TOWARDS A MORE CURRICULAR FOCUS IN INTERNATIONAL COMPARATIVE STUDIES ON MATHEMATICS AND SCIENCE EDUCATION

WILMAD KUIPER[1], KERST BOERSMA[2], JAN VAN DEN AKKER[3]

[1]University of Twente, [2]University of Utrecht, The Netherlands

ABSTRACT

From international comparative studies (TIMSS, PISA) it appears that students in lower secondary education in the Netherlands perform relatively well in mathematics and science compared to their peers from other participating countries. Policy-makers, especially, are eager to bring these positive outcomes into the limelight. However, one may wonder whether, in case of the Netherlands, there is good reason for such zeal. An evaluation study, conducted by the Netherlands Inspectorate of Education, shows that lower secondary schools do not meet the quality required in implementing a curriculum reform that started in 1993, entitled ‘basic secondary education’. So, in spite of all rhetoric on the positive outcomes of TIMSS and PISA in the Netherlands, when putting the relatively good student performance in the context of the implementation of this ambitious curriculum reform, many people become puzzled. Research findings on the quality of mathematics and science education seem to be in conflict with the results of TIMMS and PISA. This conclusion and also the observation that international comparative assessment studies have serious difficulty in meeting the goal of providing proper interpretations of student achievement, especially from a curriculum perspective, give reason to attempt to disentangle the conflicting images.

1. INTRODUCTION AND PROBLEM STATEMENT

The outcomes of international comparative studies like TIMSS and PISA get widespread attention in media and policy circles. Depending on the nature of the results, they tend to provoke a wide array of, often rhetorical, reactions. For example, the relatively poor performances of American 13 year old students in mathematics and science in TIMSS-1995 and TIMSS-Repeat 1999 gave cause to a still continuing flow of discussions, arguments, and reflections on origins of this problem (“the mathematics as well as the science curriculum is a mile wide and an inch deep”) and on possible solutions to it (‘rigorous’ and ‘demanding’ new standards). The weak performances of students in lower secondary education in Germany in TIMSS-1995, in TIMSS-Repeat 1999, and especially in PISA-2001 (with 31 participating countries Germany appeared 20th in the ranking for mathematics and science and 21st in the ranking for reading comprehension) caused a public debate that was dominated by great displeasure and concern about the quality of education in Germany. Also in the Netherlands the reactions poured into

K. Boersma et al. (eds.), Research and the Quality of Science Education, 41—54.
© 2005 *Springer. Printed in the Netherlands.*

the air, although they were quite different in nature due to the fact that – as it appears from TIMSS-1995, TIMSS-Repeat 1999, as well as PISA-2001 – students in lower secondary education perform relatively well in mathematics and science compared to their peers from other participating countries (Bos, Kuiper & Plomp, 1999; Bos & Vos, 2000; Kuiper, Bos & Plomp, 1999; Kuiper, Bos & Plomp, 1997; Wijnstra, 2001). Policy-makers are especially eager to bring these positive outcomes into the limelight.

However, one may wonder whether, in case of the Netherlands, there is good reason for such zeal, and especially for the sense of self-satisfaction that some quotes and comments seem to convey. This is not because of the low response rates in TIMSS-1995 and PISA-2001 that may have biased the good results, but rather because of the outcomes of an evaluation study conducted almost at the same time by the Netherlands Inspectorate of Education (Inspectie van het Onderwijs, 1999a-e). This evaluation study shows that secondary schools do not achieve the quality required in implementing a curriculum reform that began in 1993, entitled 'basic secondary education'. An even less favorable picture emerges when the performances of Dutch students in TIMSS are contrasted with the demanding instructional and learning goals as defined at system level in terms of attainment targets, instead of with the international mean achievement score (which is common practice in international comparative studies). Others (for example, Boersma, 2000a-b) criticize the new curriculum for mathematics, physics/chemistry, and biology, as it is overloaded and fragmented, lacks coherence and longitudinal alignment, is implemented without sufficient relevance for students, and is dominated by rather traditional modes of assessment.

So, in spite of all rhetoric on the positive outcomes of TIMSS (and also PISA) in the Netherlands, when putting the relatively good student performance in the context of the 'challenging' implementation of the curriculum reform in lower secondary education, many people become puzzled. Research findings on the quality of mathematics and science education seem to be in conflict with the results of TIMMS and PISA. This conclusion and also the observation that international comparative assessment studies have serious difficulty in meeting the goal of providing proper interpretations of variations in student achievement in view of policy implications (Bos, 2002; Kellaghan, 1996), in general and especially from *a curriculum* perspective, led us to attempt to disentangle the conflicting images. The outcomes of this attempt are described in this chapter. We start with a more in-depth analysis of conflicting images in the Netherlands as appearing from main findings from TIMSS and the Evaluation Study by the Inspectorate of Education (ESIE). This analysis is meant to clarify the debate and to articulate a curricular focus in international comparative studies like TIMSS. There is a clear need for doing the latter, as it has also been cogently substantiated by Westbury (1992) in his analysis of differences in achievement – found in SIMS – between American and Japanese secondary school students. A conceptual focus that emphasizes the "fundamental salience of curriculum" (Westbury, 1992, p.23) offers a chance for a sharper understanding of (factors influential to) mathematics and science achievement, which, in turn, is a prerequisite for more focused policy recommendations aiming at the enhancement of mathematics and science education.

2. CONCEPTUAL FRAMEWORK

As a stepping-stone in our analysis, we start with the curriculum typology known from TIMSS and other IEA studies (Robitaille et al., 1993; Schmidt et al., 2001): the intended, the implemented, and the attained curriculum. In our definition a curriculum is 'a plan for learning' that, depending on its nature and scope, may pertain to several components (van den Akker, 2003): rationale; aims, goals, and objectives; contents; teacher's role; student activities; materials and resources for teaching and learning; time allocation; location; and assessment modes and criteria. The *intended* curriculum refers to all those provisions aimed at being offered to students, including all those concepts, processes, and attitudes students are expected to study and learn. These may find expressions in formal documents (such as official attainment targets) and textbooks. The *implemented* curriculum is the curriculum as interpreted by teachers and made available to students (curriculum-in-action). The *attained* curriculum refers to that portion of the curriculum actually attained by students. This includes achievement measures as well as students' attitudes, perspectives, and values.

These three curriculum representations closely cohere. Also, there is never a linear, top-down transformation from curriculum intentions via implementation in teaching and learning settings to students' outcomes. It is a complicated process in which much elaboration and adaptation may be needed and may occur. Also a lot of 'noise' may arise. Original intentions can be blurred, distorted, or even devastated. Also other, often more powerful variables than only the intended curriculum may have an effect on the implemented and the attained curriculum (Figure 1; based on van den Akker, 1998). Some of the variables in this curriculum transformation or 'curriculum dilution' process may also be non-curricular in nature, like socio-cultural context (home, media, peers) and student characteristics (aptitude, motivation, gender).

Nevertheless, in our analysis of the Netherlands case we focus on the curriculum levels depicted at the horizontal axis in the middle of the Figure: intended – implemented – attained. Findings about (school and socio-cultural) context, student characteristics, and teacher characteristics will be left aside for the greater part, as it is not our ambition to look for an explanation of the findings of each of two studies. Instead, we try to disentangle conflicting images by comparing main findings at the three curriculum levels within and across the two studies

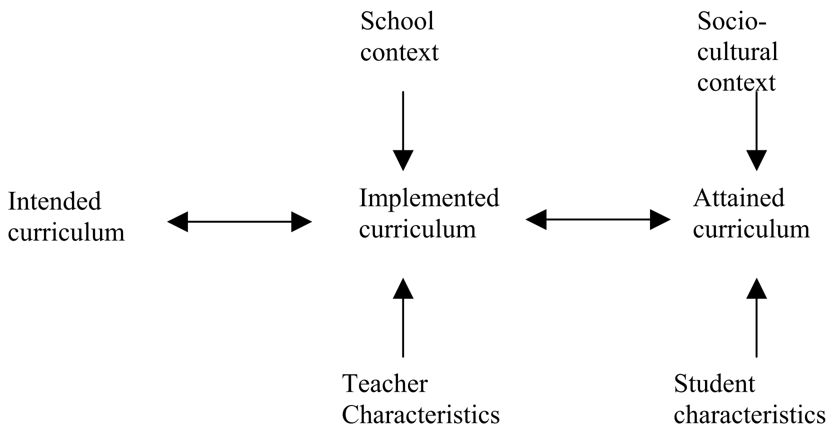

Figure 1. Curriculum typology

Nevertheless, in our analysis of the Netherlands case we focus on the curriculum levels depicted at the horizontal axis in the middle of the Figure: intended – implemented – attained. Findings about (school and socio-cultural) context, student characteristics, and teacher characteristics will be left aside for the greater part, as it is not our ambition to look for an explanation of the findings of each of two studies. Instead, we try to disentangle conflicting images by comparing main findings at the three curriculum levels within and across the two studies.

3. NETHERLANDS CASE: TIMSS AND ESIE

This section encompasses an analysis of the main findings from TIMSS and ESIE, preceded by some context information about the implementation of basic secondary education and about the goals, design, and instrumentation of ESIE and TIMSS.

Basic secondary education

The formal implementation of basic secondary education started in 1993. It aims at raising the standard of lower secondary education and at 'modernizing' the curriculum while maintaining the existing structure of four student ability tracks. It entails a core curriculum of 15 subjects (including mathematics, physics/chemistry, biology) covering the first three years of secondary education. For each subject, attainment targets have been set which indicate the expected level of achievement in terms of knowledge, understanding, and skills. The modernization not only refers to an increase of the number of subjects up to 15, but also to an intended change of both the subjects' contents (more application-oriented) and pedagogy (more activity-based and student-centered) at classroom level. As time has passed, complaints began to pour in about the reform (see earlier). As an attempt to arrive at a solution, the government has prepared a proposal for a sweeping revision of basic secondary education from 2005 onwards. It is in this turbulent reform context that Secondary 1 and 2 students perform(ed) relatively well on TIMSS tests.

ESIE

ESIE was carried out by government order during the school year 1997-1998. It was a large-scale evaluation (120 schools) of the implementation of basic secondary education four to five years after its formal start in August 1993 (Inspectie van het Onderwijs, 1999a-e). Investigated was the extent to which the attainment targets as well as general skills (e.g., conducting a simple inquiry) were part of the intended curriculum at school level. For these purposes textbooks in use, additional curriculum materials, and schools' work plans were analyzed, teacher questionnaires and interviews were administered, and lesson observations took place (Peters-Sips et al., 2000). As regards the implemented curriculum, lesson observations were conducted in order to obtain a picture of the quality of the teaching and learning at classroom level (van den Bergh, Zwarts & Peters-Sips, 2000). At the attained level secondary analyses took place, of student performances on drafts of so-called 'basic secondary education tests' constructed and administered by CITO in the spring of 1997 and 1998.

TIMSS

The TIMSS-1995 Population 2 study in the Netherlands, with the data collection in spring 1995, entailed the following components (Kuiper, Bos & Plomp, 1997):

- Attained curriculum: the administration of a written mathematics and science test in Secondary 1 and 2 (95 schools) plus a performance assessment in Secondary 2 (49 schools).
- Implemented curriculum: the administration of a student questionnaire (attitudes) and a teacher questionnaire (teaching practices; 'opportunity-to-learn' judgments, i.e. judgments on whether the content tested via a selection of items had been taught before test administration).
- Intended curriculum: an expert appraisal on the appropriateness of the items from the written test for the attainment targets for mathematics, physics/chemistry, and biology.

As part of TIMSS-1995 the Netherlands also contributed to an extensive cross-national analysis at the level of the intended curriculum, encompassing curriculum guides and textbooks in most common use (cf. Schmidt, McKnight et al., 1997; Schmidt, Raizen et al., 1997).

TIMSS-1999 (data collection, spring 1999) consisted of the same components as TIMSS-1995 (Bos & Vos, 2000), except for: (i) the written test that was administered in Secondary 2 only (126 schools), (ii) teachers who made a 'opportunity-to-learn' judgment to all (and not only a selection of) written test items, and (iii) the cross-national curriculum analysis. In addition, only in the Netherlands, the performance assessment was repeated in the spring of 2000 (27 schools; Vos & Kuiper, 2004).

Differences in research object

A complicating factor in comparing main findings from TIMSS and ESIE is that there is a major difference between the two as regards research object. TIMSS primarily focuses on the measurement of students' performances in mathematics and science: the attained curriculum. Data at the level of the implemented curriculum (e.g. teachers' judgments about opportunity to learn) or at the level of the intended curriculum (match between test items and attainment targets) are meant as context information for interpreting students' performances. ESIE has the implemented curriculum as its primary focus, with data about students' outcomes as measured by some national tests (attained) as well as data about the match between the implemented curriculum and the attainment targets (intended) as secondary sources. Due to these differences, a direct comparison between the two studies is hard to make. Nevertheless, there are findings from both studies that seem to indicate that in this regard, the two studies are consistent with each other. As a consequence, both studies seem to provide us with sufficient input for the disentangling attempt envisaged.

Main findings: Attained curriculum

As far as the written test is concerned, Dutch Secondary 2 students performed relatively well in the TIMSS comparison in both 1995 and 1999. About two-thirds of the students achieved above the international mean for both mathematics and science (41 countries in 1995, 37 in 1999). However, Dutch students did not score outstandingly well in the TIMSS 1995 performance assessment, although curriculum experts judged the practical test as matching well with the attainment targets. The students' overall mean achievement (average 61% correct), was near the international average (average 59% correct). Five years later the overall mean achievement (average 64% correct) had improved slightly but significantly, due to a better performance on the science tasks only.

In ESIE, student performances on the 'basic secondary education tests' have been compared with the standard of achieving 'above the level', 'on the level', or 'below the level'. For each student ability track and for each content area, these standards had been set by teachers and subject-matter experts. The picture for mathematics and for the science subjects didn't appear to be univocal (Inspectie van het Onderwijs, 1999c, d, and e). The results roughly showed that for biology and physics/chemistry, students from the two higher ability tracks (havo, vwo) generally did not meet the standards of performing 'at or above their level', contrary to students from the two lower ability tracks (vbo, mavo) who generally performed 'at or above their level'. For mathematics the results were the other way around. From other analyses at the aggregate level of the exact sciences (mathematics, physics/chemistry, biology as well as technology), it appeared that 67% of the students from the lowest ability track (vbo) performed 'at or above their level'

(Inspectie van het Onderwijs, 1999a). The same was true for 67% of the students from the second highest ability track (havo) and for 53% of the students from the highest ability track (vwo). This means that 33% (vbo), 33% (havo) and 47% (vwo) of the students performed *below* the standard that had been set by teachers and subject matter experts for their ability level (for readers unfamiliar with the Dutch school system, a valuable source of information is http://www.minocw.nl).

Reflective comments

In TIMSS student achievement has been measured via *international tests*. Students' performances on these tests have been expressed in a country's mean score for both mathematics and science. This mean score determined a country's position in the international ranking, and also its position relative to the *international mean*. The criterion for a country's performance is referenced to the international mean, which in turn depends on the number of participating countries, as well as their performance levels. The substantial number of developing countries participating in TIMSS makes the performance of Dutch students appear relatively good, but one may wonder what is usefulness of such a comparison.

In ESIE quite a different approach has been applied. Teachers and subject matter experts formulated standards for each student ability track. Next, it was determined to which extent students were able to meet those standards. In addition, those *absolute standards* were *based on the attainment targets*, which are a much more ambitious criterion than an international mean on a test that covers "an *internationally* consensual body of content defining mathematics (and science)" (Westbury, 1992, p. 19). Another difference with TIMSS is that the (secondary analyses of) student performances reported in ESIE were based on (drafts of) *national tests*. The administration of these laborious and time-consuming tests took some doing. As a consequence, the results may be disputed.

So, in the two studies, student outcomes have been measured using different tests. Also there are large differences in standards that have been used as a reference. Differences in standards make it possible to judge student performance in TIMSS much more positively than those in ESIE. However, due to differences in test instruments used and standards set, making comparisons between student achievements in both studies does not make sense.

Main findings: Implemented curriculum

In TIMSS-1999 it appeared that 82% of the mathematics teachers and 64% of the science teachers determined each of the items from the written test appropriate to the implemented curriculum (Vos & Bos, 2000). For the performance assessment, 58% of the mathematics teachers and 46% of the science teachers came to such a judgment (Vos & Kuiper, 2004). So, the appropriateness of the written test, the written science test, and the performance test (both mathematics and science) to the implemented curriculum was good, less good, respectively moderate. This conclusion implies that part of the content tested had *not* been taught before test administration. The finding that students, in spite of this, performed relatively well is striking and suggests an influential role of extracurricular factors like school and socio-cultural context on student outcomes (Figure 1). Unfortunately, there are no findings, either from TIMSS or from ESIE, that might offer further clues in this respect.

In order to picture the implemented curriculum in ESIE a distinction was made on five quality standards for teaching and learning at classroom level (van den Bergh et al., 2000): positive class climate, class management and teaching and learning approach, pedagogical content approach, promotion of active learning, and considering individual differences. Indicators were set for each of these standards. Based on observations, the various standards were rated in terms of: predominantly weak, weakness dominates strength, strength dominates weakness, and predominantly strong. Strengths of the implemented mathematics, physics/chemistry, and biology curricula appeared to be class climate as well as, but to a less extent, class management and instructional approach. The majority (varying from 61% to 83%) of the lessons observed were rated as 'at least sufficient' with regard to these two standards. The pedagogical content approach was rated as 'not sufficient' in half (physics/chemistry and biology) or one-third (mathematics) of the lessons. Promotion of active learning was rated as 'not sufficient' in about half of the mathematics, physics/chemistry, and biology lessons. In two-thirds or more of the lessons individual differences were not sufficiently considered.

These findings brought the Inspectorate to the overall conclusion that a teaching approach that promotes active learning – one of the key-characteristics of the intended curriculum reform in basic secondary education – is still lacking. The Inspectorate's conclusion, also reached by Kuiper (1993) in an earlier study on science teaching practices, seems to be to some extent in line with the TIMSS performance assessment findings. The performance assessments in 1995 and 2000 showed that not only student achievement, but also the appropriateness of the performance test to the implemented curriculum, turned out to be less suitable than hoped (Vos & Kuiper, 2004). The Inspectorate's conclusion can be regarded as a support for the latter, as it is plausible that a predominantly instructivist and teacher-centered approach has a detrimental effect on mastering practical skills.

Reflective comments

In order to get an understanding of the appropriateness of the international tests to the implemented curriculum, in this component of TIMSS the implemented curriculum has been conceived as 'Did the students have the opportunity to learn the content tested?'. Those content coverage findings are crucial context information for interpreting student achievement, but of course the implemented curriculum represents more than only opportunities to learn content tested. In ESIE a much broader definition of the implemented curriculum was used, much more resembling the curriculum-in-action definition given in the beginning of this chapter. To some extent this broader definition can also be recognized as a guide to the TIMSS teacher questionnaires, but the latter instruments were designed and administered merely to obtain "some information about the implemented curriculum" (Beaton, Martin &

Mullis, 1997, p.53). Getting a thorough understanding of instructional practices using only written questionnaires is a utopian situation indeed. For that purpose, a larger array of research methods and instruments is needed, similar to what was done in ESIE.

So, although there are differences in research focus between the two studies, in conceiving the implemented curriculum as well as in research methods and instruments, the findings that are available from TIMSS and ESIE as regards the implemented curriculum give support to the conclusion that the two studies are rather consistent.

Main findings: Intended curriculum

Mathematics education experts determined that, on average, 69% (1995) and 72% (1999) of the mathematics items from the TIMSS written test were appropriate for the attainment targets for mathematics at Secondary 2 level. Science education experts came to a comparable judgment for the science items (on average, 70% in 1995 and 69% in 1999). The appropriateness of the performance test was of about the same order: 9 out of 12 mathematics and science practical tasks were rated as matching the core objectives.

As part of TIMSS-1999 (written test), the item ratings by mathematics and science education experts (intended curriculum) were compared with the 'opportunity to learn' item ratings by teachers (implemented curriculum). For that purpose the mathematics and science items were split up in a set 'appropriate' for the intended curriculum (111 mathematics items, 72%; 69 science items, 69%) and a set 'not appropriate' (44 mathematics items, 28%; 44 science items, 31%). Next, these four categories were cross-indexed with teachers' ratings. Mathematics teachers generally appeared more positive in their ratings than mathematics experts. For science, however, the teachers' ratings were generally consistent with the ratings made by the experts.

ESIE showed that there was only a partial match between the intended curricula at school level – as appearing from analyses of textbooks, additional curriculum materials, schools' work plans, teacher questionnaire and interview data, and lesson observations – and the attainment targets for mathematics, physics/chemistry, and biology. The match for the 15 subjects altogether (including mathematics, physics/chemistry, and biology) varies from, on the average, 40% for the lowest ability track to 59% for the highest track. The match for biology (32% - 46%) and physics/chemistry (33% - 54%) is less than the overall average in each student ability track; for mathematics (55% - 68%) it is the other way around (Peters-Sips et al., 2000). Another finding was that the general skill 'conducting a simple inquiry' is sufficiently part of the intended curriculum at school level for both biology (65% of the lessons) and physics/chemistry (62%), but not for mathematics (21%). Another general skill, 'collaborating with peers', is only sufficiently part of the intended school curriculum for physics/chemistry (62%); this general skill is only visible in 40% of the biology lessons and 46% of the mathematics lessons. A third component investigated in ESIE was the match between, on the one hand, the core objectives and the goals pertaining to general skills, and, on the other hand, textbooks in most common use. It appeared that textbooks have a substantial match with the attainment targets, but not with goals pertaining to general skills. ESIE also showed that: (i) textbooks generally reflect the attainment targets, but aims pertaining to general skills goals are much less visible; (ii) although teachers heavily rely on textbooks in their teaching practice, the implemented curriculum is only a slight representation of how attainment targets and general skills goals have been represented in textbooks.

Reflective comments

The TIMSS finding, that in 1995 and in 1999 Dutch students performed relatively well on the written mathematics and science test, which consisted of about 30% of items *not* appropriate for the attainment targets, indicates that the relatively good scores for these students are based on an item set that partly consisted of 'policy irrelevant' items. In addition, that about 30% of the items were not covered by the attainment targets, shows that more content has been tested than is covered by the attainment targets. Unfortunately, it has not been analyzed what that 'more content' refers to. However, another observation form ESIE, that teachers still teach 'old' content, nourishes the assumption that 'more content' refers to the old curriculum (i.e. that which preceded the basic secondary education era). So, the new still seems to be blended with the old. This seems to happen not only at the level of the intended curriculum but also at the implemented curriculum level. The latter can be inferred from the ESIE finding that the implemented curriculum matches poorly with textbooks which, in their turn, match well with the attainment targets and general skills goals. Next to this, it is not unlikely that variables in the socio-cultural context, as well as student characteristics, are influential. The TIMSS finding, however, that about 70% of the mathematics and science items from the written test match with the attainment targets, raises the question whether all attainment targets have been represented in the written test and, if not, which attainment targets have been omitted. Also no data are available on this topic. However, that the TIMSS written test does not fully cover the attainment targets goes without saying.

4. DISCUSSION

In the foregoing we have tried to analyze conflicting images of the quality of lower secondary mathematics and science education in the Netherlands as appearing from two large-scale studies, TIMSS and ESIE. From this analysis a number of conclusions can be drawn, and some additional reflective comments made.

It was emphasized already that the two studies differ in regard to the *research object*. In TIMSS the attained curriculum is the primary focus; in ESIE the emphasis is on the implemented curriculum. Our analysis has made clear that this difference in research object results not only in differences in outcomes of the two studies but also in difficulties in explaining those differences. However, *research methods and instruments* also differ. As a consequence, it is impossible to make comparisons between a number of findings, a problem that becomes clearly obvious at the level of the attained curriculum. A third difference between the two studies deals, as it seems, with *conceptualizing the curriculum*. Due to its main interest in student achievement, in TIMSS the primary focus is on test-curriculum matching issues; in ESIE the researchers seem to rely on a broader curriculum concept. Notwithstanding these important conceptual and methodological differences, there seem to be findings at the level of the implemented and the intended curriculum, that give support to the conclusion that between the two studies there is also some ground in common.

The area of the common ground, though, is small. It is so small that no proper explanation can be given for the expected occurrence of the process of 'curricular dilution' in ESIE findings vis-à-vis the unexpected non-occurrence (or better: the reverse) of this phenomenon in the TIMSS written test findings. Curriculum dilution appears from the ESIE findings in terms of the following areas: (i) unsatisfactory student achievement on basic secondary education tests (attained); (ii) in the context of teaching approaches in which the promotion of active learning is still lacking (implemented); (iii) against the background of attainment targets (intended) that only partially match with the intended curriculum at school level. In these findings a dilution process is visible that is consistent with the curriculum transformation process depicted in Figure 1. However, the main TIMSS findings seem in contrast, as, again roughly speaking, Dutch students perform relatively well on the written test (attained) – the appropriateness of the written test to the implemented curriculum was good for mathematics and less good for science (implemented), while almost one third of the items were rated not appropriate to the attainment targets (intended). A good explanation for the non-occurrence of the dilution process in the TIMSS findings cannot be given. When commenting and reflecting on the results from the two studies pertaining to the intended and implemented curriculum, the TIMSS findings (with support inferred from ESIE) seem to point more in the direction of the occurrence of a process of 'curricular blending'.

A further reflection on the curriculum typology, taken as the stepping-stone in our analysis, in relation to one of the main results of TIMSS brings us to a further comment. In TIMSS a partial match was found between the international test and the attainment targets for mathematics, physics/chemistry, and biology. This finding might suggest that a full match is something for which one should strive. However, as experiences with a national option mathematics test (administered in 1995 in addition to the written test; Kuiper, Bos & Plomp, 2000) have taught us, a proper match is not a guarantee for proper student achievement. Even more important, trying to realize a full coverage of the attainment targets in tests seems to be a kind of a top-down approach that doesn't make sense in the Netherlands. It seems to be more fruitful to take the intended curriculum as a guide. A partial match between the test and the intended curriculum is not a problem. Via the attainment targets, an intended curriculum for mathematics, physics/chemistry, and biology has been framed that can and should be perceived as an area within which schools can make and account for their own choices. Such an approach fits the national government's new education policy to give schools more autonomy and responsibility in making their own curricular choices. Enlarging schools' autonomy may at the same time be a lever in creating more dynamics in the transformation process as depicted in Figure 1. Quite typical for the Netherlands is that, so far, the arrows between the three curricular appearances predominantly point to the right (from intended to attained) instead of also to the left (from attained to intended).

In conclusion to our analysis we make two final comments.

First, the linking of student performance on international tests, which are administered as part of studies like TIMSS (and PISA), also to *national* standards (ESIE) is to be preferred above making comparisons with only an international mean as reference. Such an approach puts achievement results in a nationally relevant perspective, which in turn is a prerequisite for inferring meaningful policy implications aiming at the improvement of the quality of mathematics and science education.

Second, in international comparative studies there is also needed a broader conception of curriculum than mainly 'content (to be) taught and learned' and 'goals and objectives (to be) achieved'. In presenting our conceptual framework, we have emphasized that content and goals/objectives are only two of nine components to which a curriculum can pertain. Data on the match between 'content tested' (attained), 'content taught' (implemented), and 'content to be taught' (intended) are very relevant, but a broadening of curricular focus to teaching practices, especially, will provide vital clues for interpreting student performance (see focus of Inspectorate Study). The implemented curriculum (still reflecting traditional features as ESIE clearly shows) is the link between the intended and the attained. In TIMSS, however, this curricular appearance currently it is too much like a black box to provide a worthwhile frame for interpreting (discrepancies between) the attained, implemented, and intended curriculum. In more practical terms this means that, if possible, there should be an attempt to connect future studies like TIMSS and ESIE to each other in such a way that the strengths of both are exploited. We have indicated in this paper the most salient issues to consider for this. In exploiting the strengths of several studies, it is probable that less energy will be needed to disentangle conflicting images.

REFERENCES

Beaton, A.E., Martin, M.O. & Mullis, I.V.S. (1997). Providing data for educational policy in an international context: The Third International Mathematics and Science Study (TIMSS). *European Journal of Psychological Assessment, 13* (1), 49-85.

Boersma, K.Th. (2000a). Het leerplan van de basisvorming als problem. In M. Peters-Sips, J. van der Linden & A. Wald (Eds.), *Verder werken aan de basis. Basisvorming bundelt krachten* (pp. 43-56). Utrecht: Inspectie van het Onderwijs.

Boersma, K.Th. (2000b). Oorzaken en aanpak van overladenheid van het operationele curriculum van de basisvorming. *Tijdschrift voor Onderwijsresearch, 25* (1/2), 110-117.

Bos, K.Tj. (2002). *Benefits and limitations of large-scale international comparative assessment studies: The case of IEA's TIMSS study* (doctoral dissertation). Enschede: University of Twente.

Bos, K.Tj. & Vos, F.P. (2000). *Nederland in TIMSS-1999. Exacte vakken in leerjaar 2 van het voortgezet onderwijs.* Enschede: Universiteit Twente.

Bos, K.Tj., Kuiper, W. & Plomp, Tj. (1999). Student performance and curricular appropriateness in the Netherlands. *Studies in Educational Evaluation, 25*, 269-276.

Inspectie van het Onderwijs (1999a). *Werk aan de basis. Evaluatie van de basisvorming na vijf jaar. Algemeen rapport.* Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs (1999b). *Bijlagen bij ' Evaluatierapport onderwijsleerproces basisvorming'*. Utrecht: Inspectie van het onderwijs.

Inspectie van het Onderwijs (1999c). 3: *Biologie in de basisvorming. Evaluatie van de eerste vijf jaar*. Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs (1999d). 15: *Natuur- en scheikunde in de basisvorming. Evaluatie van de eerste vijf jaar*. Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs (1999e). 19: *Wiskunde in de basisvorming. Evaluatie van de eerste vijf jaar*. Utrecht: Inspectie van het Onderwijs.

Kellaghan, T. (1996). IEA studies and educational policy. *Assessment in Education,3* (2), 143-160.

Kuiper, W. (1993). *Curriculum reform and teaching practice* (doctoral dissertation). Enschede: University of Twente.

Kuiper, W., 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.* Enschede: Universiteit Twente.

Kuiper, W., Bos, K.Tj. & Plomp, Tj. (1999). Mathematics achievement in the Netherlands and appropriateness of the TIMSS mathematics test. *Educational Research and Evaluation, 5* (2), 85-104.

Kuiper, W., Bos, K.Tj. & Plomp, Tj. (2000). The TIMSS national option test mathematics. *Studies in Educational Evaluation, 26*.

Peters-Sips, M., Zwarts, M., Van den Berg, H. & Schuurmans, L. (2000). Kwaliteit van het vakspecifieke aanbod. *Tijdschrift voor Onderwijsresearch, 25* (1/2), 40-52.

Robitaille, D.F., Schmidt, W.H., Raizen, S., McKnight, C. Britton, E. & Nicol, C. (1993). *Curriculum frameworks for mathematics and science.* TIMSS Monograph No. 1. Vancouver: Pacific Educational Press.

Schmidt, W.H., McKnight, C.C., Houang, R.T., Wang, H., Wiley, D.E., Cogan, L.S. & Wolfe, R.G. (2001). *Why schools matter. A cross-national comparison of curriculum and learning.* San Francisco, CA: Jossey-Bass.

Schmidt, W.H., McKnight, C.C., Valverde, G.A., Houang, R.T. & Wiley, D.E. (1997). *Many visions, many aims (Volume 1). A cross-national investigation of curricular intentions in school mathematics.* Dordrecht: Kluwer.

Schmidt, W.H., Raizen, S.A., Britton, E.D., Bianchi, L.J. & Wolfe, R.G. (1997). *Many visions, many aims (Volume 2). A cross-national investigation of curricular intentions in school science.* Dordrecht: Kluwer.

Van den Akker, J.J.H. (1998). *De uitbeelding van het curriculum* (orational address). Enschede: University of Twente.

Van den Akker, J. (2003). Curriculum perspectives. An introduction. In J. van den Akker, W. Kuiper & U. Hameyer (Eds.), *Curriculum landscapes and trends* (pp. 1-14). Dordrecht: Kluwer.

Van den Berg, H., Zwarts, M. & Peters-Sips, M. (2000). Kwaliteit van het onderwijsleerproces. *Tijdschrift voor Onderwijsresearch, 25* (1/2), 20-39.

Vos, P. & Kuiper, W. (2004, in press). Trends (1999-2003) in the TIMSS mathematics performance assessment in the Netherlands. *Educational Research and Evaluation.*

Westbury, I. (1992). Comparing American and Japanese achievement: Is the United States really a low achiever? *Educational Researcher, 21* (6), 18-24.

Wijnstra, J.M. (2001). *Bruikbare kennis en vaardigheden voor jonge mensen. Nederlandse uitkomsten van het OESO Programme for International Student Assessment op het gebied van begrijpend en studerend lezen, wiskunde en de natuurwetenschappelijke vakken in het jaar 2000*. Arnhem: Citogroep.