

Alignment between TIMSS Advanced and the Swedish national curriculum

Peter Nyström
Annika Kjellsson Lind

Innehåll

1. INTRODUCTION	2
2. BACKGROUND	2
2.1. SWEDISH NATIONAL TESTS FOR UPPER SECONDARY SCHOOL	3
2.2. AIM AND RESEARCH QUESTIONS	4
2.3. METHOD	4
2.3.1. <i>Taxonomy</i>	4
2.3.2. <i>Subject content</i>	5
2.3.3. <i>Thought processes</i>	5
2.3.4. <i>Demands on students' answers</i>	6
2.3.5. <i>Presentation</i>	7
2.3.1. <i>Sampling of items</i>	9
2.3.2. <i>Categorisation procedure</i>	9
3. RESULTS	10
3.1. RESULTS MATHEMATICS	10
3.1.1. <i>Mathematics content</i>	10
3.1.2. <i>Thought processes</i>	11
3.1.3. <i>Demands on students' answers?</i>	13
3.1.4. <i>Presentation</i>	14
3.2. PHYSICS	16
3.2.1. <i>Physics content</i>	16
3.2.2. <i>Thought processes</i>	17
3.2.3. <i>Demands on students' answers</i>	19
3.2.4. <i>Presentation</i>	19
4. CONCLUSIONS AND DISCUSSION	22
4.1. COMPARISON OF SUBJECT CONTENT	22
4.2. COMPARISON OF THOUGHT PROCESSES DEMANDED FROM STUDENTS	22
4.3. COMPARISON OF DEMANDS ON STUDENTS' ANSWERS	23
4.4. COMPARISON OF ITEM PRESENTATION	23
4.5. CONCLUDING REMARCS	24
5. REFERENCES	24

1. Introduction

Discussions about the credibility, relevance and usefulness of TIMSS is to a high degree related to TIMSS alignment to national curricula and students opportunity to learn. TIMSS can not be fully aligned with intentions and practice in every participating country since such an ambition would narrow the assessment framework of TIMSS to an unacceptable degree. It is however important to investigate alignment and to use it in the interpretation of the results. There is information in answers from the TIMSS questionnaires that can be used for this purpose, for example teachers' descriptions of the content taught. TIMSS international reports (e.g. Mullis, Martin, Robitaille, & Foy, 2009) present a TCMA (Test Curriculum Matching Analysis) in order to investigate the robustness of national results compared to other countries.

The study presented here is one of several studies of alignment between TIMSS Advanced 2008 and Swedish intended and implemented curriculum. One study compared the Swedish national syllabus in advanced mathematics with the content domains in TIMSS advanced as described in the TIMSS framework (Garden et al., 2006). Another study made a similar comparison between the Swedish syllabus and TIMSS cognitive framework. Ongoing studies evaluate Swedish students' opportunities to learn the content in TIMSS Advanced by analysing textbooks.

The study reported here focuses alignment between the TIMSS items in mathematics and physics and Swedish national tests. The purpose of the study is to investigate in what ways and to what extent that the items in TIMSS Advanced differ from items in Swedish national tests in mathematics and physics, with respect to a broad range of features.

2. Background

Alignment concerns how well parts of a system are in accordance with other parts of the system, and to what extent they send the same message. According to the British sociologist Basil Bernstein there are three message systems of particular interest in the realisation of formal educational knowledge: curriculum, pedagogy and evaluation (Bernstein, 1971). "Curriculum defines what counts as valid knowledge, pedagogy defines what counts as a valid transmission of knowledge, and evaluation defines what counts as a valid realization of this knowledge on the part of the taught." (p. 141). Accordingly, in the educational context alignment between curriculum, teaching and assessment is of particular interest.

Webb (1997) deals with alignment in theory and practice and focuses specifically on the alignment between learning goals and assessment. He defines the theoretical concept of alignment as "the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system towards students learning what they are expected to know and do." (p. 4) By "expectations" he refers to teaching and learning goal expressed in documents such as curricula, syllabi, etc.. In order to realize this theoretical concept into practice and actually evaluate the alignment between expectations and assessment Webb (1997) developed five criteria. The first criterion concerns subject content and emphasises that expectations and assessments should focus students' subject matter knowledge in a coherent way. The second criterion states that expectations and assessment should be founded in a common view on how students develop during their complete schooling and how school best can contribute to their learning in different developmental phases. The third criterion deals with equity and justice and states that if the expectation is that every student should reach learning goals of high standard alignment requires assessment situations to give every student reasonable opportunities to show what they know and can do. The fourth criterion concerns pedagogical implications. Expectations and assessment can and should have an influ-

ence on teaching, and should send clear and consistent messages to teachers about what to do. Webb's fifth and final criterion talks about adjustment to the system and focuses whether parents, students, teachers, and others can understand and relate expectations and assessments, and find it possible to believe that the expectations are possible to reach.

In education, alignment is particularly relevant in two aspects. The first aspect concerns ideas about how to accomplish effective schooling. Roach, Niebling, & Kurz (2008) (with reference to Baker & Linn, 2002) argue that large-scale assessment systems and accountability systems are based on a theory of action "that assumes increased information about student achievement, coupled with salient incentives for increased performance (and corresponding punishments for lack of improvement), will motivate educators and produce improved student outcomes". (p. 158). They also argue that a consequence of this is that the development and implementation of e.g. large-scale assessment is viewed by advocates of standards-based reform as one way to improve classroom instruction and increase equity across the educational system. Standards-based reform has definitely been influential in educational reform and has influenced the development of the Swedish school system during the last 20 years. With reference to Smith & O'Day (1991), Clune (2001) claims that the central starting point for the standards movement is that alignment between different means of control and teaching materials is the only way to create schools reaching high standards concerning students' achievement.

The second aspect describing the importance of alignment is connected to validity in educational assessment. In order to interpret results from large-scale assessment it is important to know the extent to which the assessment reflects what students have had an opportunity to learn. This aspect is the basis for the study presented here.

2.1. Swedish national tests for upper secondary school

The Swedish comprehensive school is characterised by a substantial empowerment of schools and teachers (Vedder & O'Dowd, 1999), but this also counts for upper secondary school. National curricula are fairly general and intended to be interpreted locally. The freedom of interpretation is quite big. Teachers are responsible for grading their students, and they are empowered to do that based on national assessment criteria describing what students should know and be able to do in order to achieve a certain grade. There are no formal examination tests in Swedish upper secondary school, but a system of national assessment serving several purposes. Teachers are expected to use the results together with results from other assessments during the school year in grading their students. Furthermore, a purpose with the national assessment system is to clarify the learning aims and operationalise national assessment criteria. This means that Swedish national tests can be seen as part of the intended curriculum, but also as an instrument leading into the implemented curriculum. At least in theory, the tests found in the Swedish national assessment system are low-stakes in the sense that they have no immediate consequence for the individual student. However, the tests are viewed as very important by teachers and students, raising the stakes substantially.

Swedish upper secondary school is organised in subjects and modular courses within subjects. In mathematics, the core of the subject is studied in five courses (A-E) with Mathematics A as the first course, Mathematics B building on Mathematics A, etc. TIMSS Advanced content can be associated to most of these courses, but you need to have studied at least Mathematics A-D in order to have a reasonable chance of solving a majority of the TIMSS items. In physics, the core of the subject is studied in two courses. Both courses are relevant in relation to the TIMSS items (and the TIMSS framework), which means that students participating in TIMSS Advanced had to have studied both courses.

In mathematics, national tests are given two times a year for each of the courses Mathematics A-D. For mathematics E, and for the two courses in physics, a national item- and testbank is accessible for teachers. In this internet-based bank, teachers can create and download tests using a large number of items. Teachers can also download ready-made tests from the website. National tests are compulsory for mathematics A and for the final mathematics course in each study programme. Using the item- and testbanks is voluntary, but national evaluations tell us that the bank in physics is frequently used by the physics teachers (Skolverket, 2005). Even though there is a formal and legal difference between national tests and tests from the national test bank, they will both be referred to as national tests for convenience.

2.2. Aim and research questions

The aim of the study is to describe and analyse alignment between the TIMSS Advanced 2008 student achievement instruments and tests from the Swedish national assessment system.

More specifically four research questions are focussed. The first question concerns the subject domains covered by each task. This is important to the alignment issue since two tests which represent different subject specific content can give different pictures of what student know and can do in the subject. The second question deals with the thought processes demanded of students in order to answer the items in the two tests. In addition to alignment with respect to subject content, cognitive demands of the items play an important role for students' possibilities to. The third question addresses the demands on students answers. What kind of formats for answering items is used in the studied materials? The requirements for giving a correct answer also differ between items where students only have to give a short answer and items where they are expected to show their line of thought. The fourth research question concerns the presentation of the item, focusing context, format and text.

2.3. Method

A taxonomy was developed and applied on the TIMSS tests and on a sample of mathematics- and physics-tests from the Swedish national assessment system. Items from both systems were categorised on the basis of the taxonomy. In this section, the taxonomy, test samples and categorisation procedure will be described in detail.

2.3.1. Taxonomy

For the purpose of comparing TIMSS and Swedish national tests a taxonomy was developed, building on a taxonomy from previous Swedish research on international comparative studies for comprehensive school (Lindström, 2006) The taxonomy has been developed in order to find meaningful categories for the advanced level national tests as well as for TIMSS. The parts of the taxonomy dealing with subject content domains and cognitive domains is based on the TIMSS Advanced 2008 Framework (Garden et al, 2006). The framework corresponds to the research questions stated above, and has the following overall structure:

Research question 1 (Q1): Subject content

Research question 2 (Q2): Cognitive aspects

Q2a: Cognitive level

Q2b: Calculations

Q2c: Tools

Research question 3 (Q3): Demands on students answers

Research question 4 (Q4): Presentation

Q4a: Text length

- Q4b: Graphical elements
- Q4c: Subject specific words
- Q4d: Out-of-school contexts

In the following sections the categories in each part of this structure will be elaborated.

2.3.2. Subject content

The categories used for classification of items with respect to subject content are based on the description of content domains in TIMSS Advanced Assessment Framework (Garden et al., 2006), but in order to make meaningful categorisations of all items from the national tests some areas have been broadened and some subdomains have been added. The content categories used in this study are presented in Appendix 1 (mathematics) and Appendix 2 (Physics). In Physics a specific category has been added to be able to identify items in physics that primarily have a mathematical content, and require no understanding of physics. In these items is the information given in the item sufficient to make it possible to answer using mathematics and a table of formulae.

2.3.3. Thought processes

Three different aspects are used in order to categorise cognitive demands in the items analysed: cognitive levels, use of calculations, and usefulness of tools.

2.3.3.1. Cognitive levels

Categorisation into cognitive levels is intended to indicate the kind of thought processes demanded in order to solve a task. Thought processes are of course individual, and categorisation of items with respect to cognitive levels can only be done in a judgmental process estimating what kind of thought processes that most students most likely will engage in when working with each item in TIMSS. In this study, categories describing cognitive level are based on the TIMSS Advanced assessment framework (Garden et al., 2006). The cognitive levels of TIMSS were reformulated to some extent in order to enable clearcut categorisations of items from TIMSS and national tests. Descriptions of the categories used are found in Appendix 2.

Items are categorised with a typical student in mind. This student has studied the courses required to be included in the TIMSS Advanced study, i.e. mathematics D or physics B. Familiarity with items is a characteristic of the cognitive levels in mathematics, and to some extent also in physics. Estimates of the novelty of items for the students have been made by identifying whether similar items are found in dominating textbooks or not.

Since there are items in TIMSS covering content that is not covered in advanced mathematics and physics in Swedish upper secondary school, a category of *prerequisites missing* was added. For items put in this category students lack the possibility of reaching a correct answer, regardless of cognitive level, because there are crucial conceptual or procedural elements that they have never met. This makes categorisation of cognitive level meaningless for these items.

For both mathematics and physics, the main cognitive levels used are *knowing*, *applying* and *reasoning* (Appendix 2). More than one of these levels can be identified as useful in working with an item, but here each item is categorised according to the highest cognitive level required in order to receive maximum score on the item. If, for example, an item requires knowing as well as applying (which all applying items do), but no reasoning, the item will be categorised as applying. Each cognitive level is described using a number of criteria for categori-

sation, and meeting one of these criteria is enough for an item to be categorised in a certain cognitive level.

2.3.3.2. Calculations

Three categories are used in order to describe and compare the types of calculations required in items.

Trivial/no calculations	Items demanding no calculations or only simple mental calculations
Straight-forward calculations	Items demanding writing or using calculators in order perform algorithms or to keep track of the calculations. No more than one more difficult calculation is demanded in order to reach an answer.
Complex calculations	Items demanding on or more partial results in order to make the final calculation

2.3.3.3. Tools

For TIMSS Advanced as well as the national tests, students are allowed and expected to use calculators and tables of formulae when working with the items. One difference between the two tests is that national tests contain a part where calculators are not allowed. Another difference is that the tables of formulae allowed are very different in TIMSS compared to the national tests. The categorisation with respect to the usefulness of these tools is made on the basis of the particular table of formulae allowed in the two tests, using four categories; no tool useful, table of formulae useful, calculator useful, and both calculator and table of formulae useful.

No tool	Neither table of formulae nor calculator give significant help in solving the task
Table of formulae	The table of formulae allowed gives definitions or relations that are useful in solving the task
Calculator	Solving the task without a calculator takes substantially longer time
Table of formulae and calculator	Both table of formulae and calculator are very helpful in solving the task

2.3.4. Demands on students' answers

2.3.4.1. Type of answer

The types of answers expected by students are described in five different categories. Some of the items give more the one score point with different kinds of answers demanded for different scores. These items are categorised based on the type of answers required in order to achieve the highest score.

Multiple choice	Items answered by choosing between given alternatives with the number of correct alternatives set out
-----------------	---

Complex multiple choice	Item answered by choosing between given alternatives. The number of correct alternatives is unknown or several different concepts must be paired to a definition or a phenomenon
Short answer	Item answered with a calculated value or a single word
Extended answer	Item answered with students describing their line of thought or presenting a statement that is motivated or validated
Draw	Item answered by drawing a picture, graf or table (a kind of short answer) or supplement a statement by drawing (a kind of extended answer)

Type of answer is also categorised with respect to whether students are expected to show their work or not. In TIMSS it is explicitly stated when students are expected to show their work, i.e. to explain how they reached the final answer. The opposite is true for the national tests. If it is not explicitly stated that only a final answer is to be presented, students are expected to explain how they reach their final answer.

2.3.5. Presentation

2.3.5.1. Text length

Text length is a characteristic of items that can signal fundamental features of items, such as construct irrelevant variance due to undue influence of reading comprehension on mathematics and science achievement scores. Differences in text length can also be seen as an alignment problem since students might not have had the opportunity to learn to solve problems posed in a particular format.

The relationship between text length in an item and students difficulties in solving it is complex, and text length is an uncertain measure of readability (Nyström, 2008). Nevertheless, test length was categorised in our study as a simple description of item format and readability demands.

In categorising the number of words per item, items are quite simply given a number representing the number of words. Numbers, signs, variables or equations are not counted as words. For example, the figure “1” is not counted as a word, but the quantification “one” is. In the cases where an item stem is relevant to several specific items, the number of words in the stem has been included in the word count for each item. Prompts, like “you don’t have to show your work”, are not included in the estimation of text length.

2.3.5.2. Graphics

Graphics refers to pictures, tables, graphs and diagrams found in the items. Items with more than one type of graphical representation are assigned to multiple categories. Graphical representations in a stem connected to several items are valued with respect to their significance for each item. The five categories used are shown below.

No graphics	Item presented without graphical representations
Decoration	Item presented with a graphical representation illustrating the task, but not supplying any necessary information

Informative picture	Item presented with a graphical representation containing necessary or clarifying information, needed in order to answer the item
Table	Item presented with table/s containing necessary or clarifying information, needed in order to answer the item
Diagram/Graph	Item presented with diagram or graph containing necessary or clarifying information, needed in order to answer the item

2.3.5.3. Subject specific vocabulary

Categorisation of items with respect to the occurrence of vocabulary is slightly different in mathematics and physics.

In mathematics the complexity of the vocabulary used in items is categorised on the basis of how late in the educational system these words are introduced and how common they are. The items are categorised into five different categories based on the highest complexity of vocabulary found in each item. The vocabulary is considered unknown if at least one word in the item is not found in the dominating textbook in advanced mathematics.

None, or used in every-day language	Item presented using no subject specific words, no subject-specific words explained in the text, and no subject-specific words used in everyday language
Basic subject-specific vocabulary	Item presented using subject specific words that students most likely have met earlier and frequently, for example <i>function</i>
Basic specialised vocabulary	Item presented using the words: <i>derivative, number sequence, integral</i>
Other specialised vocabulary	Item presented using subject-specific words that the students can be expected to have learned in the last year with the exception of <i>derivative, number sequence, integral</i>
Unknown vocabulary	Item presented using subject-specific words that the students most likely never have encountered

The category “basic specialised vocabulary” was meaningful to use in mathematics because these words have a specific role in the core of the courses in advanced mathematics.

The categories used in Physics resemble to some extent the categories used in mathematics, but they are described slightly different due to the fact that physics is only studied in two courses in upper secondary school and little subject-specific vocabulary can be expected to be introduced in comprehensive school.

None, or used in every-day language	Item presented using no subject specific words, no subject-specific words explained in the text, and no subject-specific words used in everyday language
Basic subject-specific vocabulary	Item presented using subject specific words that students most likely have met in Physics A
Specialised vocabulary	Item presented using subject-specific words that the students can be expected to have learned in Physics B

Unknown vocabulary	Item presented using subject-specific words that the students most likely never have encountered
--------------------	--

2.3.5.4. Out-of-school contexts

Setting mathematics and science tasks in an out-of-school context is a way of making the tasks more concrete and understandable, and can also hopefully point to the usefulness of what students are expected to learn. Furthermore it is claimed that the concordance between school word problems and the corresponding out-of school task situations (often denoted task authenticity) can affect both the students' engagement in the task context and the extent to which they require that their solutions make sense in the 'real-life' situations described in the tasks (see e.g. Palm & Nyström, 2009). Palm (2006) has argued that a school task with an out-of-school context can be considered authentic only if it represents some task situation in real life, and simulates important aspects of that situation to some reasonable degree. For this purpose he has developed a framework identifying a number of important aspects concerning authenticity. The categories used in the present study are inspired by the work by Palm (2006), but a much simplified version.

No reality	Item not describing a context relating to a world outside of the subjects internal concreteness
Artificial reality	Item describing a context relating to a world outside subject-specific theory and relating to practical situations, but asking questions that most likely would not be asked in these situations
Real reality	Item describing a real-world context (every-day life, society, working life, research), asking questions that can be expected to be relevant in the given context

2.3.1. Sampling of items

All of the items in TIMSS Advanced 2008 are categorised and compared with approximately equal numbers of item from Swedish national tests. In all, four national tests in mathematics D and four tests from the national test bank in physics B have been analysed.

Mathematics D and physics B are the courses in the Swedish upper secondary school most closely resembling the requirements of TIMSS Advanced, and therefore the study is restricted to a comparison with the national tests used in these courses. There is however a problem with this restriction because some of the TIMSS content is covered in courses preceding mathematics D and physics B.

TIMSS Advanced 2008 contains in total 76 items. To match this item pool, four mathematics D national tests were selected, in total 87 items. The physics part of TIMSS contains 72 items, which are analysed and compared to items from four physics tests in the national test bank, in total 86 items.

2.3.2. Categorisation procedure

Each item in the sample has been categorised by two persons. In mathematics, the categorisation of each item was made in collaboration and during this collaborative work the categories were adjusted until they were judged to give a concordant categorisation of the items. In physics, a subsample of items were categorised in collaboration in order to adjust the categories and reach a common interpretation. Half of the remaining items were categorised by one person and half by the other, and problematic cases were discussed and agreed upon.

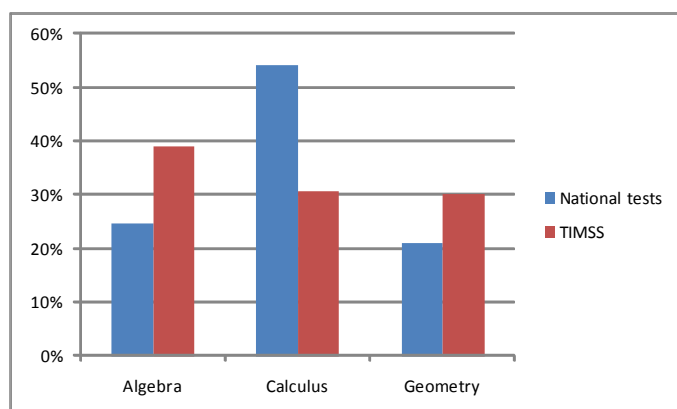
In case an item can be solved using more than one method, and these methods represent different categories, the item has been given multiple categorisations. Results are mainly presented as percentage of raw-scores connected to the items found in each category.

3. Results

3.1. Results mathematics

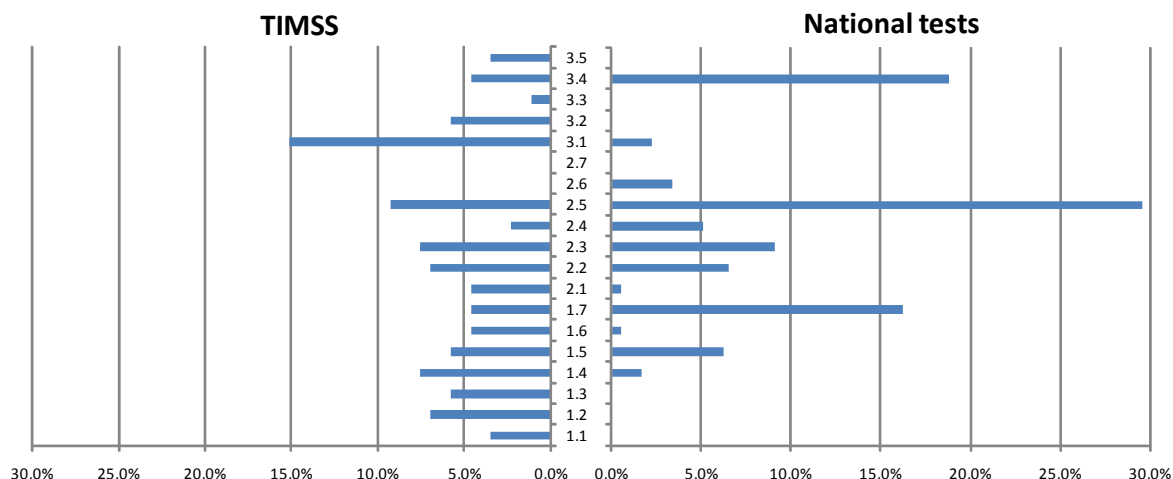
3.1.1. Mathematics content

The comparison of content domains represented by tasks in Swedish national tests and in TIMSS indicates some differences. The TIMSS tests show an even distribution of item scores over the three content domains. In the Swedish tests, more than half of the attainable score points are found in the area of Calculus and as little as 13 percent in the area of Geometry. The Algebra-scores are approximately equal in the two item sets.



Figur 1 Distribution (in %) of available score points on content domains in mathematics for TIMSS Advanced 2008 and Swedish national tests.

A comparison of score distributions for content subdomains point to other differences. Some subdomains are only represented in TIMSS: series and sums (1.2), combinations, probability and statistics (1.3), solve problems using gradients (3.2), equations and relations for the circle (3.3) and vectors (3.5). Subdomains only represented in the Swedish national tests are differential equations (2.6) and volume by rotation (2.7). Other categories showing large differences are complex numbers (1.1) which is more common in the Swedish tests, and using geometrical properties to solve problems (3.1) which is much more frequent in TIMSS.



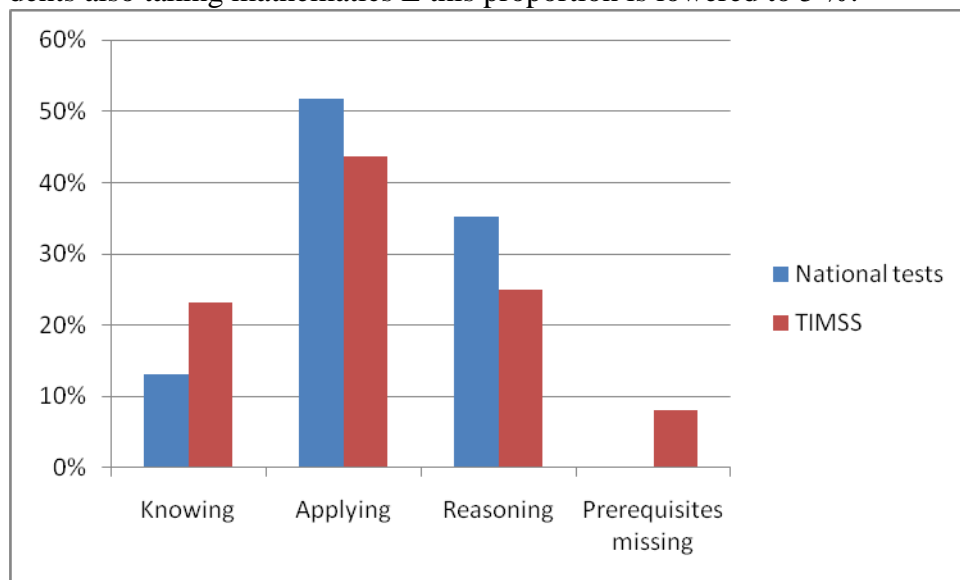
Figur 2 Distribution (in %) of available score points on content subdomains in mathematics for TIMSS Advanced 2008 and Swedish national tests.

3.1.2. Thought processes

3.1.2.1. Cognitive level

A comparison of cognitive levels in tasks from national tests and TIMSS shows that the distribution between cognitive levels is fairly similar in the two tests.

Some of the items in TIMSS have been classified as “not relevant”. These are items in TIMSS having content that the Swedish students have not met. These are not possible to categorize with respect to cognitive level in any meaningful way. For students taking mathematics D as their highest course, 8 % of the items in TIMSS are found to be not relevant. For students also taking mathematics E this proportion is lowered to 5 %.



Figur 3 Distribution (in %) of available score points on cognitive levels in mathematics for TIMSS Advanced 2008 and Swedish national tests.

The distribution of cognitive levels in TIMSS deviates from what has been reported in TIMSS international reports and in the TIMSS Assessment framework (Garden et al., 2006) because

the categorisation presented here is made on the cognitive levels required from Swedish students in solving the TIMSS items.

3.1.2.2. Calculations

Figure 4 presents the results from the categorisation of items with regard to the complexity of the calculations needed to reach a correct answer. The national tests in mathematics require more complex calculations than TIMSS Advanced.

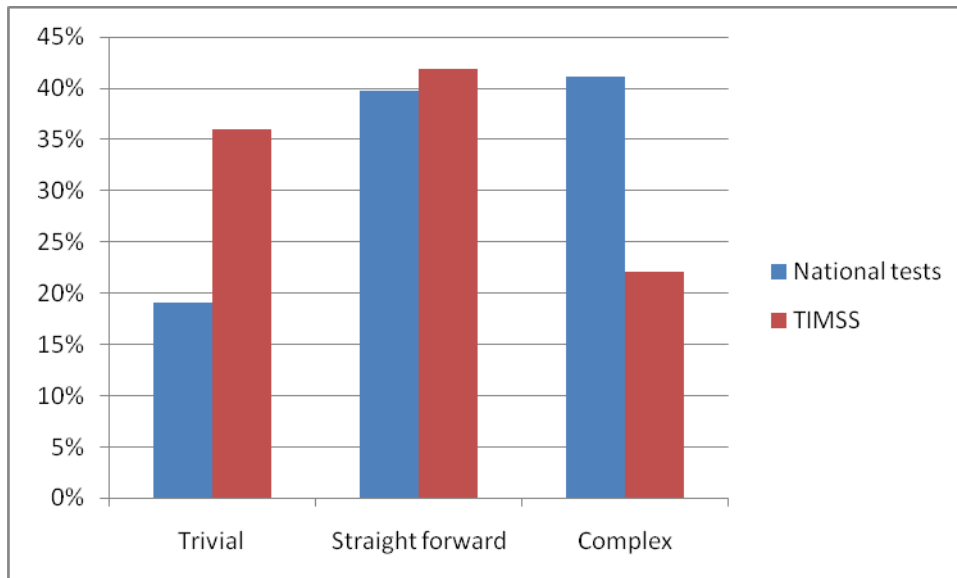
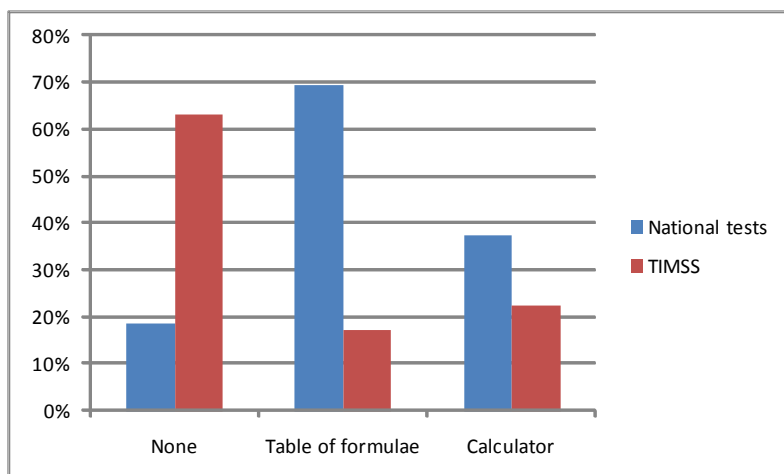


Figure 4 Distribution (in %) of available score points on types of calculations demanded in mathematics for TIMSS Advanced 2008 and Swedish national tests.

3.1.2.3. Tools

The items from Swedish national tests and TIMSS were also categorised with respect to the different tools available for each test. As previously described (see section 2.1) Swedish national tests consist of two parts, one without calculator and one with. Furthermore, the national tests a rather extensive and detailed collection of formulas is allowed.

Figure 5 shows the results from this categorisation. For a substantial number of items both calculators and tables of formulas are useful, which makes the total percentage add up to more than 100 percent. The tables of formulas used in the national tests are useful for as much as three items out of four. The much more sparse tables of formulas in TIMSS is useful for one fifth of the TIMSS items.

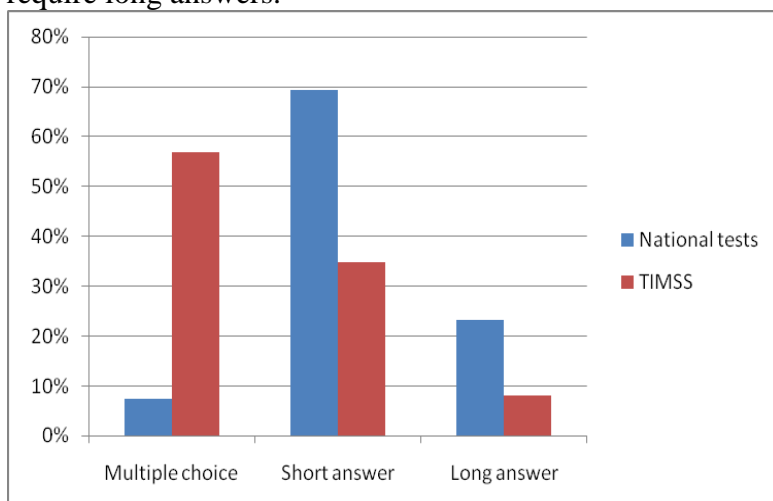


Figur 5 Distribution (in %) of available score points on usefulness of tools in mathematics for TIMSS Advanced 2008 and Swedish national tests. The category *Table of formulae* contains items for which the available table of formulae can be helpful in solving the item. The category *Calculator* contains items which would take substantially longer time to solve without the help of a calculator.

3.1.3. Demands on students' answers?

3.1.3.1. Type of answer

About half of the score points in the TIMSS items are multiple choice, which is in accordance with the prescribed target in the TIMSS assessment framework (Garden et al., 2006). Just over a third of the score points (37 %) comes from items requiring a short answer and 10 percent comes from items expecting a long answer. This distribution differs substantially from what is found in the Swedish national tests. In these tests there are hardly no multiple choice items. Almost 70 percent of the items are of the short answer type, and more than 20 percent require long answers.



Figur 6 Distribution (in %) of available score points on item format in mathematics for TIMSS Advanced 2008 and Swedish national tests.

In the Swedish national tests in mathematics, 77 percent of the items (87 percent of the available score points) require students to explain their answers. In TIMSS only 17 percent of the items (33 percent of the available score points) require these extended answers.

3.1.4. Presentation

3.1.4.1. Word count

Figure 7 shows the distributions of words per item for the national tests and TIMSS in intervals from 0 to 50 words. In TIMSS advanced 2008 there are on average 17 words per item (median 15) and there are no items with more than 50 words. In the national tests there are 31 words per item on average (median 15) and 30 percent of the available score points come from items with more than 50 words. One item contained as much as 222 words. These results show that the main difference between TIMSS and the national tests regarding number of words is that the national tests contain a few items with very long texts.

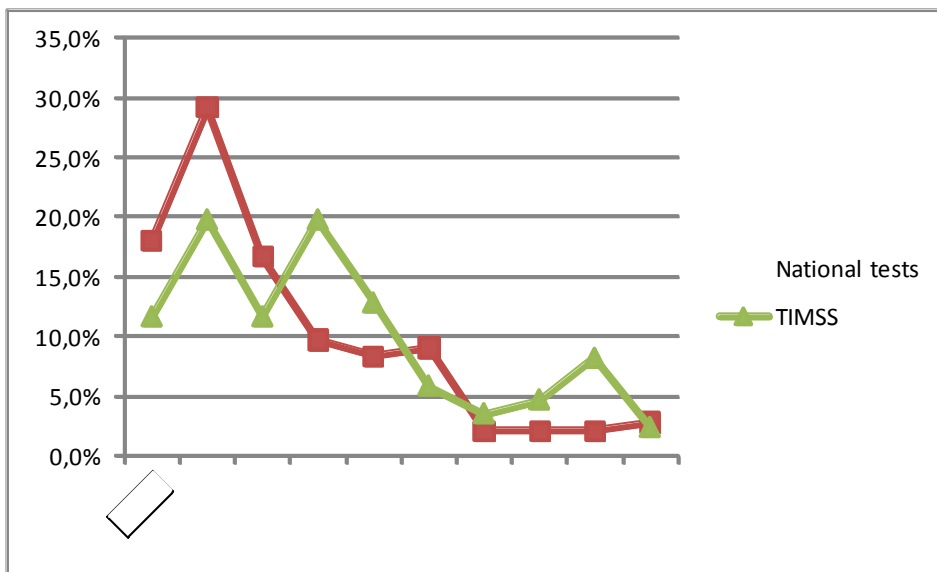


Figure 7 Distribution (in %) of text length per item, measured as number of words, in mathematics for TIMSS Advanced 2008 and Swedish national tests.

3.1.4.2. Graphics

In the sample of items from the national tests, more than one graphical object is found in four items (16 score points). In TIMSS, the corresponding frequency is two items (2 score points). In categorising the type of use of graphics, these items can be counted more than once, making the percentages in figure 8 add up to more than 100 percent. Figure xx indicates that the national tests contain more graphics, both with an informative function and as decoration.

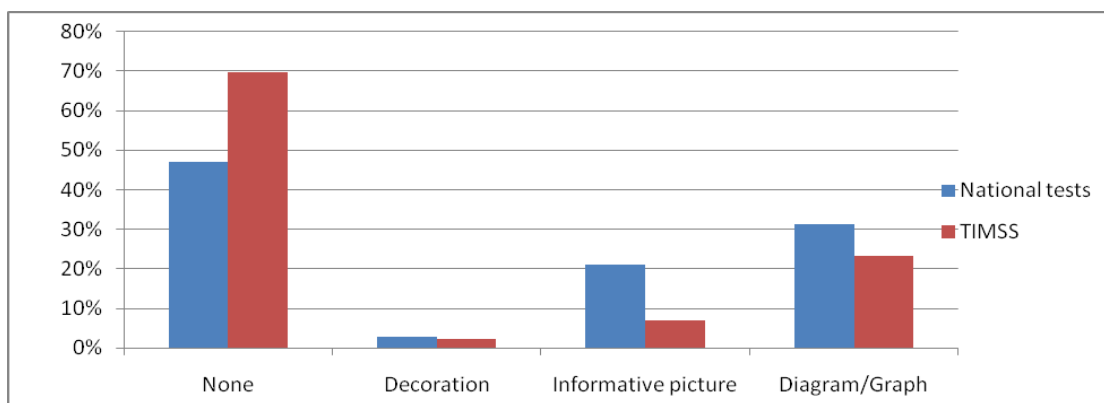


Figure 8 Distribution (in %) of available score points on use of graphical elements in mathematics for TIMSS Advanced 2008 and Swedish national tests.

3.1.4.3. Subject specific vocabulary

According to our categorisation, the national tests contain somewhat more mathematics vocabulary than TIMSS. Some words used in TIMSS are not part of the syllabus in Sweden and are therefore categorised as unknown vocabulary.

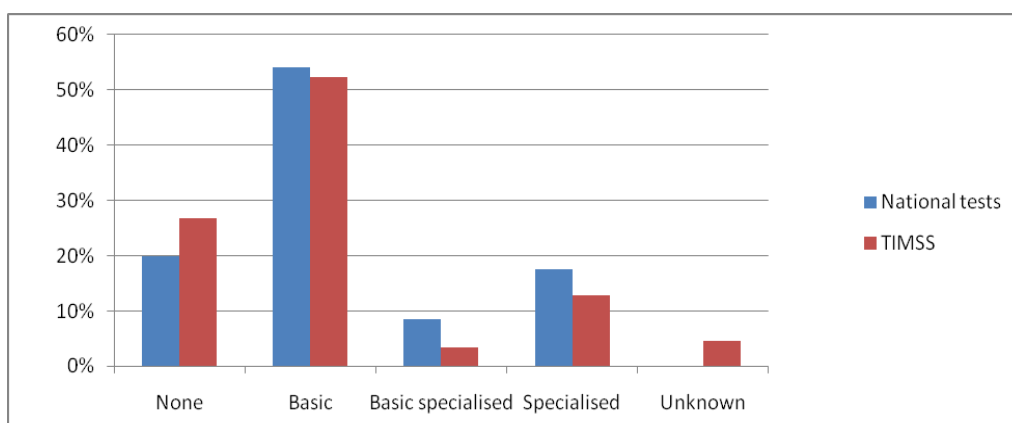


Figure 9 Distribution (in %) of available score points on occurrence of vocabulary in mathematics for TIMSS Advanced 2008 and Swedish national tests.

3.1.4.4. Reality

Items connection to the real world outside mathematics classrooms was analysed using three categories. The highest level, "true" reality, is more common in the Swedish national tests than in TIMSS.

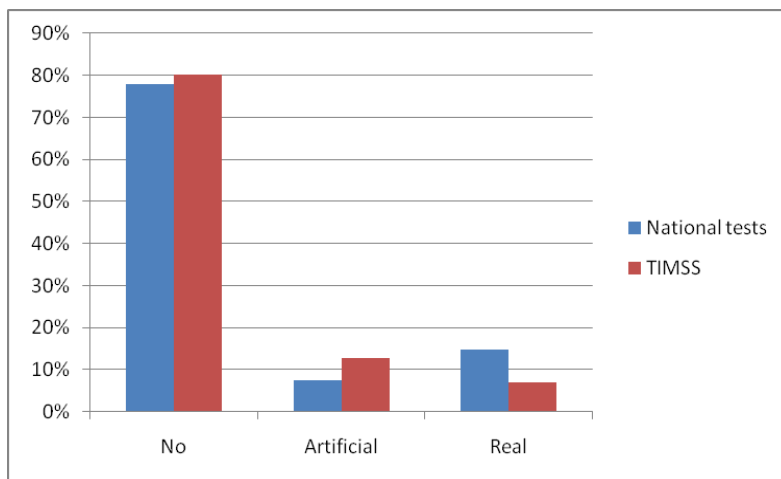


Figure 10 Distribution (in %) of available score points on type of out-of-school context in mathematics for TIMSS Advanced 2008 and Swedish national tests.

3.2. Physics

3.2.1. Physics content

Figure 11 shows the distribution of available score points over the content domains. For items containing a mixed content, the score points have been distributed over different content domains, if possible.

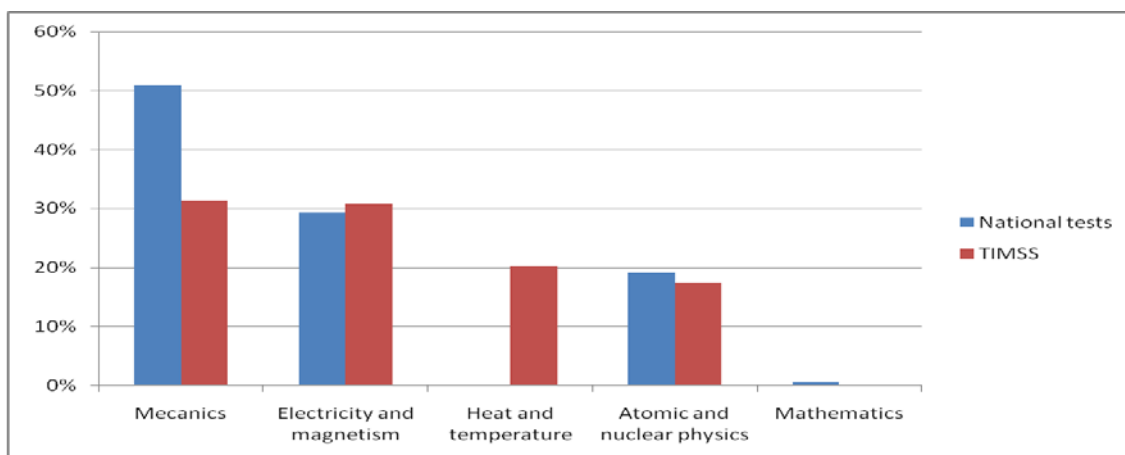


Figure 11 Distribution (in %) of available score points on content domains in physics for TIMSS Advanced 2008 and Swedish national tests.

The content domain of Heat and temperature is not covered by the Swedish tests in Physics B, but they are covered in Physics A and in Chemistry. Taking that into account the overall content coverage is fairly similar in the two sets of items, with the exception of an overrepresentation of Mechanics in the Swedish tests.

A comparison of score distributions for content subdomains point to some differences. Within Mechanics, the subdomains of waves (1.3), forces on ... (1.4), circular motion (1.5) and harmonic .. (1.8) are more well represented in the Swedish tests compared to TIMSS. Furthermore, the national tests emphasise electric and magnetic fields (2.3) and optics (2.5). As previously stated, heat and temperature is not represented at all in the national tests analysed in this study, but it is covered in preceding physics and chemistry courses. Finally, atomic structure (4.2) is a more frequent content in the national tests compared to TIMSS.

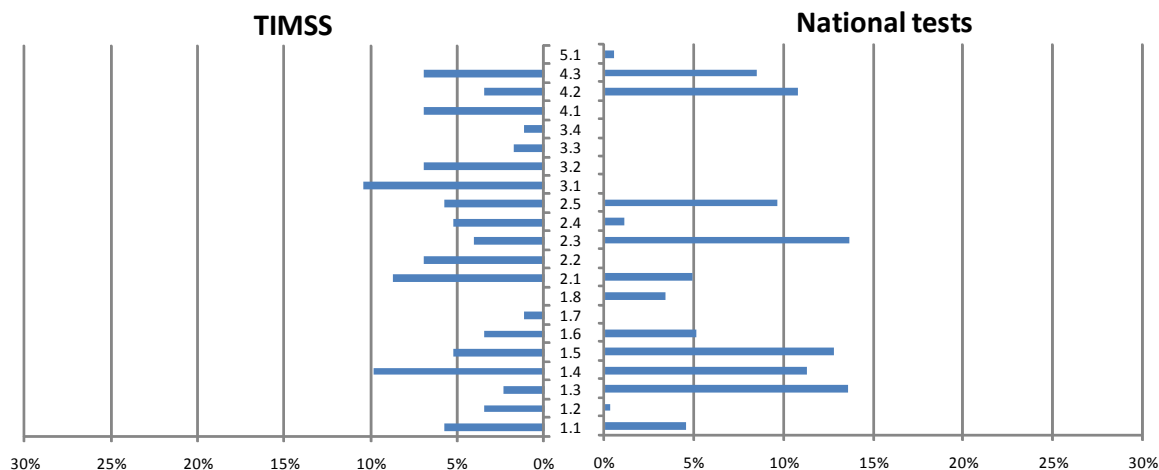


Figure 12 Distribution (in %) of available score points on content subdomains in physics for TIMSS Advanced 2008 and Swedish national tests.

3.2.2. Thought processes

3.2.2.1. Cognitive level

The comparison of cognitive levels required in the items indicates that Knowing is more focused in TIMSS, and that Reasoning is more focused in the national tests. Applying is however, equally emphasised in both systems. Swedish students miss the prerequisites to answer approximately 1 percent of the items in TIMSS, because they have not met necessary content.

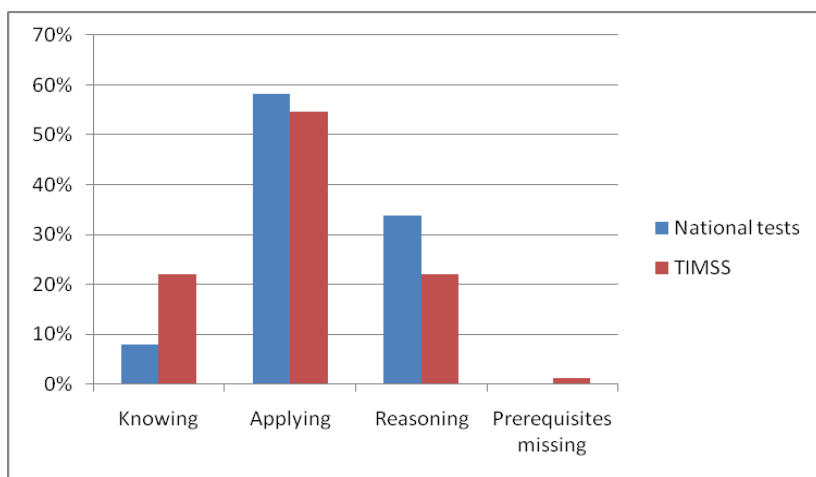


Figure 13 Distribution (in %) of available score points on cognitive levels in physics for TIMSS Advanced 2008 and Swedish national tests.

4.4.2.2 Calculations

The categorisation of need for calculations in items (see Figure 14) shows that items requiring no calculations or trivial calculations are more common in TIMSS, while the national tests focus more on complex calculations.

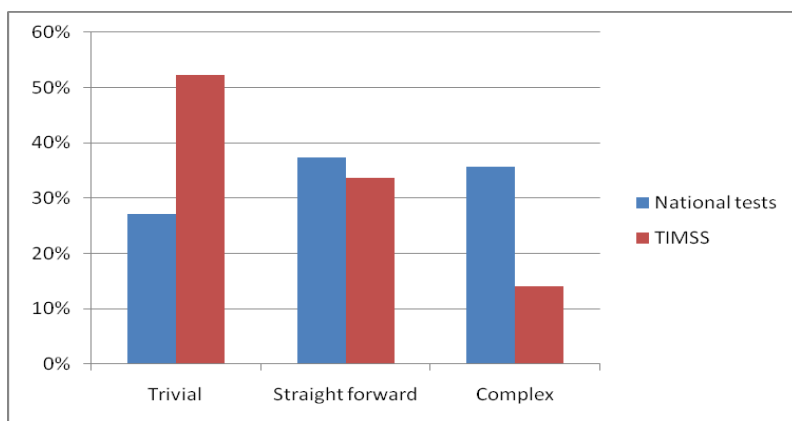


Figure 14 Distribution (in %) of available score points on types of calculations demanded in physics for TIMSS Advanced 2008 and Swedish national tests.

3.2.2.2. Tools

Figure 18 shows the extent to which calculators and available tables of formula are useful in each set of items. In many items, both tools are useful which causes the total percentage in figure 15 to exceed 100 percent. The usefulness of these tools in TIMSS is substantially lower than in the national tests. The rather comprehensive tables of formula allowed and used in the national tests makes them useful in 90 percent of the items. The tables of formula used in TIMSS are useful in less than half of the items, possibly because this set of formula is much more restricted.

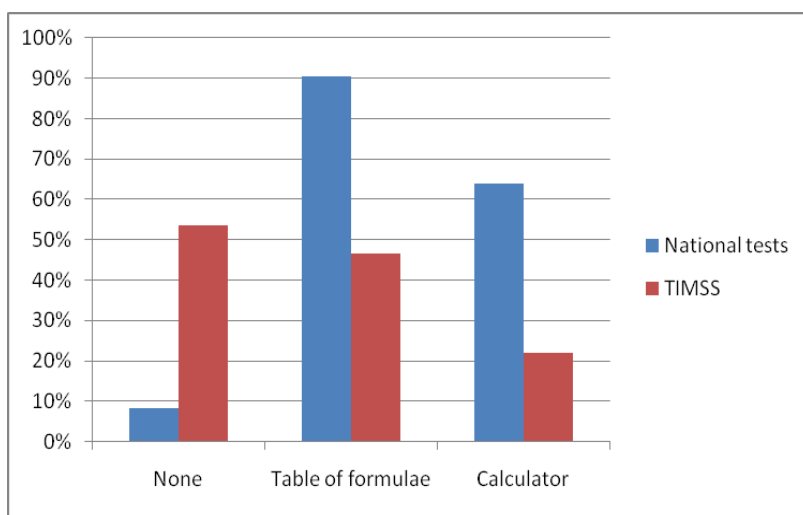


Figure 15 Distribution (in %) of available score points on usefulness of tools in physics for TIMSS Advanced 2008 and Swedish national tests. The category *Table of formulae* contains items for which the available table of formulae can be helpful in solving the item. The category *Calculator* contains items which would take substantially longer time to solve without the help of a calculator.

3.2.3. Demands on students' answers

3.2.3.1. Type of answer

In accordance with TIMSS Assessment framework, approximately half of the available score points in TIMSS come from multiple choice items. Short answers are represented by 26 percent of the available score points, and the corresponding percentage for long answers is 17. Short answers are most common in the national tests (49%) and long answers are also fairly frequent (36%). Multiple choice items are not very common (10%).

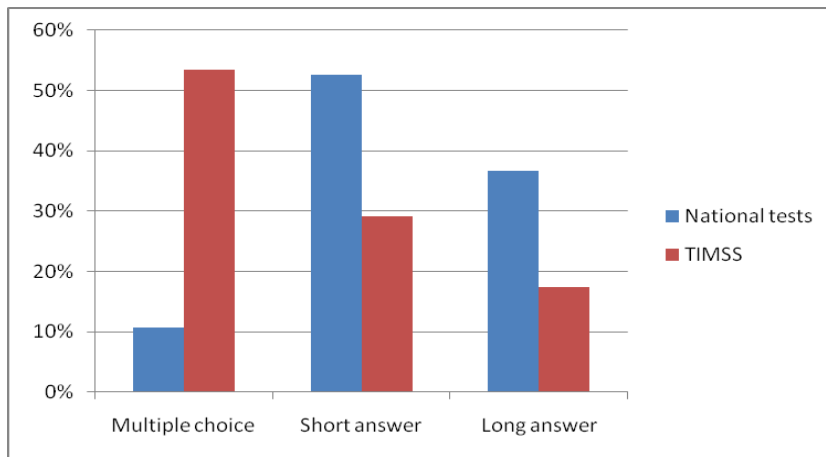


Figure 16 Distribution (in %) of available score points on item format in physics for TIMSS Advanced 2008 and Swedish national tests.

In the national tests in physics, 85 percent of the items require students to explain their answers. In TIMSS only 21 percent of the items require these extended answers.

3.2.4. Presentation

3.2.4.1. Word count

The TIMSS items contain 41 words per item on average (median 39) and the national tests contain 48 words per item (median 39). The difference between the two sets of items is very small when it comes to average number of words. Similarly to mathematics (but not to the same extent) the differences between mean value and median for the national tests indicates an overrepresentation of items with a lot of words. In the national tests, 12 percent of the available score comes from items with more than 100 words. In TIMSS there were no items with as many as 100 words. The item in the national tests with the longest text contained 166 words.

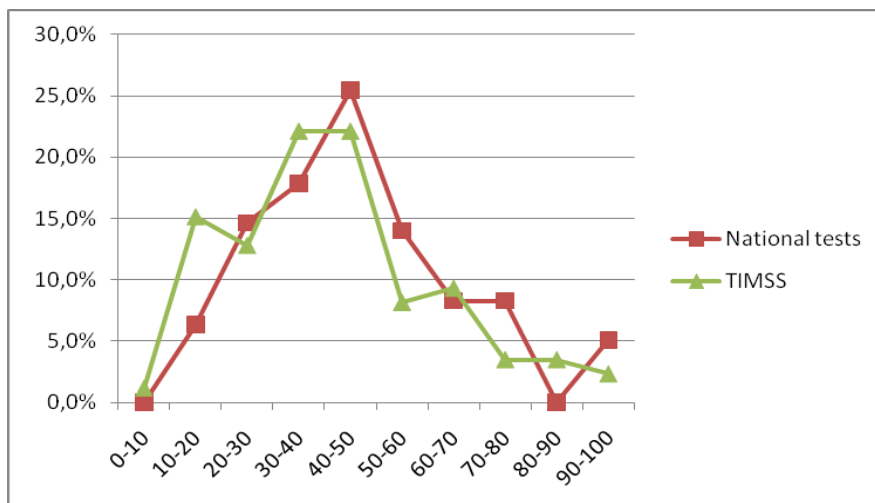


Figure 17 Distribution (in %) of text length per item, measured as number of words, physics for TIMSS Advanced 2008 and Swedish national tests.

3.2.4.2. Graphics

Figure 18 shows the percentage of available score points associated with different functions of graphical representations. Informative and decorative graphics is much more common in the national tests and the use of no graphical elements is dominating in TIMSS (60%).

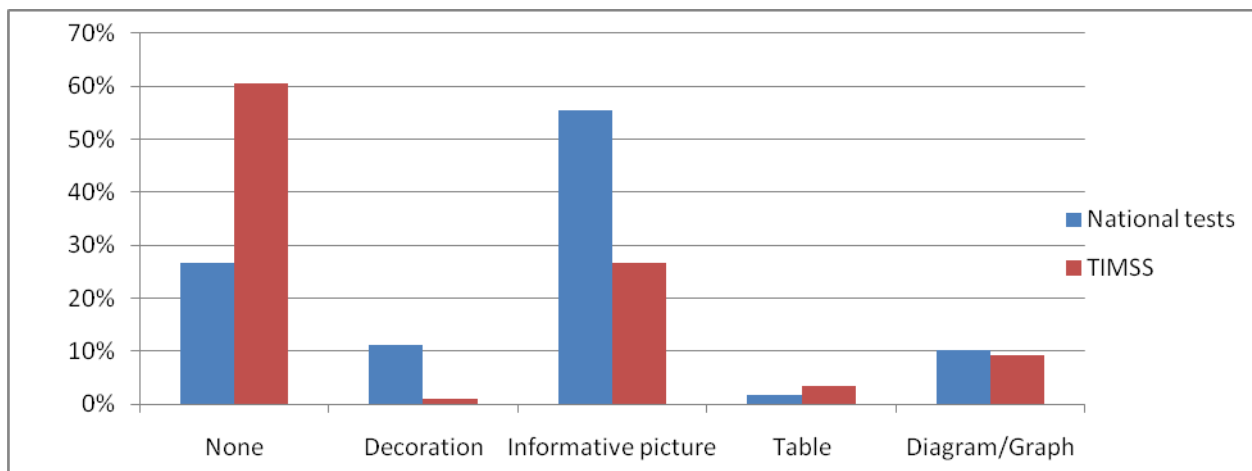


Figure 18 Distribution (in %) of available score points on use of graphical elements in physics for TIMSS Advanced 2008 and Swedish national tests.

3.2.4.3. Subject specific vocabulary

In order to be included in the population defined in the physics part of TIMSS Advanced 2008, Swedish students had to have taken both the first and the second physics course in upper secondary school, i.e. both Physics A and B. The TIMSS items contain a higher percentage of words that students can be expected to have met in Physics A, and the national tests contain more words introduced in Physics B. TIMSS also contains a somewhat larger percentage of words that the Swedish students most likely never have encountered.

Swedish students participating in the physics part of TIMSS Advanced 2008 had studied

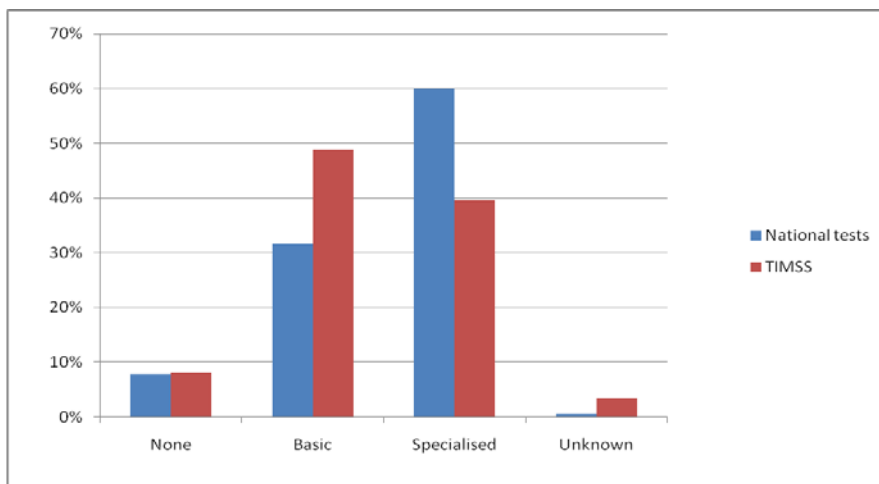


Figure 19 Distribution (in %) of available score points on occurrence of vocabulary in physics for TIMSS Advanced 2008 and Swedish national tests.

3.2.4.4. Reality

More than half of the items in both TIMSS and the national tests contain no context that connects to situations in real life outside the classroom. Dressed-up contexts, which poorly mirror any real life situation, are as common as contexts which can

Contexts connecting to real life are about as common as dressed-up contexts, i.e. items indicating a real world situation but using questions that are unlikely outside the classroom or requesting solution strategies that would not come into place in the real world situation. The distribution over these categories in TIMSS is similar to the distribution in the national tests.

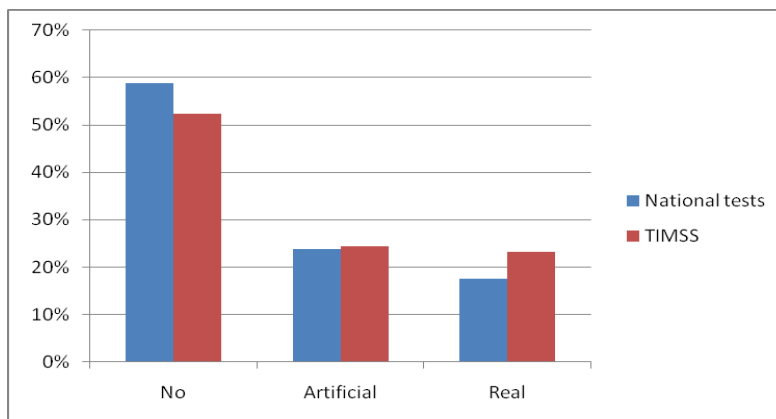


Figure 20 Distribution (in %) of available score points on type of out-of-school context in physics for TIMSS Advanced 2008 and Swedish national tests.

4. Conclusions and discussion

The study presented in this paper aimed at describing and analysing alignment between TIMSS Advanced 2008 and tests from the Swedish national assessment system. Four areas of alignment were focussed by using a framework specifically designed for this purpose. Based on the framework alignment was analysed with respect to subject domains, thought processes, demands on students' answers, and presentation of tasks.

4.1. Comparison of subject content

In mathematics TIMSS shows a fairly even distribution over Algebra, Calculus and Geometry. The national tests have a higher proportion of calculus compared to the other domains. At a first glance this could be interpreted as a major concern with respect to alignment, but it is important to keep in mind that upper secondary school in Sweden can be described as a modular system with successive courses. In mathematics there are five such courses, and the national tests analysed in this study are from the fourth of these courses, the one set as a threshold for inclusion in the Swedish TIMSS Advanced population. Most of the content covered in TIMSS is found in one or more of these courses, but it is not necessarily reviewed in later courses. An interesting issue worth further studies is to what extent this lack of cumulation is affecting what Swedish students know and can do at the end of upper secondary school.

In physics the distribution over content domains in TIMSS is different from the distribution in the national tests in that Mechanics is overrepresented in the national tests, and Heat and temperature is not represented at all. Heat and temperature is covered by the physics course preceding the one analysed in this study (the one required for participation in TIMSS), and to some extent in the chemistry course that all students participating in TIMSS also have taken. However, in the same way as for mathematics, this content might be harder for students by the time they do the TIMSS tests because this content has not been reviewed recently.

4.2. Comparison of thought processes demanded from students

Thought processes required by students in their work with the test items were analysed with respect to cognitive levels, calculations, and tools. In all three areas the patterns are very similar in mathematics and physics.

Concerning cognitive levels in mathematics, the national tests contain more Applying and Reasoning and less Knowing, compared to TIMSS. In physics the proportion of Applying is the same in the national tests and TIMSS, but the national test contain more Applying and less Knowing. As part of the categorisation into cognitive levels, a category of "prerequisites missing" was also included. Students can sometimes, using Reasoning, successfully answer an item containing unfamiliar content, but sometimes the item content is too specific to make it possible to solve even if sophisticated reasoning is used. One such example is complex numbers. If you have not encountered the idea of complex numbers, items covering that content are virtually impossible. Items of this kind have been categorised as *prerequisites missing*. In mathematics 7 % of the total attainable score comes from items in this category, indicating that there are very few items that Swedish students should be unable to solve on one of the cognitive levels defined (Knowing, Applying, or Reasoning). Some of the other items might require other cognitive levels from these students than what the international categori-

sation of items indicates. In physics there are hardly any items in the category prerequisites missing.

The analysed national tests in mathematics and physics contain relatively more complex calculations, and relatively less items requiring trivial or no calculations, compared to TIMSS. This might be consistent with the national tests having a relatively smaller proportion of items focussing the cognitive level of Knowing. It is noticeable that the Swedish national syllabus in physics is not characterised by a strong focus on the role of mathematics in physics, and yet calculations seem to have a stronger role in the national tests compared to TIMSS.

The comparison of how tools such as calculators and tables of formulae are used in TIMSS compared to the national tests in mathematics and physics shows fairly large differences. One reason for these large differences is that the tables of formulae allowed are very different for TIMSS compared to the national tests. In Swedish national tests the tables of formulae are comprehensive, making them potentially useful for a lot of items. Students are not expected to remember formulae in these tests, they have tools that make the Knowing part less important and necessary. However, this is a difference in theory. We don't know how much these tools actually are used. Another reason for differences regarding tools is that the stronger focus on Application (and possibly also the stronger focus on the more complex problem-solving included in the Reasoning category) makes calculators more useful in the national tests.

4.3. Comparison of demands on students' answers

There are obvious differences between TIMSS and the national tests with respect to the format of answering items. The multiple choice format is much more common in TIMSS than in the national tests, and the national tests are much more focused on students not only giving a short answer, but actually showing their work leading to a final answer. Expecting students to show their work enables better opportunities to assess students reasoning, but also their communicative competence. In TIMSS the prime focus is on arriving at a correct final answer to an item, while national tests often focus the process giving students partial credits for e.g. choice of appropriate method.

4.4. Comparison of item presentation

The way tasks are presented was analysed with respect to text length (number of words per item), graphical elements, subject-specific vocabulary, and out-of-school contexts.

According to this study, mathematics items in TIMSS generally have a text length similar to the text length of items in the national tests. There is however a striking difference in that the national tests contain a few items with very long texts, while TIMSS does not. The use of graphical elements in mathematics items differs to some extent in that pictures and diagrams/graphs are more common in the national tests. Concerning subject-specific vocabulary, the difference between TIMSS and national tests in mathematics is small, but a slight overrepresentation of specialised vocabulary is found in the national tests. One fifth of the available score points in mathematics comes from items with an out-of-school context in TIMSS and also in the national tests. However, out-of-school contexts that are "real" (as opposed to "artificial") are much more common in the national tests.

In physics the differences in text length per item are similar to the pattern found in mathematics, but not quite as pronounced. Generally items in TIMSS have the same text length as in the national tests, but the national tests contain some very long item texts. Concerning the use of

graphical elements in physics, the national tests contain more of both informative and decorative pictures. Specialised vocabulary is also more common in the national tests as compared to TIMSS. Though physics mostly shows the same patterns as mathematics in this study, the results differ in the comparison of out-of-school contexts. Items with no out-of-school context are more common in the national tests than in TIMSS, and TIMSS has a larger proportion of items with an out-of-school context that has been categorised as “real” (as opposed to “artificial”).

Generally, the differences between TIMSS and the national tests regarding presentation of tasks are fairly small and will most likely not be a major alignment threat. To some extent the differences can be seen as a mirror of the different functions of these tests. While TIMSS primarily is designed for a measurement purpose, the national tests also have an ambition to implement a national curriculum where for example verbal components are seen as important.

The issue of text length is particularly interesting and problematic. Items with longer texts are often assumed to put higher demands on students concerning understanding the problem and finding relevant information. While a longer text (and, even more, a complex text) can put relevant demands on students subject specific reading abilities and their understanding of subject specific vocabulary, these texts can also create irrelevant reading obstacles for students. This could imply that TIMSS, with items using fewer words, could be less susceptible to students general reading ability. On the other hand, shorter texts can also, due to their compactness, cause severe reading difficulties, in particular if students are not sufficiently socialised into the specific genres of item texts in mathematics and physics.

4.5. Concluding remarks

The aim of this study was to investigate the alignment between items used in TIMSS Advanced 2008 and items used in Swedish national tests at the same level. The degree and character of this alignment is important for the credibility and usefulness of the results from TIMSS. A deeper understanding of alignment (or lack of alignment) can also highlight characteristic features in the national system. The overall conclusion is that TIMSS Advanced 2008 and Swedish national tests in mathematics and physics have a high degree of alignment supporting valid interpretations of the TIMSS results. There are however interesting differences well worth studying further.

5. References

- Baker, E. L., & Linn, R. L. (2002). *Validity issues for accountability systems*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Bernstein, B. (1971). On the classification and framing of educational knowledge. I M.F.D. Young (Ed.). *Knowledge and Control; New directions for the sociology of education*. London: Collier-Macmillan.
- Clune, W. H. (2001). Toward a theory of standards-based reform: The case of nine NSF statewide systemic initiatives. In S. H. Fuhrman (Ed.), *From the capitol to the classroom. Standards-based reform in the States*. Chicago: National Society for the Study of Education (NSSE).
- Garden, R. A., Lie, S., Robitaille, D. F., Angell, C., Martin, M. O., Mullis, I. V. S., Foy, P., & Arora, A. (2006). *TIMSS Advanced 2008 Assessment Frameworks*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Lindström, J-O. (2006). *Med fokus på matematik och naturvetenskap* (Rapport nr. 221) [In Swedish]. Stockholm: Skolverket.
- Mullis, I.V.S., Martin, M.O., Robitaille, D.F., & Foy, P. (2009). *TIMSS Advanced 2008 international report: Findings from IEA's study of achievement in advanced mathematics*

- and physics in the final year of secondary school*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Nyström, P. (2008). *Identification and analysis of text-structure and wording in TIMSS-items*. Paper presented at the The 3rd IEA International Research Conference. Retrieved January 12 2009, from http://www.iea.nl/fileadmin/user_upload/IRC2008/Papers/TIMSS_Mathematics/Nyström.pdf.
- Palm, T. (2006). Word problems as simulations of real-world situations: A proposed framework. *For the Learning of Mathematics*, 26(1), 42-47.
- Palm, T., & Nyström, P. (2009). Gender aspects of sense making in word problem solving. *Journal of Mathematical Modelling and Applications*, 1(1), 59-76.
- Roach, A.T., Niebling B.C., & Kurz, A. (2008). Evaluating the alignment among curriculum, instruction and assessments: Implications and applications for research and practice. *Psychology in the Schools*, 45(2), 158-176.
- Vedder, P., & O'Dowd, M. (1999). Empowering teachers in times of change. The comprehensive school system. *Scandinavian Journal of Educational Research*, 43(3), 313-326.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (Research Monograph No. 6). Madison: National Institute for Science Education.

Appendix 1

Tabell 1 Mathematics domains and subdomains used in the categorisation of items from national tests and TIMSS Advanced 2008

<i>Code</i>	<i>Domain</i>	<i>Code</i>	<i>Subdomain</i>
1	Algebra	1.1	Complex numbers
		1.2	Series and sums
		1.3	Combinations, probability, statistics
		1.4	Solve equations and inequalities, including logarithmic and exponential
		1.5	Construct or recognise graphs, tables, ordered pairs and text representing a function
		1.6	Determine signs and values of functions, including rational functions, for given values and ranges of the variable. Evaluate a function of a function.
		1.7	Solve and handle trigonometric functions and equations
2	Calculus	2.1	Limits, continuity and differentiability
		2.2	Differentiate polynomials and functions. Apply the chain rule and rules for differentiating products and quotients
		2.3	Use derivatives to solve problems
		2.4	Use first and second derivatives to determine extreme values and to sketch graphs for functions
		2.5	Integrate functions, apply integration and evaluate integrals numerically
		2.6	Solve differential equations analytically and numerically, and interpret their meaning
3	Geometry	3.1	Use the properties of geometric figures to solve problems. Prove straightforward geometric propositions in two and three dimensions
		3.2	Use gradients, y-axis intercepts, and points of intersection of straight lines in the Cartesian plane in solving problems
		3.3	Know and apply the equations and properties of circles in the Cartesian plane.
		3.4	Use trigonometry to solve problems involving triangles
		3.5	Basic handling of vectors

Appendix 2

Tabell 2 Physics domains and subdomains used in the categorisation of items from national tests and TIMSS Advanced 2008

<i>Code</i>	<i>Domain</i>	<i>Code</i>	<i>Subdomain</i>
1	Mecanics	1.1	Newtons first and third law, pressure in liquids, and conditions for equilibrium
		1.2	Kinetic and potential energy, and the conservation of mechanical energy
		1.3	Wave mechanics (sound, water and strings), relationship between speed, frequency and wavelength, refraction and reflection to thicker/thinner media, standing waves
		1.4	Forces (including frictional force) acting on a body moving with constant acceleration and the forces influences on the body's movement. Calculations of speed and and acceleration.
		1.5	Relations between force, speed, acceleration and cirkling time for a body moving in a circular path. The law of gravitation
		1.6	Using the laws for impuls and conservation of momentum and energi in elastic and inelastic collisions
		1.7	Special relativity, including lenght contraction and time dilatation
		1.8	Harmonic oscillations, pendulum
2	Electricity and magnetism	2.1	Coulombs law. Forces on a charged particle in a homogeneous electric field
		2.2	Electrical circuits and components of circuits, Ohm's law, Joule's law
		2.3	Forces on charged particles in magnetic fields, relationship between electricity and magnetism, induction, Lenz' law, Faraday's law
		2.4	Electromagnetic radiation, identify different typs of radiation from wavelength and frequency
		2.5	Optics, refraction index, diffraction, transmission gratings, slits, inference, absorpction, reflexion
3	Heat and temperature	3.1	Difference between temperature and heat. Radiation, conduction and convection. Applicatgion of temperature equilibrium. Specific heat capacity, evaporation, condensation, melting/freezing
		3.2	Heat expansion for solids and liquids, simple applications of the law of ideal gases, first law of thermodynamics
		3.3	Temperature dependence of black body radiation, estimate temperature based on the wavelength of radiation
		3.4	Fundamental principle behind the green house effect
4	Atomic and nuclear physics	4.1	Structure of the atom and atomic nucleus. Apply knowledge about atomic number and atomic mass number
		4.2	Absorption- and emissionspectra of atoms. Quantized energilevels of electrons. Photoelectric effect, x-ray emission
		4.3	Fission and fusion. Radioactive decay, half-lives, and harmful effects
5	Mathematics only	5.1	Items requiring only mathematics and a table of formulae

Tabell 7 Cognitive levels and criteria for categorisation used in categorising items in mathematics and physics from Swedish national tests and TIMSS Advanced

Cognitive level (code)	Tasks requiring students to	Criteria for categorisation	
		Mathematics	Physics
Knowing (1)	account for facts and concepts or carry out basic procedures	<p>Recall definitions, terminology, notation, mathematical conventions, number properties, geometric properties.</p> <p>Recognize entities that are mathematically equivalent (e.g., different representations of the same function or relation).</p> <p>Carry out algorithmic procedures (e.g., determining derivatives of polynomial functions, solving a simple equation).</p> <p>Retrieve information from graphs, tables, or other sources.</p>	<p>Recognize and use scientific vocabulary, symbols, abbreviations, units, and scales in relevant contexts.</p> <p>Make or identify accurate statements about physical facts and relationships.</p> <p>Explain properties and relationships through descriptions of physical properties.</p> <p>Describe physical materials and processes demonstrating knowledge of properties, structure, function, and relationships.</p>
Applying (2)	use factual knowledge in order to choose and generate models or solve routine tasks	<p>Select an efficient/appropriate method or strategy for solving a problem where there is a commonly used method of solution.</p> <p>Generate alternative equivalent representations for a given mathematical entity, relationship, or set of information.</p> <p>Generate an appropriate model such as an equation or diagram for solving a routine problem.</p> <p>Solve routine tasks, (i.e., tasks similar to those students are likely to have encountered in class). For example, differentiate a polynomial function, use geometric properties to solve problems.</p>	<p>Explain physical concepts using models or diagrams (e.g. electric circuit, atomic structure)</p> <p>Choose a suitable method or strategy to solve problems that can be solved using straight-forward applications of physical relationships, equations, and formula.</p> <p>Explain observations and natural phenomena using physical laws or theories.</p>
Reasoning (3)	generalize/analyze a result or generate models for complex or unfamiliar problems, and solve them	<p>Investigate given information, and select the mathematical facts necessary to solve a particular problem.</p> <p>Determine and describe or use relationships between variables or objects in mathematical situations.</p> <p>Make valid inferences from given information.</p> <p>Extend the domain to which the result of mathematical thinking and problem solving is applicable by restating results in more general and more widely applicable terms.</p> <p>Combine (various) mathematical procedures to establish results, and</p>	<p>Analyze problems to determine the relevant relationships, concepts, and problem-solving steps.</p> <p>Make general conclusions and determine general formula in order to express physical relationships. Apply conclusions to new situations.</p> <p>Solve problems requiring knowledge and skills from mathematics as well as physics.</p> <p>Justify whether a statement is true or false with reference to physical relationships.</p> <p>Formulate hypotheses and make pre-</p>

Appendix 3

	<p>combine results to produce a further result. Make connections between different elements of knowledge and related representations, and make linkages between related mathematical ideas.</p> <p>Provide a justification for the truth or falsity of a statement by reference to mathematical results or properties.</p> <p>Solve problems set in mathematical or real-life contexts where students are unlikely to have encountered similar items, and apply mathematical procedures in unfamiliar or complex contexts.</p>	<p>dictions of physical phenomena using understanding of physics.</p> <p>Make conclusions from and detect patterns in data. Make valid inferences on the basis of evidence and/or understanding of physics concepts.</p>
Prerequisites missing (4)	Need knowledge beyond what students have encountered	
