MEASURING PROGRESS IN THE IMPLEMENTATION OF A NEW CURRICULUM: LESSONS FROM THE NETHERLANDS

PAULINE VOS
Faculty of Education, University Eduardo Mondlane, Mozambique
fpvos@hotmail.com

This paper describes a case study, measuring the implementation of a totally overhauled mathematics curriculum. The case was the Netherlands, where in 1993 a new mathematics curriculum for the lower grades of secondary school was legislated, which differed in many respects from its predecessor. The current paper describes a quantitative, large-scale trend study, measuring indicators of the extent to which Grade 8 mathematics teachers were changing towards the intentions of the new curriculum, and whether students’ performance changed accordingly. The paper has a strong methodological component on measuring teachers’ curriculum implementation on a large scale. Measurements were carried out in 1999 in such a way that data analysis allowed (a) for a trend comparison with data from 1995, and (b) for a between-country comparison with Belgium (Flandres), where the curriculum had not changed. The results show that the match between intentions of the new mathematics curriculum and its implementation improved significantly, although taking several years.

INTRODUCTION
All over the world, at regular intervals the intentions of national mathematics education are reviewed or reformulated, marking a break with former practices. For example, since the nineties in the United States, the National Council of Teachers of Mathematics (NCTM) advocates a rich approach to mathematics, offering students opportunities for understanding mathematics (NCTM 2000). By formulating their Standards, the NCTM tries to break away from an algorithmic approach of mathematics, in which rules are drummed into students’ heads by repetitive exercises. In South Africa, the new Curriculum 2005 advocates a learner-centred and integrating approach, breaking away from a racist past, and from traditional teaching approaches such as rote-learning (Chisholm et al. 2000).

When a curriculum is revised at the policy-making levels of society, it remains a question whether, when and how teachers will change classroom habits. For example on the USA, there is evidence that the NCTM Standards-based teaching is rarely implemented, and moreover, when asked for their practices, teachers tend to overreport their implementation (Ross et al. 2003). In a similar vein, Harris et al. (2003) report of Gauteng science teachers who stick to teacher-centered instruction, while they perceived their lessons as learner-centered.

The phenomenon of delayed implementation also occurred in the Netherlands after 1993, when the national intended curriculum for mathematics at junior secondary schools (grades 7-9) was changed. The new curriculum emphasized data modeling 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 areas considered relevant to daily life of the new generation of the 21st century (Kok et al. 1992). The new approach was described with three keywords: applications, skills and coherence. This Dutch mathematics curriculum has similarities with the South African C2005 curriculum, as they both intend to provide learners with opportunities to apply mathematical modeling to authentic problem situations. The new Dutch curriculum differed considerably from the prior curriculum, which was more algorithmic and abstract. Therefore a large exercise was undertaken to introduce all mathematics teachers to the new content and its approach. Three years before the date of legislation, the exercise of introducing teachers to the new approach started. For example, in autumn 1990 in all provincial capitals, two-day workshops were organized by the Dutch Association of Mathematics Teachers (Nederlandse Vereniging van Wiskundeleraren). Not one teacher had to travel more than 50 km. These introductory workshops were
attended by approximately 10% of all mathematics teachers (Verhage and Wijers 1991). In the following years, these regional information conferences were repeated. For additional inspiration, anecdotal reports of experimental lessons were published in the two Dutch journals for mathematics education, Nieuwe Wiskrant and Euclides. In addition, the experimental curriculum materials were made available to teachers, together with publications on backgrounds, instructional approaches and collections of test items (e.g. Achtergronden 1992; Kok et al. 1992).

After 1993, several studies were carried out to monitor the dissemination of the new Dutch curriculum. Classroom observations showed that many teachers stuck to their traditional routines (Inspectie van het Onderwijs 1999). In 1995, two years after the legislation of the new curriculum, still 50% of all mathematics teachers used textbooks based on the abandoned curriculum (Kuiper et al. 1997). All studies observed a low match between the intentions of the new curriculum and the classroom practice where the curriculum was meant to be implemented. An additional study is described in this paper. It researched whether there was a measurable trend in the implementation of the curriculum in the years after its introduction. The study was carried out in 1999, as a sub-study from Vos (2002). Its methodology and results could be useful to other researchers interested in measuring progress in curriculum implementation in their country.

Studying and describing the implementation of a curriculum means studying and describing processes in schools and classrooms. It can be carried out in various ways, for example through case studies by direct observation. However, classroom visits are labour-intensive, as it takes a large number of visits before generalised statements on the implemented curriculum for a whole nation can be made. Moreover, the description of classroom processes proves a complex exercise. Classroom processes can take place almost invisibly or simultaneously. The observer (or the cameraman) cannot be an invisible fly-on-the-wall, but becomes part of the classroom processes, making them less genuine. To study the implemented curriculum, alternative research methods replacing classroom observation have been sought. A low-cost alternative is the use of surveys among teachers, gathering data on the mathematical content covered at a certain level, teaching approaches or educational beliefs of teachers (e.g. Mullis et al. 2000; Travers and Westbury 1989). Yet, as teachers might tend to give socially desirable answers, the reliability of the answers remain questionable and accurate instruments are needed (Ross et al. 2003). This paper will focus on an instrument that measures what mathematical content has been covered by the teachers. Unlike the instrument from Ross et al. (2003), it does not provide data on the instructional approach (e.g. facilitating collaborative learning through groupwork, projects, etc.).

CONCEPTUAL FRAMEWORK

The study used a distinction between the macro, meso and micro level, in a similar vein as other recent studies on curriculum implementation across the world (e.g. Naidoo 2002; Schmidt et al. 2001). The conceptual framework came from the International Association for the Evaluation of Educational Achievement (IEA). It names a curriculum representation at each level (Robitaille et al. 1993):

- the intended curriculum (what society at large prescribes students to learn),
- the implemented curriculum (instruction at classroom level) and
- the attained curriculum (what is actually learnt by students).

Using these curriculum concepts, the match between the curriculum levels was termed as curricular alignment, indicating a convergence or divergence between intended, implemented and attained curriculum. When measuring in quantitative terms, the curricular alignment can be expressed as a correlation coefficient ($r=0$ means a lack of correlation, $r=1$ means a perfect linear relationship).

The study used the achievement test of the Third International Mathematics and Science Study (TIMSS), which was administered in 1999. This study will be termed TIMSS-99, to distinguish from TIMSS-95, the similar study carried out in 1995. The TIMSS-99 achievement test yielded results at the level of the attained curriculum. The same test items were also used at the other two curriculum levels, by asking for a judgement from (a) teachers and (b) curriculum experts, in order to obtain a validation of the test at the
level of the implemented and intended curriculum respectively. Like in many other countries, the TIMSS test was controversial in the Netherlands. The discussion focused on the fact that it comprised a considerable number of items (±30%), which reflected the abandoned curriculum. However, this fact proved useful for the current study, as the judgment of the teachers on these items would reveal the extent to which they implemented the new curriculum. And their judgment on the items that were remote from the new curriculum would reveal the extent to which they still taught content from the abandoned curriculum. For a satisfactory alignment of the three curricular levels, test items that received a positive judgement from the curriculum experts, were expected to also receive an average positive judgement from the teachers, and these items were also expected to be well accomplished by the students. An example of a test item, matching well with the intended curriculum is given in Figure 1. It describes a practical context, hiding notions of linear relations, requiring the comparison of prices per months with prices per area per year.

Conversely, the TIMSS tests also contained items that would only match with the abandoned intended curriculum, and not with the new curriculum. See Figure 2 for an example of such an item, carrying variables, which are detached from an interpretation in every day life. For a satisfactory alignment between the curricular levels, these items should receive a more negative judgement from the teachers, and have a relatively lower students’ score.

**Figure 1. Example of an item that matched with the intended mathematics curriculum.**

**Figure 2. Example of an item that did not match with the intended mathematics curriculum.**
For reference, the same measurement was carried out in Belgium (Flandres), where the intended mathematics curriculum had remained unchanged. It was alleged, that if the alignment in the Netherlands had comparable rates to the alignment in Belgium (Flandres), the curricular situation would be relatively acceptable. The between-country comparison was considered appropriate because of similarities in economical, social and cultural aspects. Moreover, both countries have the same language, enabling the use of identical instruments. Additionally, the measurement could be compared with earlier data from TIMSS-95.

INSTRUMENTS
The instruments in the study consisted of the TIMSS-99 achievement test to measure students’ performance. Each test item was given to a representative sample of students, and their achievement would be expressed as a percentage of students having a correct score (the percentage correct). For example, item V02 (Figure 1) yielded a percentage correct of 45, indicating that 45% of the Dutch students solved this item correctly. Similarly, item R09 (Figure 2) yielded a percentage correct of 27, indicating that 27% of the students solved this item correctly.

At the level of the intended curriculum, an instrument was available that was used in a sub-study from TIMSS, the Test-Curriculum Matching Analysis (TCMA). Curriculum experts were asked to review each item from the test, and assess whether its content matched with the intended curriculum for the majority of the target population (Mullis et al. 2000). Consequently, each item would receive a dichotomous label of whether it matched or not with the intended curriculum. For example, item V01 (Figure 1) received a “1” (meaning: “yes, it matches”), and item R09 (Figure 2) received a “0” (meaning “no, it does not match”). Aggregating these judgements over the whole test yielded the intended curriculum matching rate of 72 (the percentage of items receiving a “yes”).

At the level of the implemented curriculum, an instrument was available, developed and validated by De Haan (1992). It is item-based, asking teachers for each item whether their students had an Opportunity to Learn (OTL) the content of the item. Items offer teachers a concrete cognitive situation, stating implicitly aspects of content and skills. Not only teachers, but also students, can be given a set of test items with the question, whether these have been covered in their lessons. This method was used, among others, by Travers and Westbury (1989). In some studies, teachers were additionally asked to indicate the time spent on that content, thus differentiating the yes/no question. Travers and Westbury (1989) have argued that the question on content coverage interfered with teachers’ perception of the item’s difficulty. They reasoned that teachers would include an estimation of their students’ achievement before answering the coverage question. Therefore, Travers and Westbury asked teachers to distinguish between item difficulty and item coverage. Teachers were asked to indicate the coverage, and separately make an estimation of the percentage of their students answering the item correctly.

Considering the interference of coverage with difficulty, De Haan (1992) developed a special instrument to measure OTL, asking teachers to imagine that they were to set a test for their students, covering all content taught. They were then given a set of items, and asked to judge each item on whether they would include it into their imaginary test. This implied that either the content of the item had been covered in classroom, or that the item was proximate to the content taught and teachers considered his/her students to be ‘ready’ for the item. In both cases, teachers’ answers were an indicator of the item’s match with the implemented curriculum. The OTL rate of an item was then calculated as the percentage of teachers, indicating ‘yes’ on that item. De Haan checked the validity of her instrument, and its reliability and objectivity in a research, in which the obtained OTL rates were compared to (a) a very detailed OTL questionnaire asking for each item when, how and for which sub-group of students it was covered, (b) an estimation by teachers of item difficulty (i.e. teachers’ prediction of student scores on the item), (c) a textbook analysis, (d) students’ own OTL judgement, and (e) students’ achievement. The new OTL instrument proved to correlate well with these, especially with the detailed OTL questionnaire and with the textbook analysis (r=0.9 and 0.7 respectively, with p<0.01), testifying of the validity of the instrument.

In the present study, this cost-efficient OTL instrument from De Haan (1992) was used, yielding an OTL
rate for each test item. For example, item V02 (Figure 1) received an OTL rate of 89, indicating that 89% of the teachers thought this item fit for a test. Item R09 (Figure 2) received an OTL rate of 12, indicating that only few (12%) of the teachers thought this item fit for a test.

SAMPLES
In the spring of 1999, a random sample of n=5256 Belgian (Flemish) students and a random sample of n=2957 Dutch students were tested, according to the strict TIMSS sampling procedures (Mullis et al. 2000). These data were collected as part of the international TIMSS-99 study. Included in the TIMSS-99 study was an assessment by three curriculum experts (a curriculum developer, a researcher, an INSET-trainer) on their judgment of the test items against the light of the intended curriculum. This data collection was carried out in all participating countries. The data from the prior measurement in 1995, TIMSS-95, also included this expert appraisal. Simultaneously with the testing of the students, n=269 Belgian (Flemish) mathematics teachers, and n=112 Dutch mathematics teachers were asked to fill in the OTL instrument.

To compare the data in time between TIMSS-95 and TIMSS-99 the achievement tests of these measurements were analysed, as they were not exactly equal. TIMSS-95 contained 150 items, and half of the items from 1995 had been published and replaced by new items for the test of 1999. TIMSS-99 contained 155 items. However, many of the published items from TIMSS-95 were cloned by changing some digits. Many of the new items did not show any difference in content, nor in difficulty, and were considered equivalent to their partner in TIMSS-95. Comparing the items between the two achievement tests, there were 144 pairs of these equivalent test items (96% of the test) between TIMSS-95 and TIMSS-99, allowing for a valid trend comparison at all curriculum levels. One problem in the trend comparison remained: the data collection of 1995 had not been carried out at the level of the implemented curriculum. As a consequence, measuring the curricular alignment of 1995 was limited to correlating data from the intended and attained curriculum.

RESULTS
The obtained data were collected at item level, which meant that for each item there were findings available, at the level of the intended, implemented and attained curriculum respectively. Table 1 shows the results at each separate level, aggregated over all test items. The Dutch results of TIMSS-99 and TIMSS-95 do not differ much. The Belgian (Flemish) results show a very high appreciation of the test at the level of the intended curriculum (all but three items were judged as suitable against the light of the intended curriculum), causing a ceiling effect. Correlating any data with the Belgian (Flemish) intended curriculum would be hampered by lack of differentiation.

<table>
<thead>
<tr>
<th>Measurement (number of items)</th>
<th>Intended curr. matching rate</th>
<th>Implemented curr. average OTL rate</th>
<th>Attained curr. avg % correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL maths 1999 (n=155)</td>
<td>72</td>
<td>82 (16)</td>
<td>63 (19)</td>
</tr>
<tr>
<td>Belg maths 1999 (n=155)</td>
<td>98</td>
<td>84 (12)</td>
<td>68 (16)</td>
</tr>
<tr>
<td>NL maths 1995 (n=150)</td>
<td>71</td>
<td>----</td>
<td>61 (20)</td>
</tr>
</tbody>
</table>

Note. --- Dashes indicate data were unavailable. Standard deviations are shown in parent heses.

To measure the curricular alignment, the data were compared between the curriculum levels, by calculating a correlation coefficient. This calculation could not be fully carried out for the TIMSS-95 data, as it lacked data at the level of the implemented curriculum. The correlation between the intended and attained curriculum showed to be totally absent in 1995 (see Table 2). This meant that in 1995 students performed well on items, independently of whether these matched or not with the intended curriculum.
Table 2. Bi-variate correlations between the intended, implemented and attained curriculum

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NL maths 1999 (n=155)</td>
<td>0.44*</td>
<td>0.41*</td>
<td>0.23*</td>
</tr>
<tr>
<td>Belg maths 1999 (n=155)</td>
<td>0.11</td>
<td>0.16*</td>
<td>0.14</td>
</tr>
<tr>
<td>NL maths 1995 (n=150)</td>
<td>----</td>
<td>----</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Note. --- Dashes indicate data were unavailable. * Significant correlation at 0.01 level (two-tailed).

For the measurements of 1999, all data were available. The significant correlations of Dutch data from 1999 between intended and implemented curriculum ($r=0.44$) and between implemented and attained curriculum ($r=0.41$) show that there is a certain alignment. These translate in a significant correlation of $r=0.23$ between the intended and attained curriculum (between the judgments of the curriculum experts and the achievements of the tested students). This means that, generally, in 1999 students did better on items that were considered to match with the intended curriculum, than on items that did not match with the intended curriculum. The correlation is not high, which means that on a considerable number of items, the judgements and the achievements were contradictory. However, the correlation, which had been totally absent in 1995, is not repeated. Moreover, all Dutch correlations were higher than the Belgian (Flemish) correlations. As a result, it can be concluded that the Dutch curricular alignment has improved substantially after the 1993 innovation. With educational processes of change being very slow and taking decades, the time span of 6 years can be considered as relatively short. It can therefore be concluded, that the process of dissemination of the new mathematics curriculum has gained momentum. This can be attributed to the continuous effort of all stakeholders, such as individual teachers working on their continuous development, institutions for in-service training, teacher associations, etc. (Inspectie van het Onderwijs 1999). Hopefully other countries can implement a new mathematics curriculum at a faster pace.

REFERENCES


1024