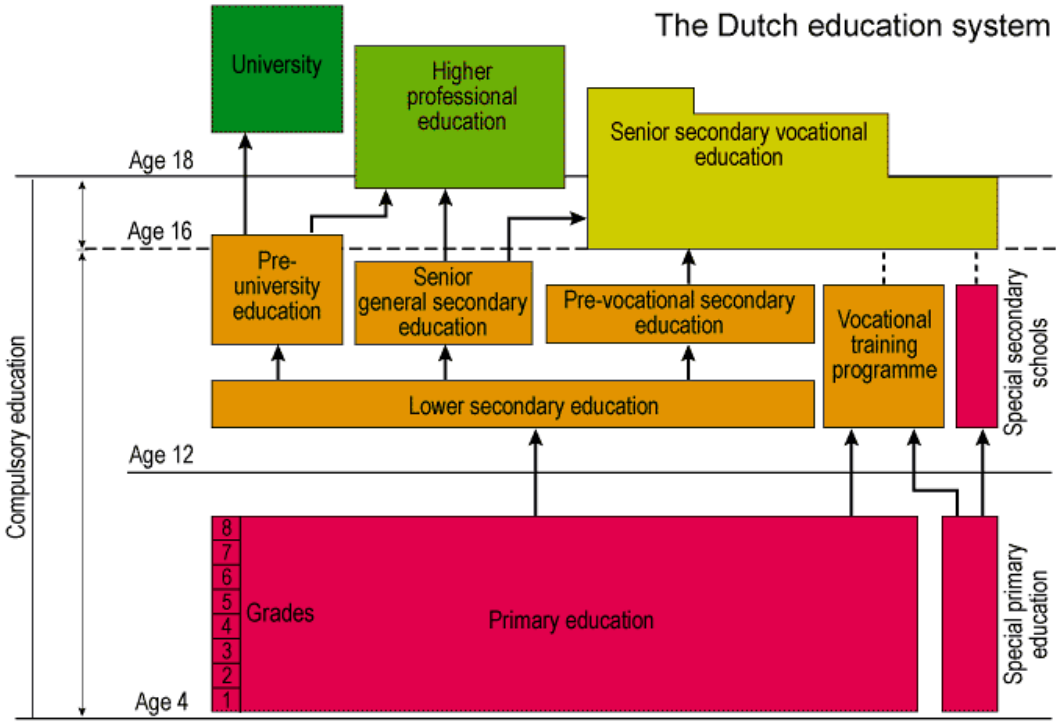
Pupil Monitoring System (PMS) for Primary Education

Marleen van der Lubbe, Cito, The Netherlands[1]

The Dutch Education System

Before explaining more about the Pupil Monitoring System (further called PMS), it is important to have some idea of the Dutch education system. In the Netherlands it is compulsory for all children to attend school full-time from the age of five. In practice, however, nearly all children start school at four. Full-time education is compulsory until the end of the school year in which the pupil reaches the age of 16. At 16 young people are still required to attend an educational institution part-time until they are 18. Generally speaking, primary education is for children aged 4 to 12 years and secondary general education is for pupils aged between 12 and 18 years. Secondary vocational education can be followed from 16 years up to 20 (see the scheme below).

Figure 1: The Dutch education system



Why should you want a Pupil Monitoring System?

Instructional guidance and a common framework of reference for assessment

For instructional guidance it is important to have a good insight into the learning progress and a good knowledge of the strong and weak sides of a pupil. For day-to-day monitoring of a learner's progress the classroom teacher can use common

[1] This text is largely based on a previous text of Fons Moelands, Cito, The Netherlands

assessment techniques such as observations, teacher-made tests, performance assessments and portfolio assessment.
Although classroom assessment is important, it has its limitations because it is subjective and person-related. Different teachers could assess the same performance differently, e.g. teacher X would see a learner as an achiever and teacher Y as a non-achiever. It is also possible that the assessment of the teacher is not consistent over time and over learners, e.g. when a teacher has to assess essay-work of different pupils it can be expected that the criteria that will be used are not interpreted in the same way for every learner: his or her assessment on Monday could be different from that on Friday. This is very likely to happen in those cases where there are no clear objective standards. If teachers do not have the same framework of reference for assessing and recording achievements, then it is very difficult to realize continuous assessment and monitor a learner from year to year and from school to school.

Monitoring the rate of progress
Another problem is the rate of progress: how does the teacher know that a learner has made sufficient progress within a certain period of time? Or: how does the teacher know that sufficient assessment criteria have been met? Or: is the pace with which we move through the teaching and learning process adequate? There is an almost inevitable danger according to international research that the pace slows down in systems that rely only on school-based assessment.
In order to help teachers in primary education obtain reliable data about the progress in the pupils' learning processes systematically, CITO has developed the PMS. The PMS complements the knowledge that the teacher has of the pupils on the basis of day-to-day progress assessment. The results of the successive assessments of the PMS are converted to the same fixed scale, so that the progress of a pupil can be monitored over a number of years. And, because there is also continuity in the recording of results, all the teachers in the school have the same frame of reference. This is of great importance for early identification of any problems. Because the PMS allows the user to combine a norm-referenced and a domain-referenced (or content-referenced) interpretation of results that are gathered nationwide, it can also be very informative for curriculum developers and policy makers.

A short outline of the Pupil Monitoring System
The PMS developed by Cito consists of a coherent set of nationally standardized tests for longitudinal assessment of a pupil's achievement throughout primary education, as well as a system for manual or automated registration of pupil progress. The PMS contains tests for measuring subject skills of Language (including decoding and reading comprehension), Arithmetic and World Orientation (Geography, History, Biology). A survey of the various tests in the system is given in figure 2.

Figure 2: Tests in the Pupil Monitoring System

	Grades (4-12 years of age)							
	1	2	3	4	5	6	7	8
Ordering	*	*						
Language	*	*						
Orientation in Space and Time	*	*						
Technical Reading			*	*	*	*	*	*
Reading Comprehension			*	*	*	*	*	*
Listening Comprehension			*	*	*	*	*	*
Vocabulary			*	*	*	*	*	*
Spelling			*	*	*	*	*	*
General Language Ability				*	*	*	*	*
Arithmetic/Mathematics			*	*	*	*	*	*
World Orientation						*	*	*
Social-emotional development			*	*	*	*	*	*
English							*	*
Science and Technology						*	*	*

Grade 1-2: Kindergarten, nursery school or reception year
Grade 3-4: Foundation Phase
Grade 5-6: Intermediate Phase
Grade 7-8: Final Phase

During the primary school period achievement tests are usually taken once or twice a year. The results of the successive assessments are converted into a fixed scale for each subject in which a pupil's progress over a number of years is monitored. The continuity in the collection of data is of great importance for early identification of any problems. In this way the PMS complements the impression that the teacher has of the pupil on the basis of day-to-day progress assessment. Moreover, the nationally standardized tests of the PMS make it possible to widen one's view beyond the classroom or the school. Thus the results of the pupils can be compared nationally with those of other children.

Working with the PMS does not merely involve testing and the registration of test results. It is an Educational System that allows teachers to make decisions about the progress of the learning process on the basis of the data collected. Should the data indicate that the pupil is not performing well, the problems will then have to be analyzed and, where needed, appropriate remedial actions will have to be taken. Therefore the PMS has been set up as a procedure that calls for a systematic, cyclic, approach.

In the systematic approach three stages can be distinguished:

1 *Identification*
   This implies all the activities that have to do with recording the pupil's achievements and interpreting the results (administrating and marking tests, recording results and first interpretations).
2 *Analysis*
   Should the results of the test show that the pupil's development is not up to standard or even stagnates, then it is desirable to collect additional data. Firstly to verify the signal and secondly to pinpoint specific problems or gaps.
3 *Actions*
   On the basis of this information a specific plan of remedial actions can be set up, carried out and evaluated. Where useful and possible, exercises and directions for use are provided for teachers.

To sum up, the PMS comprises the following elements:

- A coherent set of tests for Language/Reading, Arithmetic, World Orientation (including information processing), Social-Emotional Development, Science and Technology, and English
- A recording system, based on a measuring technique with the help of which scores are comparable on the same fixed scale in the course of time
- Means and procedures to detect the nature of the learning problems
- Didactic directions for specific help

Item Response Theory as a measuring technique

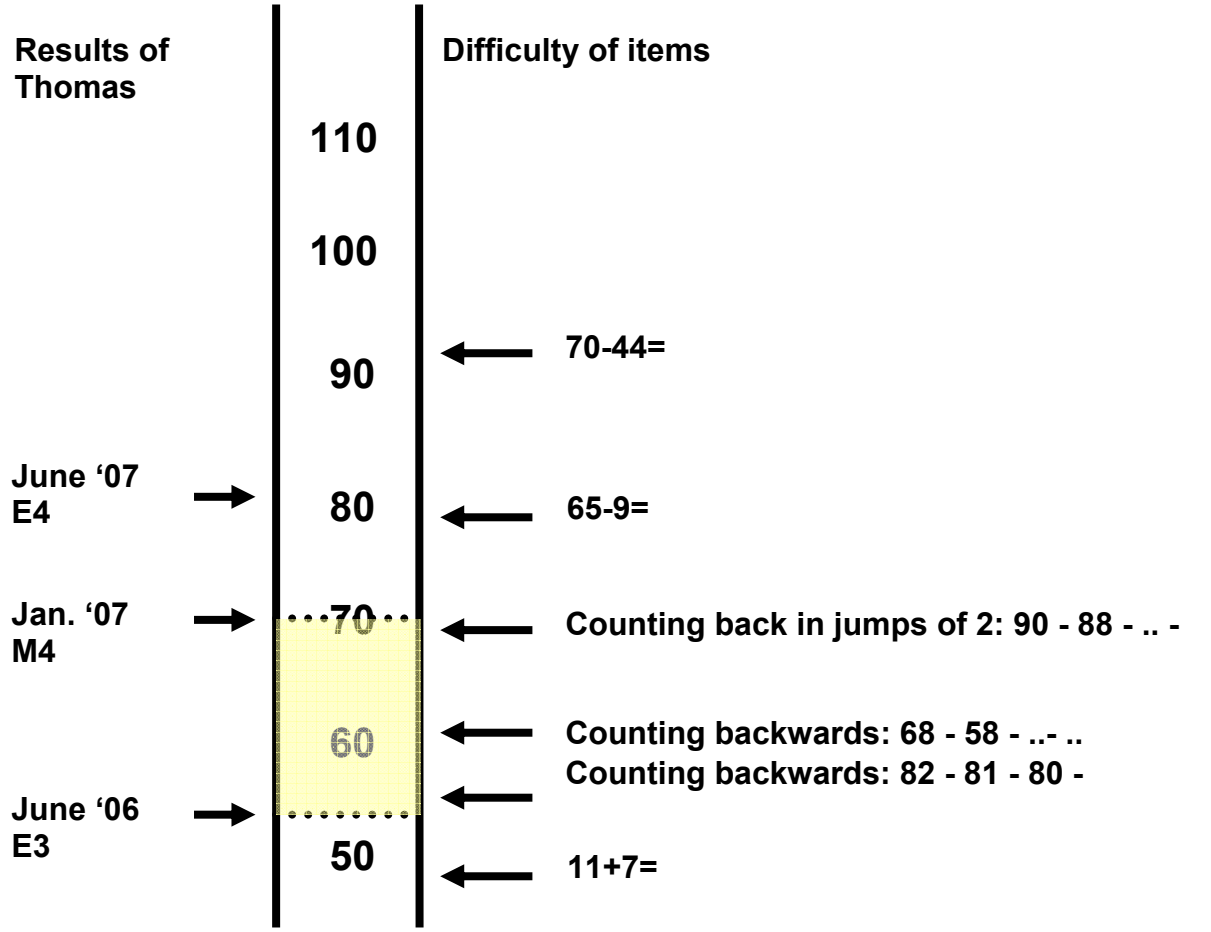
It is desirable for a system that is aimed at monitoring pupils' achievements over a number of years that the various tests of a subject matter measure the same abilities and that the results can be put on the same fixed scale. Only then it can be determined to what extent a pupil has made progress compared with a previous measurement. This possibility is offered by a measuring technique based on item response theory (IRT). IRT presents a general framework for constructing measuring instruments, validating measurements, estimating item and test characteristics, estimating individuals' abilities and spread of abilities in (sub) populations and it provides a framework for interpreting test results. In the IRT model used in the PMS the chance that an item can be solved is specified as a function of a latent one-dimensional pupil ability and one or more item characteristics (e.g. difficulty). The difficulty of the items and the latent ability can be represented on a same scale. If the model fits, the scale that measures the ability is calibrated with the help of the estimated item characteristics. This is done with the help of OPLM, a computer program developed by Cito based on a One Parameter Logistic Model.

Particularly the fact that both pupil abilities and item characteristics can be put on the same scale and can be related to each other is of great advantage to a PMS:

- The results on tests that differ according to difficulty, contents and number of items can be compared. In other words: John's results on the math tests of mid grade 4 can be depicted on the same scale as the results he obtained six months before on the math test of end grade 3, so that the degree of progress can be determined. Furthermore, the position that the pupil takes on the scale can be compared to that of other pupils nationally.
- On the basis of the position on the scale a general conclusion can be drawn about the degree of mastery of a particular subject matter.

Figure 3 gives an example of a scale consisting of several types of math items and the ability estimate of a pupil (Thomas) on the basis of the test results on Arithmetic tests that have been taken at six months' intervals (June ’06, Jan ’07, June ’07) half way the school year and at the end of the school year. Thomas’s position on the scale (June ’06, end Grade 3) indicates that he has mastered the type of items that is below his ability level (11 + 7), but that the items that are above his ability level are still too difficult for him. Items that are on the same level are partially mastered. Six months later his ability has increased (Jan ’07, medio Grade 4). Now he has mastered items that were too difficult at an earlier moment (counting backwards).

Figure 3: Part of the scale for Arithmetic



If, at the same time, for every measuring moment the spread of a (national) reference group is indicated on the scale, the relative position of the pupil compared to his 'peers' can be determined.

So we see that this technique allows three kinds of interpretations of the results:

- *Self-referenced*
  The degree of progress can be determined in relation to an earlier moment in time. After each measurement the raw score of a test is converted into a number on the ability scale, after which the difference compared to the previous scale score can be read just like measuring a child's length.
- *Norm-referenced*
  The position that the pupil takes on the scale can be compared to that of other pupils nationally.
- *Domain- or content-referenced*
  On the basis of the position on the scale a general conclusion can be drawn about the degree of mastery of a particular subject matter.

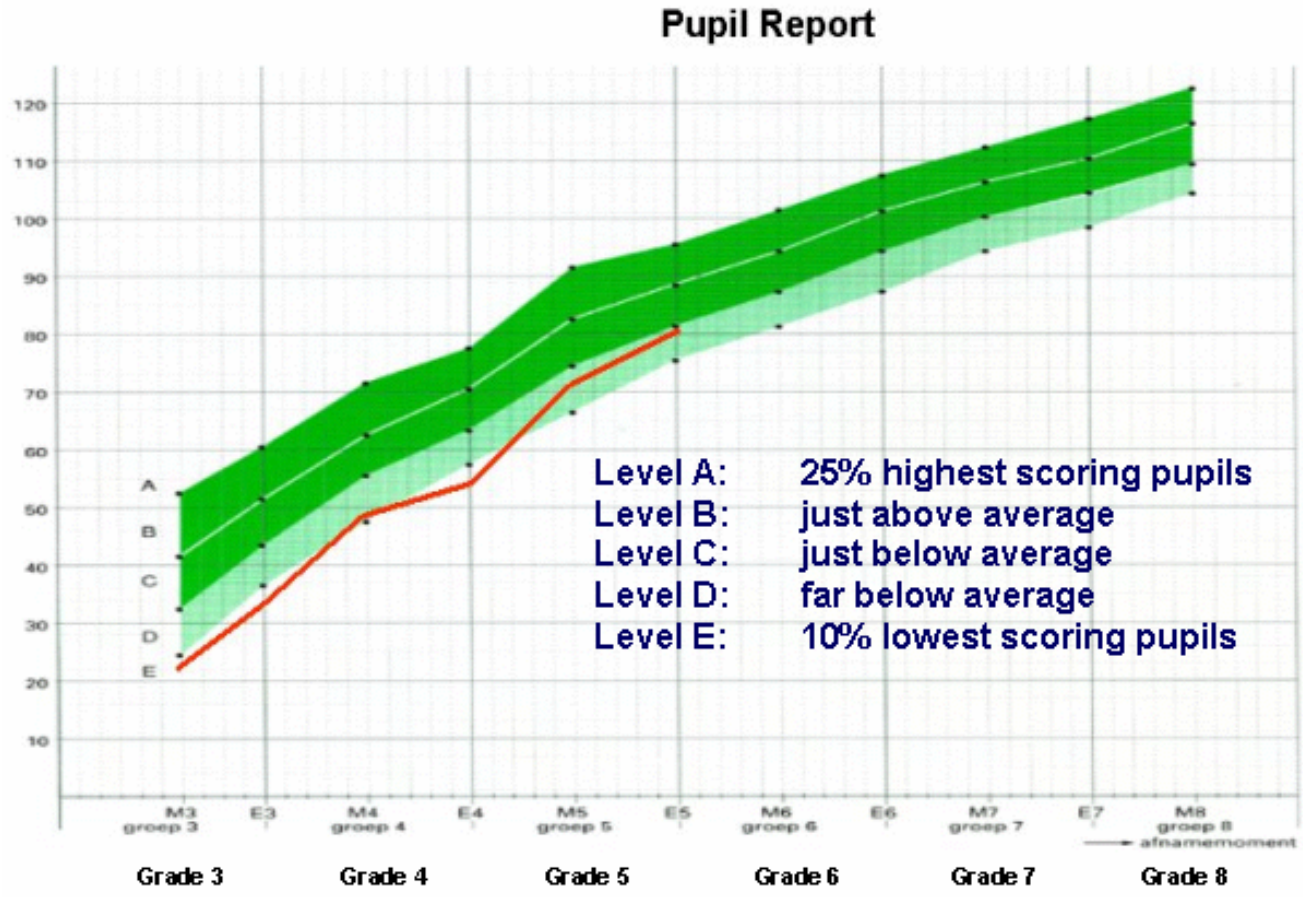
The ability-profile used for Arithmetic in the PMS is an example of a report that allows for norm-referenced and domain-referenced interpretations. The index for comprehensive reading is another example of a multi-interpretable scale. On this scale the difficulty of reading texts and the reading ability of the learner are presented on the same scale. The raw test score of the learner is transformed to a reading-index, a number on the scale. The difficulty of all kinds of reading texts can also be expressed in a number on the same scale. In this way it is possible to select texts for a learner that correspond to his/her reading ability level. A similar index has been developed for decoding.

Recording and interpretation

After each measurement the raw score of a test is converted into a number on the ability scale, after which the difference compared to the previous scale score can be read just like one measures a child's length (see figure 4). The data of the successive measurements can be recorded in two surveys: the pupil report and the group survey.
The *pupil report*, an example of which has been given in figure 4, is a graph in which the pupil's progress is visible throughout the years. In this way the teacher remains informed and can identify stagnation in the learning process at an early stage. Because the report is transparent and unambiguous, teachers of the different grades have the same frame of reference for the interpretation of the data. At a glance they can conclude to what extent the pupil has made progress. The data collected from the various subpopulations in the national survey are used as a frame of reference.

In the graph four curves have been drawn that correspond to percentiles 10, 25 and 75 and the population mean. On the basis of these data 5 levels can be distinguished:
Level A: 25% high-scoring pupils
Level B: just above average
Level C: just below average
Level D: far below average
Level E: 10% low-scoring pupils

Figure 4: Example of a pupil report

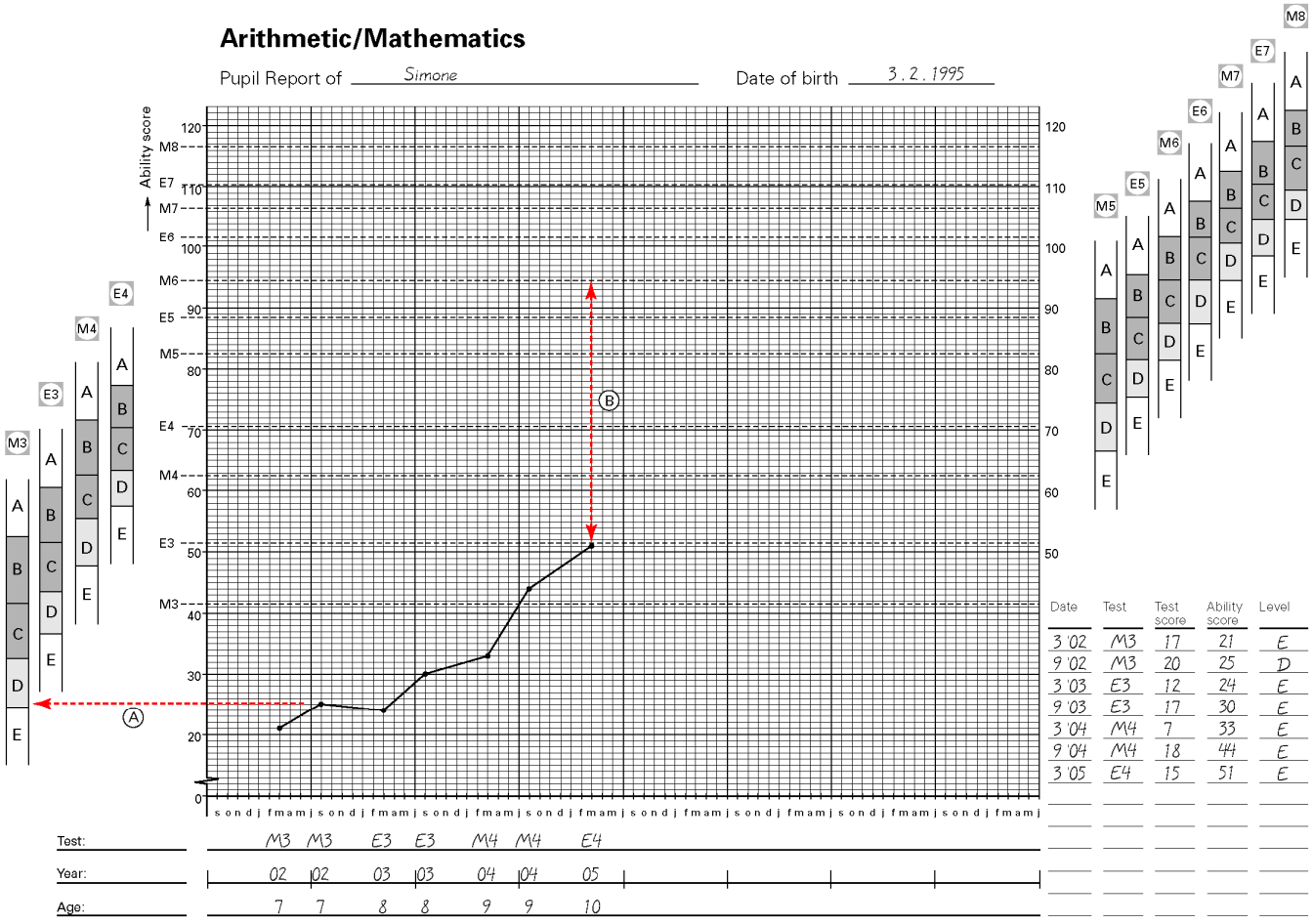


The graph shows that the pupil started out poorly compared to her peers, but is making good progress.

For children with special education needs (who attend special primary education) there is also a different kind of pupil report, the *alternative pupil report*. In this report the pupil's own progress is the central part of the report. This report also shows at what level a pupil is functioning and how to interpret the results of the pupil compared to children of the same age who attend mainstream primary education.

In figure 5 you can see that in September 2002 Simone made the Arithmetic test Medio 3 (M3). Her test score was 20 and the accompanying ability score 25. To determine Simone's level score we can draw a line from the point that represents Simone's ability score in September 2002 to the bar that belongs to the test she took at that moment, M3 (arrow A in figure 5). The ability score of Simone on the test corresponds with level D.
On the last test Simone made, E4 in March 2005, she had an ability score of 51. She then was 10 years old. Her peers in mainstream primary education at that moment made the test medio grade 6 (M6). If we want to interpret the results of Simone compared to her peers we have to draw a line between the point that represents Simone's ability score in March 2005 and the level at which her peers function (arrow B). The conclusion is that Simone has a learning disadvantage of approximately 2½ years.

Figure 5: Example of an alternative pupil report Arithmetic/Mathematics



In the *group survey* (figure 6 on the next page) the results of all the pupils from a group over a number of years are presented in a table. For each pupil the scale or ability score at the successive measuring moments is shown along with the level score.

Figure 6: Example of a group survey for Spelling

Leerlingvolgsysteem

Spelling

		25-01-93	02-06-93	23-01-94	21-06-94
	groep	M3	E3	M4	E4
Yvonne van Halst	demo	115 a	116 a	123 a	125 b
Celine Lusais	demo	110 b	114 b	116 c	118 c
David van Luyne	demo	104 c	113 b	121 b	121 c
Miranda van de Meer	demo	100 c	105 d	113 c	119 c
Laurens Plomp	demo	96 d	104 d	112 d	121 c
Anne Zandvliet	demo	93 d	95 e	105 e	111 e
gemiddelde		103.0	107.8	115.0	119.2
niveau (alle scholen)		c	d	d	c
niveau (schoolscore > 115)		b	b	c	b

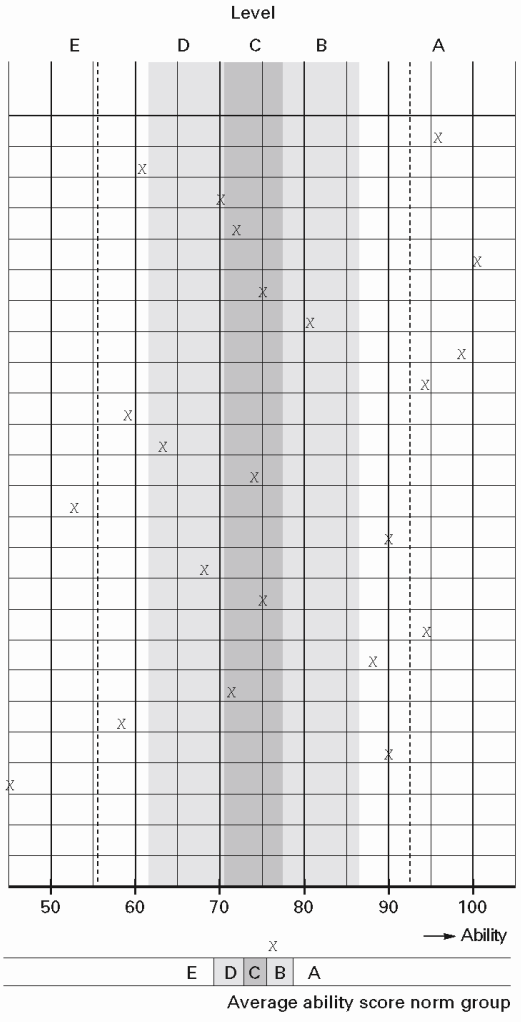
Figure 7 is an example of a group report which graphically shows the results of all the pupils from one grade. A teacher can conclude at a glance which of his or her pupils score below or above average compared to the results of other pupils nationally.

Figure 7: Example of a group report for Arithmetic/Mathematics

Groep report E5 Arithmetic/Mathematics

Date: *June 2007*

	Name	Test score	Ability score
1	*Anouk*	74	96
2	*Khalid*	35	61
3	*Eva*	48	70
4	*Max*	51	72
5	*Isa*	76	101
6	*Rik*	55	75
7	*Daan*	63	81
8	*Esser*	75	98
9	*Julia*	73	94
10	*Thijs*	32	59
11	*Michaël*	38	63
12	*Amina*	54	74
13	*Niels*	24	53
14	*Maud*	71	90
15	*Mehmed*	46	68
16	*Ruben*	56	75
17	*Sanne*	73	94
18	*Dara*	69	88
19	*Luuk*	50	71
20	*Justin*	31	58
21	*Arash*	71	90
22	*Kumari*	16	45
23			
24			
25			



Average ability score 76,2

Computer program Pupil Monitoring System

Although the tests of the PMS can be processed and recorded manually, a computer program has been developed to take over a number of the teacher's routine activities. This program is especially useful in the identification stage and in part of the analysis stage. After the test session the test results can be fed into the computer in three ways. The quickest way is to directly type up the pupils' test scores. However, it is also possible to click on the item that the pupil answered incorrectly. Both ways presuppose that the teacher has marked the test himself. In many cases a third way of feeding data into the computer is possible: directly feeding the answers given. For every item the pupil's answer is fed into the computer, after which the computer scores the test. Per pupil the most desirable way of processing data can be chosen. After the data have been fed into the computer, the computer calibrates the test-and ability scores and determines the level indication that goes with them. Then the computer can make the various reports, such as a pupil report, a group report, an answer survey, an error analysis etc. The next figure shows an example of a further analysis of the ordering test for pupils form grade 1. The red square in the figure marks the risk scores in a category.

Figure 8: Example of a category analysis for ordering

Categorieënoverzicht

Groep: **1 - 1A**
Toets taak: **Ordenen 97 - Toets M1**

Categorie	1	2	3	4	5	6	7	Tot%
Rickwin Burgers	0	17	33	17	0	50	0	17
Ibtissame Ganesh	50	50	67	17	0	67	0	36
Jessie Hendriks	17	17	33	0	50	67	50	33
Janneke Hoebers	0	0	33	17	0	0	0	7
Delano Kisters	17	17	67	33	17	33	17	29
Virgilio Speetjens	17	33	50	0	33	83	100	45
Zoë Zerouali	0	17	67	17	17	17	17	21
Gemiddelde %	14	22	50	14	17	45	26	27

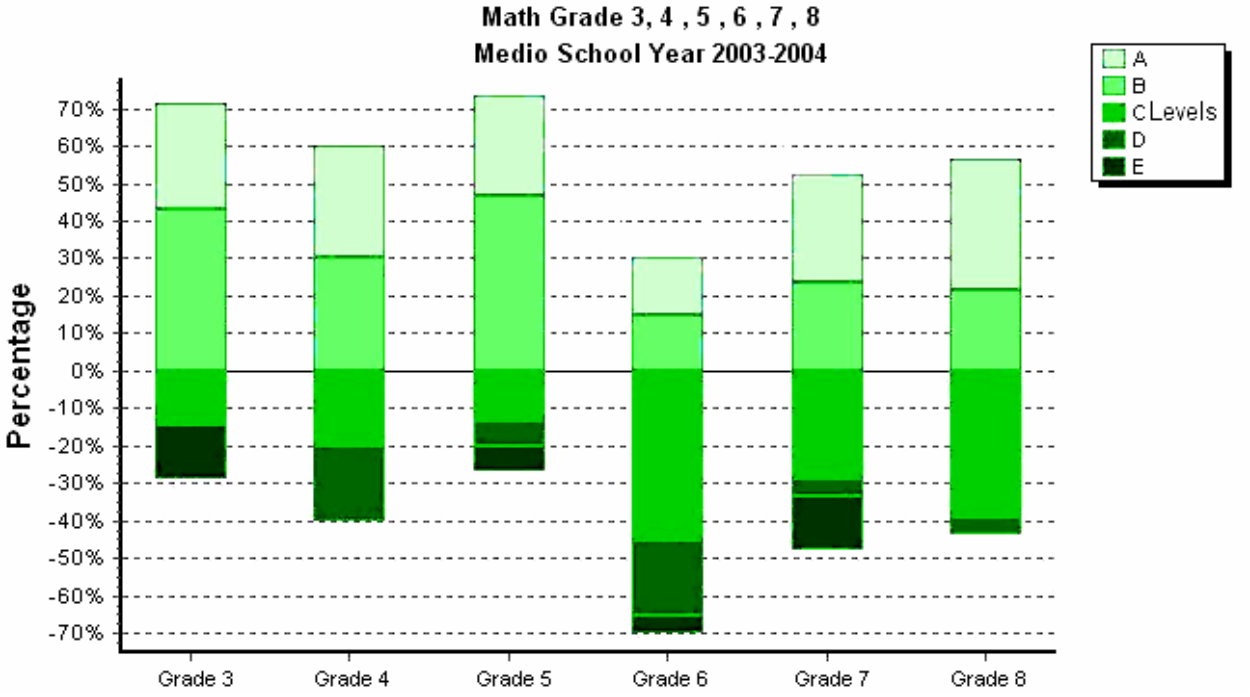
Category 1= Color
Category 2= Form
Category 3= Classification
Category 4= Size
Category 5= Seriation
Category 6= Comparing
Category 7= Counting

School self-evaluation

When the PMS is implemented in the school for a couple of years in several grades, the data gathered can also be used for school self-evaluation purposes. It is possible to fill in some reports manually, but more advanced reports can be made with a separate module of the computer program specially designed for this function. The module allows the construction of cross-section reports and trend analysis for different subjects.

A ***cross section*** shows the distribution of pupils of the different grades over the 5 levels (A to E) at a certain moment in time.

Figure 9: Example of a cross section for Arithmetic/Mathematics



The 0%-line shows the national mean. Above this line the percentage of pupils in the different grades with a level A or B are depicted. In the national reference group about 50% has an A or B-level. The other 50% has a C, D or E-level. The results of Grade 6 are eye-catching. Only 30% has an A or B-level and 70% is scoring below the 0%-line (national mean). Also compared to the results of the other grades in the same school, this result is remarkable.

The program also allows two kinds of ***trend analysis***. Figure 10 shows the results of a cohort of pupils (same group of pupils) over the years compared to the national mean in the different grades. For example the pupils from grade 7 in year 2003-2004 score almost all the years (far) below average compared to the national mean; only on the test taken in grade 7 they score around average (Medio 7). This means that they have made progress over time. The cohort of grade 4 in year 2003-2004 on the other hand started in their grade 3 (2001-2002) above average, but score in grade 4

below average (Medio 4). This is a signal for this school to analyze these results closely and investigate what might have happened in the meantime that could explain such a score.

Figure 10: Trend analysis of cohorts for Spelling

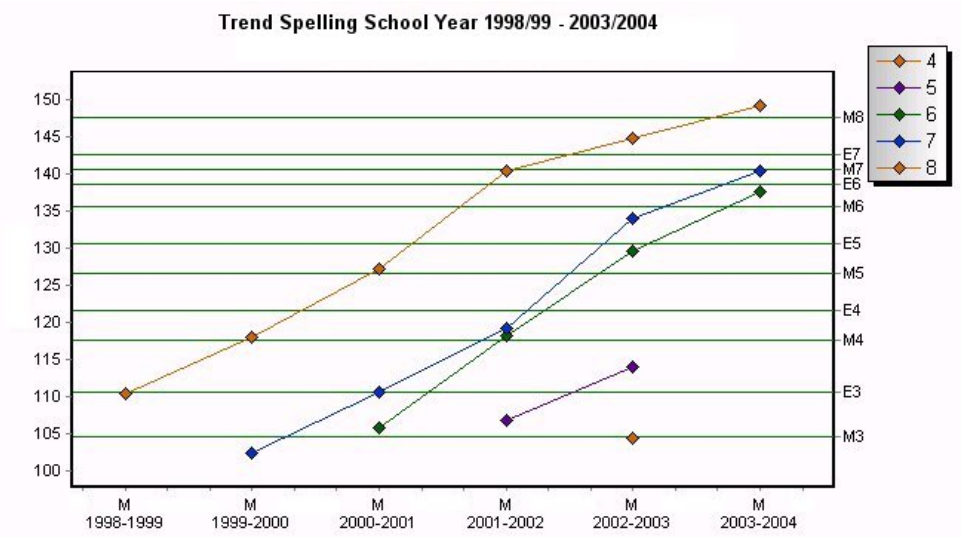
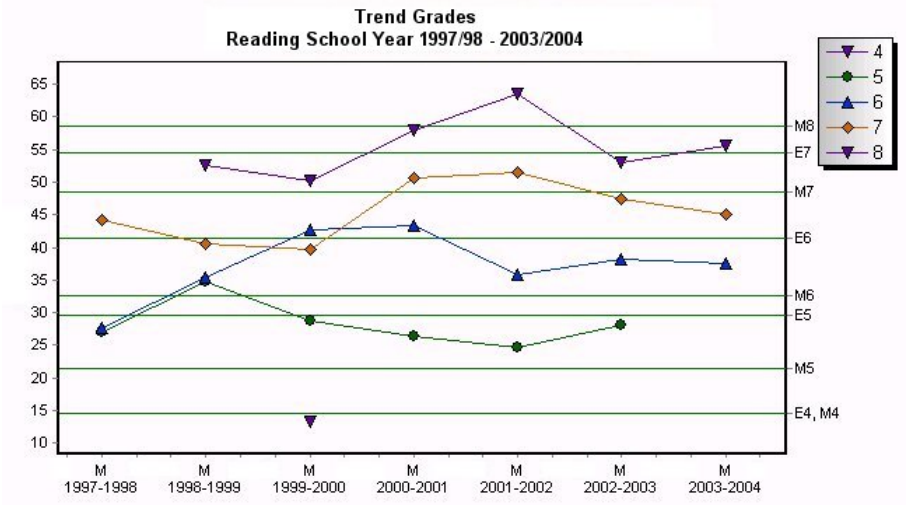


Figure 11 shows the results of different learner groups in a certain grade. The results over grades vary from year to year. It is only in grade 5 that the results are above average (at the medio-moments) throughout the years.

Figure 11: Trend analysis Grades for Reading

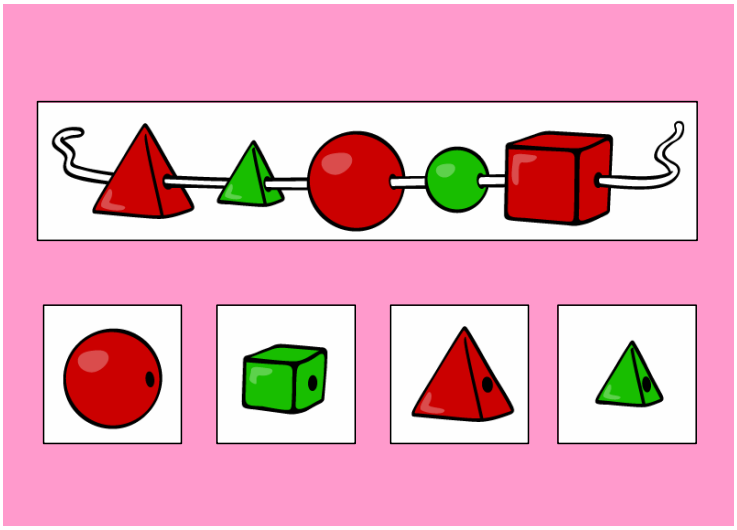


Computer-based and computer-adaptive tests

Since 2003 the PMS also contains computer-based tests. A great advantage of these tests is that they reduce the amount of test-administering time for a teacher. The teacher no longer needs to give extensive instructions to his or her pupils, does not have to mark the tests and does not have to fill in reports. Instead the computer processes the test results and reports immediately after completing the test.

Some of the computer-based tests (all the test for grades 1 and 2) are adaptive. This means that the computer selects items based on the answer given (and ability estimated) on the previous question. Apart from the advantages mentioned earlier, this also means that better information is gained in less testing time and that pupils are no longer bored or frustrated. The overall experience in schools tells us that children find it more fun to take tests on the computer.

Figure 12: Example of an item from the computer-adaptive test 'sequencing'



Unfortunately at this moment computer-based testing in the Netherlands is not yet common practice. The computer infrastructure in primary schools is still not of a satisfactory standard. Most schools do not have enough computers to organize computer-based testing for all of their pupils. With only 2 to 3 computers in every class-room computer-based testing demands a lot of organizing from the teacher. That is the reason why it is sometimes easier for schools to use paper-based tests for whole class testing. Right now the only times teachers test their pupils on the computer is, when the pupils have been absent from whole class test taking or when the pupils are working individually at their own level and at their own pace. In those situations computer-based tests are used more frequently.