AKTUELLA ANALYSER 2009

With a focus on Mathematics and Science

An analysis of the differences and similarities between international comparative studies and national syllabuses

Skolverket

With a focus on Mathematics and Science

An analysis of the differences and similarities between international comparative studies and national syllabuses

www.skolverket.se
ISBN: 978-91-85545-61-2
Form: Ordförrådet AB

Preface

The Swedish National Agency for Education (Skolverket) has, in a previous summary (*Internationella studier under 40 år* ['International studies over a period of 40 years', trans.]), described Sweden's participation in international comparative studies of knowledge. From this it is evident that the scope of Sweden's participation has expanded significantly during the past decade. During this time, the aim of these studies has shifted from describing and comparing the results of different countries – and possibly also from explaining and understanding results in light of background factors on different levels – to focusing increasingly on monitoring change in knowledge over time, i.e. trend measurements. Interest in trend studies has also increased nationally in recent years. The major national evaluation which was done in 2003 (NU-03) can be viewed as an example of this.

It is costly, however, to perform assessment studies and in many cases a complex methodology and plan is required in order to secure good quality. For this reason it is advantageous to be able to share the cost of method development and analysis with many participating countries. At the same time there is, of course, a risk that internationally developed instruments are not applicable to the Swedish situation.

Is it possible to use the international studies to measure trends from a national perspective? How well do the frameworks for the international studies agree? Do we risk losing our distinctive national character? Are our pupils' results judged on the basis of misleading principles? Do our national objectives differ from the discernible objectives of the international studies?

In 2003, three different assessments of Swedish compulsory school pupils' knowledge in mathematics and science subjects were performed. In the present study these three assessment studies have been analysed: Programme for International Student Assessment (PISA 2003), Trends in International mathematics and Science Study (TIMSS 2003) and the National Assessment of the Compulsory School (NU -03).

The purpose of this study is to illustrate the extent to which the frameworks for the international comparative studies TIMSS and PISA and the Swedish syllabuses are in agreement as regards subject content and type of knowledge examined in mathematics, biology, physics and chemistry. The results reported are intended to support the Swedish National Agency for Education's assessment of how the evaluation of knowledge is to be formulated.

The report, which has been developed within the framework for the project *Analys och syntes* ['Analysis and synthesis', trans.], was written by Jan-Olof Lindström, former project manager of TIMSS 2003. Niklas Eriksson has been involved in the data processing. Views have been expressed by Kristian Ramstedt and members of the project *Analys och Syntes* under the direction of Anita Wester.

Stockholm, 18 December 2006

Ann Carlson Ericsson	*Anita Wester*
Department Manager	Project Manager

Contents

1 Summary

In 2003, three different assessments of Swedish compulsory school pupils' knowledge in mathematics and science subjects were performed. The National Assessment of the Compulsory School 2003 (NU-03) and the Programme for International Student Assessment (PISA 2003) dealt mainly with students in the ninth year of school, while Trends in International Mathematics and Science Study (TIMSS 2003) dealt with the eighth year. The purpose of this study is to illustrate the extent to which the frameworks for the international comparative studies TIMSS and PISA and the Swedish syllabuses are in agreement as regards subject content and type of knowledge examined in mathematics, biology, physics and chemistry. The three assessments, in addition to assessing knowledge, also included the collection of data on general background factors and pupil attitudes. These aspects of the assessment are not referred to in this report, nor are aspects of the knowledge tested in respect of the other subjects included in NU-03 and PISA.

The validity concept[1] plays a central role when the quality of examinations or tests in general is discussed. The dimension of the concept pertinent in this report is the question of the conclusions that can be drawn from an individual examinee or group of pupils achieving a certain result that is expressed, for instance, in the form of a score. In a system managed by goals, it must be possible to use an assessment test to determine the extent to which the goals have been achieved. It can therefore seem important to examine whether the instruments used do test knowledge in a way that corresponds with the knowledge goals specified in the curricula and syllabuses. A discussion on validity must also refer to how well the test questions cover the domains of knowledge described in the governing documents, how the assessment is performed, how statistically certain the supporting documentation is, etc.

This report gives a clear description of the preconditions for the three knowledge assessments. All three studies were performed in randomly selected pupil groups of approximately the same size. PISA was performed with the participation of 41 countries of which 30 are members of the OECD. TIMSS had 45 participant countries but, in the Swedish report, a group of 20 countries with similar circumstances was chosen for comparison purposes. In NU, only Swedish data were collected but, as in TIMSS, there was also the possibility of studying the development of knowledge over the past decade. While in NU, only one set of test questions was given to all pupils in a certain sample of the population, in the other two assessments, question pools were used that had a lot more questions than a pupil could cope with in a reasonable time. The time allowed for the tests in the subjects in question was the same for the individual pupil in all of the assessments but, by using different versions of tests, data could be obtained for more questions in PISA and TIMSS than in NU. The reporting in NU only involved the use of simple statistical analysis, while the two international studies used complex methods for ensuring results that could be generalised.

The curricula or frameworks behind PISA and TIMSS are initially compared with the Swedish governing documents in mathematics and the subjects that are

[1] Messick (1994)

relevant for Sweden, i.e. biology, physics and chemistry. As regards mathematics, the compatibility is greatest between the Swedish syllabus and TIMSS, in both cases with goal descriptions that are based on a traditional division of subject material. As regards science, it can be said that the greatest similarity can be found between the framework in PISA and the Swedish syllabuses by emphasising processes in scientific literacy in both documents.

A system for classifying test questions has been drawn up on the basis of the frameworks studied. To describe subject content, the TIMSS system for categorisation of the material in *Mathematics* and *Science* is used. The nomenclature for the type of questions corresponds in general with that used in PISA. Otherwise, categories have been constructed on the basis of expected relevancy in the analysis of the test material, the starting point being the Swedish governing documents. The test material in NU-03, PISA 2003 and TIMSS 2003 respectively has been analysed in respect of subject content, type of question, degree of authenticity, use of calculations, amount of text and graphics, the difficulty of the questions, and the competencies that the question is intended to indicate. The reliability of the classification has been checked in a meeting between a panel of practising teachers with long experience of teaching at the school levels in question.

The results of the classification of the test questions in the respective assessment study have been contrasted with the teaching goals relevant for Sweden. With regard to both mathematics and science, TIMSS has the broadest coverage of subject matter. That test material thus has the best chances of measuring whether the knowledge in the subjects in question that is aimed at in accordance with the syllabuses actually has been achieved. In a cognitive respect, PISA and TIMSS do not differ as regards mathematics. Both leant comparatively more towards complex forms of knowledge than the analysed short-response test in NU. In the science syllabuses, the goals have been structured differently than for mathematics. The goals set up with regard to knowledge of concepts and phenomena are most comprehensively covered by TIMSS as mentioned above, while PISA aimed more at covering the goals regarding the scientific activities, as was the case regarding the use of scientific knowledge to some extent.

From a technical angle, the most obvious difference between the three knowledge assessments was in respect of text volume. The PISA questions in both mathematics and science are presented using relatively large volumes of text. Another such notable difference between the test material in science was the degree of difficulty of the questions. In NU, the biology, physics and chemistry questions are in general significantly more difficult than in the other two assessments.

The discussion refers, for instance, to the difference, supported by data and already made clear in the respective frameworks, in the goals for the two international, comparative assessments. TIMSS and PISA complement each other and together provide an assessment that corresponds to the validity requirements. The national assessment does not convey anything decisive as regards the assessment of knowledge in the form that NU-03 represents.

In connection with the discussion as to what the tests measure, the hypothesis that the order of ranking between countries can change when a selection of subject areas is made is also tested. Finally, reference is also made to methods for optimising efforts regarding the assessment of knowledge and attitudes of Swedish compulsory school pupils.

2 Purpose and outline

In this report, by way of introduction, a comparison is made of the preconditions for measuring compulsory school pupils' knowledge in mathematics and science in three assessments of the Swedish compulsory school. The National Assessment of the Compulsory School 2003 (NU-03), the Programme for International Student Assessment (PISA 2003) and Trends in International Mathematics and Science Study (TIMSS 2003) have, in part, different purposes as well as different conditions as regards carrying out the mesurement of knowledge and analysis of results.

The requirement of having to be able to say what a certain test measures can appear natural, since this would result in being able to state what conclusions could be drawn from a group of pupils having performed the test in a certain way. A certain average score achieved, for instance, apart from placing the group in a ranked list also ought to form the basis for an interpretation that certain competencies[2] described in a governing document have been achieved by the members of the group. In practice, however, this is a requirement that can never be completely fulfilled in an unambiguous way. The test questions used cannot cover anything other than a selection of the varied aspects of the knowledge goals aimed at and that can be expressed with different degrees of detail and concretion. The Swedish national governing documents, Lpo 94 with its accompanying syllabuses, are characterised by general wordings that has to be interpreted and objectified in the schools themselves by the school heads and teachers responsible. Even with national tests, problems thus arise in the interpretation of results due to the local variations in choice of material and methods referred to in the Swedish curricular system. The difficulty in designing test instruments that meet strict validity requirements is accentuated in broadly drawn up international comparative studies in which the test material has been constructed in a mainly consensus-focused process with representatives from many countries participating. The discussions on validity in such assessments is particularly important when the results, with accompanying comparisons with other countries, may often form the basis for comprehensive policy decisions on the national education systems.

In this study, the purpose has been to analyse the different assessments with regard to coverage of subject content, competencies tested etc, against a background of such goals as apply under the compulsory school curriculum and in the syllabuses for the subjects in question. Together with the presented comparison of the methods used in the three assessments, the result could possibly constitute a base on which to plan school assessments in the future.

After descriptions of background and performance for each one of the three assessments comes a short account of the Swedish syllabuses as well as explanations of the frameworks of the international comparative studies PISA and TIMSS. Thereafter, the method used for the study is presented. This includes a presentation of the taxonomy used for classification of the test instruments in NU, PISA and TIMSS. The result of the analysis of the questions is presented

[2] Niss, M. (1999)

and summarised. The report ends with a discussion and conclusions in accordance with the purpose of the study.

3 Background and general preconditions for the three assessments

The Swedish National Agency for Education is commissioned by the government to follow up and assess the quality and equity in the school system. The Agency is also commissioned to present proposals for measures to maintain and reinforce the fulfilment of the national goals. It is therefore important to obtain reference points with regard to the school's goal fulfilment, partly from an internationally comparative perspective and partly from a time-related trend perspective. Examples of measures for this purpose are the Agency's execution of the national assessment of 1992, which was followed up by NU-03. These national surveys were designed internally within the Agency with the participation of research groups from different universities. In the international assessments PISA and TIMSS, the Agency has joined external projects that have been carryed out by research centres on behalf of the Organisation for Economic Cooperation and Development (OECD) and the International Association for the Evaluation of Educational Achievement (IEA) respectively. These organisations are responsible for the development of the comparative assessments, which can be repeated at certain intervals. In these studies the Agency has the overall national responsibility for planning and analysis but has commissioned different university institutions in the country to assist with the preparations, field work and reporting.

It is natural in the context that regulations regarding the carrying out of the studies, with instructions, controls and documentation of different procedures, must be better prepared in a study that is to function for the different players when different countries' school systems are to be assessed. As regards PISA and TIMSS, there are comprehensive descriptions[3] of the frameworks that have formed the basis for both the preparation of test material and questionnaires and for the administration while, in respect of NU-03, there are no corresponding, special documents with specifications that have governed the survey. One must assume here that the test material is constructed on the basis of the curriculum and the related syllabuses, with thorough, practical knowledge of the frameworks within which these documents are interpreted in the Swedish school.

Selection of schools and performance of the knowledge assessments

The three assessments were performed in randomly selected schools in the spring term of 2003, with the participation of pupils in year 8 of compulsory school for TIMSS and year 9 for NU as also, in the main, for PISA. The basic principle for the random selection was that the probability of a pupil being selected to participate in a certain assessment should be equal for all pupils in the

[3] See e.g. OECD (2003) and Mullis et al. (2003)

age groups in question. For practical reasons, whole schools were selected in a first step in which the probability of a school being selected should be in proportion to the number of pupils in the classes in question at the school. So that no school would be unnecessarily burdened, the Agency decided to carry out the selection of schools in such a way as to avoid the assignment of more than one assessment to the same school. In TIMSS, as in NU, whole teaching groups (classes) were chosen at random, while in PISA, the selection unit was comprised of individual pupils. In the two international studies, the national centres leading the surveys checked that the statistical selection of schools, teachers and pupils was carried out in accordance with established norms.

In all three studies, mainly written tests were used for the knowledge assessment. In the two international studies, there were rigorous descriptions of the procedure to be used when carrying out the tests. In PISA, the tests were conducted by test administrators who were centrally employed in the project, while the test administrators in TIMSS were trained by the national centre but appointed by the respective school administration. In NU, the class teachers acted as administrators of the test. Marking of the pupils' responses was done by groups of specially trained assessors at the different university institutions that were responsible for the performance of the three projects.

The reporting of results in both PISA and TIMSS is based on statistical methods whereby each pupil's expected test result for every question is decided and whereby each pupil's data are weighted in relation to the proportion of the total population that the pupil represents when, for instance, an average achievement is decided for the pupils in a country. The method means, in spite of only a minor proportion of the questions being the same as in a previous assessment, that the scale for the reporting is the same. This allows analyses of trends in the pupils' performance.

The data analysis for NU can seem simpler as everyone who sits the test in a certain subject gets the same test and as it is more or less the same test as in the previous assessment. The final chapter in the report discusses the problem of using a joint test on repeated occasions.

The National Assessment of the Compulsory School 2003 (NU-03)

The purpose of the National Assessment of the Compulsory School 2003 was to give a broad view of the work and, in particular, of the pupils' knowledge in fifteen subjects. A corresponding survey was carried out in 1992. The main purpose of NU-03, as a basis for national decisions relating to the compulsory school, was to:

- give a comprehensive picture of goal achievement in the compulsory school, per subject as well as from an overall perspective
- demonstrate changes since the national assessment of 1992
- point out the need for efforts to be put in.

Mathematics

When performing NU-03, the intention was to use the result from the national subject test for school year 9, spring 2003, partly to assess goal achievement and partly to assess the pupils' knowledge development in mathematics since 1992.

This test, to a certain extent, was given content that was comparable with the test used in 1992. As the national test, due to an irregularity, was exposed prior to the official test occasion, data from this test could not be used in the assessment, and the test has not been accounted for in detail in the reporting from NU-03. The test that could be used instead for the analysis of the knowledge development was one with 21 short-response questions and that was completely comparable with the test used in 1992. For this reason, only that test is accounted for in this report. Each question was assessed with one point for an acceptable response and zero points for an incorrect or omitted response.

When the test was analysed by a group of practising teachers, they laid down a limit of at least eight points for the pupils to be deemed to have attained the ”necessary” level of knowledge according to the 2000 syllabus. The terminology is taken from Lgr 80[4] and can be presumed to correspond to the requirements for the Pass grade in accordance with the goals that the pupil should have attained at the end of year 9. It is, however, pointed out in the report[5] that the test was drawn up on the basis of Lgr 80 and that it only tests a limited part of the present syllabus.

Table 3.1 Result of the 21-point test in mathematics used in 1992 and 2003

Test result		1992	2003
0–7 points	%*	13.2	16.7
8–14 points	%*	38.2	47.2
15–21 points	%*	48.7	35.7
Average	points	13.6	12.4
Number of pupils in the sample		9 873	4 420
Drop-out	%	5	37

* Proportion of the pupils who were awarded results in the specified points interval

The average achievement in the 1992 survey was 13.6 points, and 13.2% did not reach the relevant limit for necessary knowledge. In NU-03 the average achievement is reported[6] as being 12.4 points, with 16.7% of the pupils achieving under the determined limit. See table 3.1. Via data from the national tests in mathematics in year 9 from 2001–2003, it is estimated that between 80 and 90 per cent of the pupils achieve each of the syllabus's 'goals to attain'. The approximate agreement between the two different assessments of goal achievement in the pupil group could be taken as an indication that the point limit used for a Pass grade or higher in the 21-point test was reasonable.

Science studies

In NU-03, for the science subjects, a test was used in each of the content domains, biology, physics and chemistry, in order to assess goal achievement and also to assess the pupils' knowledge development in science studies since 1992. The tests were, for the most part, the same as those used in 1992. The tests consisted of 12–13 questions. The results of the tests in each of the science subjects were reported by stating the proportion of the responses given to the individual questions that was deemed by the examiner to indicate the pupil's level of

[4] The previous National Curriculum for the Compulsary School 1984
[5] Kjellström (2003)
[6] Skolverket (2004a)

knowledge to be a Pass (G) and, in appropriate cases Pass with distinction (VG), in respect of a certain specified goal in the syllabus.

Approximately two thousand pupil responses were collected for each of the questions in the three tests. More than half of the pupils had answered the questions electronically. The absence drop-out in biology was 11 per cent, in physics 14 per cent and in chemistry 18 per cent. Only about 30 per cent of the 'open' pupil responses were assessed and categorised.

Table 3.2 Result of the NU-tests that were used in 1992 and 2003 in science

Subject		1992	2003
Biology	%*	28	25
Physics	%*	36	29
Chemistry	%*	46	36

* Average proportion of pupil responses to individual questions that were awarded Pass or Pass with distinction.

The results are presented in summarised form in table 3.2. Apart from averages, no further statistical measures are given in the report for the science subjects in NU-03.[7] The results in biology demonstrate a certain deterioration from 28 to 25 percentage points, i.e. a relative decline of approximately 10 per cent. The table also indicates that the deterioration in results between the two occasions of assessment is 19 per cent in physics while it is 21 per cent in chemistry, measured as an average of the proportion of pupil responses that were deemed to correspond to a Pass grade or higher. A reasonable interpretation of data in statistical terms is that the decline in test results ought to be deemed significant in the three subjects and be least in biology, while it can be twice as large in both physics and chemistry.

PISA 2003

The Programme for International Student Assessment (PISA) is an internationally standardised assessment of the compulsory school that has been developed through cooperation between the member countries of the Organisation for Economic Cooperation and Development (OECD). The tests included are intended to assess 15-year olds' reading, mathematical and scientific literacy. In the 2003 assessment, problem solving was also included as a separate subject. The choice of the concept of literacy that is used here can be understood on the basis of an interpretation of what is said initially in the appurtenent framework.[8] It is emphasised there that the assessment should be directed at the ability to use knowledge and skill in reading, mathematics and science in a way that corresponds to needs and requirements of today's society rather than the ability to demonstrate knowledge of specific facts or procedures. This means that the participating countries' learning objectives in the form of specifications of subject content do not necessarily have to be covered in the framework for the tests used. Instead, the starting point will be the general competencies of applying the subject's specific concepts, procedures etc. (in order to analyse, discuss, communicate etc.) that are deemed by the participating OECD countries' experts to

7 Andersson et al. (2005)

8 OECD (1999) p. 9.

be important for the young people about to leave compulsory school and who will make up the base from which the workforce of the future will be recruited.

Results from 40[9] countries were reported[10] of which 29 were OECD members that had performed the study acceptably. In Sweden, 186 out of 188 selected schools took part, with 4970 randomly selected pupils. The drop-out rate was 7.0 per cent.

Mathematics

The test material in this subject consisted of 86 questions. Of these, 20 were the same as in the PISA 2000 assessment, which was not as comprehensive with regard to mathematics. Changes in the level of knowledge in mathematics will be possible to study more definitely after PISA 2012, when focus will again be on this subject.

PISA 2003 focused, in the first hand, on mathematics. The average achievement for the Swedish pupils was, in mathematics, 509 points, which was significantly better than the total average of 500 points for the OECD countries. The Czech Republic, Iceland, Denmark, France, Austria, Germany, Ireland and the Slovak Republic achieved results that had no statistically secured deviation from the Swedish results. Finland, the Netherlands, Belgium and Switzerland were some of the countries that had better results, while Norway, Hungary and the USA were examples of countries that had significantly poorer results.

The content in mathematics is divided into four themes of which two were included in the first PISA assessment. Sweden's averages within these two themes in 2003 did not differ significantly from the results in 2000.

Science studies

In PISA 2006, a broader assessment of science studies, as per the framework for PISA, will be made. The number of test questions in science was 35 in total, of which 25, as in mathematics, could be used to link the science result in PISA 2000 with the result in 2003.

Sweden's total result in PISA 2003 was 506 scale points, which did not show any statistically secured deviation from the results for Switzerland, France, Belgium, Ireland, Hungary, Germany, Poland and the Slovak Republic. Finland, the Netherlands and the Czech Republic belonged to the group of countries with markedly better results while, for instance, Iceland, the USA, Norway and Denmark had significantly poorer results than Sweden. In 2000, Sweden's result was 512 points, which was not significantly better than the result in 2003. However, it could be noted that the standard deviation increased significantly, from 93 in the earlier assessment to 107 scale points in the later one.

According to the plans for PISA, the assessment should be performed every third year, with focus on one of the knowledge domains – reading, mathematics and science. This means that more general statements regarding knowledge development within a certain subject area can be made every ninth year.

---

[9] The United Kingdom was excluded from the reported results due to too great a drop-out.

[10] The National Agency for Education (2005a)

TIMSS 2003

Trends in International Mathematics and Science Study is an international assessment study with a history going back to the end of the 1950s. The International Association for the Evaluation of Educational Achievement (IEA) is responsible for the study, which is the fourth in which Swedish pupils' knowledge and attitudes in mathematics and science are compared with an equivalent picture in other countries and with what has come out in previous assessments. The IEA describes itself as an independent cooperative of national pedagogical research institutions and nationally responsible education agencies. The initiative for the organisation has been taken by researchers, and the starting point for the knowledge assessment is the governing documents of the participating countries. On the basis of a jointly agreed compilation of these, instruments should be developed to provide data regarding the degree to which the national goals are being achieved. Further, the results will provide a basis for analysis of possible trends and comparisons with other countries which, in turn, can affect the development of the national school systems.

In TIMSS 2003, 50 countries or regions participated in that survey, which mainly applied to pupils in their eighth year of compulsory school.[11] In the reports[12] that describe the results in Sweden, a group of 20 countries was selected to constitute the basis for international comparisons. The results from TIMSS 1995 were used to make comparisons regarding knowledge development in Sweden from the national and international perspective.

In Sweden, 159 of the total of 164 invited schools participated in a way that fulfilled the requirements of access to pupils in the correct school year, drop-out etc. The absence drop-out was a total of 11 per cent of the total number of 4790 selected pupils.

Mathematics

The test material in TIMSS 2003 covered 194 questions[13] in mathematics of which 20 were also used in TIMSS 1995.

For a comparable group of nations or regions (16 in all) that participated in both TIMSS 1995 and 2003, the average in mathematics in each respective assessment was 528 and 517 scale points, a decrease of 11 points in eight years. Sweden, which in the earlier assessment had significantly better results than the average of the group in question, dropped from 540 to 499 points in the later assessment. The decrease was, in other words, 41 scale points, and the result for Sweden was now markedly poorer than that of the reference group.

Science studies

The number of test questions[14] in science totalled 194, of which 25 were anchor questions for linking the science results in TIMSS 1995 with the results from 2003.

Sweden's total result in TIMSS 2003 was 524 scale points, which did not show any statistically secured deviation from the results for the USA, Australia,

---

[11] Sweden did not participate in a corresponding survey of pupils in year 4 that was carried out in 26 countries.

[12] The Swedish National Agency for Education (2004b, 2005b, c)

[13] Calculated from the number of answers to be assessed.

[14] Calculated from the number of responses to be assessed.

Slovenia, New Zealand, Lithuania and the Slovak Republic. Estonia, England, Hungary and the Netherlands belonged to the group of countries with markedly better results while, for instance, Belgium, Russia, Latvia, Scotland, Norway and Italy had significantly poorer results than Sweden. In 1995, Sweden's result was 553 points, which was markedly better than the result for 2003.

Curricula

Lpo 94 and the Syllabuses 2000

The pupils who participated in NU-03 and the majority of the pupils in PISA started school in 1994 and are thus the first age group who, during their entire schooling, have had the curriculum called Lpo 94.[15] Pupils in year 8 who participated in TIMSS have all attended school during the period when Lpo94 applied. The document states overall goals and guidelines for compulsory school education. Syllabuses[16] (and time plans) can be found, in contrast to the curriculum that applied previously, Lgr 80 in separate ordinances. The syllabuses were revised in 2000.

The subject matter in the syllabuses is not specified over and above general listing of the subject areas and broad descriptions of basic competencies. In the "goals to aim for" in the curriculum and syllabuses, the orientation of the teaching is indicated by expressing the knowledge qualities that should be aimed for within selected subject areas and, in "goals to attain", the minimum level of knowledge in the subject that all pupils should attain in year 5 and in year 9 is specified. It is presumed that an interpretation and concretisation of the goals takes place at every school, where more detailed, local work plans should be drawn up. This applies both to the selection of subject matter such as in which school year a certain subject area is taken up and to the importance it is awarded in the teaching.

The management by objectives model that has been chosen in Sweden can, even in a nationally limited study, seem to restrict the chances for centrally developing a high-validity knowledge assessment for evaluation of the school system. This is particularly the case in subjects such as mathematics and science, where factual knowledge of specific concepts and procedures is usually a precondition for being able to demonstrate more general competencies in conventional tests. The question is, in other words, what material can be said to be included in a normal course and that can thereby be presumed to be familiar in a test that examines factual knowledge and also the ability to understand concepts, use logical arguments etc. The way out of this dilemma is to trust that, in a small country like Sweden, there is consensus as to how the material should be limited. The author's interpretation of the syllabuses, as reflected in the generally used teaching materials, gives a probable standard value as regards what can be deemed to be included in the subject, the weight given to a subject area and at what point during the teaching it is taken up.

The most important perspective for the comparison that follows here is how NU, PISA and TIMSS, as regards subject content and competencies, correspond to what is described in the national goals for the teaching. As a basis

---

15 The National Agnecy for Education (2000). Curriculum for the compulsory school, the pre-school class and the after school centre, Lpo94

16 Skolverket (2000): Syllabuses for the Compulsory School http://www3.skolverket.se/ki03/front.aspx?sprak=EN&ar=0809&infotyp=15&skolform=11&id=2087&extraId=

for the analysis to be presented here, the content of the syllabuses in question has been organised in such a way that goals to aim for and goals to attain have been matched. The order of the syllabus's listing of the goals to attain has been retained and has governed the order in which the goals to aim for have been inserted in a table for each of the subjects.

Mathematics

The mathematics syllabus[17] begins with wordings that emphasize the important role of mathematics for pupils in future being able to find use for the flow of information and being able to follow and participate in society's decision-making processes. The two first goals to aim for that are listed are of an affective character and refer to one's attitude to the subject. The immediately following cognitive goals, which can be examined with the help of conventional knowledge tests, say that the pupil should develop different competencies. These goal formulations contain the signal words reasoning, problem solving, and modelling, but also wordings regarding the skills of communicating mathematics and using technical aids. These are followed by goals that refer to specific, main areas within the subject matter. It starts with the words "The aim should also be that the pupil develops their number sense and spatial sense and their ability to understand and use ...". Each one of these goals to aim for corresponds to a goal to attain within different subject areas for which conventional names have been introduced here: Number sense, Arithmetic, Measurements, Geometry, Statistics, Probability, Algebra and Functions. The designations M1, M2 etc. are introduced and will be used here from now on.

Science studies

The presentation of the goals in the syllabuses for the science subjects[18] takes on a different pattern than was the case for mathematics. To begin with, the picture is complicated by the three subjects biology (B), physics (P) and chemistry (C), each with its own syllabus, being included in the block subject Science Studies (S), which also has its own syllabus. As is apparent from the tables, this means in a few cases that, included among goal formulations for a specific subject, there are also goals to attain that have been taken from the syllabus for the integrated subject of science studies.

Another peculiarity is that the goals are divided into three headed categories, namely knowledge "concerning nature and Man", knowledge "concerning scientific activity" and knowledge "concerning use of knowledge". In the designations that have been introduced here, the first digit 1, 2 or 3 refers to these goal categories. Further, the division of the three science subjects into content areas is not as clear as in mathematics, which makes the matching between goals to aim for and goals to attain more diffuse. Every goal to aim for usually corresponds to several goals to attain, but it can also be the case that there is no goal to aim for that can be deemed to match a certain goal to attain. The reverse can also apply. With regard to the numbering that has been used here, the second digit refers to a goal to aim for, and the third digit refers to the goal to attain.

---

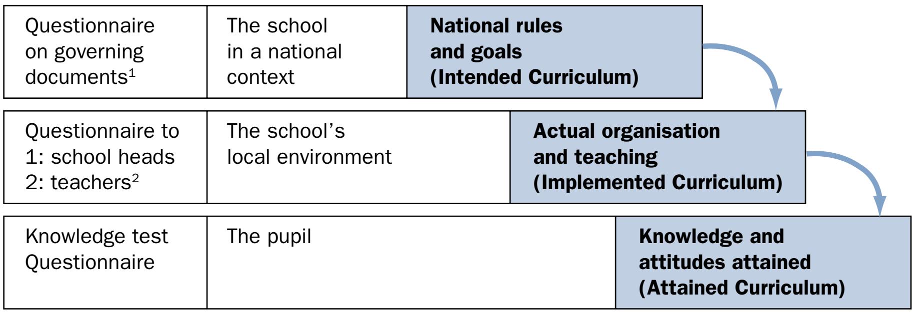
[17] Syllabus for Compulsory School Mathematics http://www3.skolverket.se/ki03/front.aspx?sprak=EN&ar=0809&skolform=11&infotyp=2&id=11

[18] Syllabuses for Compulsory School Science studies http://www3.skolverket.se/ki03/front.aspx?sprak=EN&ar=0809&infotyp=23&skolform=11&id=3878&extraId=2087

The frameworks in PISA and TIMSS

An obvious problem with international comparative assessments is that the participating countries give different priority to the general competencies aimed at, to the fact content to be covered, and to how the planning for the school year should look in a certain subject. It is therefore natural, as a starting point for this comparative analysis, to discuss the guidelines that have governed the design of the instruments used for the knowledge assessment in PISA and TIMSS.

Both PISA and TIMSS are based on multidimensional, explicit frameworks which, in the parts where the knowledge assessment is described, can be said to correspond to curricula and syllabuses in the Swedish system. The framework in TIMSS contains a model which, in general, can be said to be applicable to NU as well as to PISA and TIMSS. In it, the pupil's attitudes and knowledge are seen as results of the interaction between players at three levels. See figure 3.1. At the top of the model is society in general, the intentions of which as regards school are assumed to be expressed in national goals and guidelines. Next comes the level with the school heads and teachers who act for the implementation of the goals in school. At the base is the pupil with the knowledge and attitudes that have been attained. Knowledge tests and pupil questionnaires are the central components in all three assessments, but background data (such as in questionnaires) are also gathered in at the other levels in the model. As has already been said, this report refers in the main only to the knowledge tests in the three assessment studies.



[1] Only in TIMSS

[2] Only in NU & TIMSS

Figure 3.1 The TIMSS Curriculum Model.

As stated previously, the purpose of PISA is not primarily to evaluate how the different national goals of the school have been fulfilled. That, on the other hand, is the ambition expressed in TIMSS. It is therefore not surprising that PISA's framework is that which, in form, differs most from a conventional way of expressing goals regarding pupils' mathematical and scientific knowledge. TIMSS, however, describes the subjects in question on the basis of the respective subject's specific factual content and with a structure that is closely connected to, for instance, the Swedish syllabus.

Mathematics

According to the framework for PISA, the knowledge domain in mathematics, in which the pupils are to be tested, is described with the help of three components or dimensions:

- The situation or context in which the problem exists (daily life, education, working life, society, intra-mathematical).
- The mathematical content that can be used for solving the problem (organised in four overall themes: Space and shape, Change and relationships, Quantity and Uncertainty).
- The processes and competencies that have to be activated in the solution, such as modelling (i.e. mathematisation of a true situation), argumentation, communication, use of aids).

The framework for mathematics in TIMSS gives two dimensions for knowledge to be evaluated for pupils in the eighth school year: content and cognitive. These are further divided into domains:

- *Content domains:* Numbers (arithmetic), algebra, measurement, geometry and data (statistics).
- *Cognitive domains:* Knowing facts and procedures, using concepts, solving routine problems, and reasoning.

The two dimensions in TIMSS have direct equivalents in the national mathematics syllabus but they are also covered by PISA. An analysis of what in the three documents corresponds to the cognitive domain under TIMSS shows that all three contain clear wordings describing forms of knowledge that are aimed at in terms of reasoning, modelling, problem solving, communication and use of technical aids. It should, however, be noted in this context that Syllabus 2000 contains a comprehensive goal formulation that mentions the ability, with the help of knowledge in mathematics, to solve problems that arise at home and in society. It is only in the framework for PISA that the context in which a mathematical problem is presented is defined as a dimension. This indicates a difference between the two international studies that is possibly also evident in a comparison of the different test materials.

Table 3.3 compares the specification of the subject content of mathematics in TIMSS and PISA with that specified in Syllabus 2000. As can be seen, TIMSS and PISA contain information about content that corresponds to all of the goal areas (M1–M8). The subject area *Discreet mathematics* in PISA's framework is the only subject area that, in an analysis of content descriptions in the frameworks, cannot be matched with any topic in the chosen description of subject content under TIMSS and Syllabus 2000 as it has been interpreted in the teaching materials. In the case of TIMSS, it is doubtful whether *Symmetry and transformations* is an area that is included in a normal interpretation of the mathematics syllabus.

Table 3.3 A summary of the specification of subject content in mathematics according to the frameworks for TIMSS 2003 and PISA 2003 in comparison with Syllabus 2000.

Mathematics		
SYLLABUS 2000	Timss 2003	PISA 2000
Goals to aim for and Goals to attain by end of ninth school year	**Content domains**	**Themes/Main areas**
M1. Number sense M2. Arithmetic	Numbers Whole numbers Fractions and decimals Integers Ratio, proportion, percent	Quantity Numbers, Number sense Estimation Discrete mathematics
M3. Measurements	Measurements Attributes and units and tools, techniques and formulas	Measurements
M4. Geometry	Geometry Lines and angles Two and three-dimensional shapes Congruence and similarity Locations and spatial relationships Symmetry and transformations	Space & shape/ Geometry
M5. Statistics	Data Data collection and organization Data representation Data interpretation	Uncertainty Statistics
M6. Probability	Uncertainity and probability	Probability
M7. Algebra	Algebra Algebraic expressions Equations and formulas	Change & relationships Algebra
M8. Functions	Relationships	Functions

Science studies

In science, as in mathematics, the syllabuses for the subjects in question and the frameworks for PISA and TIMSS can be analysed on the basis of the three aspects or dimensions – subject content, thought processes or competencies, and situation or context. Detailed specifications of the first two aspects are the basis both for the construction of the questions and for the description of the pupils' performance in the two international assessments. As far as the context of the test questions is concerned, it is only in PISA that such weight is attached to this aspect that it is specified in detail in the definition of a special dimension. It is not surprising that this is the case considering the ambition in PISA to assess scientific literacy, defined as the ability to use scientific knowledge in today's society.

Matter's structure and properties, Chemical and physical changes and Human biology are examples of overall themes that define the subject content of PISA. The skills or competencies in science that, in accordance with the framework, the questions are intended to assess in PISA are:

- to be able to describe, explain and predict scientific phenomena,
- to understand scientific investigation,
- to be able to interpret scientific evidence and conclusions.

The contexts in which the questions in science are categorised are Life and Health, the Earth and Environment, and Technology.

The assessment of pupils' scientific knowledge in TIMSS should, in the same way as described in the framework in mathematics, be performed in a content dimension and in a cognitive dimension. The domains within which the tests are to assess are:

- *Content:* Biology (Life Science), Physics, Chemistry, Earth Science and Environmental Science
- *Cognitive activity:* Factual knowledge, Conceptual understanding, Reasoning and analysis.

In contrast to PISA, in TIMSS it was decided not to describe the cognitive dimension with reference to specific activities in science studies. As was the case for mathematics, TIMSS retains definitions of competencies expressed in more general terms. The corresponding situation can be noted in PISA, where the Swedish syllabuses in mathematics are compared with the science syllabuses. The mathematics goals of the pupil being able to communicate and discuss, and other general competencies, have their equivalent in science in goals relating to processes, being able to perform different activities in science (such as knowledge about the scientific way of working) or being able to use the knowledge in different, everyday contexts (such as being able to use scientific arguments in discussions).

Table 3.4 compares the specification of the dimension content and competencies in science in TIMSS and PISA with Syllabus 2000 for biology, physics and chemistry. The table provides a basis for analysing how the frameworks in PISA and TIMSS correspond to the three goal categories in the syllabuses for biology, physics and chemistry in respect of knowledge concerning nature and Man, the scientific activity and the use of science.

In biology, specifications of subject content in both PISA and TIMSS cover all of the Swedish syllabuses' goals to aim for with regard to nature and Man (B1.1.1–B1.5.2[19]). According to Syllabus 2000, more or less all the goals in physics and chemistry can be matched to equivalents in TIMSS, while the coverage in PISA is less good. In it, there is no specific mention, for instance, of sound, light or electricity in physics (P1.2.3–P1.2.5)

TIMSS's content domains Earth Science and Environmental Science contain a few topics that do not correspond to any goals in science as per Syllabus 2000. This is also applicable to that which is included in PISA in the theme ”Geological processes” and in TIMSS's topic area ”Population changes”.

The goals "concerning scientific activity" and "concerning use of knowledge" in biology, physics and chemistry are covered to a high degree in PISA, where the former is included as an independent dimension and the latter is emphasised in the basic definition of the knowledge domain that PISA is to assess. The framework in TIMSS does not, in these respects, contain correspondingly clear guidelines for what the tests should assess. TIMSS's framework, however, as a supplement to the description of the two dimensions, content and cognitive activity, has added a section on the scientific work procedures.

[19] See table 3.4.

Table 3.4 A summary of the specification of subject content in science studies according to the frameworks for TIMSS 2003 and PISA 2003 in comparison with Syllabus 2000.

Science studies		
SYLLABUS 2000	Timss 2003	PISA 2000
Goals to attain by end of ninth school year	**Content domains**	**Scientific themes**
Biology ... *concerning nature and Man*	Biology (Life Science)	
• recognise and name common plants, animals and other organisms in the local environment and recognise their environmental requirements[1]	• Living organisms: types, characteristics and classification	
• B1.5.1 ... organs in own bodies	• Structure, function and life processes of organisms	• Human biology • Physiological processes
• B1.1.2 ... how cells are built • B1.1.1 ... photosynthesis and combustion	• Cells and their functions	• Form and function • Energy conversions
• be able to give examples of life cycle of some plants and animals and their different growth processes[1]	• Development and life cycles of organisms	
• B1.3.1 ... genetic heredity • B1.3.3 ... conception	• Reproduction and heredity	• Genetics
• B1.3.2 ... development of life ... biological diversity	• Diversity, adaptation and natural selection	• Biological diversity
• B1.2.2 ... recycling and accumulation • B1.2.1 ... the worlds' ecosystems ... interrelationships between organisms	• Ecosystems	• Ecosystems
• B1.4.1 ... biology of sexual life, • B1.5.2 ... addictive substances	• Human health	• Human biology
Physics ... *concerning nature and Man*	Physics	
• P1.1.1 ... forms of energy and energy conversion	• Energy types, sources and conversion	• Energy conversions
• P1.2.2 ... pressure, heat and temperature in relation to different forms of matter	• Forces and motion • Heat and temperature • Physical states and changes in matter	• Force and motion • Atmospheric processes
• P1.2.3 ... sound is created, dispersed and recorded • P1.2.5 ...propagation, reflection and breaking down of light ... the eye	• Sound and vibration • Light	
• P1.2.4 ... electric current, voltage, electrical energy and its effects ... generation of electricity,	• Electricity and magnetism	
• P1.2.6 ... applications of physics ...		• Science in technology
• P1.2.7 ... building up of matter from elementary particles and atoms, • P1.3 ... different kinds of radiation and its interaction	See Chemistry	• Structure and properties of matter
• P1.4.1 ... structure of the universe • P1.4.2 ... solar system ... stars	Earth Science • Earth in the solar system and the universe • Earth's structure and physical features • Earth's processes, cycles and history	• Earth and its place in the universe • Geological processes
Chemistry	Chemistry	
• C1.1 ...elements, chemical compounds and chemico-technical products ...	• Classification and composition of matter	
• C1.2 ... transformation in chemical reactions,	• Chemical changes • Acids and bases	• Chemical and physical changes in matter
• C1.3 ... structure of atoms and chemical bonding ... • C1.5.2 properties of water ... its role as a solvent ...as means of transport	• Particulate structure of matter • Properties and uses of water	
• C1.4 ... thinking concerning chemistry in earlier times ... • C1.5 ...indestructibility of matter, transformation, recycling and dispersion	See Earth Science	
	Environmental Science • Changes in population • Use and conservation of natural resources • Changes in environments	

Science studies		
SYLLABUS 2000	**Timss 2003**	**PISA 2000**
Goals to attain by end of ninth school year	**Content domains**	**Scientific themes**
Science studies *concerning scientific activity* • B2.1 … scientific ways of working … • P2.2 …interaction between investigations … and development of concepts, model s and theories … • S2.3 …Man's perceptions of the world *concerning use of knowledge* • C3.1.1 … use results of measurements and experiments in discussions on environmental issues • C3.2.3 …be able to discuss use of resources in private life and in society	Scientific method of working	Scientific method • Interpretation of arguments and conclusions

[1] Goals to attain by end of fifth year in school

4 Method

Introduction

This study has been performed in three stages. A preliminary task was to go through the documents on which the design of the three assessments of compulsory school pupils' knowledge in mathematics and science was based. The result was presented in the previous section. The tests that had been used in NU, PISA and TIMSS in the spring of 2003 were studied and a taxonomy was drawn up for describing the three sets of test material. The second stage involved classification of all the questions in accordance with the chosen taxonomy, i.e. the system for categorising the questions' subject content, type of question etc. To conclude this stage, a reference group was appointed, with four active, experienced teachers who, independently of each other, examined randomly chosen questions. The third and last stage involved an analysis of the results and the writing of the report.

The three assessments have, as has already been said, been drawn up on the basis of different aspects. While the purpose of NU and TIMSS is said to be to find how well the pupils achieve the knowledge and skills described in the national syllabuses for the subjects in question, the purpose of PISA is to find out whether the pupils achieve a general scientific and mathematical level of proficiency, defined by the OECD, that is not as clearly divided into subjects nor able to be described in terms of subject content. However, it is not clear whether the differing intentions are reflected in the formulation of the questions or in the pupils' reaction to the test questions in the three knowledge assessments. The frameworks do differ in some cases in the way they are drawn up, their content and terminology but, as we have already seen, they also demonstrate many common features. The questions that were designed within one framework's special way of describing the domains examined can very well also be expected to fit into a test in another of the assessments of competencies achieved. The tests used refer to different school subjects that, apart from the question of subject matter, method of working etc, also differ with regard to the general competencies that are trying to be achieved. In a comparative study to find similarities and differences between the three tests, it has been considered necessary to examine all the questions used. It is therefore important to use one, single system for categorising the questions, to be able to answer the question regarding the extent to which the instruments in the three assessments

- focus on the competencies aimed at in accordance with the syllabuses,
- refer to subject matter that corresponds to the goals in the syllabuses and
- ask the same kinds of questions.

By examining features of the questions posed to the pupils, an attempt is made to obtain a basis for use in drawing conclusions about the extent to which the results of the knowledge tests are a measure of goal attainment in Swedish school.

On the basis of Syllabus 2000, the following aspects have been chosen in this study in order to categorise the test questions used:

- what type of answer is expected,

- how much text has been used to present the task and question,
- whether there is information provided in pictures or diagrams
- in what context is the task and question presented,
- what subject content is being tested,
- to what extent calculations are required,
- what type of thinking is entailed and
- the level of difficulty of the question

Type of answer

Table 4.1 Types of question classified in accordance with the format of the expected response.

Type of question	Code	Question answered by
Multiple choice	MC	selection of one from four or five response options
Complex multiple choice	CMC	selection of indicator that a statement is true or false or identification of gaps in text or figures by selection of given response options
Short response	SR	calculated value or a single word possibly with several possible acceptable responses
Open constructed response	OCR	pupil's own account or statement, possibly motivated by several possible acceptable interpretations and responses
... draw	_d	only drawing (SR_d) or by supplementing a statement by drawing (OCR_d).

The type of response the pupil should give to the questions is classified in accordance with the categories presented in table 4.1. When looking at the questions from the different tests, it can be seen that several possible principles could be the basis for this classification. One common breakdown is whether the questions are answered by choosing from stated alternatives or if they are answered by giving an account of something the pupil has constructed. In the tradition of American testing, the most common form of multiple choice question contains a stem followed by four or five response options, of which one should be chosen. These are designated with the code MC (multiple choice) in this classification. Several more complex variants of this type of response occur, such as questions where a number of given concepts should be paired together with a definition or phenomenon. These complex forms of response are classified with the code CMC (complex multiple choice).

NU F1K

Which phase transition is this?

In the right-hand column, for each case, write one of the words *melting, evaporation, boiling, condensation, solidification* or *chemical reaction*. Select the word that is most appropriate.

Water sputters in a frying pan	....
Ice is forming on the lake	....
The charcoal is glowing	....
A clear liquid is forming round the burning wick of the candle	....
The wet asphalt is drying up after the rain	....
A mist is forming on the spectacles	....

Figure 4.1 A question in the NU-03 physics test that has been classified as complex multiple choice (CMC).

The degree of openness (Löfdahl 1987) varies from completely closed multiple choice questions in which only one particular given response as per the assessment instructions can be judged as correct, through to open responses where different interpretations of the question, giving responses corresponding to certain conditions, can be tolerated. The chosen categories are designated SR and OCR (short response and open constructed response) as indicated in table 4.1.

Text volume/Tables/Graphics

All subjects must contribute to the pupils' language and communication development, as mentioned in the curriculum. This includes the ability to obtain as well as to produce information, expressed in different forms of text and graphic formats. There has to be a balance between these general goals and the more subject-specific knowledge goals, and it can be considered important to compare the different knowledge assessments from this aspect. The categories in tables 4.2–4.3 are intended for classification of the presentation of the questions with reference to how much text the pupils have to read and any other data they have to obtain from tables, figures and diagrams in order to familiarise themselves with the task at issue. In those cases where an introductory text refers to several of the questions that follow, for each of these questions the number of lines has been registered that is the sum of the number of lines in the question plus the introductory text's length divided by the number of questions relating to it.

Table 4.2 Text volume.

Information	Code	Question presented with
Text < 0.5 line	0	less than half a line of text
Text ≥ 0.5 and < 1.5 lines	1	half to one and a half lines of text
Text ≥ 1.5 and < 2.5 lines	2	one and a half to two and a half lines of text
Text ≥ 2.5 and < 3.5 lines	3	two and a half to three and a half lines of text
Etc.	…	… …

Table 4.3 Occurrence of pictures, tables and diagrams.

Information	Code	Question presented with
Decoration	D	picture that in some way illustrates the question but does not convey necessary information
Informative picture	P	with picture/s with necessary and/or explanatory information to allow the question to be answered
Table	T	table with necessary information to allow the question to be answered
Diagram	D	diagram/graph with necessary information to allow the question to be answered

Subject content

As has already been said, it is plain that TIMSS has the most detailed document as regards the way in which subject content is described in the Swedish syllabuses and in the two international studies. We have also previously seen that the definitions of subject content in TIMSS cover what is stated in the goals in our Swedish syllabuses for the subjects in question. This is confirmed by the result of a review of one of the most used textbooks in each respective subject area, where it is apparent that the content can acceptably be described with the help of the subject areas defined in TIMSS. It therefore seems possible to let the "syllabus" in TIMSS constitute the basis for a classification of the tests as far as subject content is concerned. See tables 4.4 and 4.5, which briefly describe the basis for the classification used in the analysis of the subject content in the different sets of test material. The topic areas are defined in further detail (topics) in TIMSS Assessment Frameworks and Specifications 2003 (Mullis et al. 2003).

Table 4.4 Content domains and topic areas in mathematics according to TIMSS.

Content domaim	Topic area
Number sense and Arithmetic:	Whole numbers Fractions and decimals Integers Ratio, proportion and per cent Discrete mathematics[1]
Algebra and functions:	Patterns Algebraic expressions Equations and formulas Relationships
Measurement:	Attributes and units Tools, techniques and formulas
Geometry:	Lines and angles Two and three-dimensional shapes Congruence and similarity Locations and spatial relationships Symmetry and transformations
Statistics:	Data collection and organization Data representation Interpretation of data Uncertainty and probability

[1] This subject area is not represented in the framework for TIMSS

Table 4.5 Content domains and topic areas in science studies according to TIMSS.

Content domain	Topic area
Biology (Life Scence):[1]	Types, characteristics of and classification of living things Structure, function and life processes in organisms Cells and their functions Development and life cycles of organisms Reproduction and heredity Diversity, adaptation and natural selection Ecosystems Human health
Chemistry:	Classification and composition of matter Particulate structure of matter Properties and uses of water Acids and bases Chemical change
Physics:	Physical states and changes in matter Energy types, sources and conversions Heat and temperature Light Sound and vibration Electricity and magnetism Force and motion
Earth Science:	Earth's structure and physical features Earth's processes, cycles and history Earth in the solar system and the universe
Environmental Science:	Changes in population Use and conservation of natural resources Changes in environments

[1] Förklaring till fotnot saknas

Context

An implication of the link of the science study subjects to human activity and, thus, to a reality experienced by pupils up till the present time or likely to be in the future is implicitly given in the syllabuses but is not mentioned explicitly in any general formulations in the curriculum or syllabuses. As a direct consequence of the character of these subjects, it can therefore be considered natural that test questions in science are most often suitably presented in a context referring to experiences in nature and human activity in general.

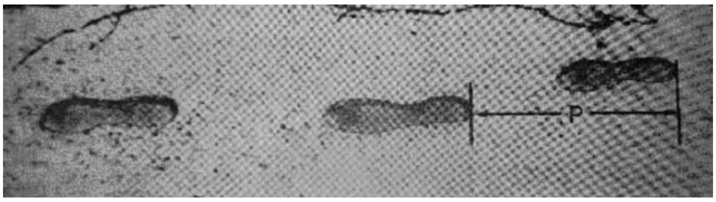
In mathematics, being a subject more aimed at abstractions, the situation is somewhat different. It is declared in some contexts that mathematics has no need for the surrounding world, but that the reverse applies. Such a notion, however, is in contrast to the general spirit of the national governing documents. In the curriculum's "Goals to attain" it says that every pupil shall have "a command of basic mathematical thinking and be able to apply it in everyday life". The areas of application that are then mentioned in the syllabus are everyday life, society and scientific activity. The conclusion is that the context in which tasks and questions in NU, PISA and TIMSS are presented is a relevant aspect for the analysis of the knowledge tests and the assessment of their validity in relation to the school's governing documents. In TIMSS's framework there is no explicit discussion of any context dimension while, in PISA, an attribute is used that is designated 'situation'. The categories that are used there to classify the context in which the problem or task is presented are "Personal, Education, Professional life, Society and Science" for the questions in mathematics as well as in science.

In accordance with the intentions of the curriculum, a first step is thus to classify whether the task refers to a world outside the pure subject theory or the laboratory. A task that does not fulfil this condition is given the code 0. Instead of then following the model in PISA and actually classifying the context in which the task is placed, a taxonomy was selected here that is aimed more towards the quality of the connection to reality. That the context in which a task is presented belongs to a possible world of experience outside the school for the pupils does not always mean that the question thus framed is authentic, i.e. that it occurs in the real situation that the question is trying to simulate. (Palm, 2002). Classification is therefore made here of those tasks that are placed in a 'real situation' with regard to the relevance of the question that the pupil is asked to answer. The categories selected are presented in table 4.6. If the task refers to a real life situation, and the question to be answered also corresponds to something that could occur in a real situation, the task then belongs to sub-category 2. Otherwise it belongs to sub-category 1. Figure 4.2 gives an example of a task that has been judged to belong to the latter category (1).

Table 4.6 Categories for classifying connection to reality and the relevance of the problem presented.

Connection to reality	Code	Question is placed in a context
None or special	0	that does not refer to a world outside the subject's internal reifications
Situation but not question	1	that is based on a reality outside the subject theory and refers to practical situations but with questions that probably would not be asked in reality
Situation and question	2	that is connected to reality (everyday life, society, professional life, research) but problems that can be assumed to be relevant in the given context

PISA M124
WALKING



The picture shows the footprints of a man walking. The pacelength P is the distance between the rear of two consecutive footprints.

For men, the formula $\frac{n}{P} = 14$ gives an approximate relationship between n and P where

n = the number of paces per minute and P = the pacelength in metres.

Question Q1
If the formula applies to Heiko's walking and Heiko takes 70 paces per minute, what is Heiko's pacelength? Show your work.

Question Q3
Bernhard knows his pacelength is 0.80 metres. The formula applies to Bernhard's walking.

Calculate Bernhard's walking speed in metres per minute and in kilometres per hour.
Show your working out.

Figure 4.2 Questions in the mathematics test in PISA 2003, both of which have been coded in category 1 as regards connection to reality.

Calculations

With regard to calculation requirements, the classification is based on whether the calculations that occur are considered to constitute a decisive difficulty in the question. The categories selected are presented in table 4.7.

Table 4.7 Requirements for calculations to answer the question.

Requirements for calculations	Code	The question is answered
None	0	without prior calculations or otherwise numerically based conclusions.
Less decisive	1	with numerically based statements where the occurrence of calculations is of the simplest type (simple mental arithmetic) and/or only slightly affects the difficulty in solving the question acceptably.
More decisive	2	numerically based statements and occurring calculations are more demanding and/or decisive if the question is solved correctly.

Competencies

One classification of a question in a cognitive respect, i.e. with regard to the mental activity that is required to answer a question, is based on a number of assumptions about the pupil who is to answer the question. The pupil is assumed, for instance, to have followed a normal course in the subject, with normal experiences from the teaching and from life in general. It can be realised directly that the marking of a question, particularly in this respect, has to be subjective and vary between different markers. The categorisation of the questions should therefore be done by experienced teachers in a group and in a process where consensus is reached, for instance, through discussions on a selection of typical cases of the questions to be categorised.

Many different taxonomies have been developed to describe the cognitive dimension in tests and test questions. The most established is surely Bloom's taxonomy[20] for educational objectives. This taxonomy has often been regarded as a one-dimensional, hierarchical division into levels that normally does not produce reliable results in a classification of test questions.[21] Bloom's taxonomy was recently presented in a revised form.[22] The immediately noticeable difference is that the latter uses two dimensions, the objects of knowledge and the cognitive process, and claims to be adapted for use in analysis of target documents for pupil's learning as well as of the teaching, test and assessing activities. This situation can seem promising at the prospect of this task of finding a classification system for the cognitive properties of test questions.

The goals in the syllabuses for science and mathematics are expressed in the form of competencies or knowledge levels that the pupils should attain in order to receive a Pass after the ninth year in school. As regards the higher grades, there are grade criteria that, in combination with goals to aim for and goals to attain, describe expected performance of pupils who have attained a higher level of grade. All in all, these guidelines for marking can be said to correspond to the categorisation in a cognitive respect that occurs in different forms in TIMSS and PISA. It is desirable that a simple basis be found for classifying the questions with regard to the degree of complex thought that is required of the pupils in order to answer the question.

The concept of reasoning stands out as a central aspect in the learning process. In the mathematics syllabus, "the ability to follow, understand and use mathematical reasoning" constitutes one of the three aspects to be taken into account during marking. In science studies, under "Science as a human and social activity", it is stated that "The pupil's ability to argue on the basis of scientific as well as ethical and aesthetic perspectives is taken into account during assessment and marking.". As a basis for the classification of competencies in mathematics in PISA, the ability to argue is an important component while "Reasoning" is the designation of one of the cognitive domains in both mathematics and in science in TIMSS.

Faced with the task of classifying, with acceptable reliability, the approximately 500 test questions in the three assessments, it has been necessary to make step-by-step simplifications. The end product, which in its essential features corresponds to PISA's cognitive taxonomy, is a three-step scale for making the classification possible to perform with reasonable effort and reliability of judgement. The categories selected are presented in table 4.8.

TIMSS M032233

A computer club had 40 members and 60% of the members were girls. Later, 10 boys joined the club. What percent of the members now are girls? Show the calculations that lead to your answer.

Answer: ....

Figure 4.3 Question in TIMSS 2003 that was deemed to belong to category 2 as regards cognitive level.

[20] Bloom (1956)

[21] Jansson (1980)

[22] Anderson & Krathwohl (2001).

Table 4.8 Categories for classifying the expected mental activity that answering a question requires.

Cognitive level	Code	The question requires that the pupil
Reproduce	0	in mathematics and science can reproduce facts and concepts and, in mathematics, can conduct a procedure that can be deemed normal in the school year in question.
Apply	1	in mathematics, can routinely apply elements in the subject in question that are needed to solve the question.
	-	(Not applied in science studies.)
Reason	2	in mathematics, can mathematicise and, in two or more steps, interpret, analyse and/or evaluate a new problem situation.
		in science, can break down a new problem into sub-questions, the answers to which lead to the solution of the original question.

Difficulty

As has previously been mentioned, the degree of correctness (designated p here) in the interval 0.0 to 1.0 states how great a proportion of the maximum attainable points total for a question that has been awarded to a group of examinees whose test included that question. The mean value of the degrees of correctness in a test provides an overall measure that can be assumed to provide evidence of the extent to which the complexity of the solution procedure experienced by the examinees is high or low and the extent to which the problems in the questions in a test are new or familiar. In norm-related tests, a mean value of close to 0.5 of the degrees of correctness is aimed at for a test, and a reasonable spread from easier to more difficult questions, in order to be able to use the whole scale from 0.0 – 1.0 to describe the pupil groups' achievements.

The result of the assessment of the pupil answers is reported in mathematics in NU, and in both the mathematics and the science questions in PISA and TIMSS, in a similar way, using zero for an unacceptable response and a maximum of one or two points for an acceptable response (solution). With regard to the science subjects in NU, a quantitative scale is not used for the assessment. To create a basis for comparing the degree of difficulty in the three sets of test material, the data reported for NU have been processed so that "Fail" answers have been awarded zero points, while questions with responses that could have been marked on two levels have given one point for a ”Pass” answer and two points for an answer marked ”Pass with distinction”. If an acceptable answer to the question has been deemed to be markable only on one level ("Pass" or "Pass with distinction"), the question has been given the maximum point of one.

5 Results

Mathematics

Type of answer

The most common type of questions in TIMSS was where the pupil should select one of four or five given response options. Two-thirds of the 194 mathematics questions in TIMSS were multiple choice questions (MC), while this was the case for only 14 per cent (3 questions) of the 21 questions in NU. See figure 5.1. In NU, the remaining 86 per cent were short-response questions (SR and SR_d). In PISA too, short-response was the most relatively commonly occurring classification, referring to 41 per cent of the 86 mathematics questions. Complex multiple choice questions (CMC) together with multiple choice questions (MC) occurred in PISA in 35% of the questions. In PISA, 23 per cent were of the type where the pupil should construct their own response (OCR). In TIMSS, only 5 per cent were open constructed response questions that required the pupil's own, constructed answer, including those where the pupil was expected to supplement their answer with a drawing (OCR_d).

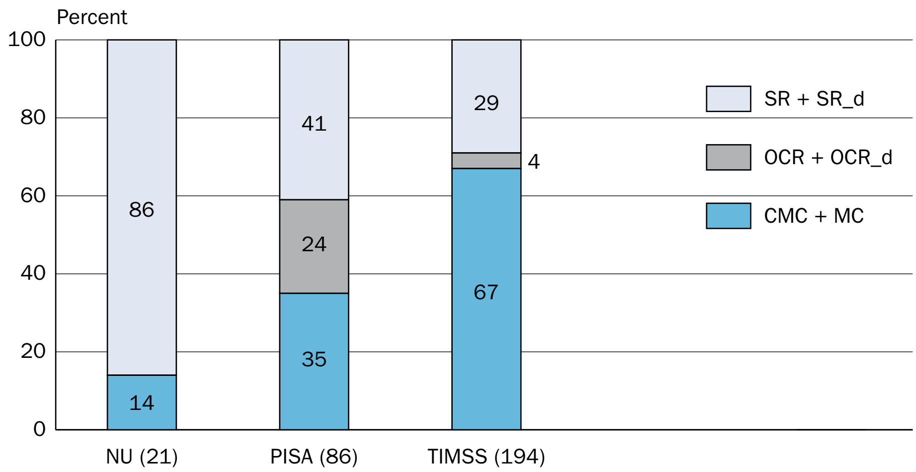


Figure.5.1 Distribution of questions in mathematics in NU, PISA and TIMSS 2003 based on type of answer.

Text and graphics

There are considerable differences in the length of the texts used to present the mathematics questions in the three sets of test material. The mode for both NU and TIMSS is one line, while for PISA it is three. See figure 5.2. The corresponding mean values are 1.2, 2.3 and 3.7 lines. See table 5.1. Almost one-third of the questions in PISA (30%) are presented in five lines or more, while texts of that length do not occur at all in NU and, in TIMSS, are so infrequent as to occur in only eight percent of the questions.

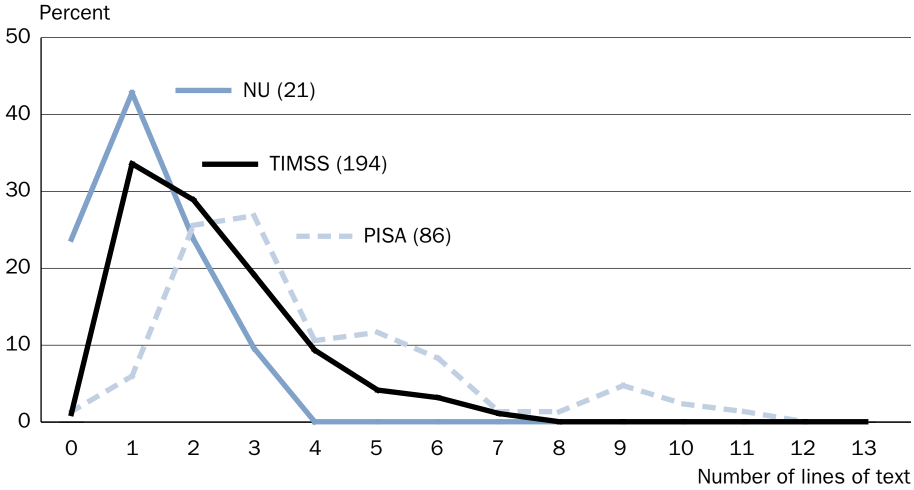


Figure 5.2 Distribution of questions in mathematics in NU, PISA and TIMSS 2003 based on length of text in the question.

Table 5.1 Length of text in mathematics questions.

	Number of questions	Mean number of lines
NU	21	1.2
PISA	86	3.7
TIMSS	194	2.3

PISA has the greatest proportion of questions containing illustrations, tables and diagrams. See table 5.2. About every third question in PISA contains a table or a diagram, while this is only half as common in TIMSS. Two of five questions in PISA contain illustrations, while the corresponding proportion in TIMSS is one out of four questions, not counting the questions in geometry. (This can be reasonable when comparing the test instruments, as PISA does not have any questions in geometry, which is a subject area where the questions almost per definition include a figure.)

Table 5.2 Distribution of questions in mathematics in NU, PISA and TIMSS 2003 based on presence of illustrations, tables and diagrams in the question.

		Illustration				
	Number	Decoration	Support	Necessary	Table	Diagram
NU	21	0%	0%	10%	0%	0%
PISA	86	8%	13%	22%	15%	19%
TIMSS	194 (163[1])	0%	2%	34% (21%[1])	9%	8%

[1] If geometry is excluded

Subject content

Tthe questions in mathematics in the three sets of test material have been possible to be categorised in accordance with the selected taxonomy for subject content. The reliability, independently assessed by two or three examiners, has been over 90%. On the basis of these data and in accordance with the purpose of this study, different aspects can be placed on the three assessments, and answers to different issues can be sought. One question can be about the degree of coverage of the tests, in other words, how well the tests cover the subject content described in the syllabus. The most detailed level for description of the subject

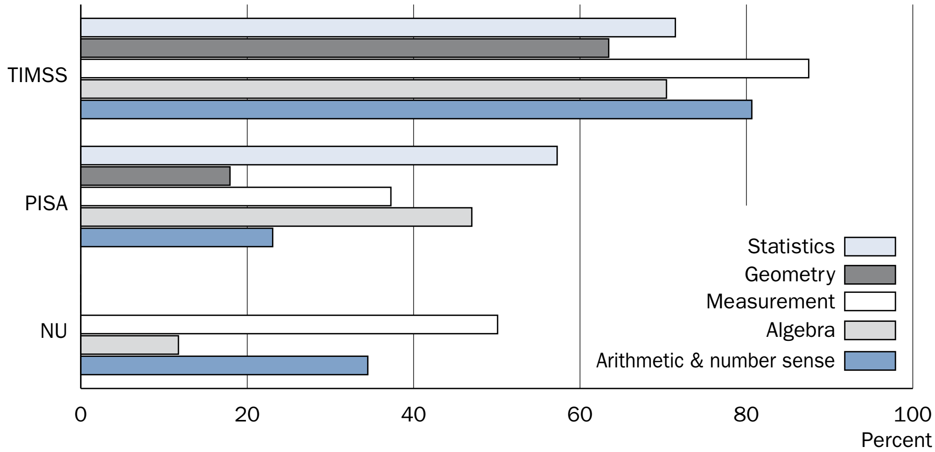


Figure 5.3 Relative coverage in NU03, PISA 2003 and TIMSS 2003 of the topics in five content domains in mathematics.

content is given by the topics in the selected framework. Figure 5.3 presents the way in which the subject content is covered in the tests in mathematics, defined in the content domains (Arithmetic and Number Sense, Algebra etc.). This is done by stating the proportion of the topics in a content domain represented by at least one question.

In a comparison between the three assessments, TIMSS has easily the broadest coverage of subject content. All the content domains in TIMSS are covered up to 64 per cent (Geometry) or more, counted in that proportion of the total number of topics that are represented by questions. Measurement, which has the highest value in TIMSS, has a coverage of seven of a total of eight topics (88%). As regards PISA, statistics has the greatest share of coverage with 57 per cent. In total, the description of mathematics contained 87 different topics of which TIMSS covered 74 per cent and PISA 33 per cent. The short-response test in mathematics in NU covered 18 per cent of the topics.

The different topic areas and topics in mathematics can be brought in under the eight goal categories M1–M8. One important aspect that can be established is the relationship between the weight "normally" attached to a certain content domain during teaching and the orientation of the evaluating test. As has been said previously, the syllabuses do not give any clear information as to what should be regarded as normal. Partly to get a measure of the scope a certain goal takes up in the teaching and partly to obtain a breakdown per school year of the subject matter, a commonly used teaching aid has been used as a source of information.[23] In a series of textbooks for years 7–9 in compulsory school[24] the content has been categorised in a way equivalent to that applied in the tests. The number of pages used to deal with matter that belongs to the 87 mathematics topics has been registered for years 7–8 and 7–9 respectively. It was then possible to match the content to the goals in mathematics in accordance with Syllabus 2000 and to the result of the analysis of the content in the questions in NU, PISA and TIMSS. Table 5.3 presents the result of how the goals M1–M8 are covered by the questions in NU, PISA and TIMSS. In less than 3% of the

[23] The teachers follow the textbooks to a higher degree in mathematics than in science subjects. The National Agency for Education (2003).

[24] Björk et al. (1997)

Table 5.3 Distribution of questions in mathematics on the basis of relevant goal to aim for.

	develops their number and spatial sense and ability to understand and use	NU number	PISA number	TIMSS number
M1. Number sense	- basic number concepts	5	1	15
M2. Arithmetic	and calculation with real numbers, approximate values, proportionality and per cent,	8	30	42
M3. Measurements	- different methods, measuring systems and instruments in order to compare, estimate and decide the size of important quantities,	6	4	31
M4. Geometry	- basic geometrical concepts, properties, relations and theorems,	0	13	31
M5. Statistics	- basic statistical concepts and methods for collecting and handling data and for describing and comparing important properties of statistical information,	0	16	20
M6. Probability	- probability thought processes in particular random situations.	0	7	8
M7. Algebra	- basic algebraic concepts, expressions, formulas, equations and inequalities,	2	8	43
M8. Functions	- properties of some different functions and corresponding graphs,	0	7	4

questions assessed in each of the PISA and TIMSS surveys, it has been considered impossible to attribute any of the goals for the school years 7–9 to the questions.

TIMSS has 194 questions in mathematics while PISA has a total of 86 questions. The test that was used in NU contained 21 questions. In PISA and NU, arithmetic was the most common subject content while, in TIMSS, there were about as many questions in algebra as in arithmetic.

Figures 5.4–5.6 present comparisons of how in the three assessments the questions were distributed over the eight goal areas in mathematics. In each of the diagrams, the result of the content analysis of the series of textbooks for the school years in question are also presented.

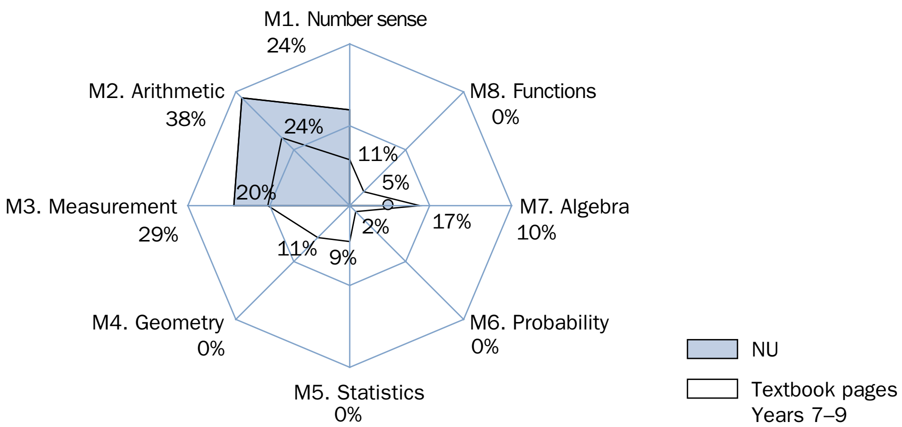


Figure 5.4 Relative distribution of the questions in mathematics in NU-03 compared with the distribution of the content in a typical series of textbooks for years 7–9. The text in the outermost ring gives the goal areas and their shares in the test material while, in the inner ring, shares in the textbooks for years 7–9 are given.

The main subject content in the NU questions has been assessed to belong to four of the eight goal areas. See figure 5.4. Arithmetic, measurements, number sense and algebra are the areas that, in the analysis of the textbooks for years 7–9, received the highest entries for number of pages (24, 20, 11 and 17%), which corresponds to those particular areas that the mathematics test in NU contains (38, 29, 24 and 10%).

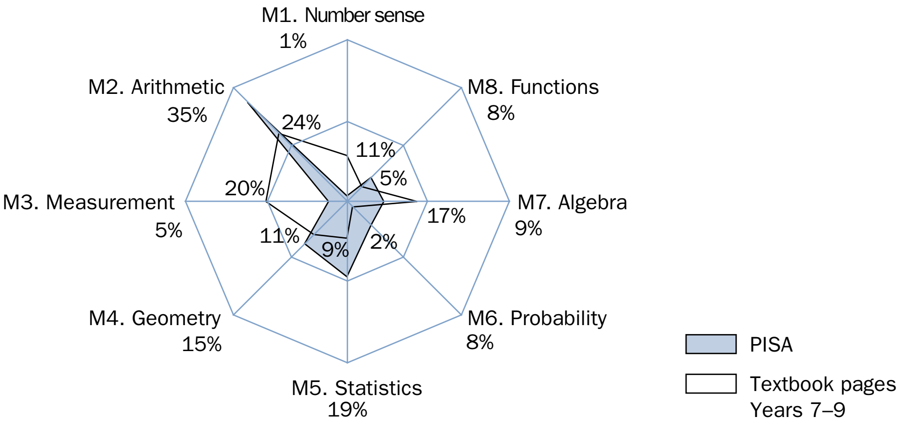


Figure 5.5 Relative distribution of the questions in mathematics in PISA 2003 compared with the distribution of the content in a typical series of textbooks for years 7–9. The text in the outermost ring gives the goal areas and their shares in the test material while, in the inner ring, shares in the textbooks for years 7–9 are given.

Of the questions in PISA, 35% have been judged mainly to measure knowledge in arithmetic. See figure 5.5. Statistics is the next most common (19%). Only one and five per cent respectively have been categorised as questions in number sense and measurement, both of which nevertheless constitute areas that are dealt with on a relatively large number of pages in the textbooks (11 and 20% respectively of the total number of pages).

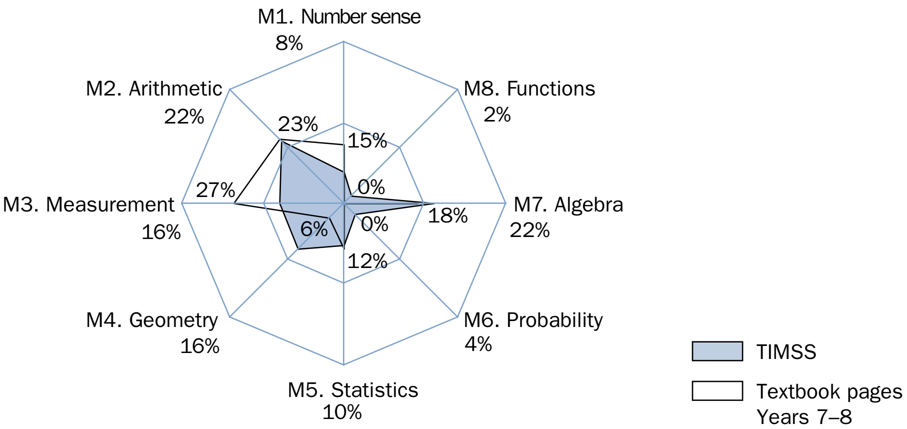


Figure 5.6 Relative distribution of the questions in mathematics in TIMSS 2003 compared with the distribution of the content in a typical series of textbooks for years 7–8. The text in the outermost ring gives the goal areas and their shares in the test material while, in the inner ring, shares in the textbooks for years 7–8 are given.

Of the three sets of test material, the distribution of the questions in the eight goal areas agrees best in TIMSS with the distribution of the number of pages in these areas in the textbooks for years 7 and 8. See figure 5.6. It can be noted, for instance, that algebra questions are more common in TIMSS than in PISA, while the reverse applies for statistics.

Competencies

In practice, all mathematics questions require some kind of reasoning in order to arrive at an answer but, in the analysis of the questions and in the reference group's validation, an attempt was made to separate the questions requiring complex thought processes of the pupil, in accordance with the chosen system for categorisation.

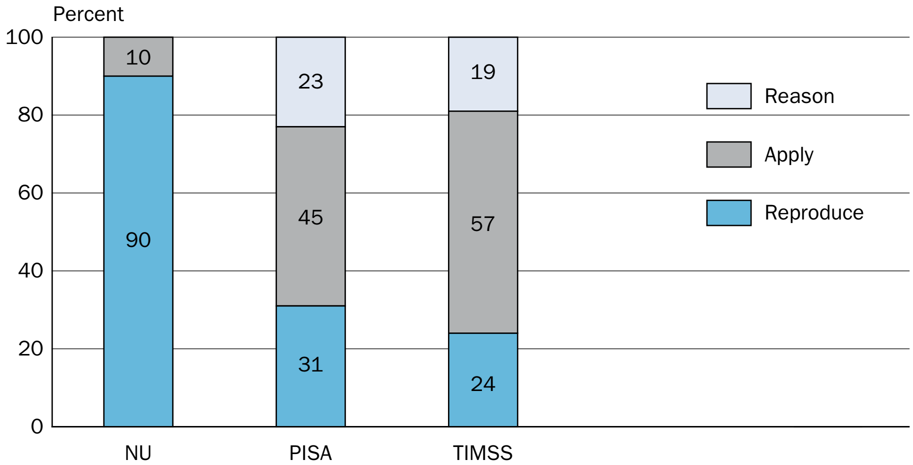


Figure 5.7 Distribution of questions in mathematics in NU-03, PISA 2003 and TIMSS 2003 on the basis of assessed cognitive level.

The distribution of the number of questions in the three categories eventually used is presented in figure 5.7. The results in PISA and TIMSS are similar, with about twenty per cent being classified as "Reasoning". In the analysed short response test in NU, nine out of ten questions were such that the pupils were presumed to follow a standard procedure when solving the questions while, in one of ten, the pupils were presumed to routinely apply mathematical knowledge and understanding.

The assessment of calculation requirements did not produce any remarkably differences between the three testing materials. NU had the greatest proportion of questions (approximately every third question) where the calculations that the pupil should make represented a decisive problem in the question. The corresponding proportion was just more than every fourth question in PISA and TIMSS.

Connection to reality

It is natural that some of the attributes used for classifying the properties of the questions become particularly dependent on varying interpretations of definitions and, thus, subjectivity. The connection to reality, with the two aspects chosen here, proves itself, as expected, not to give such high reliability coefficients[25] but the results are nevertheless very clear. As could be expected on the basis of the framework's construction, PISA has the greatest proportion of questions with varying degrees of connection to reality. As many as 84 of the 86 mathematics questions (98%) in PISA were connected with reality. One half,

[25] Number of corresponding assessments divided by the total number of assessments with mode = 0.7 for questions with codes 1 and 2.

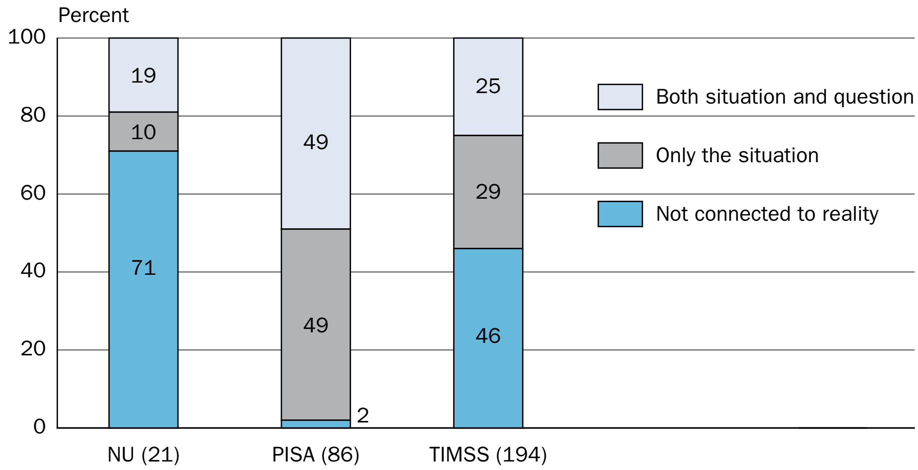


Figure 5.8 Distribution of questions in mathematics in NU-03, PISA 2003 and TIMSS 2003 on the basis of assessed connection to reality and authenticity.

49%, were in the category 'connection only of situation to reality' (code 1) and the other half, 49%, had connection to reality of both situation and problem (code 2). See figure 5.8. In TIMSS, the combined proportion is 54%, with 25% having been coded in the highest category (code 2).

Difficulty

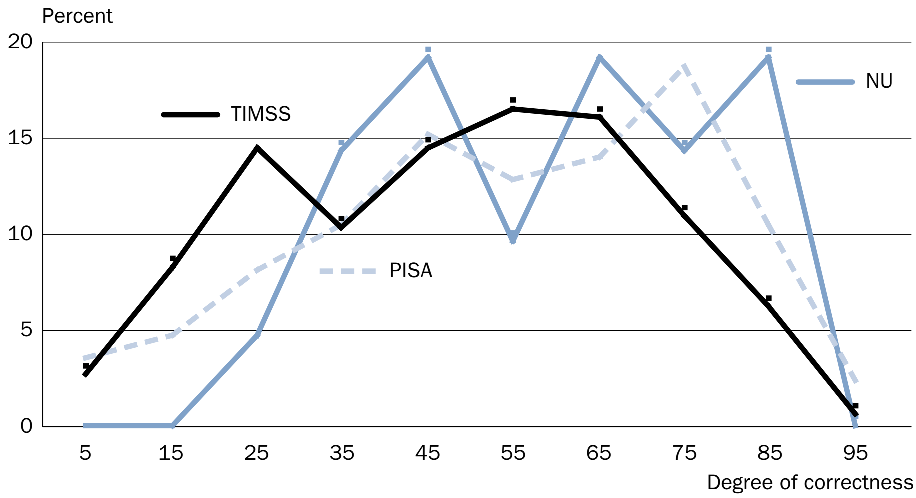


Figure 5.9 Difficulty of the mathematics questions in NU-03, PISA 2003 and TIMSS 2003.

The question pools in the three assessments NU-03, PISA 2003 and TIMSS 2003 do not differ significantly as regards difficulty, neither in respect of the distribution's general look nor its mean value. See figure 5.9. The questions in NU have on average a degree of correctness of 0.59, in PISA the average is 0.54, and the questions in TIMSS are insignificantly more difficult for the pupil group, with an average degree of correctness of 0.49. The questions in all three question pools are distributed over the whole interval of difficulty 0.05–0.95.

Science studies

Type of response

Multiple choice questions in science occur about equally frequently in the three assessments. See figure 5.10. In PISA, 57% of the questions are either multiple choice or complex multiple choice (MC or CMC). This should be compared with 51 and 58% respectively in NU and TIMSS. NU has the test material in science that has the greatest proportion of complex multiple choice questions (32% CMC) where, for instance, it is a case of choosing which one of several statements is right or wrong. In all three studies, questions with open responses to be constructed by the pupil (OCR) constitute one third, or a little more, of the questions.

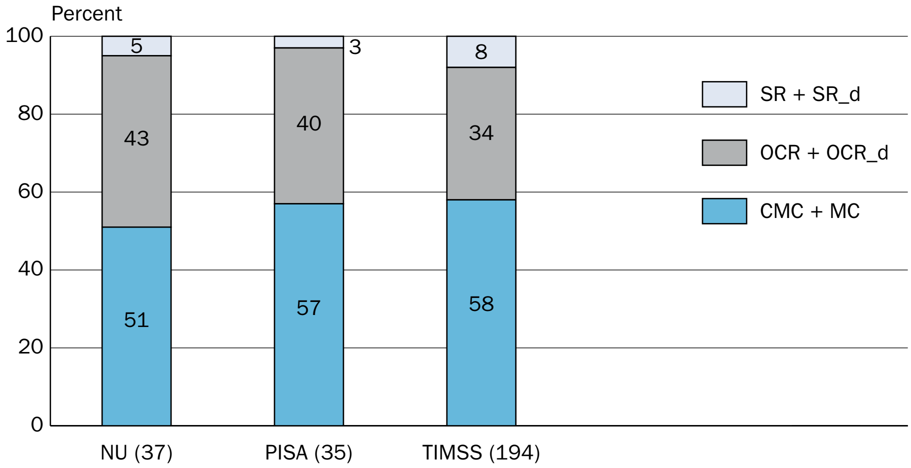


Figure 5.10 Distribution of science questions in NU-03, PISA 2003 and TIMSS 2003 based on type of response.

Connection with reality

As in mathematics, PISA has the greatest proportion of science questions that, to different degrees, are connected with reality. None of the 35 questions is classified as completely lacking connection to a world outside the subject theory or

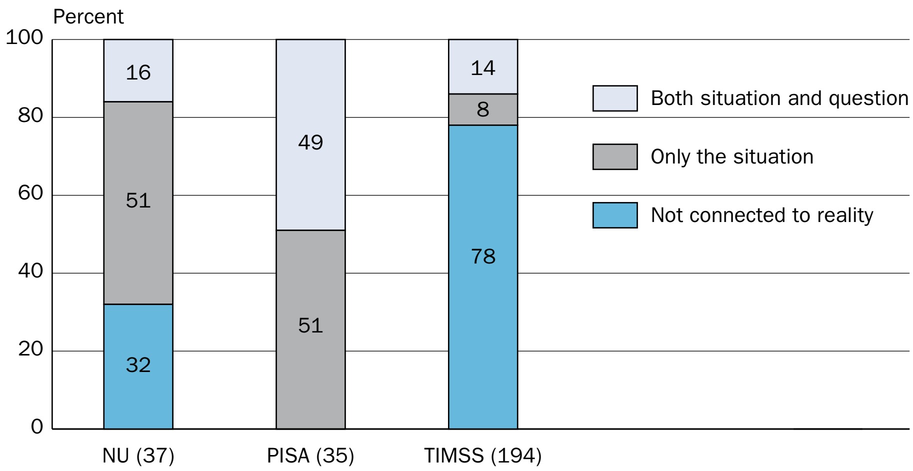


Figure 5.11 Distribution of science questions in NU-03, PISA 2003 and TIMSS 2003 based on assessed connection with reality and authenticity.

the laboratory (code 0). Just less than half of the questions have been assessed as being in a context and also as containing a problem considered to have possible authenticity. (See figure 5.11.) In NU, many of the questions are also connected to reality to some extent. (67% or 25 of 37 questions in total in science.) In TIMSS, on the other hand, 78% of the questions (152 of a total of 194) are of a kind not connected with reality.

Text and graphics

As in mathematics, there are great differences in the length of the texts used to present the science questions in the three sets of test material. TIMSS has the shortest texts while PISA is once again characterised by longer descriptions of context and problem. See figure 5.12.

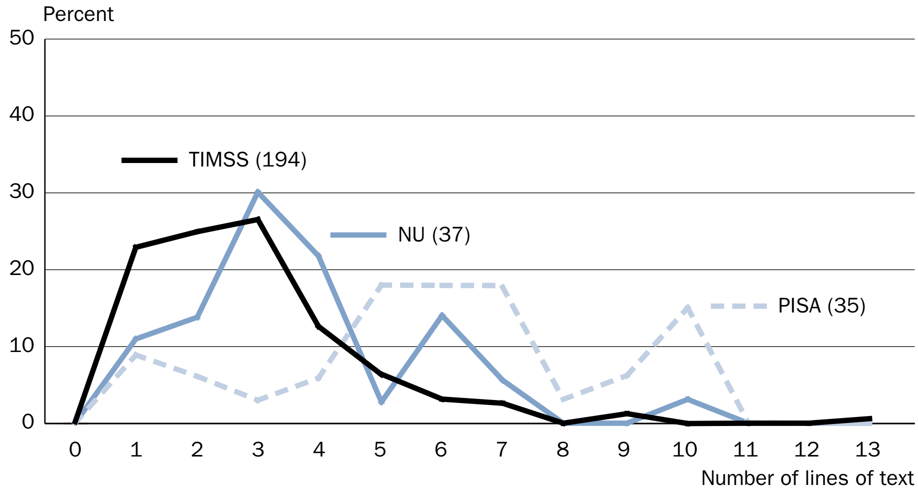


Figure 5.12 Distribution of questions in science in NU, PISA and TIMSS 2003 based on length of text in the question.

The mean values for NU and TIMSS are 2.8 and 3.7 lines, while a question in PISA usually has about double the length of text as in TIMSS, with a mean value of 5.8 lines. See table 5.4.

Table 5.4 Text length in science questions.

	Number of questions	Average number of lines
NU	37	3.7
PISA	35	5.8
TIMSS	194	2.8

As regards illustrations in the form of direct reproductions, tables and diagrams, the differences are not as marked between the three pools of questions. However, the variation between different graphic forms of expression is significantly greater in PISA than in NU and TIMSS. See table 5.5.

Table 5.5 Distribution of questions in science in NU03, PISA 2003 and TIMSS 2003 based on occurrence of pictures, tables and diagrams in the question.

		Illustration				
	Number	Decoration	Support	Necessary	Table	Diagram
NU	37	0%	11%	14%	0%	0%
PISA	35	6%	11%	9%	26%	9%
TIMSS	194	0%	7%	21%	9%	7%

Subject content

The question of the breadth of the three sets of test material, with regard to subject content in science, has been examined using the same method as in mathematics. The proportion of topics in the different content domains and topic areas that are represented in the respective test material by at least one question has been determined. The detailed data is summarised in figure 5.13 for the topics under the science syllabuses. It is apparent that, in PISA, biology has the broadest coverage (39%), while in NU that role is taken on by chemistry (50%). TIMSS has the broadest coverage of factual content in science, with almost complete coverage of the topics in the classification system. If one only examines the topics belonging, as per the syllabuses, to biology, physics and chemistry, TIMSS covers 96 per cent, NU 35 per cent and PISA 23 per cent of the number of topics.

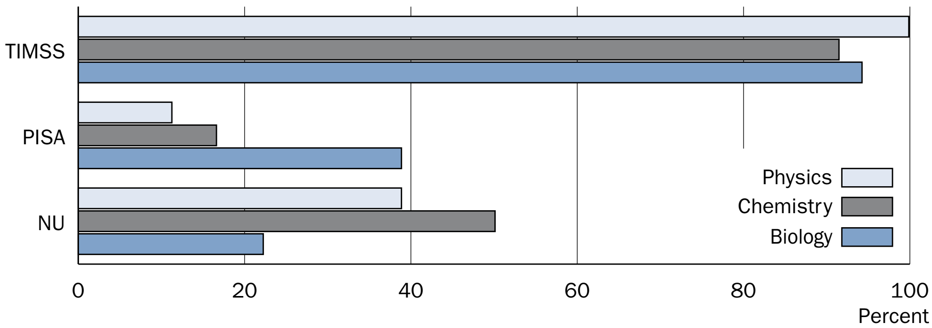


Figure 5.13 Relative coverage in NU03, PISA 2003 and TIMSS 2003 of the topics in the three science subjects.

Of the total of 68 topics, 52 correspond to the subject content that can be considered to be included in science under the Swedish syllabuses, including certain topics that, in the classification system, are part of earth science and environmental science. The remaining 16 topics that, for instance, are within the topic areas Earth processes and Population changes, have been deemed not to belong to the subjects biology, physics and chemistry. Certain individual questions in such areas, however, can be devoted to indicating goal achievement in respect of knowledge use in accordance with the third goal category in the science syllabuses (see below).

It has been possible to categorise the questions more or less unequivocally in TIMSS and NU in accordance with the chosen taxonomy for subject content in science. PISA includes science questions that could not be categorised so easily in terms of the scientific matter to which they refer. Knowledge of the scientific process, which in PISA has been defined as the second of the three dimensions introduced, can be tested without at the same time testing specific knowledge of concepts and phenomena within a related subject. Therefore, questions that in fact do not test factual knowledge have nevertheless been classified according to the content domain, topic area and topic to which the content of the question theoretically belongs. The reason is that, even if knowledge of facts is not presumed within the subject area in which the question has been placed, questions of this kind ought to be easier to solve if, during teaching, the pupil has come across the concepts and phenomena described in the question. A supplement to the categorisation has been made in such cases, that school-specific factual knowledge within the subject is not a condition for being able to answer the question. Identification of what questions should be deemed to belong to this special group has taken place with almost one hundred per cent agreement between the assessors, while the reliability of the following classification into topic areas and topics has been lower than for other questions (0.75).

Table 5.6 Occurrence of science questions in NU-03, PISA 2003 and TIMSS 2003 where only general knowledge is presumed to be able to answer the question.

		Requirement for answering the question	
	Total number of questions	specific factual knowledge	general knowledge only
NU	37	100%	0%
PISA	35	57%	43%
TIMSS	194	96%	4%

As shown in table 5.6, all of the questions in NU presume specific scientific factual knowledge while, in TIMSS, a few test questions (4 per cent) occur where this does not apply. In PISA, on the other hand, almost half (43 per cent) of the 35 questions are such where either the required facts are provided in the text and illustrations, or where the problem context is of a more general nature. An example of this is given in the figure below. This situation is dealt with further at a later stage in this section.

Since the coverage of subject content in science is most comprehensive in TIMSS, it could be expected that the syllabus's goals would also be covered best there. This was the case in mathematics. As regards science, however, there is a different situation both in relation to the design of the syllabus and the practice in the schools. Here, goals to aim for and goals to attain, cannot be grouped together on the basis of theoretical content as in mathematics,. Goals to aim for in the science subjects, and goals that the pupils should have attained by the end of the ninth school year, which have been divided here into three categories in the three different subjects, do not define content and sequencing in teaching in the same way as in mathematics. Throughout the material, goals to aim for are expressed very generally, with accompanying goals to attain that, to a varying degree, specify the topic and concretise the minimum level to be achieved during the teaching.

From what has already been said in general about the science syllabuses, it could be expected that selection of content to be studied and time planning vary between different schools, and that the textbooks and teaching aids in biology, physics and chemistry, in contrast to mathematics, do not define what

PISA S128 CLONING

Read the newspaper article and answer the questions that follow.

A copying machine for living beings?

Without any doubt, if there had been elections for animal of the year 1997, Dolly would have been the winner. Dolly is a Scottish sheep, that you see in the photo.

But Dolly is not just a simple sheep. She is a clone of another sheep. A clone means: a copy. Cloning means copying ”from a single master copy”. Scientists succeeded in creating a sheep (Dolly) that is identical to a sheep that funtioned as a “master copy”.

It was the Scottish scientist Ian Wilmut who designed the “copying machine” for a shep. He took a very small piece from the udder of an adult sheep (sheep 1).

From that small piece he removed the nucleus, then transferred the nucleus into the egg –cell of another (female) sheep (sheep 2). But first he removed from that egg-cell all the material that would have determined sheep 2 characteristics in a lamb produced from that egg-cell. Ian Wilmut implanted the manipulated eggcell of sheep 2 into yet another (female) sheep (sheep 3). Sheep 3 became pregnant and gave had a lamb: Dolly.

Some scientists that that within a few years it will be possible to clone people as well.. But many governments have already decided to forbid cloning of people by law.

Question Q1

Which sheep is Dolly identical to?

A. Sheep 1 B. Sheep 2 C. Sheep 3 D. Dolly’s father

Question Q2

In lines 11-12, the part of the udder that was used is described as “a very small piece”. From the article tex you can work out what is meant by “a very small piece”.

That “very small piece” is

A. a cell B. a gene C. a nucleus D. a chromosome

Question Q3

In the last sentence of the article it is stated that many governments have already decided to forbid cloning of humans by law. Two possible reasons for this decision are mentioned below. Are these reasons scientific reasons?

Circle either ”Yes” or ”No” for each.

Argument	Scientific?
Cloned people could be more sensitive to certain diseases than normal people.	Yes / No
People should not take over the role of a Creator.	Yes / No

Figure 5.14 Question from PISA 2003. The sub-questions have been classified 1.5.1 (... sexual or asexual reproduction ... the offspring’s similarity to the parents), 1.5.2 (... hereditary characteristics ... transfer of genetic material) and 1.5.1 with addition that Q1 and Q3 do not presuppose knowledge of type facts within these areas.

weight the subject content is given in the science teaching.[26] The analysis of textbooks[27] that has been performed correspondingly to the analysis for mathematics thus only provides indications as to whether it is probable that different

[26] The National Agency for Education (2002, 2003)

[27] E g Andréasson et al (2002)

areas are taught in normal schooling, but says nothing about the year in which this is done or about the scope given to the concepts or phenomena in question. A brief account of the result of this analysis of textbooks is that all topic areas and topics that occur in the system used for categorisation are normally dealt with in years 7–9 of compulsory school. The examination of the science questions carried out by the teacher group shows, moreover, that only isolated TIMSS science questions could be deemed to fall outside the subject topics of which the pupil group in year 8 should have command. The result of the teachers' questionnaire in TIMSS,[28] where questions about the subject areas covered were asked, contradicts the results of the reference group to a certain extent. Subject matter that received low values there, expressed in the form of proportion of students who have teachers who, up until the occasion in question in year 8, had dealt with a certain area, turns out only to have a possible, weak connection with the values of the degrees of correctness for questions within these areas.

Competencies

Table 5.7 Distribution of questions in science in NU03, PISA 2003 and TIMSS 2003 based on whether calculations and/or reasoning are required to answer the question.

	Number	Calculation	Reasoning
NU	37	0%	11%
PISA	35	0%	26%
TIMSS	194	2%	19%

Some isolated science questions did occur in TIMSS with calculations that could be decisive for whether the pupil who was responding was awarded any points for the question or not. There were no corresponding questions in NU or PISA. PISA had the highest proportion of questions that were deemed to indicate ability to reason as per the system for classification that was used. Every fourth question in PISA was deemed to be of that type, while the corresponding proportion in TIMSS was a fifth and, in NU, a tenth.

The three goal categories that express what the pupil should grasp concerning "... nature and Man", "... the scientific activity" and "... use of knowledge" were used for a supplementary classification of which competencies the science questions in the three assessments can measure. Goals under "nature and Man" (NV1) could be applied to the majority of the questions that, in the classification, partly could be ascribed a topic area and topic, and partly the characteristic that factual knowledge was presumed in order to be able to solve the question. If the question was intended to test understanding of the scientific method, it was classified as NV2 (the scientific activity) and, finally, if a question tested the ability to argue about scientific applications, it was classified as NV3 (use of knowledge). The latter categories were usually applicable to those questions that, in the general classification, were deemed to be such that specific factual knowledge was not presumed. In some isolated cases, however, a question was classified as belonging both to the NV1 category and to one of the other two goal categories.

[28] Mullis et al (2004)

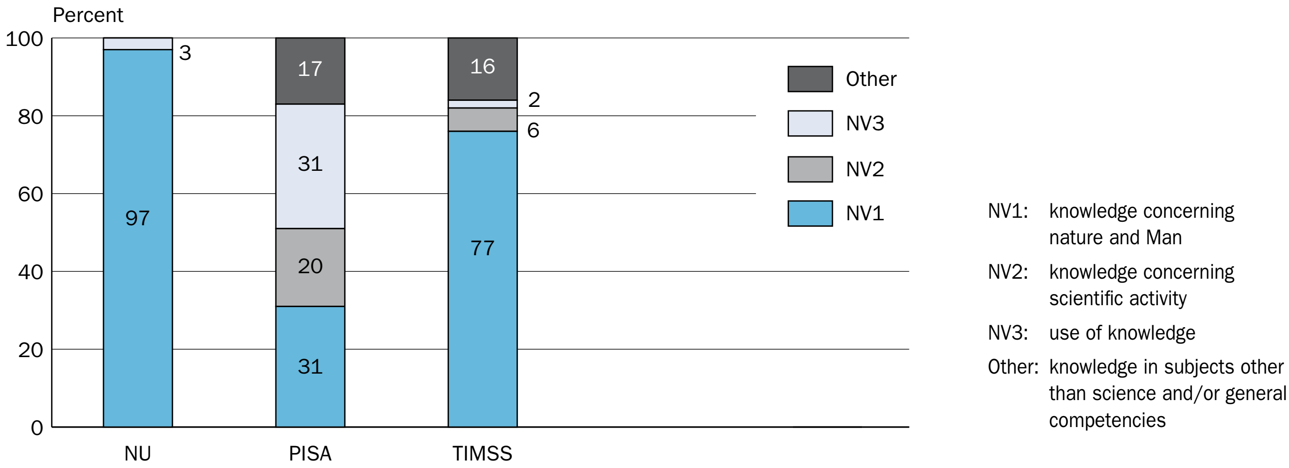


Figure 5.15 Proportion of science questions in NU03, PISA 2003 and TIMSS 2003 that were classified in the three goal categories.

The result is presented in figure 5.15. In both TIMSS and PISA, approximately one-sixth of the questions are outside the knowledge domain defined by the science study syllabuses. If, on the other hand, one looks at the distribution of the questions between the three goal categories, the picture is quite different. PISA occupies a place apart here, with the most even distribution between the three goals. One-fifth of the questions are such that are intended to indicate the attainment of goals within NV2, and almost one-third of the questions within the two other goal categories, NV2 and NV3. TIMSS still has a clear focus on goals that are in the NV1 category (nature and Man), but six per cent of the questions (11 questions) test understanding of the scientific method (one of the goals under NV2) and four questions are classified as NV3. NU does least well in the comparison, with no questions in category NV3, only one in category NV2, and the remainder of the 37 questions referring to scientific concepts, phenomena etc, that are classified as NV1.

Difficulty

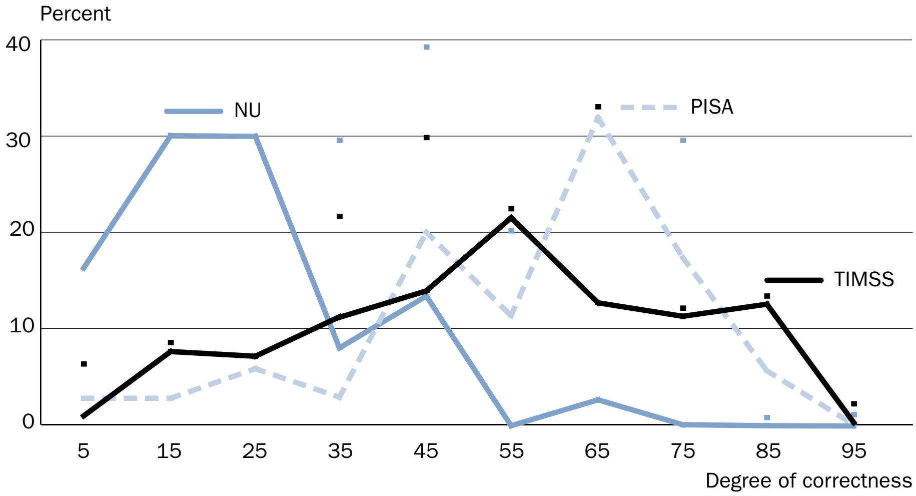


Figure 5.16 Difficulty of the science questions in NU-03, PISA 2003 and TIMSS 2003.

It is apparent from the reported data that the science questions in PISA had on average the highest degree of correctness (0.56) while the average in TIMSS was 0.52. It was slightly easier for the pupils in year 9 to get points for the questions in PISA than it was for the pupils in year 8 for the science questions in TIMSS. In both these assessments, questions occurred that had degrees of correctness down to 0.05 (see figure 5.14), which means that very few pupils solved these questions, but there are also questions with degrees of correctness close to 0.95, with almost all pupils solving them correctly. In NU, the average was 0.24, with only the odd, isolated question being solved by more than half of the pupils.

6 Discussion and conclusions

Some comments

Comments about the purpose

By way of introduction, it can be worth pointing out that the classification system used here was drawn up to allow comparisons between the instruments for knowledge assessment in NU-03, PISA 2003 and TIMSS 2003, but not to make general quality judgements of the three assessments. In mathematics and also in the science subjects there were clear differences with regard to the characteristics examined. Several of them can be derived from the differences in the basic ideas and intentions as well as the general preconditions on which the three projects are founded. It is almost certain that there are also differences and similarities in the design of the three test instruments that are not apparent from the picture that is presented of the attributes selected here. However, the respective purposes of both of the international studies, as with the national one, correspond to each other from a Swedish perspective. The three projects are to form the background data for an evaluation of goal fulfilment in school and of the knowledge development in mathematics and science as regards the Swedish pupils. In PISA and TIMSS, the comparison with other countries is also a central theme. The knowledge assessments should thus be able to serve as supporting data on which to base decisions regarding the development of compulsory school, with reference in particular to mathematics and science. This means that the material presented here about how and what the tests measure can result in completely different conclusions, depending on what idea one has about the school of the future.

Comments about mathematics in NU and science in PISA

As mathematics is a subject involving national tests, it has a unique standing in the school regarding the preconditions for an assessment of whether the teaching goals have been attained. Data can be gathered in year after year, and possibly form the basis for statements about goal attainment and trends. For the particular reasons that have been described briefly, data from the National Test in mathematics for year 9, spring term 2003, could not form a basis for the assessment and, for that reason, it has not been included in this comparative study. The extent to which a test system with the function of guidance regarding grades is suitable as an assessment instrument has often been questioned.[29] When designing a test system to be used for producing supporting material for statements about knowledge development, parallelism[30] in the instruments used is essential.

In NU, only the extra test that was unique to that survey has been analysed. In the reporting, that test formed the basis for statements about trends in knowledge development. The comparisons between the three assessments, NU, PISA and TIMSS, thus falter in respect of mathematics – the preconditions differ too much and the discussion as regards the mathematics test will mainly fo-

[29] Wedman (1988), Popham (1999)

[30] Crocker & Algina (1986)

cus on the two international assessments. In the case of science, however, there is no equivalent restriction and, in relevant cases, all three projects have been included in the discussion.

In respect of PISA, as already said, focus in the 2003 assessment is on mathematics. This means that science, in contrast to TIMSS, does not take up as much room as mathematics in the tests. In the approaching PISA 2006, it is science that instead receives the greatest attention. This report, however, only discusses the version of test material used in the study that has already been performed.

Comments about statistical strength

With regard to the selection of schools and disposition of pupils in respect of time and number, the stipulations for the three surveys have been the same, and the preconditions to that extent are the same for all three. PISA is alone in referring to individual pupils in the second phase of the selection process. The random selection of whole teaching groups, as in TIMSS, resulted in simpler administration[31] for implementation out in the schools. With weighting, that method produces the same statistical material as the previously mentioned procedure for commenting on the whole population (e.g. all pupils in a certain age group). The fact that a randomly selected teaching group (e.g. class) does not normally give a selection that is representative for the school means that the method involves limitations. It is a case of making statements about entire schools, so the possibility of making certain types of analysis is thus restricted. In TIMSS as, to a certain extent, in NU, the simpler administrative situation has been used in carrying out a teacher's questionnaire, so that data for individual pupils can be linked with data for those particular pupils' teachers. It has not be practicable to do this in PISA as all of the classes with 15-year olds in the chosen school are normally involved, meaning that a large number of teachers would need to answer the questionnaire in order to match the pupil selection. It can be said about PISA as well as TIMSS that both the statistical strength in the data and, thus, the ability to generalise in respect of the populations in question (all of Sweden's 15-year old pupils and pupils in year 8 respectively) fulfil strict requirements,[32] while a corresponding assessment cannot be presented here for NU on the basis of the information available in the reports.

Frameworks

Expectations

The situation for NU-03, in contrast to the international projects, is that the Swedish curriculum and syllabuses partly constitute the basis for this study and are partly assumed to be the starting point for the construction of the test material. As has already been said, TIMSS should as its starting point have the goal descriptions that have been agreed between many nations, while PISA's should be based on guidelines drawn up by the OECD. It should be possible for the

[31] Such as in the performance of the test session in already established groups in the school. It can, however, be noted that the dropout in TIMSS is not, with this simpler organisation, lower than in PISA. That survey was performed where, in one session, pupils came from different classes, thus requiring more administrative work on the part of the school administration.

[32] See, e.g. Martin et al. (2003)

expectations, that tests should be adapted for Swedish governing documents, to be higher for a national assessment than for the two international surveys in which the frameworks result from compromises between the countries.

Mathematics for everyday life and for study

Mathematics can be said, in general, to be a subject with a collection of knowledge about basic procedures and definitions of concepts that, in the main, are independent of national characteristics. This situation was one of the reasons for mathematics being the subject on which the IEA focused its first international, comparative study.[33] From a cognitive point of view, there is a good level of agreement between the three frameworks compared. The syllabus's descriptions of the subject matter have, in general, equivalents in the framework in PISA as well as in TIMSS, with only a few topics that are difficult to suppose can be included in a normal interpretation of the Swedish governing documents. On the other hand, there is no topic in the interpretation that the chosen textbook represents that is not included in the classification system based on TIMSS's framework.

Everyday mathematics is a concept that has occurred frequently in the didactic interpretations and discussions that have taken place on the question of the orientation that mathematics should have in compulsory school. This agrees with the weight PISA attaches to the problems in the test questions occurring in situations or contexts that can be felt to occur in the pupil's present situation or in the pupil's perspective of their future life. On the other hand, the syllabus says that the teaching should provide a good foundation for further education and lifelong learning, which presupposes, for instance, knowledge and proficiency that continued studying of mathematics requires. There, TIMSS can be said to provide a clearer intra-mathematical orientation of what should be assessed that also corresponds to the list of mathematical skills included in the syllabus's goals to aim for and goals to attain.

Subject matter in science studies

The subject matter in science in the different countries' school systems is not as distinctly given as in mathematics, where the demarcations and specialisation of the subject are more obviously determined by national conventions and preferences. It can thus be expected that the science syllabuses, as compared with mathematics, will show fewer similarities to the frameworks in science in PISA and TIMSS. As the definition of the school subject is not obvious, it can possibly be expected that TIMSS, with its comprehensive, detailed specification of concepts and phenomena, will run the greatest risk of falling outside the Swedish demarcation of the subject. However, it has been shown that science in TIMSS agrees, in the main, with a Swedish interpretation of normal subject content. In PISA, with its limited specification of subject topics, the structuring of the different competencies aimed for is more in agreement with the Swedish syllabuses in the three science subjects. It is perhaps possible to claim that the latter is a more important aspect in an evaluation and that this is completely in line with a basic philosophy behind the design of syllabuses in Sweden. The selection of subject matter in science, according to this opinion, can be extremely

---

33 Husén (1967).

free without damaging the quality of the science teaching as long as the pupils are given the conditions for attaining the goal to aim for concerning nature and Man, "that they believe in and develop their ability to see patterns and structures which make the world understandable, as well as strengthen this ability through oral, written and investigatory activities."

Test questions' technical format

More to read in PISA?

The result of the comparison of the test questions' outer characteristics demonstrated, in some cases, considerable differences between NU, PISA and TIMSS. The most striking difference relates to the stated volume of text per question where, in mathematics as well as science, the volume of text on average is clearly greater in PISA than in NU and TIMSS.

When making the comparison, the total volume of text the pupil has to read during the course of a test can also be taken into account. On the basis of an estimation of the time a pupil spends on mathematics and science in PISA, a comparative figure of two lines of text per minute is arrived at.[34] What is surprising is that, in TIMSS, the same comparative figure is reached. Correspondingly, it can be discussed whether the observation about the differing occurrence per question of illustrations, diagrams and tables is actually relevant. The requirements on the pupil to gather information, provided in different forms, within a certain time are quite similar in the two international knowledge assessments. The justification in concluding that reading ability influences the results to a greater extent in PISA than in TIMSS can, in other words, be questioned. In NU, the volume of text per minute is markedly less, meaning that the result a pupil achieves in that study should not be as dependent as in the other two studies on the ability to quickly assimilate the information presented in the question in text and other forms.

Multiple choice questions

On studying the whole test that the pupil performs, TIMSS has the greatest proportion of multiple choice questions, even if PISA and NU have equally large shares of that type of question in science as TIMSS. The use of multiple choice questions means advantages in the administration of tests and examinations, but the issue is usually called into question of whether multiple choice questions are suitable for testing more complex forms of knowledge other than remembering and recognising facts, concepts and procedures. A closer study of the mathematics questions in both PISA and TIMSS shows that approximately 20 per cent of the questions were deemed to test reasoning. Almost half of these (nine and eight percentage points respectively) were multiple choice. In comparison, the number of short response questions in PISA made up 41 per cent

---

[34] The composition of the test booklets in the two international studies differ both in respect of the projects and also in respect of the separate test booklets within a project. An estimation with regard to PISA, supported by the mean values, is that, in addition to two-thirds mathematics and science, with an average of 160 lines of text, a booklet also included questions involving reading or problem solving. Of the whole test period of two hours, it is estimated that two-thirds, i.e. 80 minutes, are spent on mathematics and science, giving a comparative figure of two lines per minute. As TIMSS has approximately double the number of questions in mathematics and science per booklet for a 90 minute test period, the same comparative figure is arrived at.

(35 questions) of the total number of questions (86) but, of these, only 2% (2 questions) were classified as reasoning. The conclusion is that in both PISA and TIMSS an effort has been made to include multiple choice questions that test complex forms of knowledge as well.

The advantages, in relation to scoring, are obvious, but the most significant benefit is in the tests being able to contain a large number of questions if the person taking the test only has to choose between a selection of prepared answers. The short time it takes to produce the answer means that a pupil has time to answer several questions within a limited time. This is particularly evident in TIMSS, where the fact that multiple choice questions have been used to a greater extent means there are more questions within each knowledge domain, resulting in better coverage of the content and statistical reliability.

Among the disadvantages involved in multiple choice questions, the problems of guessing can be mentioned. It is always risky to claim, on the basis of the result of a question, that specific knowledge goals have been attained, particularly if such a claim is based on the result of a single, more difficult multiple choice question in an area with proportions of correct responses that are close to the guessing probability. The solution is to increase the number of questions intended to indicate the extent to which the examined goal has been attained, which can be difficult in a study such as NU, where the number of questions is limited to only one set of questions used for all pupils. Another aspect is that there can be specific gender differences in respect of the tendency to guess the answer to a multiple choice question, an issue that is discussed in penetrating analyses of TIMSS 2003.[35] There, it can be seen that boys have a higher degree of correctness in multiple choice questions than girls, if a comparison is made of pupils who are otherwise of the same ability. This indicates the possibility of a distortion of knowledge assessment results obtained with the help of multiple choice questions only. Another comment is that the use of this type of question can hardly be said to be in line with the overall goal in the curriculum, that the pupil should be able "to express ideas and thoughts orally and in writing". Taken together, these doubts constitute factors that must be considered in a discussion about the validity of the assessments examined here. A point of view that is often put forward in a didactic context, internationally as well as in Sweden, is that the trend in assessments should be towards using a greater proportion of questions that require the pupil's own constructed response.

Difficulty

There can be many different purposes for performing an assessment of pupils' knowledge, and the assessment norms will differ too, depending on what is being examined. If the intention is to report results on the basis of a statistical norm and based on results of the whole test, it is of course an advantage to have several questions in each of the domains to be studied. There should be a distribution of degree of correctness across the whole scale, but the majority of the questions should lie in the middle of the interval zero to one. It is possible to identify that ambition in PISA and TIMSS in both subject areas, and also in mathematics in NU. In the biology, physics and chemistry tests in the national assessment, the measured degree of difficulty, however, is significantly

[35] Eriksson (2005)

higher and, in the majority of questions, a comparatively small proportion have responses that are given a pass mark in accordance with the interpretations of the syllabus determined by researchers. A detailed analysis of pupils' answers can certainly provide valuable didactic information, but the question is whether the main purpose of the assessment is thus achieved. Too few pupils get the opportunity to demonstrate their knowledge in a test that is too difficult if the result of such an assessment is to provide information about the whole pupil group's level of knowledge.

Assessments in line with syllabuses

Solutions to the problems

As has already been touched on, it can be thought that, in the national assessment, it is easier to design tests and questionnaires that conform to the national frameworks, when it is only necessary to keep in line with the relevant Swedish goal formulations. The multinationally based assessments are designed by a body that includes only one Swedish voice among many from the other participating countries. No one country has ideal agreement between their national syllabus and the framework for the assessment.[36] Most of the national requests for the inclusion of specific topics are met, while the reverse does not always apply. Most countries find that the framework contains one or some topics within the subject that are not included in the national syllabuses for the pupil age groups in question.

Similar problems can also occur in a study of only national scope. If the syllabus changes between the assessment occasions, the same technical problems can occur in the national assessment as in PISA and TIMSS. The instruments should be designed so that measurement is made against relevant goals for the two separate test occasions and using the same, absolute scale as provides the opportunity to draw relevant conclusions about knowledge development.

In NU, the problem was solved with the scale in science[37] by using the same test as in 1992. Thus the possibility of allowing the new syllabuses to influence the content and design of the tests was rejected. This put the situation on the same footing as the conditions in PISA and TIMSS where, in mathematics and in science, the target area is consciously allowed to lie partly outside the national goals but where, through the breadth of the coverage and use of many questions, there is less sensitivity regarding shifting of the weight for different target areas.

In PISA and TIMSS, some of the questions are also the same for the different assessment occasions; these are what are known as the anchor questions. These, however, constitute a relatively small proportion of the total number and can be chosen so that they lie within areas that are central to the subject. Reference points are thus obtained that can be used to calibrate the scale for the new assessment. The rest of the questions are new and this, together with the fact that the total number of questions used is large, means that the risk of the testees having had the questions previously is eliminated. In the national assessment, in the 2003 evaluation there was no awareness in the participating schools that

[36] In TIMSS, this is examined in a special study called Test Curriculum Matching Analysis. See, e.g. Mullis et al (2004).

[37] In mathematics, only the "one point test" analysed here could be used for the reasons mentioned initially.

the relatively few questions would be the same as in 1992, and the probability is minimal that any pupil would have been prepared for them. This will, however, not be the case in a future assessment if the same conditions are given. The risk cannot be ignored of the result being affected by, for instance, a teacher adapting their teaching on the basis of information about what a specific test has contained in the past.

Validity

The question of the validity of the international assessments has been dealt with in a number of reports and papers. Different aspects of validity are examined and various methodical procedures are used. In the US, several studies have been reported that include comparisons between different issues of TIMSS, PISA and the national assessment NAEP.[38] In certain respects, several of these studies correspond to parts of this particular study, with comparisons of technical data such as the occurrence of different types of questions, subject content etc.[39] One essential difference is that, in the case of the USA, it has not been possible to make a comparison with a national curriculum, as the latter does not exist there.

In a Danish study, the correlation between school leaving certificate grades and results in PISA 2003 was examined.[40] A proven correlation was taken as justification for both tests, to a great extent, assessing the same targeted competencies despite differences in starting points and goals between the national school-leaving test and PISA. There are also technical preconditions for similar studies in the Swedish data, as those participating in the national tests in mathematics are registered with their personal identity numbers. This was also the case in PISA 2003 for about half of the pupils taking part. The decisive part of a discussion about the validity of the tests under comparison is, however, the result of an accompanying survey of the link with the syllabus document.

In the international main reports in TIMSS a study[41] is presented that examines whether the tests are in line with the mathematics and science courses of the different countries. Each country has initially had to report which of the questions in mathematics and science that can be said to relate to the course that the country's pupils have gone through up until the time of the test. Thereafter, for every country, the mean values of the average performances in each one of the national courses so defined are calculated. Sweden's result as regards questions that were deemed to be covered by the implemented syllabus were, for example, two hundredths greater than the mean value of 0.47 that was obtained for all the questions in TIMSS. The conclusion that was arrived at in the report was that the relative order between the participating countries' average performances was only marginally affected when different countries' selection of questions was used as a basis for calculating the average performance. A positive interpretation of this claim is that TIMSS keeps itself within central parts of the courses in question.

A difference noticed between the result of different assessments is interesting from different points of view and could, in this context, provide a basis for test-

[38] National Assessment of Educational Progress

[39] McLaughlin et al (1997), Nohara et al (2001), Scott et al (2005), Ginsberg et al (2005)

[40] Mejding et al (2006)

[41] Test-Curriculum Matching Analysis (TCMA). See Appendix C in Mullis et al (2004).

ing the relevance of the indicators of validity that are under discussion here. The three assessments examined here, however, are reported on different scales, making a direct comparison impossible. For PISA and TIMSS it is nevertheless possible to use the reference system that the comparison between other countries involves. We assume that the 2003 assessments, which were performed mainly with pupils in year 8 (TIMSS) and year 9 (15-year olds in PISA), can be placed on an equal footing with two assessments at different times but with the same population. In both cases we expect that the results give a general indication of effectiveness as regards the mathematics and science teaching in compulsory school in the participating countries.

Tabell 6.1 Average Scale scores in mathematics and science in PISA 2003 and TIMSS 2003 (8th grade).

	PISA Math.			PISA Science			TIMSS Math.			TIMSS Science		
Countries	Average score	S.E.	Ranking	Average score	S.E.	Ranking	Average score	S.E.	Ranking	Average score	S.E.	Ranking
Australia	524	2,1	5	525	2,1	4	505	4,6	9	527	3,8	6
Hong Kong, SAR	550	4,1	1	539	4,3	2	586	3,9	2	556	3,1	2
Indonesia	360	3,9	16	395	3,2	16	411	2,4	16	420	4,1	16
Italy	466	3,1	14	486	3,1	13	484	3,2	13	491	3,1	14
Japan	534	4,1	4	548	4,1	1	570	2,1	3	552	1,7	3
Latvia	483	3,7	11	489	3,9	11	508	3,2	6	512	2,6	12
Netherlands	538	3,1	3	524	2,1	5	536	3,8	4	536	3,1	5
Norway	495	2,4	9	484	2,9	14	461	2,5	15	494	2,2	13
New Zeeland	523	2,1	6	521	2,4	6	494	3,3	12	519	4,9	9
Russian federation	468	4,2	13	489	4,1	11	508	3,7	6	514	3,7	11
Serbia	437	3,8	15	436	3,5	15	477	2,6	14	468	2,7	15
Slovak Republic	498	3,3	8	495	3,7	9	508	3,3	6	517	3,2	10
Sweden	**509**	**2,6**	**7**	**506**	**2,7**	**7**	**499**	**2,2**	**11**	**524**	**1,6**	**8**
Korea, Rep. of	542	3,2	2	538	3,5	3	589	2,2	1	558	1,6	1
Tunisa	359	2,3	17	385	2,6	17	410	2,2	17	404	2,1	17
Hungary	490	2,8	10	503	2,8	8	529	3,2	5	543	2,8	4
United States	483	2,9	11	491	3,1	10	504	3,3	10	527	3,1	6

The average performances in mathematics and science for the countries that took part in both of the international assessments of compulsory school in 2003 are presented in table 6.1. The table also states the ranking to which the results correspond in lists in order of size of the average scale scores for the two subjects. As regards science, it can be seen that the difference is relatively small between the ranking a country gets in PISA and in TIMSS (usually zero or one step, and four steps at most). One interpretation of the relatively strong link between the two measurements of the pupils' performances can be that the base of good factual knowledge in science that, according to TIMSS is attained in certain countries, is a condition for the same countries' pupils succeeding in demonstrating the general competencies that are tested in the questions in PISA.

A corresponding comparison in mathematics shows a greater variation (up to seven steps) between the countries' placings in the two assessments. This is not what one would expect on the basis of what has already been discussed here

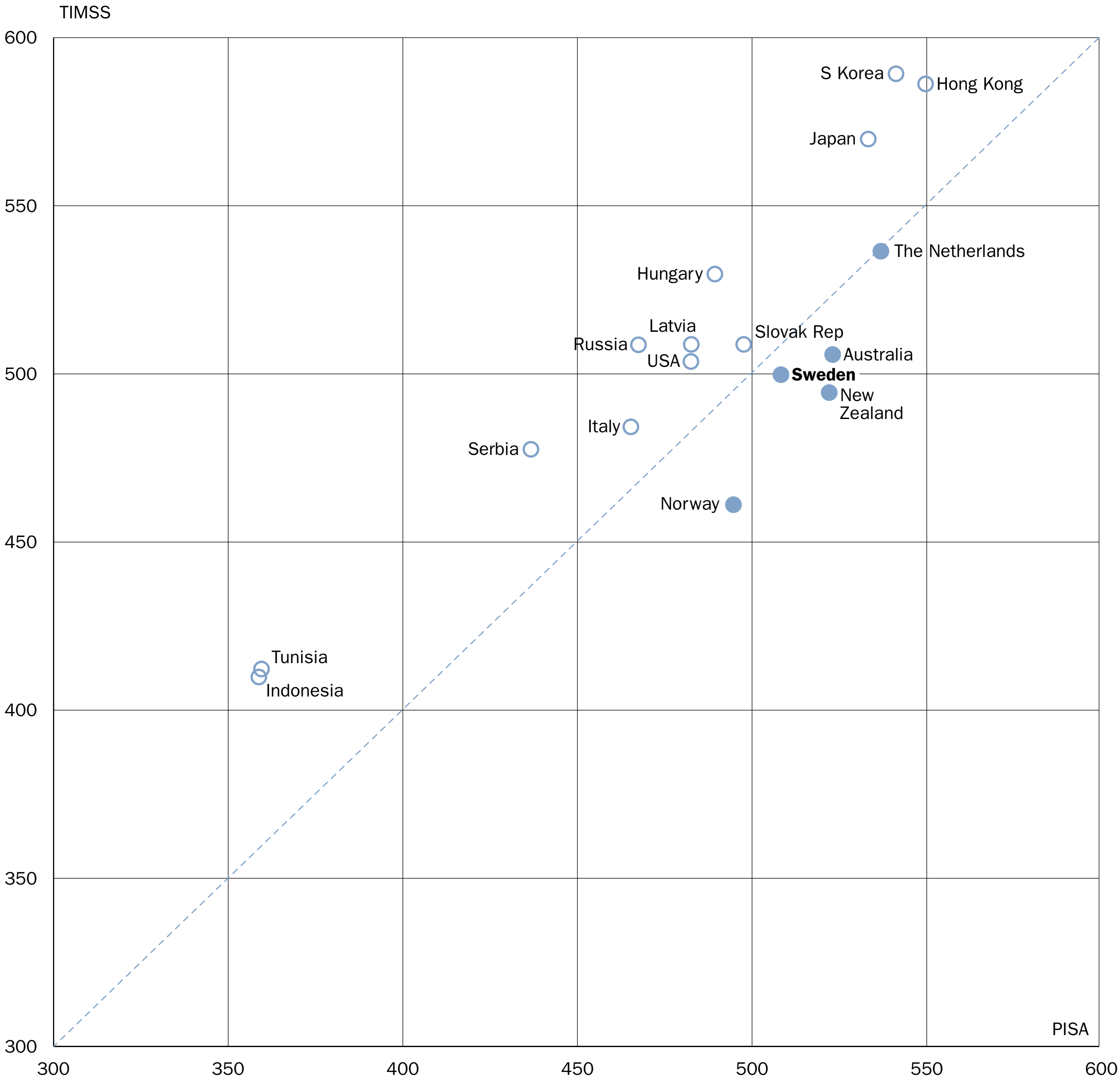


Figure 6.1 Mean values for pupils' performances in mathematics in PISA 2003 and TIMSS 2003 for those countries that participated in both studies.

about the greater differences in focus on different knowledge dimensions in science compared to mathematics that were noted as occurring between PISA and TIMSS. It could thus possibly have been expected in mathematics that PISA and TIMSS would produce more similar results so that a country with a higher average performance than another in the one assessment would also have a higher one in the other. The reason for this would be that the properties of the questions in the two question pools in mathematics are more similar than they are in science.

The apparently weaker link between ability in mathematics for a country's pupils as measured in PISA and in TIMSS is clarified in figure 6.1. Australia, Norway, New Zealand and the Netherlands, together with Sweden, comprise a group that tends to perform better in PISA than in TIMSS. Russia is one example of a country where the opposite applies. The pupils there, in comparison with Sweden, had a slightly higher average performance in mathematics in TIMSS but had significantly poorer results in PISA. Different explanations for these results have been put forward. PISA's emphasis on applications in society has an equivalent in Swedish methodical discussion while, in many countries, a higher value seems to be placed on formal mathematics.

According to this present study, however, as far as mathematics is concerned, there was not much support for this description of differences between the question pools examined. The coverage of the subject content, however, was noted to be different in PISA and TIMSS. Instead, the hypothesis can be put forward that the weight given to different subject content plays a significant role for the combined result and can, for instance, give different orders of ranking of the countries' results.

The classification of the questions that has been carried out in this study can be used here to test the hypothesis. A new value for the countries' average performance can be calculated by only using data for the questions within the topics in mathematics that are represented in both PISA and TIMSS. As supporting data in TIMSS for such a comparison, 100 questions of a total of 194 fulfil the condition that has been set up. In reverse, the majority remained in PISA, i.e. 73 questions of a total of 86.

A detailed picture of the two countries' corrected results is obtained from profiles for a country that show the results divided up into mathematics' main areas M1–8 in the syllabus. See figure 6.2, where the degrees of correctness are presented for those questions that remained in the different main areas. It can be seen there that the statistics area has a greater significance for the end result in TIMSS, while the effect of the results in algebra and geometry is reduced in comparison with the result for the original composition of the tests.

If we compare the results thereafter, we find the same difference in ranking between Sweden and Russia in TIMSS and in PISA. Sweden had a better placing than Russia in both of the assessments. The result is in agreement with the observation that both sets of test material were similar from the cognitive angle, and this can be taken as a pretext for the classification having served the purpose

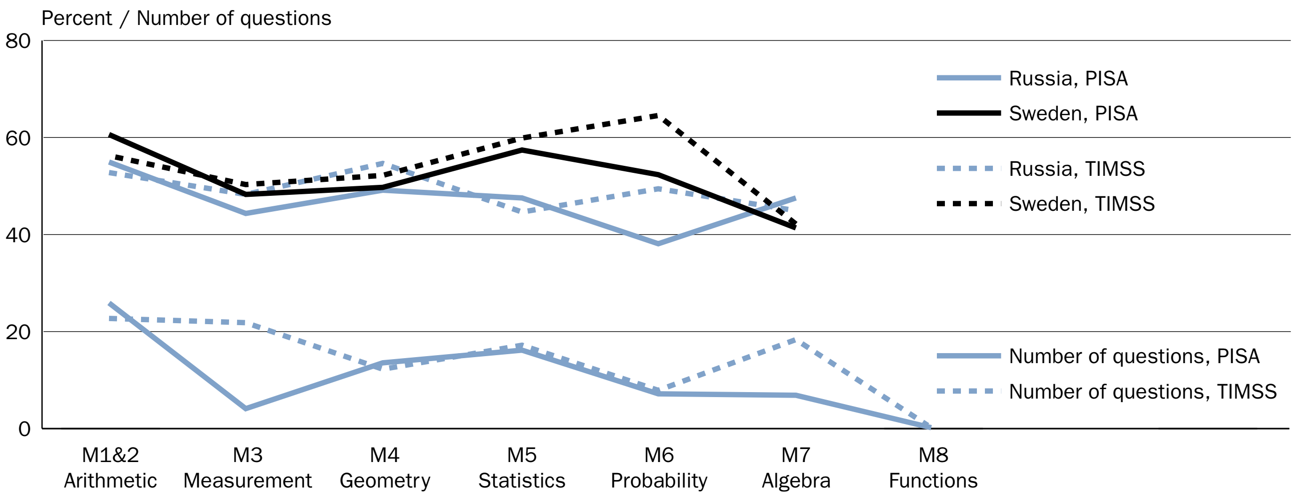


Figure 6.2 Number of questions and average performances within the main areas in mathematics for the selection of questions related to topics that occur both in PISA 2003 and in TIMSS 2003 for Russia and Sweden.

of providing a relevant basis for a validity assessment. Another conclusion is that the total mean value for degrees of correctness should not be treated without criticism. It constitutes a total measurement of the country's pupils' performance in a test with a certain composition. The design of the test must be problematised and data must be explained from different perspectives.

Formative evaluation in summative tests

The different roles that tests and assessment have in the teaching are summarised by classifying the degree to which a summative or a formative evaluation is involved. An external evaluation has mainly a summative function, meaning that it is intended to be used to decide how one group performs in relation to all others and/or the degree to which the testees have fulfilled certain established requirements. The tests that the teacher uses locally are or should be of a more formative character; in other words they should constitute an integral part of the teaching, guiding the teacher and reinforcing the pupil's learning. The supporting documentation for evaluation that NU, PISA and TIMSS produce is usually used in the first role, fully in accordance with the purpose of these tests. As is the case for all tests, however, these tests also have the potential to work formatively and on a national level. In that case it is a question of how the data are used to obtain different information, and how this information is fed back into the system.

The three natural science components in NU-03, both in the design of the tests and in the design of the reporting, have been regarded here as demonstrating certain limitations in respect of delivering a broad evaluation of how the national goals are attained. The data that has been produced has instead provided the basis for a didactically inspired analysis with more specific conclusions as to how the teaching should be performed in order to achieve an understanding of concepts etc.[42] There are technical preconditions both in PISA and in TIMSS for similar analyses based on the classification of how pupils answer the open constructed response questions and the method of using a two-digit code for registration.[43] Neither in the international nor in the national reports in PISA and TIMSS has there been any pursual of the analysis of pupil responses that, apart from the number of points that the response is worth, have been registered with a code showing which category of different acceptable or non-acceptable responses the pupil has presented.

Figure 6.3 shows a coding key in its simplest form for a question in TIMSS with a response that the pupil should construct. With that assessment instruction, possibilities are indicated for developing the analysis of data that already exists, technically, in PISA and TIMSS 2003. The open constructed responses with accompanying assessment instructions that are included in the international studies have not, however, been designed with a view to the possibility of deeper studies into, for instance, the understanding of concepts. PISA and TIMSS, however, offer possibilities for national supplements in the assessment instructions, completely within the normal framework, and also special arrangements with a national choice of questions. With a marginal increase in the selection of pupils, supplements in the form of a national test booklet and some

[42] Andersson (2005)

[43] Lie et al (1996)

extra questions in the pupil questionnaire can be permitted without affecting the quality of the international picture of results. In that way, the comparative international assessments could also include specific, national surveys of the type performed in the science section of NU.

TIMSS M032233

Code	Response
	Correct response
20	48% with calculations presented
	Part response
10	24 girls
11	Acceptable method but error in calculation
12	48% without calculations presented
	Incorrect response
00	50 %
09	Other incorrect responses (including crossing out / erasures / haphazard notes / illegible or outside the subject)

Figure 6.3 Assessment instruction for a two-point question of the open constructed response type in TIMSS.

Robust evaluation instruments

The syllabuses for the science subjects are under discussion in most European countries.[44] The debate refers to a variety of issues such as time plans, recommended or imposed methodology etc. A trend can be identified, that an increasing number of nations, like Sweden, have chosen or plan to choose not to control the subject content in detail but rather to design syllabuses that define competencies or general abilities to be aimed for in the pupils' learning in the science subjects. From the comparison between NU, PISA and TIMSS, it has emerged that PISA is the science studies assessment that has been regarded as being most clearly in line with such thinking, which seems to be characteristic of our time. The knowledge assessment in PISA is less sensitive to differences between choice of subject content in the courses within one country as well as between countries. NU and TIMSS have been directed more at specific factual knowledge, but also at the understanding of concepts in the question of scientific phenomena and processes. The broader coverage of subject content in TIMSS ought nevertheless to provide an evaluation that, for the most part, is in line with normal practice in the teaching of science in schools. The conclusion is that a more all-round evaluation of a teaching system should contain both of the main components that PISA, TIMSS or NU respectively represent.

As regards mathematics, there is an ongoing international didactic debate about school, with different answers to the questions about why, what and how. Polarisation is much more obvious in the mathematics than the scientific community, between advocates of the reformation of syllabuses and followers of the more traditional kind.[45] In the US, the concept The Math Wars has been minted for describing the conflict caused by the proposals for syllabuses that were presented by NCTM[46] at the beginning of the 1990s. There, the general com-

[44] Eurydice (2006)

[45] Schoenfeld (2004)

[46] National Council of Teachers of Mathematics (2006)

petency within the areas of understanding of concept, problem solving, modelling and communication was held up as being a core goal in mathematics. This necessitated a toning down of the training of the skill of mastering mathematics' algorithms that is linked with traditional syllabuses. The influences of NCTM's "standards" in the Swedish mathematics syllabuses are clear, but normal practice in our country corresponds more to a balanced point of view, somewhere between a traditional and a reformist attitude.

Conclusions

The frameworks in PISA and TIMSS agree, both in their main features and in detail, with aims stated in the Swedish governing documents. The difference, clarified in the frameworks, between the goals for PISA and TIMSS emerges also as a result of the analysis of the test material. The difference is most apparent in science. PISA aims at the processes while TIMSS aims at facts. Accordingly it can be said that PISA 2003 and TIMSS 2003 complement each other as regards the evaluation of whether the goals of the syllabus are achieved in Swedish schooling.

Future assessments

On the basis of the results and conclusions that have been presented here, the following are proposals for approaches to future assessments.

Participate in assessments

- that give all pupils the chance to prove that they have learned something within the subject areas studied.
- that give high validity by covering as many aspects as possible of and being in line with the national goals for the teaching.
- that are robust in that they tolerate changes in the goals being assessed without losing out on reliability.
- that give international reference points for school assessment.
- that guarantee statistical competence for selection, weighting and scaling and that thus meet strict requirements for generalisability.
- that are such that the executive consortia for the assessments permit and support national choices in questionnaires and tests.

In connection with the work on the international assessments, conduct work with

- recurrent analysis of the validity of the test instruments used.
- focused, in-depth analyses of results, against background of curricula and syllabuses.
- continuous discussions with those responsible for development and implementation of governing documents.
- a forward-thinking contact with the research and teacher-training communities, with discussions about didactic questions that should be answered with the help of the evaluation instruments.

7 The Agency's assessment

Sweden is at present participating in a number of international studies (PISA, PIRLS, TIMSS years 4 and 8, TIMSS advanced and ICCS). These are controlled, to a great extent, by international organisations and consortia. The possibilities for Sweden to exert any influence on the design of such studies are naturally limited. It is at the same time important from a national perspective to be able to assess pupils' knowledge and knowledge development. An alternative to participation in international studies is that Sweden conducts its own, national survey and this has taken place through, for instance, the national assessment of 2003 (NU 03).

The realization of studies at a national level means that they can be adapted to Swedish circumstances and that the results can be clearly related to the goals for Swedish schools. This is one advantage of nationally designed studies. Disadvantages are that the costs for the development of instruments and the analyses have to be borne by Sweden alone, and that the results cannot be compared with those of other countries. This favours participation in the international studies. On the other hand, there is a risk that the Swedish goals will be secondary to the goals and frameworks formulated for the international studies and that the Swedish goals will be neglected.

One possible alternative could be that the result of the national test be used to assess knowledge level and knowledge development. A closer examination of results of different national tests shows, however, that they do not suit such tasks particularly well. This is because, quite simply, they are not constructed for such purposes but rather to support teachers in their interpretation of course goals and grading criteria (if grades are to be awarded), thus promoting fair, impartial grading. The choice is between specially developed, national instruments or participation in international studies, or a well-balanced combination of both. The international studies also provide possibilities for national supplements both with regard to background information as well as cognitive aspects.

Yet another important aspect to consider when weighing participating in international studies against the performance of national assessments is the international studies' higher level of statistical strength and higher degree of generalisation, due to the rotating design that is characteristic of them. Such design can theoretically be transferred to national assessments, of course, but due to the extremely high costs involved in such design, it would in reality not be possible for an individual country to develop sophisticated studies of that type.

So, how well do the frameworks for the international studies agree with the Swedish curricula and syllabuses? Do we risk losing our distinctive national character? Or are our pupils' results assessed on incorrect grounds, such as that they have never been given the opportunity in school of learning what is asked for in the different studies because Swedish goals differ from what is being tested?

To find answers to such questions, the types of analyses presented in this report are important in helping us to understand and interpret the results from the national as well as the international studies. The analyses are also of decisive significance as regards the decisions that should be made in respect of participation in different international studies and also as regards any planning of national assessments.

The results presented here for mathematics and science are interesting in many ways and lead to the following conclusions:

Firstly, there seems to be a fairly high degree of international consensus – Sweden included – as to what a good level of knowledge/ability in mathematics and science means. This implies that international studies in mathematics and science can be said to have high relevance to the extent that their frameworks as well as their concretisation in the form of questions give a good reflection of the Swedish syllabuses.

In the discussion regarding international studies' possible effect on national governing documents, misgivings have been expressed about the risks of international organisations taking over, formulating, in the future, a kind of global syllabus and curriculum. The results show that the concordance between Swedish syllabuses and their concretisation in national tests and (modern) international frameworks is rather good. This is not particularly surprising considering that we live in a time of globalisation and internationalisation. Nonetheless, the empirical evidence is important.

Secondly, the analyses show that both the TIMSS and PISA studies complement each other extremely well. The mathematics framework in TIMSS and the Swedish governing documents in mathematics correspond well in respect of the classical grouping of subject matter, while the science framework in PISA corresponds well with the Swedish syllabuses through the emphasis on processes within scientific knowledge in both documents.

The analysis on the question level shows that, with regard to content, TIMSS has the broader coverage in relation to the factual knowledge sought in both the mathematics and science syllabuses. From the *cognitive* point of view, PISA and TIMSS do not differ with regard to mathematics, while this is not the case in science since the goals there have been structured in a different way from the mathematics goals. As regards knowledge of concepts and phenomena, the goals are covered more comprehensively by TIMSS, while PISA to a greater extent covers the goals for the scientific activity as well as, to a certain degree, the use of scientific knowledge.

Thirdly, it can be stated that with regard to studies of mathematics and science/scientific knowledge, the results in this report indicate that the international studies fulfil their purpose well, of testing the level of knowledge and knowledge development in the subjects in question, even from a Swedish national perspective.

8 References

Anderson, L.W., Krathwohl D. R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., Wittrock, M.C [Eds.], (2001). *A Taxonomy for Learning, Teaching and Assessing.* New York: Addison Wesley Longman Inc.

Andersson, B., Bach, F., Olander, C., Zetterqvist, A. (2005). *Nationella utvärderingen av grundskolan 2003. Naturorienterande ämnen.* Ämnesrapport till rapport 252. Stockholm: Skolverket.

Andréasson, B., Bondesson, L. Boström, K. (2002). *Puls Grundbok i Kemi för grundskolans senare del.* Stockholm: Bokförlaget Natur och Kultur.

Andréasson, B., Bondesson, L. Gedda, S., Johansson, B., Zachrisson, I. (2002). *Puls Grundbok i Biologi för grundskolans senare del.* Stockholm: Bokförlaget Natur och Kultur.

Björk, L.-E., Björksten, C., Brolin, H., Larsson, B, Lundström, K. (1995). *Matte till 1000.* Stockholm: Bokförlaget Natur och Kultur.

Bloom, B.S. [Ed.], Englehart, M.D., Furst, E.J., Hill, W.H. and Krathwohl, D.R. (1956). *The Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain.* New York: David McKay Company Inc.

Crocker, L., Algina, J. (1986) *Introduction to Classical & Modern Test Theory.* Orlando: Harcourt Brace Jovanovich, Inc.

Eurydice (2006). *Science Teaching in Schools in Europe. Policies and Research.* Brussels: Eurydice

Ginsberg, A., Cooke, G., Leinwand, S., Noell, J. & Pollock, E. (2005). *Reassessing U.S. International Mathematics Performance: New Findings from the 2003 TIMSS and PISA.* American Institutes for Research.

Hambleton, R. K. (1989). Principles and applications of item response theory. In R. L. Linn (Ed.), *Educational measurement. Third edition.* New York: American Council on Education/Macmillan Publishing Co.

Husén, T. (1967). International Study of Achievement in Mathematics. Vol II. New York: Wiley.

Jansson, S. (1975). *Undervisningsmål som utgångspunkt vid konstruktion av målrelaterade prov. Metodproblem och empiriska resultat.* Akademisk avhandling. Umeå: Pedagogiska institutionen, Umeå universitet.

Kjellström, K. (2005). *Nationella utvärderingen av grundskolan 2003. Matematik årskurs 9.* Ämnesrapport till rapport 251. Stockholm: Skolverket.

Lie, S., Taylor, A, Harmon, M. (1996).: Scoring Techniques and Criteria. In Martin, M. O., Kelley, D. (Eds): *Third International Mathematics and Science Study Technical Report, Volume I: Design and Development.* Chestnut Hill, M.A: Boston College.

Löfdahl, S. E. (1987). *Fysikämnet i svensk realskola och grundskola. Kartläggning och alternativ ur fysikdidaktisk synvinkel.* Acta Universitatis Upsaliensis. Uppsala Studies in Education 28. Uppsala: Uppsala universitet.

McLaughlin, D., Dossey, J. and Stancavage, F. (1997). *Validation Studies of the Linkage between NAEP and TIMSS Eighth Grade Mathematics Assessment.* Educational Statistical Services Institute.
http://nces.ed.gov/pubs98/linking/appx-a.asp

Mejding, J., Reusch, S., Andersen, T. Y. (2006). Leaving Examination Marks and PISA Results – Exploring the Validity of PISA Scores. I Mejding, J., Roe, A. (Eds.). *Northern Lights on PISA 2003 – a reflection from the Nordic Countries.* Copenhagen: Nordic Council of Ministers.

Messick, S. (1989). Validity. I R. l, Linn (Ed.). *Educational Measurement (Vol. 3).* New York: American Council on Education.

Mullis, I.V.S., Martin, O.M., Gonzalez, E.J. and Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report.* Boston: The International Association for the Evaluation of Educational Achievement.
http://timss.bc.edu/timss2003

Mullis, I.V.S., Martin, O.M., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., Chrostowski, S.J. and O'Connor, J.M. (2003). *TIMSS Assessment Frameworks and Specifications 2003.* (2nd Ed.). Boston: The International Association for the Evaluation of Educational Achievement.
http://timss.bc.edu/timss2003i/frameworks.html

National Council of Teachers of Mathematics. (2006) *Principles and Standards for School Mathmatics.* http://www.nctm.org/standards/

Niss, M. (1999). Kompetencer og Uddannelsebeskrivelse. *Uddannelse, 9.* pp 21–29 Köpenhamn: Undervisningsministeriet.
http://udd.uvm.dk/199909/udd9-3.htm?menuid=4515

Nohara, D. and Goldstein, A.A. (2001). *A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R) and the Programme for International Student Assessment (PISA).* Working Paper No 2001-07 June 2001 U.S. Department of Education. Washington DC: Office of Educational Research and Improvement.

OECD. (2003), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills.* Paris: OECD.
http://www.pisa.oecd.org/dataoecd/46/14/33694881.pdf

Organisation for Economic Co-operation and Development. (1999). *Measuring Student Knowledge and Skills. A New Framework for Assessment.* Paris: OECD.

Organisation for Economic Co-operation and Development. (2004). *Learning för Tomorrow's World. First Results from PISA 2003.* Paris: OECD.

Palm, T. (2002). *The Realism of Mathematical School Tasks.* Doctorial Thesis No 24. Umeå: Department of Mathematics, Umeå University.

Popham, W. J. (1999). Classroom Assessment: What Teachers Need To Know. 2nd Edition. Needham Heights, MA: Allyn & Bacon.

Schoenfeld, A., H. (2004). The Math Wars. Educational Policy, Vol. 18 No. 1:253–286

Scott, E. (2005). *Comparing NAEP, TIMSS and PISA in Mathematics and Science*. U.S. Department of Education. http://nces.ed.gov/nationsreportcard

Sjöberg, S., Ekstig, B. (2002). *Puls Grundbok i Fysik för grundskolans senare del.* Stockholm: Bokförlaget Natur och Kultur.

Skolverket. (2000). *Grundskolan Kursplaner Betygskriterier.* Stockholm: Skolverket. http://www3.skolverket.se/ki03/front.aspx?sprak=EN&ar=0809&infotyp=15&skolform=11&id=2087&extraId=

Skolverket. (2002). *Curriculum for the compulsory school system, the pre-school class and the after school centre – Lpo 94.* Stockholm: Skolverket.

Skolverket. (2003). *Lusten att lära – med fokus på matematik.* Rapport nr 221. Stockholm: Skolverket.

Skolverket. (2004a). *Nationella utvärderingen av grundskolan 2003. Sammanfattande huvudrapport.* Rapport 250. Stockholm: Skolverket.

Skolverket. (2004b). TIMSS 2003. *Svenska elevers kunskaper i matematik och naturvetenskap i skolår 8 i ett nationellt och internationellt perspektiv.* Rapport 255. Stockholm: Skolverket.

Skolverket. (2005a). *PISA 2003. Svenska femtonåringars kunskaper och attityder i ett internationellt perspektiv*, Rapport 254. Stockholm: Skolverket.

Skolverket. (2005b). *Matematikuppgifter i TIMSS 2003.* Uppgiftsrapport till Rapport 255. Stockholm: Skolverket.

Skolverket. (2005c). *NO-uppgifter i TIMSS 2003.* Uppgiftsrapport till Rapport 255. Stockholm: Skolverket.

Wedman, I., (1988). *Prov och provkonstruktion.* Stockholm: Utbildningsförlaget.

Skolverket

Swedish National Agency for Education

www.skolverket.se