TRENDS (1995 – 1999) IN PERFORMANCES OF DUTCH GRADE 8 STUDENTS IN TIMSS AGAINST THE BACKGROUND OF THE REALISTIC MATHEMATICS CURRICULUM

Pauline Vos & Klaas Bos

University of Twente, Faculty of Educational Science and Technology, Netherlands

In 1993, a new curriculum was established in junior secondary schools in the Netherlands, based on the principles of Realistic Mathematics Education (RME). Yet, teachers and students needed time to adjust to the new curriculum. In this article, data from the TIMSS test are studied, by distinguishing between test items that match and those that do not match the RME curriculum. Trend data (1995–1999) of Dutch students' achievement on these two distinct sets of items and data of teachers' approval of the test items suggest that the attained curriculum is approaching the intended curriculum at a very slow pace.

The RME core-curriculum in the Netherlands for junior secondary schools

Three decades ago Hans Freudenthal and his colleagues started to transform the mathematics curriculum with a treatise, which is generally known as *Realistic Mathematics Education* (RME). It is characterised by the understanding that mathematics is an integral part of real-life. Another component is the importance of enabling students to make mental images (Freudenthal, 1973, de Lange, 1987, van den Heuvel-Panhuizen, 1996).

In 1993 the "W12-16-project" established a core-curriculum based on RME for all students at Dutch junior secondary schools. The new curriculum emphasised data modelling and interpreting (through tables, graphs and word-formula), visual 3-d geometry, approximation and rules of thumb, the use of calculators and computers and other topics considered relevant to daily life of the new generation of the 21st century (Kok, Meeder, Wijers & van Dormolen, 1992). National assessment was adjusted to the new content approach. Generally, test items in the RME core curriculum describe an appealing daily life situation (often with authentic photographs to enliven imagination) followed by questions that integrate different mathematical content areas. The test items contain *horizontal mathematization* (Treffers, 1987) whereby realistic situations are modelled and reversibly the model is interpreted in its context. Several integrated mathematics topics can be combined and any test item is expected to keep students' concentration alive for approximately 15 minutes (Dekker, 1993). As for the format of questioning, multiple choice items do not match RME, because the world of real-life hardly ever offers four ready-made alternatives from which to choose.

With TIMSS data, implementation aspects of this new curriculum are studied below.

The TIMSS items matching a heterogeneous set of curricula

TIMSS (Third International Mathematics and Science Study) is an international comparative study of education in mathematics and science with the important question: *what can we learn from other countries?* It was conducted at grade 8 level in 1995 and again in 1999. The conceptual framework for this large scale curriculum study is based on the distinction of three curriculum levels (figure 1): the intended curriculum (what society at large prescribes students to learn, curriculum experts' opinions), the implemented curriculum (instruction at classroom level, teachers' opinions) and the attained curriculum (what is actually learnt by students).

To study the implementation process of the RME core curriculum at the level of junior secondary schools in the Netherlands, this framework was considered useful.

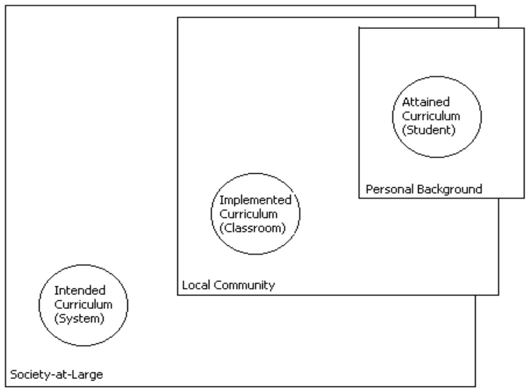

Figure 1. Framework for curriculum levels (Robitaille, 1993)

At the level of the attained curriculum, the TIMSS achievement test was carefully constructed in a process that is well-documented (Garden & Orpwood, 1996). Several experts in the field developed items testing for cognitive and procedural knowledge in a wide range of mathematics topics. With 40 different nations participating, a cross section of contents and levels had to be found that would be *equally unfair* for all.

At the level of the intended curriculum, a Test Curriculum Matching Analysis was carried out. In each participating country curriculum experts were asked to review each test item and assess whether its content was covered by the intended curriculum for the majority of the target population. Results from this analysis are presented in table 1. For each nation participating in TIMSS-95 (testing grade 8), it shows what percentage of the TIMSS mathematics items were considered appropriate to the intended curriculum.

The Netherlands lingers at the bottom, with 71% of the TIMSS test being covered by the intended curriculum of the target population. The complement of 29% of the 157 TIMSS items was considered remote from the national curriculum. Kuiper, Bos & Plomp (1997 & 1999) already analysed this problem and their research (with a slightly different set of 150 TIMSS-95 mathematics items considered) displays a percentage of 69% of all test items that are reasonably well covered by the Dutch RME curriculum. Four years later in 1999, when curriculum experts again were asked to judge the TIMSS mathematics items, the percentage was 71% (out of 155 items), showing that no substantive change had occurred in the judgement (Bos & Vos, 2000). The minimal discrepancies (69% or 71%) can well be associated with the differences in the assessed sets of items.

It remains to be noted, that the approximately 70% portion of items that were considered to be in line with the national RME curriculum would have been much smaller if the experts also had considered the question format. With approximately 75% of all test items having a multiple choice format, there would remain very little to consider.

Table 1.
Nations and their percentage of test items from the TIMSS-95 international test matching the intended mathematics curriculum

Nation	% items addressing natl. curr. (n=157)	Nation	% items addressing natl. curr. (n=157)
Hungary	100	Singapore	90
United States	100	Ireland	89
Latvia (LSS)	99	Romania	88
Israel	98	France	86
Spain	98	Belgium (Fl)	86
Germany	96	Kuwait	86
Lithuania	96	Belgium (Fr)	85
Australia	95	Denmark	84
Japan	94	Switzerland	83
Slovak Rep	94	Iceland	83
Portugal	94	Colombia	82
Slovenia	93	England	81
Hong Kong	92	South Africa	80
Norway	92	Sweden	78
Korea	92	Russian Fed.	78
Czech Rep	92	Scotland	76
Iran, Isl. Rep	92	Cyprus	76
Canada	91	Bulgaria	74
Austria	90	Netherlands	71
New Zealand	90	Greece	46

Dutch students' performances on TIMSS-95 and TIMSS-99

When TIMSS-95 was carried out, serious doubts arose whether TIMSS would do justice to the Dutch target population. Additional research established that, although Dutch students were not prepared for the full set of TIMSS items by their curriculum, their abilities were well measured by TIMSS (Kuiper, Bos & Plomp, 1997, 1999). In other words: the approximately 70% of TIMSS items that matched with the curriculum gave them enough room to display their abilities. It was assumed that when learning mathematics with real-life contexts and integrated topics, students would still be able to display their abilities on isolated questions and the multiple-choice format was not to obstruct their performance.

Moreover, although it might not officially be intentional, somehow Dutch students were knowledgeable about the *remote* items and could attain reasonable scores on these items as well. It could be, that teachers would still follow the abandoned curriculum or mix forthcoming content of higher grades in their present teaching.

Another reason could be that students acquired their knowledge outside the mathematics classrooms, or that they just attempted the unknown tasks with an open mind.

In 1995 a sample of n=1921 students was tested, in 1999 a sample of n=2878 students was tested, according to the strict TIMSS sampling procedures (Beaton, Mullis, et.al., 1996; Kuiper, Bos & Plomp, 1997; Mullis, Martin, et.al., 2000; Bos & Vos, 2000). The tests from 1995 and 1999 were of comparable level, with half of the items being identical, and the other half being mostly *clones.* In the process of replacing items, only minor adjustments were made. For example item N19 in TIMSS-95 would read "*shade in 5/8 of the unit squares in the grid*", and its substitute in TIMSS-99 would read "*shade in 3/8 of the unit squares in the grid*".

Overall, Dutch grade 8 students performed well on both TIMSS mathematics achievement tests. In 1995 they scored an average percentage correct of 63% on the mathematics items, and in 1999 this was 65%. This small, though statistically insignificant improvement is confirmed in the international TIMSS-99 report, in which a different scale for measurement of country performances is used (Mullis, Martin, et.al., 2000). Both in TIMSS-95 and TIMSS-99 Dutch grade 8 students rank well above the international average. The new curriculum seems to have had a positive impact.

Do Dutch students show a better performance on the ± 70% portion of the TIMSS items that were considered to match their curriculum? In table 2 a trend for the average percentage correct on different sets of items is summarised. Looking at the data for 1995, the performances on the two complementary sets of items show no difference with the overall performance. On average 63% of Dutch students answered any item correctly, whether it matched the intended curriculum or not. The students performed just as proficient on the RME-matching items as on the set of items NOT covered by the curriculum.

Table 2.
Average percentage correct by Dutch students in TIMSS-95 and TIMSS-99, on subsets of items

	Average percentage correct	
	TIMSS-95 (n=1921)	**TIMSS-99 (n=2878)**
All TIMSS items (100%)	**63**	**65**
Non-RME-items (± 70%)	**63**	**57**
RME-items (±30%)	**63**	**68**

In 1999 there is a small (though not statistically significant) gap between achievements on items that match and do not match the curriculum. A larger percentage of students (68%) perform well on the test items that match the RME curriculum than on the items that are remote from it (57%). A reason for the slight discrepancy in the columns of table 2 could be that time was needed for the implementation of the new curriculum. 1993 was the year of introduction. Thus, in 1995, two years after the introduction of the RME curriculum, there was still a period of curriculum transition and the new curriculum had not yet established itself well.

Teachers might still incorporate topics from the abandoned curriculum. Four years later, in 1999, the implementation of the new curriculum was starting to show.

Dutch teachers' approval of the TIMSS test items

With students' achievement at the attained curriculum level, and the curriculum experts' judgement at the intended curriculum level, there is still an intermediate level in the conceptual framework to analyse: the implemented curriculum. The question is whether the grade 8 mathematics teachers had covered the content of the test items and whether students would have had an *opportunity to learn* this content. Would teachers also cover non-RME content?

The instrument for this research question consisted of a questionnaire, disseminated to the mathematics teachers of the 126 tested classes in which they were asked to judge the TIMSS test items by the following question: *if you were to set a test on all mathematics content which has been taught so far to the concerning class (tested in TIMSS), would you consider this item from its content to be suitable for this test?* It was explicitly stated that teachers were to ignore the format (multiple choice) and the difficulty level, and only indicate whether the content was taught.

The instrument of 1995 covered 16 mathematics test items out of 150. This small selection was based on the intended curriculum (curriculum experts had selected these 16 to match well with the RME curriculum) and proved to be too small to provide an analysis of teachers' coverage of the whole test. In 1999 the instrument was re-developed in a way that all 155 items were scrutinised. Yet, this number was considered too large to be included into one questionnaire, and the judgement of all items would become a tedious job for the teachers concerned. It could negatively affect their responses. Therefore, three mixed sets of items were created (of 52, 52 and 51 items), and each teacher would randomly receive one set. In this way all items would receive a judgement, although each item would only be seen by one-third of the teachers.

The response reached a satisfactory level of 89% (126 teachers were approached, 112 teachers responded). Further details on the methodology of this instrument can be found in Bos & Vos (2000). As no data from 1995 would be available to make a trend analysis, comparative data were created into two other dimensions:

- internationally, with mathematics teachers from Belgium (Flanders) judging the same mathematics items from TIMSS-99.
- cross-curriculary, with Dutch physics/chemistry teachers judging the 70 items from TIMSS-99 of their subject.

For the international comparison, Belgium (Flanders) was considered appropriate, because of similarities in economical, social and cultural aspects. Moreover their intended curriculum is very similar to the pre-RME-curriculum of the Netherlands and closely matches the TIMSS test. As can be seen in table 1, in 1995 the Belgian curriculum experts chose 86% of the TIMSS items to match their curriculum. In 1999 this percentage had increased to 98% (Mullis, Martin, et.al., 2000).

For the cross-curricular comparison, items from science were selected. The complete TIMSS science item set (n=135 items in 1995, n=143 items in 1999) covers the subject areas physics, chemistry, life science, earth science, environmental science and nature of sciences. In the Netherlands, we do not teach *integrated science* like in many other countries at grade 8 level, but instead there is the combined subject of physics/chemistry, the separate subject of biology, and the separate subject of geography. From the TIMSS test, we selected 70 items that could possibly be covered in the lessons of physics/chemistry. The selection consisted of items on physics (n=38), chemistry (n=19), together with items on environmental science (n=3) and the nature of science (n=10). The curriculum experts in this field had chosen two thirds of these 70 items to match with the intended curriculum, which is just slightly less than their counterparts did for mathematics (±70%, compare table 1).

In the analysis, for each item the percentage of approving teachers was calculated, by taking those who had indicated "yes" on the question whether they would include this particular item into a test covering all taught content. It was stipulated that if an item had a high teacher' approval, many students would have had an opportunity to learn its content.

In table 3 the percentage of approving teachers is divided into categories with ranges of 20%. Items in the first approval category (0-20%) have a very low teacher approval. Items in the last category (80-100%) have ample teacher approval. In three columns the percentages of items in the five approval categories is given, for the Belgian (Fl.) and Dutch mathematics teachers, and for the Dutch physics/chemistry teachers. To visualise the comparison, figure 2 displays the same data in a bar chart.

Table 3.
Approval rates of TIMSS-99 test items of Belgian (FL.) mathematics teachers, and Dutch mathematics and physics/chemistry teachers.

Percentage of approving teachers	Belgium (Fl.)[1]	Netherlands	
	% of math items (n=155)	% of math items (n=155)	% of physics/chem. items (n=70)
0-20	0	1	3
20-40	1	3	10
40-60	4	6	37
60-80	24	24	33
80-100	71	66	17

The general profile of answers by Belgian and Dutch mathematics teachers is fairly similar. Despite differences in the intended mathematics curriculum, there seems to be a tacit cross-border understanding in approval of the TIMSS test items. Both mathematics teacher groups display a high approval of the items, with the Belgian (Fl.) teachers displaying a slightly higher approval. According to the answers in both

[1] We thank the Flemish TIMSS researchers, Prof dr. J. Van Damme and drs. A. Van den Broeck (K.U.Leuven), for making these data available (Departement Onderwijs Vlaanderen, 2000).

groups, more than two-thirds of the TIMSS mathematics test items had been taught by more than 80% of the teachers, and more than 90% of the items had been taught by more than 60% of the teachers. When looking at the 70% of items matching the Dutch RME curriculum, an average 87% of the mathematics teachers indicated that it was covered in their lessons (data not in table 3). The other 30% of non-RME items were covered by an average of 71% of the teachers. This could explain, why a majority of Dutch students proved proficient on the items that were not matching the intended curriculum: their teachers would still cover these topics.

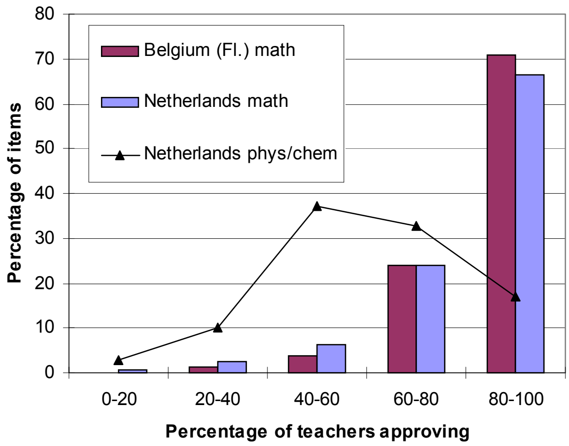

Figure 2.
Approval of TIMSS-99 test items by Belgian (Fl.) and Dutch mathematics teachers, compared to Dutch physics/chemistry teachers

Comparing the patterns of answers from the Dutch physics/chemistry teachers and the mathematics teachers proves difficult. Most physics/chemistry items receive an approval of half of the teachers, far less than the mathematics items. There are few physics/chemistry items with a high teacher' approval percentage, while in mathematics these outnumber all other items. To reach better understanding of this pattern, further cross-curricular research is needed, especially with regards to the physics/chemistry curriculum.

With the similar profiles of item approval by Belgian (Fl.) and Dutch mathematics teachers, it appears as if Dutch mathematics teachers still maintain characteristics in their instruction from the pre-RME era. Considering that their average years of teaching experience is 17 years (Bos & Vos, 2000), this means that the bulk of mathematics teachers in the Netherlands matured in their profession before 1993. It could mean that they still teach topics from the abandoned curriculum. As a consequence, it will take decades for the implemented curriculum to approach the new intended RME curriculum.

Conclusion

From the comparatively high performances of Dutch grade 8 students in the tests of the international comparative studies TIMSS-95 and TIMSS-99, it is obvious that the mathematics curriculum provides them with a solid foundation for doing mathematics. Yet, the implementation of the new RME-based curriculum at this level is still in progress, six years after its introduction in 1993. To analyse the process of implementation, a distinction in TIMSS items was made with the criterion of matching the intended RME curriculum or not. The performances of Dutch students on these complementary sets of items were calculated separately. Between 1995 and 1999, a slight improvement had taken place in the performance on the RME-matching items. Dutch students also did well on items that were not part of their intended curriculum, although their score decreased from 1995 to 1999.

To analyse the intermediate curriculum level of teacher instruction in the classrooms (implemented curriculum), an instrument was developed in which mathematics teachers were asked to judge all TIMSS mathematics items. A large majority of these teachers indicated that their students had an opportunity to learn about the content covered in the TIMSS test, whether the items matched the intended curriculum or not. Comparison with their Belgian (Fl.) colleagues displayed a fair agreement between the two groups. This could imply that the process of implementing the new RME curriculum is still proceeding slowly and Dutch teachers stay attached to the abandoned curriculum and that in Dutch mathematics classrooms a mix of two curricula is carried out. Further research into the implementation of the RME curriculum is advised.

References

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzales, E.J., Kelly, D.L. & Smith T.A. (1996). Mathematics Achievement in the middle school years. IEA's Third International Mathematics and Science Study. Boston, MA: Boston College.

Bos, K.Tj, & Vos, F.P. (2000). *Nederland in TIMSS-1999, exacte vakken in leerjaar 2 van het voortgezet onderwijs [The Netherlands in TIMSS-1999, mathematics and science in grade 8].* Enschede, Netherlands: University of Twente, OCTO.

Dekker, T (1993). De basisvorming getoetst [The core curriculum assessed]. *Nieuwe Wiskrant, 13.2*. Pp 5-9.

Departement Onderwijs Vlaanderen (2000). *TIMSS Repeat Onderzoek in 2000 in Vlaanderen.* Brussel: Ministerie van de Vlaamse Gemeenschap.

Freudenthal, H (1973). *Mathematics as an Educational task.* Dordrecht: Reidel Publishing Company.

Garden, R.A. & Orpwood, G (1996). Development of the TIMSS Achievement Tests. In: M.O. Martin & D.L. Kelly (Eds.), *Third International Mathematics and Science Study, Technical Report, Volume I: Design and Development* (pp 2-1 – 2-19). Boston, MA: Boston College.

Heuvel-Panhuizen, M. van den (1996). *Assessment and realistic mathematics education*. Utrecht, Netherlands: CD-beta-press, Freudenthal Institute, Utrecht University.

Kok, D., Meeder, M., Wijers, M. & Dormolen, J. van (1992). *Wiskunde 12-16, een boek voor docenten.* Utrecht: Freudenthal Instituut RU Utrecht / Enschede: SLO.

Kuiper, W.A.J.M., Bos, K.Tj. & Plomp, Tj. (1997). *Wiskunde en de natuurwetenschappelijke vakken in leerjaar 1 en 2 van het voortgezet onderwijs. Nederlands aandeel in TIMSS populatie 2* [*Mathematics and the science domains in secondary 1 and 2. Dutch participation in TIMSS population 2*]. Enschede, Netherlands: University of Twente, OCTO.

Kuiper, W.A.J.M., Bos, K.Tj. & Plomp, Tj. (1999). Mathematics Achievement in the Netherlands and Appropriateness of the TIMSS Mathematics test. In: *Educational Research and Evaluation, 5(2),* pp. 85-104.

Lange, J. de (1983). *Mathematics, Insight and Meaning.* Doctoral Dissertation. Utrecht: IOWO.

Mullis, I.V.A., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'connor, K.M., Chrostowski, S.J., Smith, T.A. (2000). *TIMSS 1999 International Mathematiccs Report, Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* Boston, MA: Boston College.

Robitaille, D.F., Schmidt, WH., Raizen, S, Mc Knight, C, Britton and E., Nicol, C (1993). *Curriculum Frameworks for Mathematics and Science, Timss monograph nr. 1*. Vancouver Canada: Pacific Educational press.

Treffers, A. (1987). *Three Dimensions. A model of goal and theory descriptions in mathematics instruction - the Wiskobas Project*. Dordrecht: Reidel Publishing Company.