Ida Friestad Pedersen
Department of Teacher Education and School Research
University of Oslo

Is TIMSS Advanced an appropriate instrument for evaluating mathematical performance at the advanced level of Norwegian upper secondary school? An analysis of curriculum documents and assessment items

Abstract
The results of international, large-scale achievement studies attract a lot of attention and may affect educational policies. With this in mind, the primary aim of the present study was to examine the appropriateness of the mathematics tests developed for the two cycles of TIMSS Advanced for the evaluation of the mathematical performance of Norwegian upper secondary school students. The study utilized the methodology for alignment studies developed by Porter (2002), which entails analyzing and describing the mathematical content of the Norwegian curriculum documents and the TIMSS Advanced assessment items and calculating the agreement between these. The analyses showed a moderate alignment between the different iterations of TIMSS Advanced and the curriculum followed by the participating students, implying that the emphasis in the Norwegian mathematics curriculum differs somewhat from that in the TIMSS Advanced tests. This does not mean that the TIMSS Advanced tests were inappropriate instruments for assessing the mathematical performance of participating Norwegian students. However, when interpreting the results of this large-scale survey, the differences in emphasis between the different iterations of the TIMSS Advanced and the curriculum should be taken into account. Finally, as alignment is evaluated by contrasting alignment indices calculated for different combinations of the intended and assessed curriculum, additional research is needed to make more firm judgments. An additional contribution of the present study is demonstrating a powerful methodology for conducting this kind of research.

Keywords: TIMSS Advanced, Mathematical content, Curriculum alignment, Validity

Introduction

TIMSS (Trends in International Mathematics and Science Study) 1995 was the first large-scale international comparative achievement study to assess Norwegian students' performance in mathematics. This was in part followed up by PISA (Programme for International Student Assessment) in 2000 and by subsequent cycles of both these studies. In many cases, the results have been disappointing. TIMSS 2003 revealed a marked decline in Norwegian primary and lower secondary school students' mathematics performance (Grønmo, Bergem, Kjærnsli, Lie, & Turmo, 2004), and this was supported by data from PISA 2003 and 2006 (Kjærnsli, Lie, Olsen, Roe, & Turmo, 2004; Kjærnsli, Lie, Olsen, & Roe, 2007). The ability of schools to adequately prepare Norwegian students for the demands of post-secondary education and employment was publicly debated. The results of these studies played an important role in the process leading up to the implementation of the Knowledge Promotion Reform (K06) in 2006 (Bergesen, 2006; Udir, 2011; UFD, 2003). In 1995, TIMSS included an assessment of final-year upper secondary school students enrolled in advanced mathematics courses. Norway postponed this part of the study until 1998 (Angell, Kjærnsli, & Lie, 1999). For simplicity, it will be referred to here as the 1998 study. The final-year TIMSS study was repeated in 2008 (TIMSS Advanced), and the results again revealed a pronounced decline in the achievement of Norwegian students in mathematics (Grønmo, Onstad, & Pedersen, 2010). The weak performance of Norwegian students was a cause of concern for teachers and teacher organizations (Norsk lektorlag, 2009), for educational researchers (Grønmo et al., 2010) and for educational authorities (KD, 2010).

Later cycles of TIMSS (2007, 2011) and PISA (2009) revealed progress in the mathematics performance of Norwegian students (Grønmo et al., 2012; Grønmo & Onstad, 2009; Kjærnsli & Roe, 2010), and relatively positive TIMSS 2011 results were taken as an indication that K06 had been effective (KD, 2012). A new round of TIMSS Advanced is scheduled for 2015. If Norway chooses to participate, this may give some indication of whether the negative trend has also been reversed in upper secondary school.

The results of these large-scale international studies in mathematics receive a great deal of attention and influence educational policies in Norway. They employ well-documented and methodologically sound procedures for defining and sampling student populations, translating items into the different languages of the participating countries, monitoring how students' responses are scored (including consideration of inter-rater reliability) and evaluating aspects of the items' properties (Arora, Foy, Martin, & Mullis, 2009; Olsen, 2005). However, irrespective of the quality of the studies, they have attracted a fair amount of criticism regarding their credibility, relevance and usefulness, with the criticism to a large extent related to issues of agreement between the assessment instruments and the national curricula (Nyström & Lind, 2010; Sjøberg, 2005).

It has been claimed that such studies make no allowance for "different aims, issues, history and context across the mathematics curricula of the systems being studied" (Keitel & Kilpatrick, 1999, p. 243). Given such objections, it may be desirable to investigate the validity of the assessment instruments with respect to the Norwegian instructional situation.

This article focuses on the TIMSS Advanced mathematics study. TIMSS is chosen because it, unlike PISA, is a curriculum based study. Thus, while the assessment frameworks developed for TIMSS reflect consensus across the participating countries on the core content of school mathematics in relation to the respective countries' curricula, the PISA frameworks are, to a larger extent, detached from the curricula and focus on what mathematics students are expected to need in their present and future life *(mathematical literacy*) (Grønmo & Olsen, 2006). As TIMSS in essence measures how well an educational system has implemented the mathematics curriculum, a satisfying degree of correspondence between national curricula and the assessment instruments seems to be especially important for this study. Furthermore, I focus on the *Advanced* version of TIMSS because the differences in mathematics curricula across countries are more pronounced in the final year of upper secondary school than in the early years of schooling (Mullis, Martin, Robitaille, & Foy, 2009). Hence developing an assessment instrument that is reasonably matched to the curricula of the participating countries is thought to have been especially challenging.

The use of high-stakes testing, especially in the U.S., has led to substantial interest in methods for determining the degree of agreement, or *alignment*, between an educational jurisdiction's curriculum and its assessment system (Bhola, Impara, & Buckendahl, 2003). Consequently, there are several methods available for determining the alignment between the content of curriculum objectives and assessment instruments (see e.g. La Marca, Redfield, Winter, Bailey, *&* Despriet, 2000; Porter, 2002; Webb, 1997), ranging from low to high levels of complexity. Methods of low complexity identify alignment as the extent to which the items on a test match relevant curricular objectives (Bhola et al., 2003). TIMSS Advanced 2008 utilizes this type of low complexity alignment method. In a test-curriculum matching analysis, 92% of the items (score points) were in accordance with those of the Norwegian curriculum (Mullis et al., 2009). Although this analysis answered some questions pertaining to the validity of the TIMSS Advanced tests, it did not reveal whether some aspects of mathematics are emphasized in curriculum documents but not in the assessment instrument. To address this potential threat to validity, more complex alignment models are needed.

The aim of this article is to discuss whether the mathematics tests developed for the two cycles of TIMSS Advanced are appropriate instruments for evaluating the mathematical performance of Norwegian upper secondary school students enrolled in advanced mathematics courses. The study utilizes the method-logy for alignment studies developed by Porter (2002, 2006), which entails analyzing and describing the mathematical content of the Norwegian curriculum documents and the TIMSS Advanced assessment items and calculating the alignment between these.

Content language – a tool for describing the curriculum

When seeking to describe and compare the mathematical content specified by curriculum documents or contained in assessment instruments, one needs a framework that distinguishes between different types of content. The classical conception of mathematical content represents this as the sum of the topics to be covered in a given course (Niss & Højgaard, 2011; Schoenfeld, 1994). According to this conception, two curricula will be equivalent with respect to mathematical content if the same topics are included in the curricular objectives.

However, the topic-oriented perspective has been criticized by mathematics educators who argue that mathematical thinking involves more than knowing facts, theorems, techniques, etc. (Schoenfeld, 1994). Developing the ability to build mathematical models, assess chains of arguments, generalize methods and results and find and justify new results are all important components of learning mathematics (Niss, 2003; Niss & Højgaard, 2011; Schoenfeld, 2007), and this perspective should be taken into account when analyzing instructional material. For instance, a curriculum objective that states that students are expected to identify numerical patterns and use known formulas to sum finite arithmetical series may give students other learning experiences than a curriculum objective asking students to find, analyze and prove formulas for terms in sequences or sums of series. Although both hypothetical objectives include the same topic (numerical patterns), the former example seems to direct the focus towards using procedures, whereas the latter invites students to generalize patterns, derive formulas and justify their results. Identifying the mathematical content of curriculum documents, textbooks or assessment instruments with a list of the included topics would, however, make it impossible to distinguish between these example objectives.

To obtain more complete descriptions of the content specified by curriculum documents or contained in assessment instruments, different taxonomies or *content languages* have been developed (e.g. Garden & Orpwood, 1996; Garden et al., 2006; Porter, 2002; Webb, 1997; Krathwohl, 2002). A common feature of such content languages is that they are two-dimensional, with *topics* forming the first dimension and *cognitive demands or processes* forming the second dimension. Content is thus defined as a combination of topics and cognitive demands/processes.

Table 1. Content languages of the TIMSS Advanced assessment frameworks (Garden & Orpwood, 1996; Garden et al., 2006)

	Topic dimension	Cognitive dimension
1998	• *Numbers and Equations,* • *Calculus,* • *Geometry,* • *Probability and statistics,* • *Validation and structure*	• *Knowing,* • *Using routine procedures,* • *Using complex procedures,* • *Solving problems,* • *Justifying and proving,* • *Communicating*
2008	• *Algebra,* • *Calculus,* • *Geometry*	• *Knowing* - including recalling and computing • *Applying* - including selecting appropriate strategies and solving routine problems • *Reasoning* - including analyzing, generalizing, justifying and solving non-routine problems

Table 1 presents the content languages of the TIMSS Advanced mathematics assessment frameworks, where the topic dimension specifies the subject matter to be assessed within mathematics and the cognitive dimension specifies the thinking processes or performance expectations to be assessed. Note that the content of these dimensions changed somewhat from the first to the second cycle of the study.

The group of researchers responsible for the TIMSS Advanced study used the content languages in Table 1 to characterize the items developed for each of the associated mathematics assessments (this information is available at http://timssandpirls.bc.edu). However, for the present study, a content language that is suitable for classifying the Norwegian curriculum documents, as well as items from both cycles of TIMSS Advanced needs to be developed. With respect to the topic dimension, this can easily be obtained by including all topic areas contained in the assessment frameworks and the curriculum documents. The cognitive dimension, however, requires more careful consideration. As noted above, the description of the thinking processes or performance expectations in the two assessment frameworks differs somewhat. There is a division between routine and complex procedures, or between routine and non-routine problems, in both frameworks. This is problematic because it will depend on the teaching that students have received: a routine problem for one group of students may be a non-routine problem for another group of students. To resolve this, the cognitive dimension of the TIMSS Advanced 2008 framework (chosen because it provides the most detailed description) was modified by including elements of the cognitive process dimension in Bloom's revised taxonomy (Krathwohl, 2002). The revised Bloom's taxonomy was selected because (i) Bloom's taxonomy is well known and recognized in educational research, (ii) this taxonomy has several elements in common with the cognitive dimension in the 2008 TIMSS Advanced framework but avoids the problematic distinction between routine and non-routine procedures/problems and (iii) empirical studies have shown that using Bloom's taxonomy as an alignment tool yields higher levels of inter-rater reliability than some of the other taxonomies found in the literature (Näsström & Henriksson, 2008). The resulting content language is presented in Table 2.

Table 2. Content language based on the TIMSS Advanced assessment frameworks (Garden & Orpwood, 1996; Garden et al., 2006) and Bloom's revised taxonomy (Krathwohl, 2002)

Topic dimension	**Cognitive dimension**
Algebra • Number systems • Equations and inequalities • Functions • Sequences and series	***Know*** Recalling, retrieving or being aware of relevant information • Define, identify, list, memorize, remember, recall
Geometry • Euclidean geometry • Analytic geometry • Trigonometry (including trigonometric functions) • Vectors (including vector functions)	***Understand*** Determining the meaning of written, graphical or oral communication • Interpret, exemplify, summarize, explain, infer
Calculus • Limits, continuity and differentiability • Differentiation and using derivatives to determine maxima, minima, turning points etc • Integration • Differential equations	***Apply*** Carrying out or using a procedure in a given situation • Execute, compute, solve, represent, graph, ***Analyze*** Investigate given information, determine and describe relationships between variables or objects (and use these relationships) • Compare, derive, classify, organize
Probability and statistics • Combinations and permutations • Probability and probability distributions • Descriptive statistics • Inferential statistics	***Evaluate*** Make judgments based on criteria and standards • Judge, assess, discuss, defend, critique ***Create*** Combine elements (results and/or procedures) to form a coherent, novel whole • Prove, plan, construct, formulate, model, combine, generalize, produce

The cognitive dimension considered here is a loose hierarchy, in the sense that the six major categories are believed to differ in their complexity. Hence, *know* is considered to be less cognitively complex than *understand*, and *analyze* is similarly less cognitively complex than *evaluate*. That being said, some cognitive processes associated with the category *understand* (e.g. explain) are arguably more complex than some of the cognitive processes associated with *apply* (e.g. execute). This was acknowledged by Krathwohl (2002), who nevertheless argued that if “one were to locate the ‘center point’ of each of the six major categories on a scale of judged complexity, they would likely form a scale from simple to complex” (Krathwohl, 2002, p. 215).

Therefore, in the present study, mathematical content is defined as the combination of both the topics and the cognitive processes outlined in Table 2. Using this content language will enable us to distinguish between both curricula and assessment instruments that include different topics and between curricula and assessment instruments that include the same topics but emphasize different cognitive processes.

Empirical material and methods of analysis

The students that participated in the 1998 TIMSS Advanced study followed the curriculum known as Reform 94. The description of the advanced mathematics program in this curriculum was subject to revisions in 2000, and the students who took the 2008 TIMSS Advanced mathematics test were the last cohort to follow this revised curriculum. If Norway chooses to participate in the next cycle of TIMSS Advanced (2015), it will most likely involve students belonging to the (advanced) mathematics for the natural sciences program in the current K06. In light of this, I also include the K06 curriculum in the analysis, as it will be interesting to examine to what extent the TIMSS Advanced assessments are in agreement with the most recent curriculum documents. To summarize, the empirical material for this study consists of three versions of the Norwegian (advanced) mathematics curriculum documents—the original Reform 94 (R94-o), the revised Reform 94 (R94-r) and the Knowledge promotion (K06)—and two iterations of the TIMSS Advanced mathematics tests (used in 1998 and 2008, respectively).

The following section provides a more detailed description of the empirical material. Following this, the classification of units (curricular objectives and assessment items) with respect to the content language in Table 2 and the method for calculating the alignment are discussed, and issues pertaining to ambiguity, interpretation and reliability of the classification are addressed.

Empirical material

Curriculum documents

Norwegian upper secondary school is a three-year program, where students can choose between different advanced subjects (e.g. Mathematics, Physics, English language, Social Studies) in the final two years. The advanced mathematics program in R94 consisted of the courses 2MX and 3MX (KUF, 2000), while the (advanced) mathematics for the natural sciences program in the K06 comprises the courses R1 and R2 (KD, 2006). In both cases, the final year courses (3MX, R2) build on the subject matter contained in the second-year courses (2MX, R1). TIMSS Advanced assesses the performance of school-leaving students and includes mathematical content that may not be exclusive to the final-year mathematics course. It is thus reasonable to include both the second and final-year courses when analyzing the content of the advanced mathematics program in Norwegian upper secondary schools.

For all the listed courses, the mathematics curriculum is structured into main goals or topic areas, which are further broken down into detailed objectives. The first two columns of Table 3 show examples of curricular goals from R94-o[1], R94-r and K06, including excerpts of the associated (detailed) objectives.

Table 3. Example of curricular goals, with detailed objectives (KD, 2006; KUF, 2000; Sandvold et al., 1995)

Curriculum	Goals and objectives	Sub-objectives
R94 – o (2MX)	**Goal 6: Limits and derivatives** *Students should* 6a know the concept of limit, and be able to calculate the limit of simple functions (...)	6a1 know the concept of limit, 6a2 and be able to calculate the limit of simple functions
R94 – r (Common goal)	**Goal 2: Modeling, experimenting and problem solving** *Student should* 2a be able to formulate and analyze simple mathematical models, and evaluate their validity (...)	2a1 be able to formulate simple mathematical models, 2a2 and be able to analyze simple mathematical models, 2a3 and evaluate their validity
K06 – R1	**Geometry** *The aims of the studies are to enable pupils to* Ga use lines and circles as geometric loci together with congruence and the inscribed angle theorem in geometrical analysis and calculations Gb execute and analyze constructions defined by straight lines, triangles and circles in the plane, with and without the use of dynamic software (...)	Ga1 use lines and circles as geometric loci together with congruence and the inscribed angle theorem in geometrical analysis and calculations Gb1 execute constructions defined by straight lines, triangles and circles in the plane, with and without the use of dynamic software, Gb2 and analyze constructions defined by straight lines, triangles and circles in the plane, with and without the use of dynamic software

Note: The sub-objectives described in column 3 were devised by the author to facilitate the classification of the objectives.

Both versions of the R94 mathematics curricula included two goals in common for both grade levels that were not directly related to specific topic areas. These goals describe general competencies that students are expected to exercise while encountering a range of mathematical topics. They add a distinct contribution to the content of the curriculum documents and are, to the greatest extent feasible, included in the analyses in this paper. There are no such common goals in K06, but rather a set of basic skills that should be developed in all subjects: *ability to express oneself orally and in writing*, *ability to read*, *ability in numeracy* and *ability to use digital tools*. According to the curriculum documents describing the content of R1 and R2, basic skills are integrated into the competence aims/objectives of these courses (KD, 2006). Therefore, the focus of this article was on analyzing the topic-specific objectives in K06 rather than including the basic skills in the analyses.

The first step of the analysis involved classifying the curricular objectives using the content language described in Table 2. Inspection of the examples of curricular objectives in Table 3 (column 2) reveals that there may be several cognitive processes involved in one single objective. To facilitate the classification, the curricular objectives were divided into sub-objectives formed by a single verb/verb phrase that can be linked to a cognitive process (see the third column of Table 3).

Some of the curricular objectives (e.g. "Students should know the multicultural history of mathematics and be aware of the significance of mathematics with respect to science, technology, society and culture" (KUF, 2000, common objective 1e) were deemed to be too vague to be reliably classified in terms of the present content language. This is clearly a weakness with the chosen analytical procedure, as failure to classify a subset of the curricular objectives may lead to misleading results regarding the content of the curriculum documents. However, objectives that are too vague to be classified in terms of specific topics and cognitive processes could also be difficult for teachers and textbook authors to include in the instructional material, thus marginalizing their role in the implementation of the curriculum. Nevertheless, this weakness of the analytical procedure should be kept in mind when interpreting the resulting descriptions of the curriculum documents

Removing 11 unclassifiable objectives (eight From R94-o and three from R94-r) from the analysis resulted in 82 sub-objectives from R94-o, 93 sub-objectives from R94-r and 68 sub-objectives from K06.

TIMSS Advanced

The TIMSS 1998 Advanced mathematics test comprised 65 assessment items (67 if two-part tasks are considered as separate items). Following the publication of the 1998 results, half of these items were released, whereas the other half remained undisclosed to be used in future studies. TIMSS Advanced 2008 comprised 72 mathematics items; 27 of these were trend items (items also used in the 1998 study), and the remaining 45 items were new material. Examples of released items included in both cycles of TIMSS Advanced are shown in Figure 1.

Figure 1. Examples of TIMSS Advanced items from both cycles. Released items are available at http://timssandpirls.bc.edu/timss_advanced/idb.html

M1_04

$\lim_{x\to+\infty} \frac{(2x+1)(x+1)}{3x^2-2}$ is equal to

Ⓐ $-\frac{1}{2}$

Ⓑ $\frac{2}{3}$

Ⓒ 1

Ⓓ 6

Ⓔ ∞

M1_08

Triangle *PQR* is an isosceles right triangle with a right angle at *P*. If *PT* is a median of the triangle, then *PT* has the same length as

Ⓐ *PR*

Ⓑ *PQ*

Ⓒ *QR*

Ⓓ *QT*

M3_04

$\int_1^2 \left(x - \frac{1}{x^2} \right) dx$ is equal to

Ⓐ $-3\frac{1}{8}$

Ⓑ 1

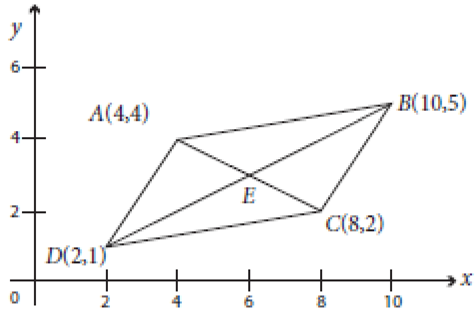
Ⓒ $2\frac{5}{8}$

Ⓓ 4

Ⓔ $4\frac{1}{2}$

M3_09

In the quadrilateral *ABCD* below, diagonals *AC* and *BD* intersect at point *E*. PROVE that *E* is the midpoint of *AC* and *BD*. Show all your work.



As Figure 1 indicates, the assessment instruments include both multiple-choice (MC) and constructed response items, with the majority being MC-format. The items are intended to measure a range of topics (here exemplified with limits, integration and Euclidean and analytical geometry) and cognitive processes (here exemplified with applying procedures, analyzing and proving).

As part of the TIMSS Advanced project group, I had access to all the undisclosed items from the two mathematics tests. Hence, complete versions of both tests are analyzed in this paper.

Classification of curricular objectives and assessment items

Using the content language described in Table 2, any curricular sub-objective or assessment item can be represented in two dimensions: the topic dimension and the cognitive process dimension. This is exemplified in Table 4, which shows how the sub-objectives listed in table 3 and the assessment items presented in figure 1 are classified in a topic by cognitive process matrix.

Table 4. Classification of example units

		Know	Understand	Apply	Analyze	Evaluate	Create
Curricular sub-objectives							
Algebra	Number systems						
	Equations and inequalities				2a2	2a3	2a1
	Functions				2a2	2a3	2a1
	Sequences and series				2a2	2a3	2a1
Geometry	Euclidean geometry			Ga1, Ga2	Ga1, Ga3		
	Analytic geometry						
	Trigonometry				2a2	2a3	2a1
	Vectors						
Calculus	Limits, continuity and differentiability	6a1		6a2	2a2	2a3	2a1
	Differentiation				2a2	2a3	2a1
	Integration				2a2	2a3	2a1
	Differential equations				2a2	2a3	2a1
Probability and statistics	Combinations and permutations						
	Probability and distributions						
	Descriptive statistics						
	Inferential statistics						
Assessment items							
		Know	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
G	Euclidean geometry				M1_08		
	Analytic geometry						M3_09
C	Limits, continuity and differentiability			M1_04			
	Integration			M3_04			

Note: To make the display more compact, unused topic categories were removed from the section of the table showing the classification of the assessment items.

When classifying units, the cognitive dimension is considered a (loose) hierarchy (Krathwohl, 2002). Hence, if a unit requires more than one cognitive process, it will be classified according to the most advanced of these processes. Consider, for example, item M3_09 in Figure 1. Here, students need to *know* the definition of a quadrilateral, and they may need to *apply* procedures for calculating the slopes of straight lines or arithmetical rules for vectors. However, as the item requests students to prove a statement, this is classified as *create*. The classification according to topics is slightly more problematic, as the categories in this dimension do not form a hierarchy. If we again consider item M3_09 above, this may be solved both by using vector geometry and analytic geometry, and students may have chosen either approach. Such ambiguities were dealt with by (a) focusing on the mathematical topic most prominent in the item stem (for M3_09, vectors are not mentioned and hence it is classified as analytic geometry) and (b) using the classification already done by the TIMSS Advanced international project group as a guide. Following this line of reasoning, all items are placed in only one cell of the table.

However, in relation to the curricular sub-objectives, multi-categorization (i.e. placing a sub-objective into more than one cell) was necessary. Consider for instance the common goals 2a1, 2a2 and 2a3, all of which are not directly related to a topic area. As per popular textbooks (see e.g. Oldervoll et al., 1996; Sandvold et al., 1995), it appears that mathematical modeling most commonly occurs in algebra, calculus and trigonometry. Hence, these objectives are placed in multiple cells in the topic dimension. Similarly, sub-objective Ga1 contains one verb phrase (use), indicating that this asks students to apply knowledge. However, as this knowledge is to be used in geometrical *analysis* and in *calculations*, sub-objective Ga1 has been classified as both *apply* and *analyze*.

Calculating content distributions and alignment indices

For each curriculum version (R94-o, R94-r and K06) and each cycle of TIMSS Advanced (1998 and 2008), the sub-objectives/items were coded into separate topic-by-cognitive process matrices[2]. Following the method proposed by Porter (2002, 2006), the number of sub-objectives categorized into each of the cells was first counted. In this process, multi-categorized objectives were distributed over the number of cells to which they belong. For example, if a sub-objective is placed in two cells (like Ga1 in Table 4), it counts as half of a unit in each of these cells. Next, the percentage of the categorized units in every cell of each matrix was calculated. Note that this quantification rests on the assumption that all units are equally weighted. One may reasonably argue that some curricular objectives are more extensive than others (e.g. sub-objectives 6a1 and Ga1 in Table 3). Assigning different weights to the different curricular objectives may have resulted in a truer description of the intended curricula. However, as this process would rely on the qualitative judgments of the coder, it would probably have reduced the reliability of the coding.

The process described above yielded *content distribution matrices*, with each cell in the two-dimensional topic-by cognitive-process matrix now containing the proportion of categorized units in the given cell. Across the cells, the proportions sum to 1 (Porter, 2002). In essence, these content distribution matrices describe the mathematical content of the different curriculum documents and mathematics tests and provide some insights into which aspects of mathematics are emphasized in the curricula and in the assessment instruments. To quantify how similar the content distribution matrices for the curriculum documents are to those calculated for the TIMSS Advanced mathematics tests, I used the index developed by Porter (2002, 2006) for comparing the distribution of objectives with the distribution of assessment items in alignment studies. This alignment index is given by:

$$A = 1 - \frac{\sum |x_i - y_i|}{2},$$

where x_i denotes the cell proportions in cell i for the content distribution matrix x (e.g. describing the content of R94-r) and y_i denotes the cell proportions in cell i for the content distribution matrix y (e.g. describing the content of TIMSS Advanced 2008). The possible values of this index range from 0 to 1, with $A = 1$ indicating that the distributions are the same (100% of the content in common) (Näsström, 2009; Polikoff, Porter, & Smithson, 2011; Porter, 2002).

Using an equivalent index, Webb (2007) set the lower boundary for acceptable alignment to 0.7. However, Porter (2006) argued that such boundary values are set somewhat arbitrarily and concluded that there is no absolute criterion for how high the alignment index should be for the alignment to be termed acceptable (Porter, 2006). In this article, alignment was therefore judged comparatively, by contrasting the alignment indices calculated for different combinations of curriculum documents and assessment instruments.

Reliability

In order to obtain a sense of how reliable the categorization is, a set of units (45 sub-objectives and 22 assessment items) have been categorized by two independent researchers. To ensure they were representative of the full set, the curricular objectives were selected to span all mathematics courses in the different curriculum documents, as well as a range of topic areas. The assessment items were selected to cover both cycles of TIMSS Advanced.

The agreement between the two coders' categorization was estimated using Porter's alignment index. This was calculated both at the *coarse-grained level* (where topic subcategories (e.g. Euclidean geometry, analytic geometry, trigonometry and vectors) are merged into a main topic category (e.g. geometry)) and the *fine-grained-level* (where the full set of topic categories presented in Table 4 is retained). The result is shown in Table 5.

Table 5. Alignment between coders

	Fine grained	Coarse grained
Curricular objectives	0.86	0.92
Assessment items	0.64	0.73

Table 5 indicates that there was a higher degree of agreement regarding the classification of curricular objectives than regarding the classification of the assessment items. A rule of thumb for evaluating the quality of inter-rater reliability based upon consensus estimates is that they should be higher than 70%, which here corresponds to an index value of 0.70 (Stemler, 2004; Näsström, 2009). If we adopt this, the agreement for the fine-grained classification of assessment items is too low, although the agreement is higher for the coarse-grained classification. Thus, some of the ambiguity in the classification of the items according to the topics is resolved by merging the topic subcategories into main categories (consider again item M3_09 in Figure 1, which could be classified as both vector and analytic *geometry*). Therefore, this article only reports results from the coarse-grained analyses, where the agreement in both cases may be termed acceptable.

After coding the material separately, the two researchers met and discussed their results, which resolved most of the initial disagreements. This discussion resulted in some clarifications on how the classification system should be understood, and these adjustments were transferred to the classification of the remaining units.

Results and discussion

The alignment between the different iterations of TIMSS Advanced and the associated curricula are reported in Table 6. As the alignment is judged comparatively, the table also includes indices of alignment between hypothetical combinations of assessment instruments and curricula to increase the basis for comparison.

Table 6. Alignment indices (coarse-grained analysis), with the highlighted indices showing the alignment between the assessment instruments and the curriculum documents followed by the participating students

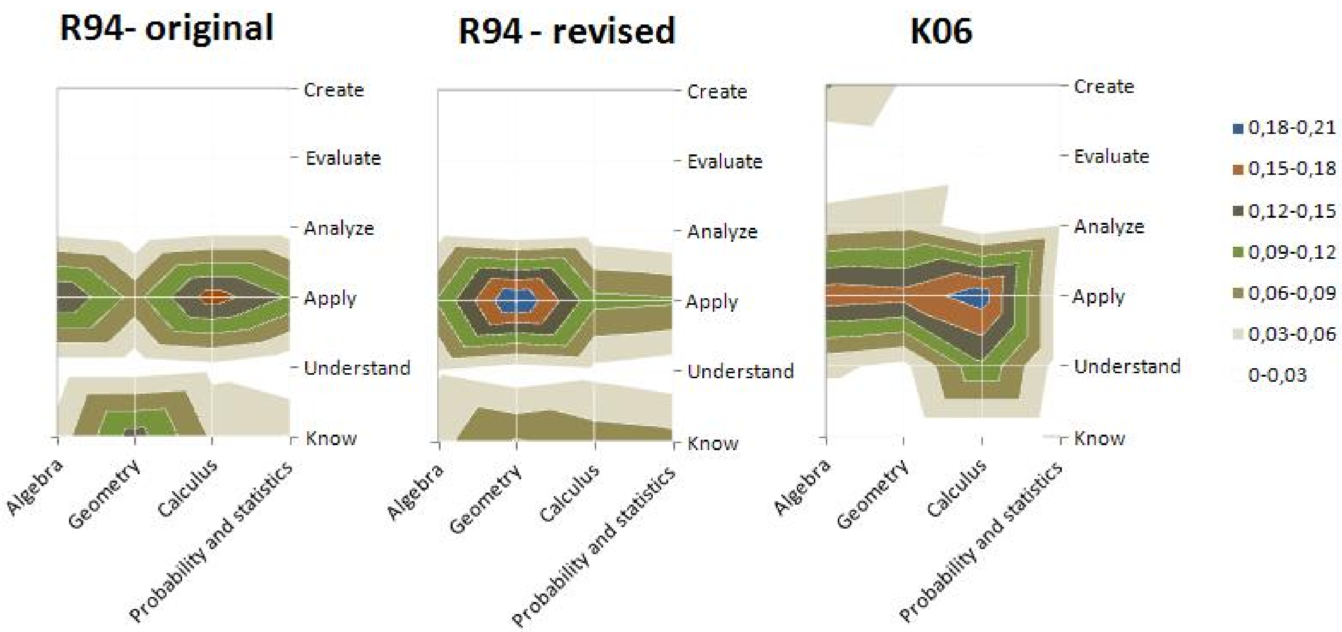
	R94 – original	**R94 - revised**	**K06**
TIMSS Adv 1998	**0.66**	0.73	0.65
TIMSS Adv 2008	0.55	**0.55**	0.70

As shown in Table 6, the alignment between the TIMSS Advanced 2008 mathematics test and the revised R94 curriculum is 0.55, which means that this pair of documents shares 55% of their content. Compared to some of the other indices in Table 6, this appears to be rather low. The alignment between the TIMSS Advanced 1998 mathematics test and the original version of the Reform 94 mathematics curriculum is somewhat higher (0.66), indicating that the match between the mathematical content of the TIMSS Advanced test and the advanced mathematics program followed by the participating students declined somewhat between the two iterations of this study.

The highest alignment indices in Table 6 are those between the hypothetical combinations of the 1998 TIMSS Advanced test and the revised R94 curriculum and between the 2008 TIMSS Advanced test and K06. This may indicate that (i) the TIMSS Advanced 1998 mathematics test would have been more appropriate for assessing the performance of the students participating in 2008 and that (ii) the TIMSS Advanced 2008 mathematics test would have been a better match to the current K06 than to either of the R94 curricula. However, one should be careful not to read too much into the relatively small differences in the alignment indices presented in table 6. The degree of reliability in the coding has to be considered when discussing alignment, since the theoretical upper boundary of 1 for the alignment index assumes perfect reliability. Closer inspection of the reliability coding reveals that the agreement between the two coders was somewhat lower for the TIMSS Advanced 2008 items than for the 1998 items, which may have contributed to the variations in Table 6.

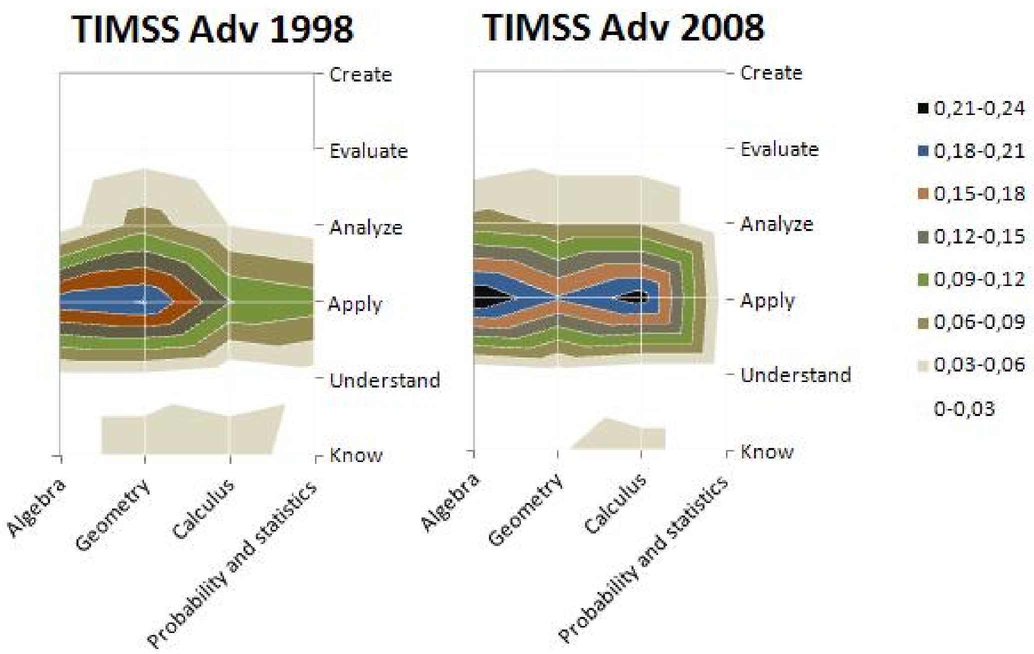
To further explain and discuss the results in Table 6, we now turn to the content distribution matrices. Figure 2 shows a topographical map of the mathematical content (at the coarse-grained level) emphasized in the original and revised version of R94, as well as in K06. Similarly, Figure 3 shows topographical maps of the mathematical content emphasized in the two cycles of TIMSS Advanced. These maps are all graphical displays of the content distribution matrices, where the shading represents the relative content emphasis (Porter, 2002). Before discussing these plots, it should be noted that although both the topic dimension and the cognitive dimension are formed by nominal variables, the graphing procedure interprets topics and cognitive processes as points on an underlying continuous scale. Hence, while the maps accurately represent each of the points of intersection between a given topic area and a given cognitive process, the areas between the points of intersections are not really meaningful, i.e. the graphing software has smoothed the data as though the variables formed a continuum (Porter, 2002, 2006). Thus, although the graphical displays are powerful, one should be careful not to read too much into them.

Figure 2. Coarse-grained display of the mathematical content of the Norwegian curriculum document, with the different colors representing the proportion of total content coverage



In both the original and the revised version of the R94 curriculum documents, all the main topic areas were included. However, the relative emphasis on these topics shifted, with the focus on *geometry* seemingly stronger in the latter. This appears to have occurred at the expense of *algebra* and *calculus*.

Figure 3. Coarse-grained display of the mathematical content of the TIMSS Advanced assessment instrument, with the different colors representing the proportion of total content coverage



In relation to the assessment instruments, Figure 3 reveals that the TIMSS Advanced 1998 test also covered all the topic areas, with a somewhat larger emphasis on *algebra* and *geometry*. In 2008, the focus shifted to *algebra* and *calculus*, and the topic area *probability and statistics* virtually disappeared from the test. This was the result of changes made to the assessment framework. According to the group responsible for the TIMSS Advanced study, these were due to changes in countries' advanced mathematics curricula since 1995, as well as the need to concentrate the assessment on those topics for which valid and reliable measurements could be made (Garden et al., 2006). However, as shown in Figure 2, *probability and statistics* was still a main topic area in the revised R94 curriculum. Hence, the changes made to the TIMSS Advanced assessment framework between the two cycles of the study may have made the 2008 test a better match to some of the participating countries' mathematics curricula, but it reduced the agreement between this test and the Norwegian curriculum documents. As already pointed out, the 1998 test was better aligned with the revised R94 curriculum, and this is also evident by comparing Figures 2 and 3. Less importance was also assigned to the topic area probability and statistics in K06, which may have contributed to the relatively good match between the TIMSS Advanced 2008 test and K06.

However, the topic areas covered are only one dimension of the mathematical content. In terms of the cognitive process dimension, Figure 2 reveals that the objectives in both versions of R94 placed the greatest emphasis on *applying* procedures or methods. There is some focus on *knowing* in these curriculum documents, but the most advanced cognitive processes (*analyzing*, *evaluating* and *creating*) seem virtually absent. The K06 curriculum covers a wider range of cognitive processes, including a small element of *creating* in algebra, but the main emphasis is nevertheless on *applying*. Mathematics educators have argued that mastering mathematics consists of more than knowledge of, and fluency in applying, facts, concepts and procedures. From this perspective, learning mathematics includes developing the capacity to engage in the processes of mathematical thinking: building models, looking for patterns, generalizing methods and results, assessing and challenging chains of arguments, proving conjectures and so on (see e.g. Niss, 2003; Schoenfeld, 1994, 2007; Stein, Grover, & Henningsen, 1996). If this is considered the desired outcome of mathematics instruction, then instructional environments (including curriculum documents) should encourage students to engage in the more advanced cognitive processes of *analysis*, *evaluation* and *creation*. The advanced mathematics program in upper secondary school is designed to prepare students for further studies in mathematics and fields where mathematics is used as a tool, and it seems unfortunate that the emphasis on the higher-level cognitive processes appears to be so low. However, it should be noted that some curricular objectives describing general competencies that students are expected to exercise while encountering a range of mathematical topics were considered too vague to be reliably classified in terms of the present content language. Excluding these objectives from the analysis may have contributed somewhat to the low emphasis on higher-level cognitive processes in Figure 2.

Critics of large-scale comparative achievement studies such as TIMSS Advanced have expressed concerns that these tests primarily measure lower learning outcomes by means of multiple-choice items, as this format is more suited for measuring procedural knowledge than higher-level thinking processes (Cai, 1997; Wang, 2001). Figure 3 shows that both of the TIMSS Advanced assessment instruments place a great deal of emphasis on *applying* procedures. The focus on analyzing may be slightly greater in the TIMSS Advanced mathematics tests than in the three curricula, but the more advanced cognitive processes seems to be downplayed here as well. In this sense, the criticism may be justified. This does not, however, imply that these mathematics tests are inappropriate instruments for evaluating the mathematical performance of Norwegian upper secondary school students because the focus of the Norwegian curriculum documents also appears to be on lower-level thinking processes.

Concluding remarks

Large-scale international comparative achievement studies such as TIMSS Advanced provide educational researchers and policymakers with valuable information about the quality of their country's educational system. In this respect, it is important to discuss the content (e.g. frameworks, instruments) of large-scale surveys to continually improve their quality. The primary aim of the present study was to discuss the appropriateness of the mathematics tests developed for the two cycles of TIMSS Advanced for evaluating the mathematical performance of Norwegian upper secondary school students.

The alignment between the different iterations of TIMSS Advanced and the curriculum followed by the participating students appears to be moderate (Table 6), implying that the emphasis in the Norwegian mathematics curriculum differs somewhat from that of the TIMSS Advanced tests. This seems to support the criticism of TIMSS offered by Keitel and Kilpatrick (1999), who problematized the underlying assumption of an international 'common' mathematics curriculum and argued that international comparisons make too little allowance for differences in the mathematics curricula of the systems being studied. One may reasonably argue that spending a substantial amount of instructional time on mathematical content not included in the assessments will likely have a negative effect on students' performance. For instance, removing the topic area *probability and statistics* from the 2008 TIMSS Advanced study appears to have reduced the agreement between this mathematics test and the Norwegian curriculum documents. Considering that students performed particularly well on items assessing *probability and statistics* in 1998 (Angell, Kjærnsli & Lie, 1999), the pronounced decline in the Norwegian students' mathematics achievement between 1998 and 2008 may to some extent be related to the somewhat lower agreement between the curriculum documents and the assessment instruments in 2008. Thus, when interpreting the validity of the results, the alignment between the different iterations of the TIMSS Advanced and the curriculum followed by the participating students should be taken into account.

These results do not mean that the TIMSS Advanced tests were inappropriate instruments for assessing the mathematical performance of Norwegian students. First, there is no generally agreed upon lower boundary for acceptable alignment (Porter, 2002, 2006). Thus, one may not firmly assert that the mathematics tests were not reasonably aligned with the advanced mathematics program followed by the participating students. Second, the TIMSS Advanced test-curriculum matching analysis (see Mullis et al., 2009) revealed that, for the most part, the Norwegian students had been exposed to the mathematical content of the test. Furthermore, the degree of reliability in the coding has to be considered since the theoretical upper boundary of 1 for the alignment index assumes perfect reliability.

In conclusion, the results of the present study are somewhat inconclusive. The alignment was evaluated comparatively by contrasting alignment indices calculated for different combinations of the intended and assessed curriculum. Additional research is needed to make more firm judgments regarding the agreement between the curricula and the assessment instruments. Such research could involve studying the alignment between curriculum documents and national examinations in mathematics, between other countries' advanced mathematics program and the TIMSS Advanced assessment or between the 4th grade/8th grade TIMSS mathematics test and the curriculum followed by the participating students. The study demonstrates the power of the Porter (2002, 2006) methodology for conducting this kind of research, although more work may be needed on developing a content language that can better capture the curriculum.

References

Angell, C., Kjærnsli, M., & Lie, S. (1999). *Hva i all verden skjer i realfagene i videregående skole?* Oslo: Universitetsforlaget.

Arora, A., Foy, P., Martin, M. O., & Mullis, I. V. S. (Eds.). (2009). *TIMSS Advanced 2008 Technical Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. http://timssandpirls.bc.edu/timss_advanced/tr.html

Bergesen, H. O. (2006). *Kampen om kunnskapsskolen*. Oslo: Universitetsforlaget.

Bhola, D. S., Impara, J. C., & Buckendal, C. W. (2003). Aligning Tests with States' Content Standards: Methods and Issues. *Educational Measurement: Issues and Practice*, *22*: 21–29. http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.2003.tb00134.x/abstract

Cai, J. (1997). Beyond Computation and Correctness: Contributions of Open-Ended Tasks in Examining U. S. and Chinese Students' Mathematical Performance. *Educational Measurement: Issues and Practice*, *16*: 5–11. http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3992.1997.tb00580.x/abstract

Garden, R.A., & Orpwood, G. (1996). Development of the TIMSS Achievement Tests. In M.O. Martin and D.L. Kelly (Eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College. http://timssandpirls.bc.edu/timss1995i/TechVol1.html

Garden, R. A., Lie, S., Robitaille, D. F., Angell, C., Martin, M. O., Mullis, I. V. S., Foy, P., & Alka, A. (2006). *TIMSS Advanced 2008 Assessment Frameworks*. Boston: TIMSS & PIRLS International Study Center, Boston College. http://timssandpirls.bc.edu/timss_advanced/frameworks.html

Grønmo, L. S., Bergem, O., K., Kjærnsli, M., Lie, S., & Turmo, A. (2004). *Hva i all verden har skjedd i realfagene? Norske elevers prestasjoner i matematikk og naturfag i TIMSS 2003*: Acta Didactica, ILS, UiO. http://www.timss.no/timss05_03.html

Grønmo, L. S., & Olsen, R. V. (2006). TIMSS versus PISA: The case of pure and applied mathematics. *The Second IEA International Research Conference: Proceedings of the IRC-2006*. http://www.timss.no/publications/IRC2006_Gronmo%26Olsen.pdf

Grønmo, L.S. & Onstad, T. (Eds.) (2009). *Tegn til bedring. Norske elevers prestasjoner i matematikk og naturfag i TIMSS 2007*. Oslo: Unipub.

Grønmo, L. S., Onstad, T., & Pedersen, I. F. (2010). *Matematikk i motvind. TIMSS Advanced 2008 i videregående skole*: Oslo: Unipub. http://www.timss.no/rapporter%202008/Matematikk%20i%20motvind.pdf

Grønmo, L. S., Onstad, T., Nilsen, T., Hole, A., Aslaksen, H., & Borge, I. C. (2012). *Framgang, men langt fram Norske elevers prestasjoner i matematikk og naturfag i TIMSS 2011*. Oslo: Unipub.

KD. (2006). *Matematikk for realfag – Programfag i studiespesialiserende utdanningsprogram*. http://www.udir.no/grep/Lareplan/?laereplanid=168732

KD. (2010). *Realfag for framtida. Strategi for styrking av realfagene 2010–2014.* http://www.regjeringen.no/upload/KD/Realfagstrategi.pdf

KD. (2012). *De internasjonale undersøkelsene TIMSS og PIRLS: Betydelig framgang for norske elever*. Pressemelding 11.12.12. http://www.regjeringen.no/nb/dep/kd/pressesenter/pressemeldinger/2012/-de-internasjonale-undersokelsene-timss-.html?id=709617

Keitel, C., & Kilpatrick, L. (1999). Rationality and irrationality of international comparisons. In G. Kaiser, E. Luna, & I. Huntly (Eds.), *International comparisons in mathematics education.* London: Falmer.

Kjærnsli, M., Lie, S., Olsen, R.V., Roe, A., & Turmo, A. (2004). *Rett spor eller ville veier? Norske elevers prestasjoner i matematikk, naturfag og lesing i PISA 2003.* Oslo: Universitetsforlaget. http://www.pisa.no/pdf/publikasjoner/PISAHovedrapport2003.pdf

Kjærnsli, M., Lie, S., Olsen, R. V., & Roe, A. (2007). *Tid for tunge løft. Norske elevers kompetanse i naturfag, lesing og matematikk*. Oslo: Universitetsforlaget. http://www.pisa.no/pdf/Rapport/Tid_for_tunge_loft.pdf

Kjærnsli, M., & Roe, A. (Eds.) (2010). *På rett spor. Norske elevers kompetanse i lesing, matematikk og naturfag i PISA 2009.* Oslo: Universitetsforlaget. http://www.pisa.no/pdf/publikasjoner/Paa_rett_spor.pdf

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An Overview. *Theory into Practice*, *41*(4), 212–218. http://www.tandfonline.com/doi/abs/10.1207/s15430421tip4104_2

KUF. (2000). *Læreplan for videregående opplæring, Matematikk. Studieretningsfag i studieretning for allmenne, økonomiske og administrative fag.* Kirke-, utdannings- og forsknings departementet. http://www.udir.no/Lareplaner/Grep/Lareplanverket-for-videregaende-opplaring-R94/#Allmenne,%20%C3%B8konomiske%20og%20administrative%20fag

La Marca, P., Redfield, D., Winter, P., Bailey, A., & Despriet, L. (2000). *State standards and state assessment systems: A guide to alignment. Washington, DC: Council of Chief State School Officers.* http://eric.ed.gov/PDFS/ED466497.pdf

Mullis, I. V. S., Martin, M. O., Robitaille, D. F., & Foy, P. (2009). *TIMSS Advanced International Report.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. http://timssandpirls.bc.edu/timss_advanced/ir.html

Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM project. In A. Gagatsis & S. Papastavridis (Eds.), *3rd Mediterranean Conference on Mathematical Education* (pp. 115–124). Athens, Greece. http://w3.msi.vxu.se/users/hso/aaa_niss.pdf

Niss, M., & Højgaard, T. (Eds.) (2011). *Competencies and Mathematical Learning. Ideas and inspiration for the development of mathematics teaching and learning in Denmark*. Roskilde: Roskilde University, Denmark. http://milne.ruc.dk/ImfufaTekster/pdf/485web_b.pdf

Norsk Lektorlag. (2009). Urovekkende svake resultater. Retrieved 02.06.2012 from http://www.norsklektorlag.no/nyhetsarkiv-2009/timms-advanced-2008-urovekkende-svake-resultater-article333-197.html

Nyström, P., & Lind, A. K. (2010). Alignment between TIMSS Advanced and the Swedish national curriculum. *The fourth IEA International Research Conference: Proceedings of the IRC-2010*. http://www.iea.nl/irc-2010.html#c2360

Näsström, G. (2009). Interpretation of standards with Bloom's revised taxonomy: a comparison of teachers and assessment experts. *International Journal of Research & Method in Education*, *32*(1), 39–51. http://www.tandfonline.com/doi/full/10.1080/17437270902749262

Näsström, G., & Henriksson, W. (2008). Alignment of standards and assessment: A theoretical and empirical study of methods for alignment. *Electronic Journal of Research in Educational Psychology*, *6*(3), 667–90. http://www.investigacion-psicopedagogica.org/revista/articulos/16/english/Art_16_216.pdf

Oldervoll, T., Orskaug, O., & Vaaje, A. (1996). *Sinus 3MX Grunnbok. Oslo:* Cappelen.

Olsen, R. V. (2005). *Achievement tests from an item perspective: An exploration of single item data from the PISA and TIMSS studies, and how such data can inform us about students' knowledge and thinking in science.* (Doctoral dissertation). Oslo: University of Oslo, Oslo. http://folk.uio.no/rolfvo/Publications/Rolf_Olsen_DrScient_new.pdf

Polikoff, M.S., Porter, A.C., & Smithson, J. (2011).How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal*, *48*(4), 965–995. http://aer.sagepub.com/content/48/4/965.full

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31(7),* 3–14. http://edr.sagepub.com/content/31/7/3.full.pdf+html

Porter A. C. (2006). Curriculum assessment. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research*. American Educational Research Association, Washington, DC, 141–159.

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, 158–176. http://onlinelibrary.wiley.com/doi/10.1002/pits.20282/pdf

Sandvold, K. E., Øgrim, S., Jasper, P., & Nordseth, T. (1995). *Matematikk 2MX Grunnbok.* Oslo: Gyldendal Norsk Forlag.

Schoenfeld, A. H. (1994). What do we know about mathematics curricula? *Journal of Mathematical Behavior, 13*(1), 55–80. http://bear.berkeley.edu/faculty/AHSchoenfeld/Schoenfeld_WhatDoWeKnow.pdf

Schoenfeld, A. H. (2007). What is Mathematical Proficiency and How Can It Be Assessed? In A. H. Schoenfeld (Ed.), *Assessing Mathematical Proficiency* (Vol. 53). Berkeley, California: Mathematical Sciences Research Institute Publications. http://library.msri.org/books/Book53/files/05schoen.pdf

Sjøberg, S. (2005). PISA, TIMSS og norske læreplaner. *Bedre Skole*, *1*, 2005. http://folk.uio.no/sveinsj/PISA_TIMSS_Bedre_Skole_05.htm

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455–488. http://aer.sagepub.com/content/33/2/455.full.pdf+html

Stemler, S. E. (2004). A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability. *Practical Assessment, Research and Evaluation*, *9*(4). http://pareonline.net/getvn.asp?v=9&n=4

UFD. (2003). *St.meld. nr. 30 (2003–2004) Kultur for læring. Oslo.* http://www.regjeringen.no/nb/dep/kd/dok/regpubl/stmeld/20032004/stmeld-nr-030-2003-2004-.html?id=404433

Udir. (2011). *Internasjonale studier om norsk skole.* Temanotat 2011:2. http://www.udir.no/Upload/Rapporter/temanotat/Internasjonale_studier_om_norsk_skole_temanotat.pdf?epslanguage=no

Wang, J. (2001). TIMSS primary and middle school data: Some technical concerns. *Educational Researcher*, *30*(6), 17–21. http://edr.sagepub.com/content/30/6/17.full.pdf+html

Webb, N. L. (1997). Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. http://facstaff.wceruw.org/normw/WEBBMonograph6criteria.pdf

Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, *20*(1), 7–25. http://www.tandfonline.com/doi/full/10.1080/08957340709336728

[1] The original R94 curriculum documents are virtually impossible to find, but a complete list of objectives for the courses 2MX and 3MX can be found in the appendices of popular textbooks (Sandvold et al., 1995; Oldervoll, Orskaug, & Vaaje, 1996)

[2] One item from the 1998 test was deemed to be un-categorizable, and one item from the 2008 test has been removed from the international database and has therefore not been included in the present study.