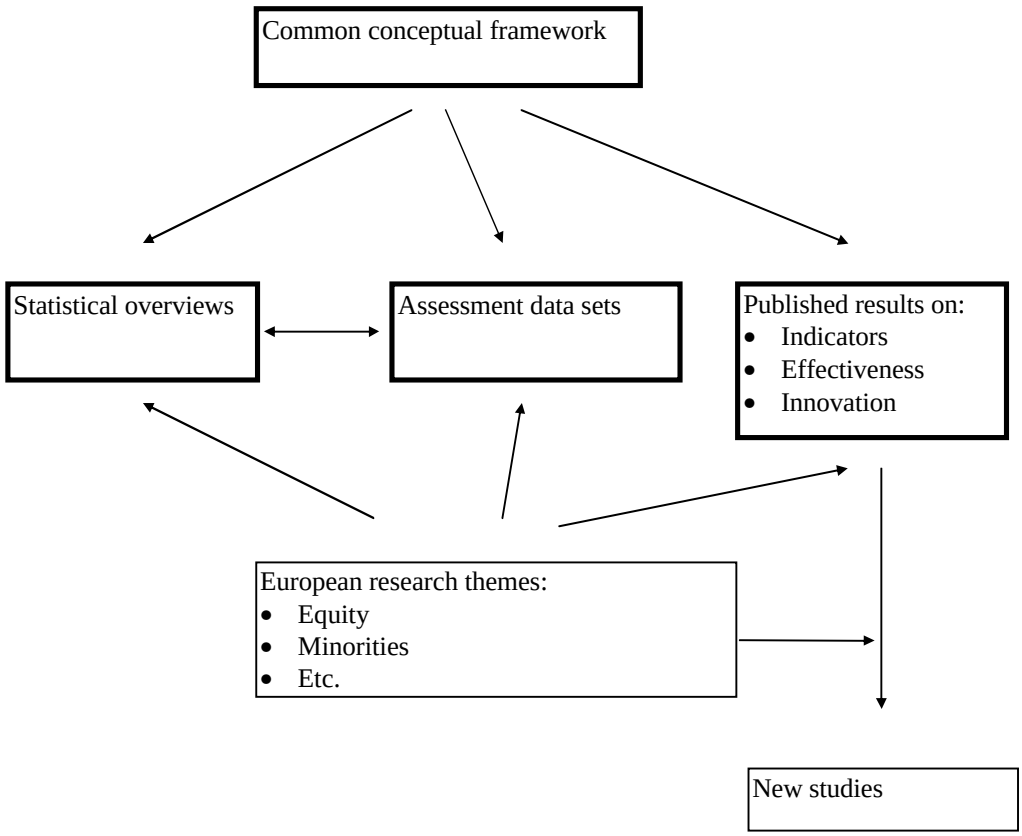
In the following sections the results of this work will be summarised.

The EU-AEI Overall Web Site

This site, which is available at address http://www.to.utwente.nl/prj/euaei/index.htm, contains project documents, announcements of meetings, and links to the member institutes.

The EU-AEI WWW-Statistical Overview

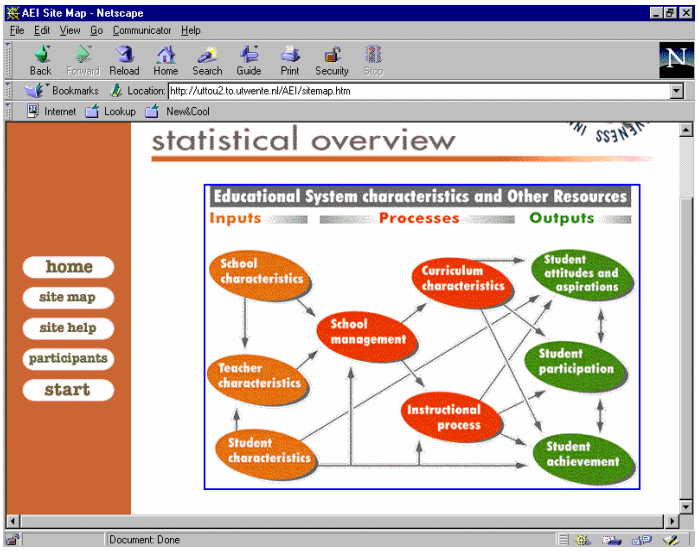
The work on the EU-AEI WWW-statistical overview (http://uttou2.to.utwente.nl/AEI/index.htm) is embedded in a structure of activities, which can be visualised as follows:



The figure above shows how the different project components are interrelated. The conceptual framework offers the basic categorisation principle for accessing the statistical overview, the assessment data sets, and published results (not yet implemented). The last three components may be linked to European research themes and stimulate the initiation of new studies.

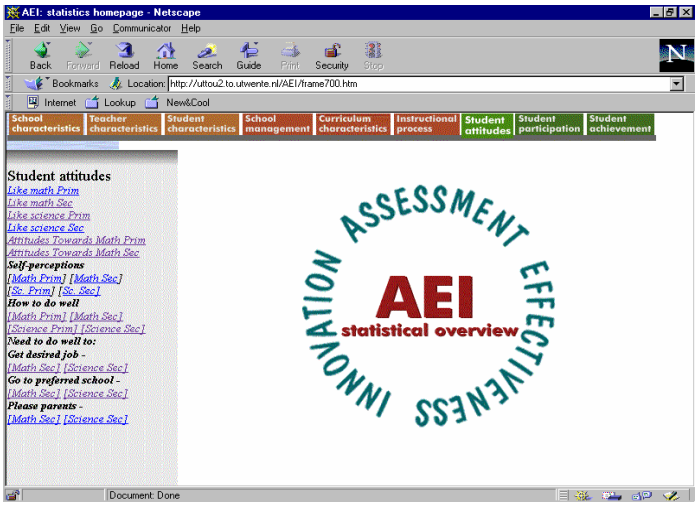
On the home page of the EU-AEI Statistical overview (see below) it is explained that the current site is seen as a pilot version which needs to be reviewed and further elaborated in terms of functionality and content. Amongst others, it is pointed out that also potential overlap with WWW statistical information from Eurostat and Eurydice needs to be explored and where appropriate, links to the sites of these institutes need to be established.

The statistical information at the site is accessible via the conceptual framework of EU-AEI, which was developed by Scheerens and Hegarty (1997). By clicking the category “Site Map” at the home page, the user is linked to a clickable version of this conceptual framework (see below).

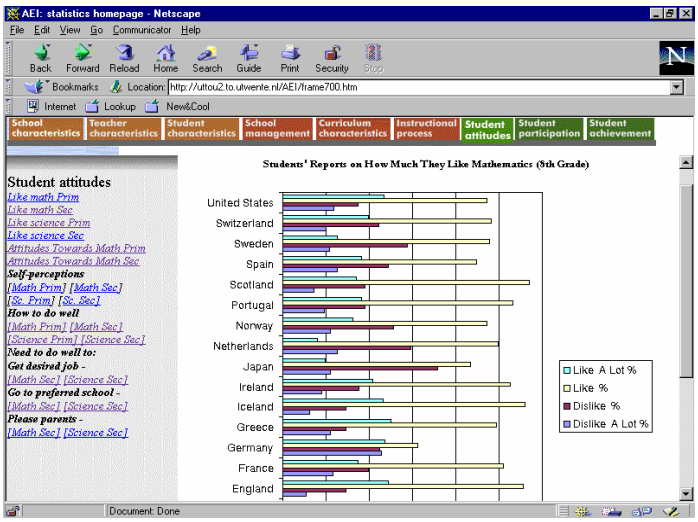


The pilot version of the site contains at this moment a total of 136 comparative statistics for the EU-countries (and Japan and the USA), from the categories school characteristics (4), teacher characteristics (18), student characteristics (17), Instructional processes (39), Student attitudes (22), and. student achievement (36). These statistics are available for primary and lower secondary schools.

By clicking a particular category on the Site map, like for instance student attitudes the user receives a list of more concrete underlying categories (see left column on screendump below), which shows for which attitudinal aspects comparative statistics are available from representative samples of students from EU-countries, for which educational level (Primary or Secondary education), and for which subject areas.



Finally by clicking the most concrete category (for instance “Like math Sec.”) the statistics can be accessed (see below).



The pilot version of this WWW-site was demonstrated and discussed at the plenary meeting of EU-AEI on September 24 in Frankfurt.

At that meeting it was decided that the current version of this statistical overview might be a first step towards the creation of a full-site, containing:

- Trend statistics;
- Effectiveness estimates;
- Help desk;
- Remote data processing;
- Links with training projects for educational researchers.

Furthermore, the members concluded that for the short term the following actions should be undertaken, in order to evaluate the need for, the relevance of, and practicality of the current pilot version.

- Review by members of the network;
- Evaluation of the current site by 2 educational journalists from each country;
- Consultation of policy-makers network;
- Consultation of Eurydice and Eurostat.

Inventory of National Education Assessment Databases

Whereas the partners in the network have cooperated to collect international comparative data on educational achievement, they also are involved in national assessment projects. National assessments are very costly operations, due to the fact that usually quite large sample sizes are needed and also because instrument construction is a very labour intensive component. Researchers are constantly searching, on the basis of theoretical insights and empirical evidence, for improvement of the measures, which are collected in national assessment. In this respect the EU is a large laboratory in which constantly innovations with regard to the design and instrumentation of national assessments are taking place. However, there has been a serious lack of coordination of these efforts. In order to facilitate the search for databases, which may be relevant during the design period of national assessments, the network partners decided to make an inventory of databases to which each of the partners had direct access. This inventory was conducted on the basis of a standardised questionnaire, which enabled the partners to describe their databases in terms of design characteristics and content.

The result of this inventory will be made available on the Web. Table 1 shows in a condensed form the content of all 48 data sets. In this table, each column represents one data set. The overview shows in which areas potential cooperation on investigating subsets of indicators may be further explored. For instance, this overview shows which indicators are quite common across the countries. Probably more interesting is the fact that developmental work with regard to some indicators (for instance, school climate indicators) is concentrated in a few countries. Especially this last category may be important for researchers involved in national assessments in other countries.

Table 1

Condensed Selection of Contents of 48 Data Sets ***(each column represents one national data set)***

	OVERVIEW DATABASE INVENTORY: selection of content categories																																															
Subject areas:																																																	
☐ Mathematics						x				x		x	x	x	x		x											x	x			x			x		x		x,	x	x			x	x			x	x
☐ Science							x	x							x		x	x										x	x			x							x	x	x			x					x
☐ Mother language	x		x	x					x		x			x	x	x	x			x			x	x	x	x		x	x	x		x					x		x	x	x			x		x		x	x
☐ Other		x			x									x			x	x	x		x						x	x	x			x	x	x		x		x	x	x	x	x	x	x			x		
☐ **Student**																																																	
☐ Background							x		x	x	x	x	x					x		x	x						x					x	x	x	x				x	x		x	x	x	x	x	x	x	x
☐ Age	x	x	x	x	x	x	x		x	x	x	x	x		x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x		x		x	x	x	x	x	x	x	x	x	x	x
☐ Gender	x	x	x	x	x	x	x		x	x	x	x	x		x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x		x		x	x	x	x	x	x	x	x	x	x	x
☐ Ethnicity		x	x									x			x	x	x	x	x	x							x							x			x		x	x	x	x		x				x	x
☐ Parental support/stimulation	x	x	x		x	x	x		x	x	x	x	x						x		x									x	x	x	x	x					x	x					x	x	x		x
☐ Home characteristics	x	x	x	x	x				x	x	x	x	x				x	x	x		x									x	x	x	x	x			x		x	x	x	x	x		x	x	x		x
☐ Activities outside school	x	x	x	x	x		x		x	x	x	x	x					x	x		x						x			x		x	x	x													x		x
☐ Achievement																																																	
☐ Multiple choice tests	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				x					x					x	x	x		x	x				x	x			x	x	x	x			x
☐ Open answers	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x		x		x			x	x	x	x													x	x				x					x
☐ Performance tests				x		x				x		x	x		x		x																x				x		x	x	x			x	x	x			
☐ Attitudes																																																	
☐ Towards subject matter	x	x	x	x	x	x	x		x	x	x	x	x				x	x			x					x	x			x		x	x	x					x	x		x	x		x	x	x		x
☐ Towards school	x		x	x	x				x	x	x	x	x				x	x	x		x						x			x		x	x		x				x	x		x	x				x		x
☐ School climate	x		x	x					x	x	x	x	x				x		x		x									x		x	x				x		x	x		x	x		x	x			x
☐ Towards learning in general	x	x			x				x	x	x	x	x				x	x			x													x					x			x	x				x		x
☐ Instruction																																																	
☐ Content	x	x		x	x	x			x	x	x	x	x					x												x	x			x								x							x
☐ Homework	x	x	x		x	x	x		x	x	x	x	x								x									x		x		x								x			x	x	x		x
☐ Use of information technologies		x	x	x	x	x				x		x						x								x	x							x								x							x
☐ Other major categories of student variables:																											x															x							x
☐ **Teacher**																																																	
☐ Background																																																	
☐ Gender	x	x	x	x	x	x	x		x	x	x	x	x																	x	x	x	x	x	x	x		x	x		x	x			x	x			
☐ Age	x	x		x	x		x		x	x	x	x	x																	x		x	x	x	x	x		x	x		x	x			x	x			
☐ Training	x	x	x		x	x	x		x	x	x	x	x																	x	x	x	x	x	x	x	x	x	x		x	x			x	x			
☐ Instructional strategies	x	x	x	x	x	x	x		x	x	x	x	x										x							x	x	x	x	x	x	x	x		x						x	x			
☐ Assessment practices	x	x	x	x		x			x	x	x	x	x																	x	x	x	x	x	x		x		x						x	x			
☐ Perceptions of work climate	x	x	x		x	x			x	x	x	x	x																	x		x	x	x					x			x			x	x			
☐ Curriculum coverage	x	x	x	x		x			x	x	x	x	x					x												x	x			x	x	x			x			x			x	x			
☐ Use of new information technologies		x			x					x		x	x																					x	x	x						x							
☐ Other major categories of teacher variables:															x	x																				x	x		x			x							
☐ **School**																																																	
☐ Organisation/management	x	x	x		x	x	x		x	x	x	x	x					x				x					x			x		x	x	x			x		x		x				x	x	x		
☐ Principal leadership	x	x	x	x			x		x	x	x	x	x																	x		x	x	x					x		x								
☐ Support	x	x				x			x	x	x	x	x					x			x									x		x	x	x											x	x			
☐ School climate	x		x			x			x	x	x	x	x						x											x		x	x												x	x			
☐ Material conditions	x	x	x	x	x	x			x	x	x	x	x				x	x	x			x								x		x	x	x	x						x				x	x			
☐ Finances			x	x		x												x				x																			x								
☐ Self evaluation practices		x					x	x										x																x															
☐ Absenteeism							x	x																																	x								
☐ Other major categories of school variables:																									x														x										

Pilot Projects

From the experience of the network members in conducting international comparative assessments, it appeared that in many societies an increased need is felt for objective empirical data on educational progress. However, given the relatively short history of educational assessments, researchers are still continuously searching for improvements of measures, which can serve as stable indicators of educational processes and outcomes at the macro-level. Moreover, it was also felt that a better insight is needed in the extent to which particular indicators can explain the malfunctioning of (parts of) educational systems. Therefore, it was decided to start –via pilot projects- activities to explore how to share the knowledge base on the quality and explanatory power of indicators collected by the member institutes. In order to realise this goal, the coordinating center devised a task list from which members could choose which contributions they might be able to deliver on the basis of their assessment experiences.

The results from the pilot projects showed that across the network there was a rich variety of experiences with many different educational process and outcome indicators. However, it should also be noted that in quite some cases, essential information about the quality of indicators was lacking. Especially information about validity (is the indicator reflecting what was intended to be measured?) was not available. This observation should not be interpreted as criticism, but rather be understood as reflecting the difficulty (and sometimes impossibility) of running (time-consuming) validity studies, given the constraints of budget and timelines which usually are imposed on national educational assessment projects. Such constraints seem to be occurring in many of the projects. Although such observation may lead to pessimistic thoughts about the possibility to systematically improve the quality of national assessments, it may also be taken as a signal that cooperative sharing of work throughout Europe may be beneficial for all parties involved.

One may argue that such efforts would not be profitable for measures, which are culturally sensitive. On the other hand, it may also be said that by systematically investigating the quality of indicators (reliability, validity, as well as explanatory power) would considerably improve the knowledge base of what are promising generic indicators and which are (culturally) specific.

Another observation from the explanatory analyses is that in many cases all variance is subject of explanation. However, it may be argued that assessments are part of a cycle of steps aimed at improving the education system or subsystems. Assessment results in displaying variance in the measures used. Variance will always exist and are in itself no reason for worry. What counts is if students are achieving below expectation. For instance, people who do not speak/understand the language of instruction hardly can be expected (ceteris paribus) to function at the same level as native speakers.

The overall conclusion from the pilot projects is that, throughout Europe, researchers are trying to find appropriate instruments for measuring indicators of factors such as the ones shown in the network's conceptual framework (see previous section). An important component of this search is the identification of indicators which can explain why unexpected low achievement is occurring in (parts of) the population of students under study.

Realising such task on the basis of single projects is (in the absence of huge investments) hardly possible. Therefore, it may be argued that an increased European effort of sharing the developmental and research tasks may be beneficial for all parties involved. Such Utopia may only be reachable if a number of conditions are satisfied, such as:

- Rigorous standardisation of reporting indices of indicators (and underlying operationalisations), which need to be easily accessible via instrument banks (including indices of reliability, validity, and explanatory power).
- Systematic replication in many contexts of potential promising indicators. Results should be fed back to the bank.
- Relaxation on the side of researchers (involved in educational assessments) with regard to the goal of trying to be original. Such attitude can be rather detrimental (because of its lack of orientation in replication and/or improvement of existing measures) for the future of educational assessments.
- Middle and long term plan for European cooperation on assessing educational progress and investigating educational effectiveness and its underlying explanatory factors.

Evaluation of the Achievements of the Network

The work performed in the past two years can be conceived as the construction of the fundaments for an information structure, which builds upon the constantly growing knowledge base of European researchers with regard to the quality and functioning of educational systems in the EU. Further work is needed in order to facilitate access to up-to-date and high quality methodologies for conducting educational assessments as well as for analysing strengths and weaknesses of educational systems. Moreover the current interface, which exists in the form of the AEI Statistical Overview Web site (see above), between researchers and interested audiences needs further elaboration in order to increase its functionality and user-friendliness.

Reflections on Policy Implications

The network's main mission in the period 1996-1998 was primarily focussed on improving the infrastructure for comparative assessments in a European context. As such this is not a substantive activity addressing policy issues. Nevertheless, from the discussions and exchanges of information, which took place in the past two years, a few observations can be made which have implications for education policies at the EU-level. From the meeting documents that were summarised in Chapter 2, it appears that the network members agree that at the EU policy level, there is not yet a strong movement towards strengthening the European research activities in order to collect on a regular basis standardised comparative data on student outcomes (and instructional processes) that allow countries to evaluate the development of their educational systems over time. On the other hand, it can be observed that there is an increased interest in comparative data (for

instance in the Socrates call for proposals from Spring 1998) and the support for pilot projects for school self evaluation. As such, these activities suggest that an increased need and awareness of the necessity to develop tools that allow for rational judgements of strengths and weaknesses of educational systems (at different levels: micro, meso and macro) is developing. However, from the point of view of the EU-AEI network, the most important basis for further stimulation of the development of a strong comparative assessment research work force would be the political intention and willingness to start an intensive cooperation between countries in order to share the burden of creating high quality systems for comparative assessment and evaluation of educational progress within a European context. In the words of the European Round Table of Industrialists (1995): *It has become a matter of urgency to set up a European Education Information and Monitoring System in order to record experience, evaluate results and improve the quality of education on a European level*. Until such system is realised one will need to rely too much on speculation for making decisions on the way ahead, or in the words of Black (1995), it will be difficult "*to avoid ideology and to identify and tackle real problems*".

Maybe it is a matter of time that such statement ultimately will also be expressed at the EU political level. Once this willingness exists, European institutes (such as Eurostat, Eurydice) can start to develop an infrastructure for systematic evaluation of Europe's educational progress, with the help of a workforce of researchers specialised in international comparative assessment and with the guidance of the European network for policy-makers for the evaluation of educational systems.