

Textbook based analysis of TIMSS Advanced – Item familiarity

Peter Nyström

Annika Kjellsson Lind

Abstract

As a national option, teachers were asked to report which textbooks they used in teaching mathematics and physics to the students participating in TIMSS Advanced 2008. This paper reports a study using textbooks as a clue to the implemented curriculum in Sweden. Familiarity with the items in TIMSS was studied based on the occurrence of similar items in the textbooks. The results indicate that the mere existence of similar items does not affect students probability of getting the answers right.

Keywords: TIMSS advanced, student achievement, implemented curriculum, cognitive domains

Introduction

Educational or academic significance of the study

A deeper understanding of the TIMSS results requires a deeper look at the items used and their features. In particular, the cognitive categorisation can give important insights into the strengths and weaknesses of students in a particular country with respect particular kinds of knowledge and understanding. The regular categorisations of items made in TIMSS is to some extent limited by the cross-cultural characteristics of the study, i.e. the categorisation of items in relation to cognitive domains may be culturally biased. For example can something that is considered to be recall or routine in one country be reasoning and non-routine in another. We cannot assume that these cognitive aspects of items are the same in different countries. Furthermore we cannot even assume that they are the same for all students within a country. The classroom experiences of the students participating in TIMSS would of course be useful here, but to get access to classroom data on a large scale is of course very hard. Since textbooks are a vital part of the implemented curriculum we can by knowing which textbooks that have been used by the students, and analysing them, learn a great deal about the implemented curriculum. Thereby we can both reanalyze the TIMSS data in order to describe for example students achievement on familiar and non-familiar content and cognitive domains. This will contribute to a richer use of TIMSS data and also contribute to the validation of TIMSS Advanced.

From the knowledge of which textbook that each student has been learning from, a more valid cognitive categorisation can be made, and more accurate descriptions of achievement in the different cognitive domains.

The concept of item familiarity

Familiarity with test items is relevant for several reasons. One reason is that familiarity is assumed to lower the difficulty of the test item (measured as frequency of correct answers). If different students have different familiarity with an item, this can be seen as differential item functioning since different types of learning are evoked and used by different students when they try to answer the item. Another reason is that familiarity is used as a criterion for discrimination between different cognitive levels and aspects in assessment.

For TIMSS Advanced 2008 the assessment framework (Garden, et al., 2006) refers to aspects of item familiarity when defining different cognitive levels. The difference between applying and reasoning in mathematics is described as problem settings for items being more routine in the applying domain, and “will typically have been standard in classroom exercises designed to provide practice in particular methods or techniques”(p. 19). Furthermore it is stated that “each of these type of “textbook” problems is expected to be sufficiently familiar to students that they will essentially involve selecting and applying learned procedures” (p. 19).

In the definition of the applying domain “solve routine problems” is included, with routine problems defined as “problems similar to those students are likely to have encountered in class” (p. 20). In the reasoning domain, the corresponding subdomain is “solve non-routine problems”. This is defined as “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. 22).

This particular difference between applying and reasoning is also found in the physics part of TIMSS Advanced 2008. The applying domain talks about “contexts likely to be familiar in the teaching and learning of physics concepts”, and reasoning is associated with unfamiliarity: “In addition to the more direct applications of physics concepts exemplified in the applying domain, problem-solving situations in the reasoning domain involve unfamiliar or more complicated contexts that require students to reason from scientific principles to provide an answer.”

The TIMSS Advanced assessment framework seems to emphasise familiarity as a criterion for distinguishing between the cognitive aspects of applying and reasoning more strongly in mathematics compared to physics. Furthermore, while familiarity in mathematics is primarily connected to standard exercises in the classroom, for physics it is only spoken of as familiarity of contexts. In addition, while the framework in physics clearly states that a range of difficulty levels should be expected for items developed for each of the cognitive domains, i.e. we should not expect items categorised as reasoning to generally be more difficult than items categorised as applying. No such statement is made concerning mathematics.

From a text-analysis of the TIMSS Advanced Assessment Framework, several different conceptualisations of familiarity can be interpreted. Familiarity is emerging as

- Routine problems are familiar since they are rooted firmly in the implemented curriculum
- Facts, concepts and procedures can be familiar to students.
- contexts can be familiar in the teaching and learning of physics concepts (physics)
- mathematical problems are denoted familiar if they are standard problems similar to those encountered in class
- familiarity concerns whether or not students have encountered similar items

The difference between the last two lies in the difference between similarity between items in general and the idea of well-known, standard problems. The list above shows several examples of the possible interpretations and operationalisations of item familiarity, a diversity which is also seen in research on this topic.

Previous research on item familiarity

Research on item familiarity seems to be highly focused on memory and language. One example is Ingleby (1973) who studied the effect of item familiarity on memory using recall of surnames. The author concludes that several studies over a long period of time have shown (not surprisingly) that common items or words are more likely to reproduce than uncommon ones. Based on the results from this study, Ingleby (1973) concludes that unfamiliar items are actually better preserved or retrieved than familiar ones, but are more reluctantly emitted as responses.

Another example is Tunnie and Fernie (2007) who conclude that recognition memory is composed of at least two functionally distinct processes, recollection and familiarity. This theoretical position known as Dual process theory has been substantiated with a lot of research. Proponents of this theory argue that recollection and familiarity give rise to distinct phenomenal states. Recollection-based memory includes awareness of episodic aspects of the study event, such as the context, while familiarity-based memory is characterised as a person feeling that an item was studied but no specific information about the study event is re-experienced. In one of their experiments they instructed participants to report their experience of remembering words presented on a computer screen using the categories remember, know or guess, whenever they indicated having seen an item before. The difference between remembering and knowing is described to the participants using the example of face recognition. Recollection parallels when “a face is recognized along with contextual details of the meeting such as the name of the person and the topic of conversation”. “Familiarity is described as when a face is recognized but without any of the accompanying contextual details such as the name”.

Item familiarity has also been explored as a possible cause of differential item functioning. Stricker and Emmerich (1999) conclude that examinees’ familiarity with the content of the items, which is variously referred to as exposure, experience, or cultural loading, is the most common and most widely discussed explanation to DIF. In their study of a psychology examination test a rating scale was used in order to measure familiarity. Examinees were asked to rate each item by answering the question: How familiar is what this test question asks about? Ratings were given from Very Familiar (4) to Not Familiar at All (1). The study focused specifically on gender differences, and found no gender differences for familiarity. The authors argue that the absence of such differences may be due to the use of a disclosed test, while the students were asked to rate the familiarity of the topics covered by the items, not the items themselves.

Concerning the more specific issue of item familiarity in large-scale assessment, Ruddock, Clausen-May, Purple, & Ager (2006) performed a study of items from TIMSS and PISA. They used teams of consultants who were very experienced in assessment at the relevant levels. The number of consultants in each panel (mathematics, science and reading) is not explicitly stated in the report, but in passing it is mentioned that there were nine consultants altogether.

The mathematics and science consultants were asked to rate the familiarity of the *concept*, the *context* and the *format* of each part of each question.

The rating scale and the codes used were

	Not in Curriculum	Percentage of Students Estimated to be Familiar with the Concept/ Skill, Context/Text type, or Format/Task				
		Very few	Fewer than half	About half	More than half	All or almost all
	None	about 5%	about 25%	about 50%	about 75%	about 95%
Code	0	1	2	3	4	5

Familiarity with respect to concepts, contexts and format was further elaborated in the study. In order to be considered familiar with a mathematical or scientific *concept*, students should have studied it in class and they should feel confident that they understand it. Concerning mathematical or scientific *context*, familiarity requires students to understand any ‘story’ and/or graphics associated with it, and any mathematical/scientific context. They must feel confident that they know what the question is asking. With respect to *format* of a question, familiarity is defined as having met similar questions in the same format, presented in the same sort of language. With respect to mathematics it is noteworthy that the consultants performing the rating exercise experienced that the rating was affected by whether students are accustomed to questions where the difficulty has been significantly increased by testing diverse skills within the same question.

Ruddock et al. (2007) found similar ratings for TIMSS and PISA in Science. They conclude that this result indicates that different levels of performance in PISA and TIMSS are unlikely to be caused by differential familiarity. For mathematics, on the other hand, they found higher ratings for familiarity with context and format for TIMSS than for PISA. Typically, items in TIMSS were familiar to 50% of the students. Ruddock et al. conclude that this moderate level reflects the differentiation present in the English national curriculum as taught to these age groups.

While item familiarity in the psychology literature presented above concerns remembering identical items, the study of Ruddock et al. (2007) conceptualises familiarity as something closer to the more general concept of alignment. Items are here considered familiar if students have been given the opportunity to learn the concepts found in the items, are able to identify with the context, and having met the same kind of questions (format).

Aim

While Ruddock et al (2006) and others (e.g. Lindström, 2006) have studied the overall familiarity of items in TIMSS and other international comparative studies, our focus is on each student’s familiarity with the items and the relation between familiarity and achievement on an individual level.

The overall aim of the study is to investigate the relationship between item familiarity and achievement in TIMSS Advanced 2008. The study reported here is limited to the TIMSS Advanced items that were released after the 2008 study. The prime reason for this limitation is that the released items are accessible and possible to discuss openly. Furthermore, the study is based on a definition of item familiarity. A TIMSS item is considered familiar for a particular student if at least one almost identical item is found in the textbook used by the student. The hypothesis is that the existence of identical items will benefit student performance and be reflected in the p-values of familiar items.

Method

As a national adaptation, questions about which textbook each teacher was using in the TIMSS class were added to the Swedish teacher questionnaire in TIMSS Advanced 2008. It is known from other studies, and confirmed in TIMSS Advanced, that textbooks are very important for teachers in mathematics and physics. Therefore, data about which textbook that is used in the classes participating in TIMSS Advanced, is a way to look into the classroom and to get additional information about the implemented curriculum.

The questions about textbooks were developed through a thorough scan of the text-book market. All available text-books were listed and for each text-book the teacher was asked to mark whether they used that particular text-book or not. Five text-books were listed in mathematics, and five in physics. This made it possible for teachers to mark more than one text-book if they used more than one in the TIMSS class. Teachers were also given the opportunity to indicate if they used any other text-book than the ones in the list. If so, they were asked to name the text-book.

As a first approximation, the existence of items identical or almost identical to TIMSS items is used as an indicator of item familiarity. The most frequently used textbooks were analysed in order to identify items that resemble the released items from TIMSS Advanced. Probability correct values (p-values) for students who have used different textbooks were analysed.

Results

Textbooks used in classrooms of Advanced mathematics in Swedish upper secondary school

Table 1 presents how frequently different mathematics textbooks were used by students participating in TIMSS Advanced 2008 in Sweden, according to their teachers. The table also presents mean plausible values for each group of students.

Table 1 Textbooks used in Swedish upper secondary education in advanced mathematics according to responses to the teacher questionnaires in TIMSS Advanced 2008, and mean achievements for each group of students.

	n	totwgt	Mnpv	Mnpv_se	Pct	Pct_se
Exponent	128	896	415,19	10,00	5,45	2,59
Matematik 3000	1831	12826	414,80	6,61	78,07	4,90
Matematik från A till E	26	163	439,54	8,18	0,99	0,75
Nya Delta	11	154	342,82	19,49	0,94	0,93
Optima	46	415	428,99	13,07	2,53	1,88
Pyramid	81	565	387,29	9,05	3,44	2,16
More than one textbook	99	643	406,23	41,04	3,92	2,44
Missing	148	766	394,71	19,12	4,66	1,96

The results in table 1 show that one textbook dominates in advanced mathematics classrooms in the Swedish upper secondary school. The other textbooks are used by small number of students. There are no significant differences in achievement between students who have experienced different textbooks.

The dominating role of one textbook makes the design and aim of this study difficult to pursue in mathematics, comparing performance of students who have learned mathematics using different textbooks. However, this situation also opens opportunities for other studies. If, for example, we can

validate that the same textbook series dominated the market in 1995 as well, we can analyse to what extent differences between the two editions can be used to understand the large differences in Swedish results between 1995 and 2008.

Description of the most frequently used textbook in mathematics

The main structure of the most widely used textbook is a division into chapters, sections and subsections. A typical subsection starts with a short introduction of theory, definitions, or procedures, often with the help of an example. This introduction is followed by exercise examples, which are solved and sometimes commented. The subsection is completed by a number of exercises, sorted in three types (A,B, and C) based on complexity and difficulty. These exercises are mainly of the same type as the ones shown as exercise examples in the beginning of the subsection.

In some subsections no new content is introduced. Instead the subsection focuses on the application of the particular content. In these subsections an introductory problem is solved and commented, followed by a number of exercises. Here and there in the book a page focusing the history of mathematics is found. At the end of each chapter there is a set of homework tasks, a set of problemsolving tasks and a set of tasks that should be solved without using a calculator. The final chapter contains a number of extended tasks in mathematics, because the national syllabus requires students to independently analyse and carry through an extended task, and present their solutions in writing and orally. At the end of the book we find a number of tasks intended for revision of the complete book, correct answers to all of the exercises in the book, and finally an index.

Occurrence of items in mathematics textbooks, similar to the released TIMSS items

Out of the 40 released mathematics items from TIMSS Advanced 2008, 29 were considered relevant for the Swedish curriculum. Among these, only five identical (or near identical) items were found in the dominating textbook. One example of such an item is shown in figure 1.

The functions f and g are defined by $f(x) = x - 1$ and $g(x) = (x + 3)^2$.

$g(f(x))$ is equal to

(A) $(x - 1)(x + 3)^2$

(B) $(x + 3)^2 - 1$

(C) $(2x - 2)^2$

(D) $(x + 2)^2$

(E) $x^2 + 8$

Figure 1 An item (MA13001) from TIMSS Advanced 2008 with similar items found in the dominating Swedish textbook in mathematics.

There are several similar items in the textbook, using similar but not exactly the same polynomials. The notation in the TIMSS item was also consistent with the notation in the textbook.

Another interesting example of a TIMSS item with close to identical counterparts in the textbook is shown in figure 2.

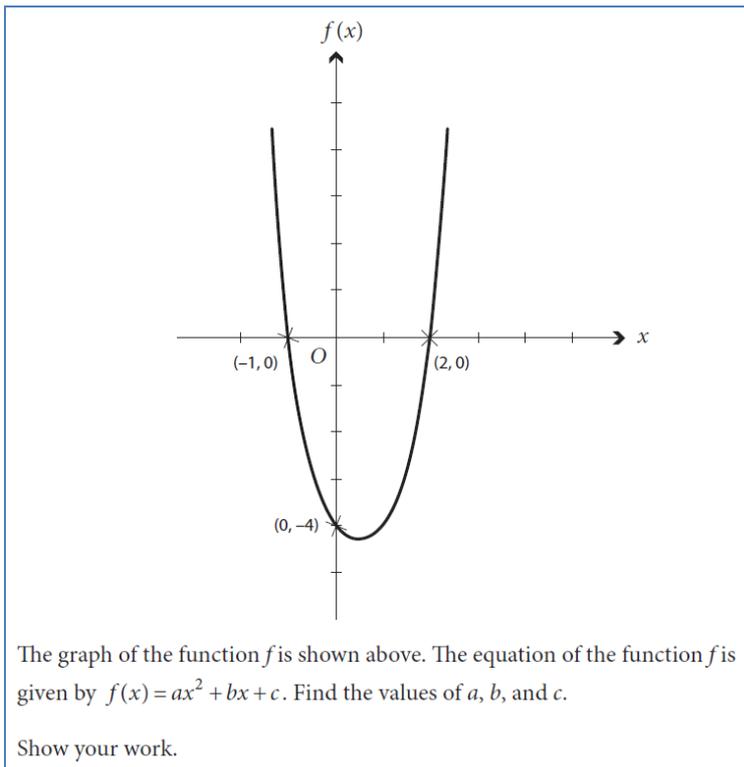


Figure 2 An item (MA23141) from TIMSS Advanced 2008 with similar items found in the dominating Swedish textbook in mathematics.

Students should have had ample opportunities to practice this because there are several items of this kind in the textbook. Despite this, less than 10 percent of the Swedish students got the item right.

An example of an item which has a mathematical content that is definitely covered in the textbook, but still has no counterpart as an item in the textbook is shown in figure 3.

$f(x) = e^{\cos x}$

What is $f'(x)$?

(A) $e^{\cos x}$

(B) $e^{-\sin x}$

(C) $e^{\cos x} \cdot \sin x$

(D) $-e^{\cos x} \cdot \sin x$

Figure 3 An item (MA23039) from TIMSS Advanced 2008 with no similar item in the dominating Swedish textbook in mathematics, even though the mathematical content is covered in the textbook.

57,5 % of the Swedish students got this item right, even though there is no combination of exponential and trigonometric functions of this kind found in the textbook. The skill required to answer this item correctly is by nature fairly general and applies to a wide range of combination.

A final example from mathematics is an item which is not found in the Swedish textbook, and the complexity of it can require reasoning if you have not encountered the specific challenges in this particular problem.

A car starts braking as it approaches a road junction. After braking for t seconds, the car has traveled a distance of $s(t)$ meters, where $s(t) = -t^2 + 20t$. How far does the car travel from the time the brakes are applied until it stops?

(A) -20 m
 (B) 10 m
 (C) 50 m
 (D) 100 m

Figure 4 An item (MA23158) from TIMSS Advanced 2008 with no similar item in the dominating Swedish textbook in mathematics.

The item example in figure 4 shows that there are items categorised as Applying, which for Swedish students most likely will require Reasoning. Does this mean that there is a lot of reasoning required by Swedish students when working with items from TIMSS Advanced? Not necessarily. Mathematics items can be varied a lot, and this variation actually captures the generality of the mathematics learned. Of course mathematics has the same problem of transfer as all other kind of learning, but the subject itself influences this to a high degree.

Textbooks used in classrooms of Advanced physics in Swedish upper secondary school

Table 2 presents how frequently different physics textbooks were used by students participating in TIMSS Advanced 2008 in Sweden, according to their teachers. The table also presents mean plausible values for each group of students.

Table 2 Textbooks used in Swedish upper secondary education in physics according to responses to the teacher questionnaires in TIMSS Advanced 2008, and mean achievements for each group of students.

	N	Totwgt	Mnpv	Mnpv_se	Pct	Pct_se
Heureka	665	4599	501,24	12,45	31,55	5,62
Ergo	636	3999	488,28	9,36	27,44	5,49
Nexus	501	2773	509,39	8,01	19,03	3,42
Quanta	255	1514	496,04	20,74	10,39	3,66
Orbit	25	155	462,13	24,39	1,06	0,82
Other textbook	54	270	489,54	35,36	1,85	1,14
More than one textbook	32	76	483,47	10,90	0,52	0,52
Missing	221	1188	493,62	11,64	8,15	2,58

In physics, four different textbooks are used relatively frequently, and further analysis is therefore focused on these four. The only significant achievement difference is found between students taught by teachers using Nexus and Ergo, with Ergo connected to the lower performance.

A brief description of textbooks in physics

In the first course in physics in upper secondary school (Physics A) three of the four textbooks start with a basic description of the subject physics, covering topics like: *What is a model? How do we know things?* and *How do we measure things?* Basic understanding like the SI-system and the relations between different units are also included in these introductory parts. The only book that does not start with a chapter covering this topic is Quanta, which instead starts directly with the traditional physics topic of *Optics*.

All four books are divided into chapters which in turn are divided into sections. Solutions to problems are shown in every chapter, mixed with theory throughout all four books, and they all have sections

trying to connect the physic-topic of the chapters to real world phenomena. At the end of every chapter there are exercises for the pupils to work with.

There are however also differences between the textbooks. Based on these differences the textbooks can be grouped with Heureka and Quanta in one group and Nexus and Ergo in another group. The order of the chapters is the same in Heureka and Quanta, even though the chapters in Quanta are wider, with more content in each chapter and a larger number of sections, compared to Heureka. The authors of Heureka are unique in their choice to divide the subject of forces into three different chapters separated from each other. The chapter concerning basic forces like gravity and friction is covered in one chapter at the beginning while linear momentum and composition and decomposition of forces are covered in two chapters at the end of the book. Heureka and Quanta has small squares throughout the chapters covering topics like historical persons and history of the subject.

Nexus and Ergo has as a termination of every chapter, after the regular problems to solve, a section which cover *Tasks to think about*, like *Why is it that your hands get warm when you rub them together?* Ergo is the only book with a separate section about laborative work in physics.

In the second course in physics in upper secondary school (Physics B) the four books do not show any common pattern regarding the structure of the chapters with the exception that all four books end with a chapter covering the topics of cosmos, astrophysics and the universe. The internal structure of the chapters follows the same patterns as respective book for the first course, meaning that Nexus and Ergo has a separate section for *Tasks to think about* at the end of every chapter. Ergo does not have a section with things to examine by laborations in this book. Heureka is the book that is missing most of the TIMSS-items.

All books are, at the end of every chapter, summing up the most important formula covered, and all of the textbooks have grouped the exercises into two levels based on their conceptual difficulty and complexity.

Occurrence of items in physics textbooks, similar to the released TIMSS items

Based on the identification of similar items in TIMSS Advanced and the different textbooks, the items have been divided into five groups. In the first group we identify items which are not found in any of the analysed textbooks. The second group contains items found in one of the four textbooks, etc. The results for each group of items will be presented separately in the next sections.

Weighted p-values for the group of students who were administered each item, and had used a specific textbook, are presented in tables, together with their standard errors. In the tables, textbooks with near identical TIMSS items were found are marked with a grey shading.

One of the released items from TIMSS Advanced 2008 was excluded from the analysis because of extremely low p-values. This item is shown in figure 5.

A particle with charge q and mass m moves at speed v in a uniform magnetic field B at right angles to the direction of the field. The particle moves in a circle.

Show that the period T of the particle's revolution does not depend on v .
Show your work.

Figure 5 An item (PA13026) from TIMSS Advanced excluded from the analysis because of extremely low p-values.

The item is one of the most difficult problem for the Swedish pupils to solve. Almost none of the pupil succeeded regardless of what book they used.

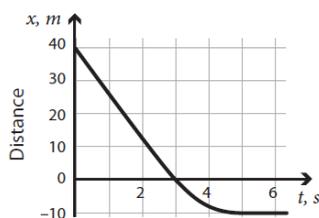
Items not found in any of the textbooks

Among the released items from TIMSS Advanced 2008 there are nine which are not found in the four analysed textbooks. Most of these items should be covered in Physics A, the first physics course in Swedish upper secondary school.

Table 3 P-values and standard errors for TIMSS items found in none of four physics textbooks.

Item	Nexus	Ergo	Quanta	Heureka
PA23110	0,71 (0,05)	0,68 (0,05)	0,73 (0,04)	0,67 (0,04)
PA23025	0,12 (0,04)	0,08 (0,02)	0,10 (0,05)	0,14 (0,03)
PA23028	0,65 (0,07)	0,54 (0,05)	0,60 (0,07)	0,66 (0,04)
PA13006	0,24 (0,07)	0,16 (0,04)	0,20 (0,11)	0,29 (0,06)
PA23034	0,07 (0,04)	0,09 (0,02)	0,07 (0,04)	0,24 (0,04)
PA23078	0,22 (0,02)	0,27 (0,02)	0,19 (0,03)	0,27 (0,03)
PA13004	0,87 (0,02)	0,86 (0,02)	0,87 (0,02)	0,86 (0,02)
PA13009	0,28 (0,03)	0,28 (0,03)	0,20 (0,01)	0,34 (0,02)
PA23084	0,15 (0,03)	0,09 (0,03)	0,12 (0,03)	0,13 (0,02)

Among the items described in table 3 there are both easy and difficult items. There are also items which seem to have been easier to solve for students using one of the textbooks, even though similar items were not found there. One of the TIMSS not found in any of the textbooks is shown in figure 6. This item is answered correctly by more than half of the students even though their textbooks do not contain a similar item.



The graph shown above represents a cyclist approaching and passing the finishing line in a race. If the cyclist weighs 60 kg, what is her momentum as she crosses the finishing line?

- (A) 2400 kg · m/s
- (B) 800 kg · m/s
- (C) 600 kg · m/s
- (D) 0 kg · m/s

Figure 6 An item (PA23028) from TIMSS Advanced 2008 with similar item/s found in none of the four most frequently used textbooks in physics in Sweden.

Items found in one of the four textbooks

Two items were identified in only one of the four textbooks. The content of both of these items should be covered in Physics A.

Table 4 P-values and standard errors for TIMSS items found in one out of four physics textbooks. The shaded cells represent textbooks where the items were found.

	Nexus	Ergo	Quanta	Heureka
PA13001	0,57 (0,03)	0,62 (0,04)	0,67 (0,05)	0,60 (0,03)
PA23128	0,27 (0,03)	0,36 (0,06)	0,49 (0,08)	0,36 (0,05)

In both cases, the p-values are not significantly higher for students using a textbook where these items are found. One of these items is shown in figure 7.

Ultraviolet light is responsible for sunburns. Explain why you don't get sunburned while sitting behind a glass window.

Figure 7 An item (PA23128) from TIMSS Advanced 2008 with similar item/s found in one of the four most frequently used textbooks in physics in Sweden.

The item in figure 2 contradicts the hypotheses that students who have seen an item before in the textbook will have easier to solve it for the second time seen. This must be something that is learned anyhow. Perhaps this is due to the relevance of the subject. Every summer this topic is covered in the news although not from a physics perspective.

Items found in two of the four textbooks

Four items were found in two of the textbooks. These items could not be assigned to just one of the two courses physics A or B.

Table 5 P-values and standard errors for TIMSS items found in two out of four physics textbooks. The shaded cells represent textbooks where the items were found.

Item	Nexus	Ergo	Quanta	Heureka
------	-------	------	--------	---------

PA13027A	0,07 (0,01)	0,11 (0,01)	0,05 (0,01)	0,08 (0,01)
PA23113	0,59 (0,06)	0,61 (0,07)	0,56 (0,05)	0,54 (0,04)
PA23050	0,43 (0,05)	0,58 (0,04)	0,30 (0,11)	0,51 (0,04)
PA23056	0,82 (0,03)	0,80 (0,04)	0,84 (0,04)	0,76 (0,04)

The item shown in figure 8 (PA23050) is an example of the items in this group. For this particular item p-values were lower for students using the textbooks where a similar item was found.

A table with metal legs and a wooden top is inside a room with a temperature of about 20 °C. Which statement explains why the metal legs feel colder than the wooden top?

(A) The heat capacity of the metal legs is lower than the wooden top.

(B) The metal has a lower temperature than the wooden top.

(C) The metal conducts heat better than wood.

(D) The molecules move faster in metal than in wood.

Figure 8 An item (PA23050) from TIMSS Advanced 2008 with similar item/s found in two of the four most frequently used textbooks in physics in Sweden.

Again there is no support for the hypothesis that having seen the item before will increase the achievement.

Items found in all but one textbook

Eight of the items were found in three of the textbooks. These items are found in both Physics A and B.

Table 6 P-values and standard errors for TIMSS items found in three out of four physics textbooks. The shaded cells represent textbooks where the items were found.

Item	Nexus	Ergo	Quanta	Heureka
PA23072	0,20(0,03)	0,09 (0,02)	0,16 (0,04)	0,18 (0,03)
PA13023	0,58 (0,04)	0,53 (0,03)	0,57 (0,08)	0,65 (0,04)
PA23022	0,06 (0,02)	0,05 (0,02)	0,04 (0,02)	0,12 (0,03)
PA13024	0,20 (0,03)	0,23 (0,03)	0,15 (0,07)	0,22 (0,02)
PA23082	0,56 (0,03)	0,46 (0,05)	0,57 (0,07)	0,50 (0,04)
PA23142	0,64 (0,04)	0,51 (0,10)	0,67 (0,05)	0,55 (0,05)
PA23140	0,49 (0,05)	0,37 (0,04)	0,56 (0,08)	0,40 (0,04)
PA23058	0,53 (0,04)	0,62 (0,05)	0,64 (0,07)	0,58 (0,06)

The list of items in table 9 shows one example of low percentage correct for students using a textbook where an item was not found. This item (PA23072) is shown in figure 9. For the rest of the items no such pattern can be detected.

A block of mass 2.0 kg travels horizontally at a speed 2.5 m/s towards a massless spring with spring constant 800 N/m. After the block collides with the spring, its speed decreases and the spring compresses. What is the maximum distance that the spring will compress? (Ignore friction and air resistance.)

Show your work.

Figure 9 An item (PA23072) from TIMSS Advanced 2008 with similar item/s found in three of the four most frequently used textbooks in physics in Sweden.

Items found in all four textbooks

Thirteen of the items were found in all textbooks.

Table 7 P-values and standard errors for TIMSS items found in all four physics textbooks.

Item	Nexus	Ergo	Quanta	Heureka
PA13003	0,78 (0,03)	0,76 (0,07)	0,74 (0,03)	0,73 (0,03)
PA13005	0,86 (0,04)	0,78 (0,03)	0,82 (0,06)	0,76 (0,03)
PA23014	0,39 (0,05)	0,33 (0,05)	0,27 (0,04)	0,41 (0,05)
PA23030	0,27 (0,07)	0,25 (0,06)	0,25 (0,10)	0,37 (0,05)
PA23044	0,24 (0,05)	0,27 (0,05)	0,26 (0,11)	0,25 (0,06)
PA13021	0,52 (0,05)	0,43 (0,04)	0,56 (0,04)	0,46 (0,04)
PA13022	0,14 (0,02)	0,06 (0,02)	0,08 (0,02)	0,14 (0,02)
PA23138	0,88 (0,02)	0,80 (0,05)	0,90 (0,05)	0,79 (0,05)
PA23059	0,74 (0,04)	0,82 (0,04)	0,91 (0,04)	0,76 (0,05)
PA23137	0,49 (0,06)	0,39 (0,07)	0,46 (0,08)	0,41 (0,06)
PA13025	0,25 (0,03)	0,11 (0,02)	0,21 (0,06)	0,25 (0,04)
PA13002	0,72 (0,03)	0,79 (0,02)	0,70 (0,06)	0,81 (0,02)
PA23115	0,36 (0,05)	0,35 (0,05)	0,49 (0,08)	0,39 (0,07)

Many of these items have high p-values and for some of the items the p-values vary between groups of students using different textbooks. One of the items in the group (PA13025) is shown in figure 10.

The table shows the work-function energy (W) for the photoelectric effect in three different metals.

<u>Metal</u>	<u>W</u>
Ca	$4.60 \times 10^{-19} \text{ J}$
Li	$4.65 \times 10^{-19} \text{ J}$
Zn	$6.94 \times 10^{-19} \text{ J}$

Which of these metals will emit electrons when hit by visible light with wavelength 400 nm? Explain your reasoning.

Figure 10 An item (PA13025) from TIMSS Advanced 2008 with similar item/s found in all four of the most frequently used textbooks in physics in Sweden.

Even though the students have had the opportunity to solve items like this in their textbooks, less than one out of four students gets the answer right. This can be due to the fact that this topic is not covered in class, although it is covered in the books. Furthermore, students using one of the textbooks perform significantly lower on this item even though the item is identified in all four textbooks.

Conclusions and discussion

This study has investigated the effect of item familiarity on student achievement in TIMSS Advanced 2008. Item familiarity is a complex concept, and the definition used in this study is but one of many possible

Familiarity with items used in for assessment of student achievement in TIMSS is defined as the existence of similar items in the textbook used by individual students. An item is under this definition expected to be familiar if students most likely have met at least one similar item in class.

Results from this study do not support the hypothesis that the achievement will increase if the pupils have seen the items once before. Despite the fact that the item has appeared in the textbook used by the students they do not manage very well to solve the item correctly. The study does not reveal any pattern of relationship between the existence of items similar to the TIMSS-items in the textbooks used and students' probability of answering the items correctly.

More than one of the items that did not appear in the textbooks has a very high p-value, which indicates that the pupils have gained the subject knowledge anyway and that the pupils are able to use their knowledge and transform that knowledge to more unfamiliar contexts.

The definition of item familiarity used in this study, and the underlying assumption that having seen an item affects the possibility of solving it, are of course limiting factors in this study. In particular for TIMSS, which is characterised as a "low-stakes" assessment the occurrence of similar items in textbook might be seen as a too narrow perspective. We know from ongoing studies that student motivation for participating in TIMSS is low in Sweden and we also know that students hardly do any revision of what they have worked with earlier, for example by reviewing text-book exercises. Compared to e.g. a final examination paper, students can be expected to be less influenced by recall of similar items. For further research

on item familiarity it is therefore important to explore more sophisticated and elaborated conceptualisations of familiarity in this context.

However, the line of research presented in this paper can also be supplemented by an analysis of more items from TIMSS. In order to detect patterns of this kind it is known from other studies that a large number of items are required. The study could be expanded from the limited number of released items, to all items used in TIMSS Advanced 2008. This would substantially increase the data and increase the possibility to reveal conceivable hidden patterns. In addition, the results can be validated by using a panel of judges for the categorisation of the items.

Finally, the study of item familiarity on a national and classroom level can contribute to a more comprehensive categorization of items with respect to cognitive aspects, which in turn can be used to deepen the analyses of the achievement results.

References

- Garden, R. A., Lie, S., Robitaille, D. F., Angell, C., Martin, M. O., Mullis, I. V. S., et al. (2006). *TIMSS Advanced 2008 Assessment Frameworks*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Ingleby, J. D. (1973). A test of current explanations of the effect of item familiarity on memory. *Quarterly Journal of Experimental Psychology*, 25(3), 378-386.
- Lindström, J-O. (2006). Med fokus på matematik och naturvetenskap. En analys av skillnader och likheter mellan internationella jämförande studier och nationella kursplaner. (In Swedish). [Focus on mathematics and science. An analysis of differences and similarities between international comparative studies and national curricula]. Stockholm: Skolverket.
- 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/Nystrom.pdf.
- Ruddock, G., Clausen-May, T., Purple, C., & Ager, R. (2006). *Validation study of the PISA 2000, PISA 2003 and TIMSS-2003 international studies of pupil attainment*. (NFER Research Report no. 772). Nottingham: DfES Publications.
- Stricker, L. J., & Emmerich, W. (1999). Possible determinants of differential item functioning: Familiarity, interest, and emotional reaction. *Journal of Educational Measurement*, 36(4), 347-366.
- Tunnie, R. J., & Fernie, G. (2007). Repetition priming affects guessing not familiarity. *Behavioral and Brain Functions*, 3:40.