HOW CAN LARGE INTERNATIONAL COMPARATIVE STUDIES CONTRIBUTE TO THE QUALITY OF SCIENCE EDUCATION?

SVEIN LIE

University of Oslo, Norway

ABSTRACT

In this paper the two international comparative studies IEA TIMSS and OECD PISA have been discussed by comparing their similarities and differences. A number of examples have been presented to demonstrate how findings in various areas are relevant to help improve science education. Focus are on students' conceptual understanding, gender and school differences, relations to home background factors, and on what characteristics of instruction that seem to be related to high achievement. Furthermore, the assessment frameworks for the two studies are argued to be of influential importance in its own terms, but that any influence on national aims and curricula should be carefully considered only in a national context.

1. TWO LARGE-SCALE COMPARATIVE STUDIES

This contribution focuses on the large comparative studies, i.e. the IEA TIMSS (Trends In Mathematics and Science Study) and the OECD PISA (Programme for International Student Assessment), and discusses some of their features that can contribute to increased quality of science education around the world. More and more countries are taking part in these international studies, as many as around 50 in the 2003 versions of these two studies. One of the main goals for these studies is to present reliable comparisons between countries in the simple form of league tables. Such information can make good headings in the media, but do not in itself represent any educational improvement. However, even such simple ranking measures can inform on national strengths and weaknesses, thus help defining problematic areas that could be the focus for increased attention or further investigations. And more important, in the process of planning structural, pedagogical or content reforms in science education there is rich information in the data from these studies to learn from other countries. The aim of the present paper is to point to some important areas in this respect.

There have been many critics of international assessment studies from researchers in science education in the last years (e.g. Orpwood, 2000; Jenkins, 2000). However, such sound critiques mainly concern the restricted emphasis that is put on issues central for science education. They do not challenge the technical

K. Boersma et al. (eds.), Research and the Quality of Science Education, 27—40.

quality or the validity of the insight that can be accomplished provided the data are carefully interpreted in national contexts. One of the aims of the present paper is to argue for more use of the rich databases for the two studies. Data for earlier studies have been released for public use, and databases for TIMSS 2003 and PISA 2003 will be publicly available shortly. By referring to and giving some examples of various types of results, I have a hope that more researchers in science education will put interest and effort into exploring the rich source of data from their own research perspectives. There are lots of relevant data from all around the world waiting to be further explored: on responses to individual achievement items, on student attitudes towards science and their motivation and further education plans, on teaching styles and learning strategies, on pedagogical climate in schools and classrooms, on teachers' education and pedagogical beliefs, and much more.

The two studies at hand have much in common, but they also differ in certain important respects. Both studies have a science component and assess scientific knowledge and skills according to what may be called a two-dimensional content-by-behaviour grid. That is, one dimension is specifying the content domains and subdomains, while the other dimension specifies the types of competencies that are being measured for each domain. Furthermore, the two studies have similar main research questions: Countries and national groups of students are compared by their mean and spread in proficiency along the cognitive scales, and these scale scores are further related to data from student and school questionnaires. Thus a focus is on the relationships between cognitive outcomes and various affective, home background and contextual student, classroom and school variables. The studies are repeated every third (PISA) or fourth (TIMSS) year, and since many elements in the questionnaires and assessment instruments are kept constant, important trend indicators are being measured.

The most profound difference between the two studies concerns how the scientific competencies are defined. In principle, TIMSS is measuring the "achieved" curriculum, that is, how well students have learned the core material common to the curriculum in the majority of participating countries. The framework (Mullis et al, 2001 for TIMSS 2003) describes this "material" and also makes clear how it is based on information on what is expected to be covered by instruction during the years up to and including the actual grade of assessment. PISA defines a very different point of departure in its framework (OECD 2000) by its focus on measuring students' "preparedness for life", or how well students are prepared to meet challenges of future knowledge societies. Obviously, it is no simple task to make predictions concerning what will be important in the future, and further, the response to this question also depends on what is meant by "important" and in what perspective importance is judged. Whereas the present national curricula are regarded as irrelevant as descriptors of such criteria, the actual principles for competencies to be measured are based on "what is regarded as important in a perspective of life-long learning" (ibid), according to a consensus among the OECD countries at the political level.

TIMSS can be said to be "curriculum driven", in the sense that the assessment framework is based on curricular considerations. PISA, on the other hand, may be characterized as "utility -" or "relevance driven" due to the focus on