CURRICULUM EVALUATION AND MULTILEVEL ANALYSIS: EFFECTS OF COOPERATIVE LEARNING IN MATHEMATICS

Jan Terwel* and Pieter van den Eedent†

*Graduate School of Teaching and Learning, University of Amsterdam, 1016 BV Amsterdam, The Netherlands
†Department of Methodology, Free University of Amsterdam, 2075 AP Amsterdam, The Netherlands

Introduction

In (secondary) education the learning process takes place in the context of a nested structure with several levels, namely, pupils, classes, teachers and schools. The basic idea behind this article is that schools, classes and teachers can give a specific character to pupils' learning environments. In other words, the learning results of a given pupil may be influenced by school, teacher and class.

An important question concerns the differential effects on low, medium and high aptitude pupils. We use the term "aptitude" for the prior achievement of a student as measured by an achievement test in mathematics. The term "aptitude group" (low, medium, high) concerns the achievement level of the pupils, prior to the implementation of the curriculum. The central question in this article is: do pupils of different aptitude differ in sensitivity to their learning environment (school, teacher and class)?

Our general hypothesis is that the learning process of the individual student does not take place in a vacuum but in an environment in which variables of different levels are at work.

In this article we will test a series of hypotheses for each aptitude group (low, medium and high). The results of our investigation show that differential effects do, in fact, exist. The learning processes of pupils in the highest aptitude group in the main proceed independently of the learning environment, whereas pupils in the middle aptitude group show a relatively high sensitivity to their learning environments.

One of the variables concerns the extent to which the curriculum has been implemented. In many curriculum evaluation studies the relation between curriculum implementation and student learning is determined by correlations between the degree of implementation and the mean class gains in learning scores (Fullan, 1983). Another way of relating variables of the class level to variables of the individual level is to assign class variables to each individual student and subsequently to apply a fixed regression analysis in which the dependent variable is regressed on the assigned variable. There are serious problems with both procedures (i.e., correlation and fixed regression analysis), which we hope to avoid in this article by means of a multi-level analysis model.

The Curriculum

Our evaluation is directed to a mathematics curriculum in the second grade of secondary education. The curriculum is based upon ideas of Freudenthal and Van Hiele (Freudenthal, 1973a and 1973b; Van Hiele, 1986), the so-called "Freudenthal model". On the basis
of this model a curriculum was developed by the Dutch National Institute for Curriculum Development (SLO).

The central idea behind the model and the curriculum is 'Mathematics for All'. In the context of this motto the curriculum developers tried to reconcile mathematics for the majority with more advanced mathematics for the minority. The discussion about 'Mathematics for All' should be seen in the framework of the more general, international debate about 'individual differences and the common curriculum', which has been going on for some time now (Fenstermacher & Goodlad, 1983). In The Netherlands plans are made and experiments are running in which a common curriculum for all 12 to 15 year old students plays an important role.

The implementation of the SLO-curriculum involves a teaching-learning situation with the following main characteristics:

1) Whole-class instruction by the teacher;
2) Guidance of small groups and individual pupils by the teacher;
3) Cooperative learning in small groups;
4) Mathematics set in real-life situations;
5) Different levels in the learning process.

The lesson (or series of lessons) generally consists of the following three successive parts.

a) Introduction

The teacher introduces the problem to the whole class. He motivates the pupils by locating the mathematical problem in their personal world (real-life situation). He presents the general outline of the problem, explores the various aspects of it and may suggest strategies for the problem-solving process.

b) Cooperative learning

The pupils work in small groups. The teacher observes and guides the problem-solving process and encourages each individual in the group. He tries to solve problems that originate in differences of speed or level between the pupils. When required, he helps pupils with individual learning problems.

c) Reflection and evaluation

Following the group work, the results and the actual work-processes in the small groups are discussed with the class. Special attention is paid to the different strategies and solutions to the problem in hand (i.e. to learning process levels).

On the basis of Van Hiele's level theory, Freudenthal proposes a new approach to individual differences. He points out that the levels in the learning process which occur in classroom practice can be used for the benefit of all. Differences between pupils are seen not as a hindrance but as an advantage. If pupils could reflect on the differences (e.g., differences in cognitive levels of solutions, in problem solving strategies) they would all be able to reach a higher level of understanding. According to Freudenthal, mathematics exercised on a lower level becomes mathematics observed on a higher level. Under these conditions pupils would, according to the theory, apply rules subconsciously until they became aware of them. The results of such reflections would be the formulation of ideas and rules in more abstract terms. Cooperative learning in heterogeneous classes and in small heterogeneous groups is seen as instrumental to this process of reflection.
Against the background of the 'Freudenthal model' and 'Mathematics for All' the research question about (differential) effects becomes relevant. Does this new curriculum have observable effects and what, if any, are the benefits for low and high aptitude pupils?

Data, Design and Research Questions

The multilevel analysis was carried out with the database of an earlier research project 'Mixed Ability Teaching in Mathematics in Secondary Education' (Terwel, Herfs, Dekker & Akkermans 1988, Terwel, 1989). The aim of that project was to assess the effects of the new curriculum. The data were collected at three secondary schools, with a total of 482 pupils, 12 teachers and 22 classes. The curriculum was implemented in the second year of secondary education with a majority of pupils in the ages of 13 and 14 years. The research design was pretest-posttest. Differences in implementation were found, in particular, in the extent to which the Freudenthal model had been implemented with regard to schools, teachers and pupils.

In this article we investigate the effects of the curriculum by relating these differences in implementation to differences between students' learning results. Given the basic idea behind 'Mathematics for All', the research question about the differential effects of this curriculum is important. Which pupils benefit most from the implementation of the new curriculum and how can we explain the differences in learning results?

Variables, Hypotheses and Theoretical Framework

Variables at the individual (student) level were measured by a pretest (TOT1) and a posttest (TOT3). The content of both tests concerns mathematical problems about relations and functions as applied to real-life situations or 'contexts'. The effect of TOT1 upon TOT3 is defined as 'learning result' and was measured by the intercepts and the slopes of the regressions concerned.

The basic idea behind the following hypotheses is that 'learning results' will be influenced by variables at three levels, namely, the school, the teacher and the class.

The school hypothesis implies that differences in intercepts and in slopes between schools exist and that these differences can be explained post hoc by the characteristics of the schools.

The teacher hypothesis implies that differences in intercepts and in slopes of the regressions of TOT3 on TOT1 between teachers exist and that these differences can be explained by the characteristics of the teacher's instructional behavior as measured by LPERCIA.

The variable LPERCIA scores indicate the extent to which the curriculum has been implemented by the teacher. They are obtained by calculating the mean of pupils' assessment scores. Pupils' assessments are registered by means of a questionnaire which measures assessment of the learning environment along a scale called the PERCIA-scale (PERception of the Curriculum In Action).

The research hypothesis is that degrees of curriculum implementation by the teacher correlate positively with learning results. The questionnaire is a curriculum-specific version of the learning-environment measurement scale (Fraser, 1981; Fraser & Tobin, 1989). The PERCIA questionnaire contains questions about the teacher's instruction and guidance, cooperation between pupils, task orientation and mathematics set in real-life situations. All these questions are derived from the characteristics of the 'Freudenthal model' as well as from theories and experimental research into effective instruction (Freudenthal, 1973, 1980; Brophy 1979, 1986; Terwel 1984, 1989).

The class hypothesis implies that differences in intercepts and in slopes of the regressions of TOT3 on TOT1 between classes exist and that these differences can be explained in terms of the following four variables: NTAKOB, DEVV, MPERCIA and MTOT1. For each variable a description, an operationalisation and a hypothesis is given below.
The NTAKOB variable is the percentage of time spent off-task by the pupils during classes. This variable was measured by systematic, quantitative classroom observation. The hypothesis is that the less time spent off-task during class the better the learning result will be (Brophy, 1986).

The DEVV variable concerns the increase in differences between pupils in the class. This class-level variable is determined by the difference between the standard deviations for TOT3 and TOT1. DEVV is the standard deviation increase from pretest to posttest. This variable indicates an instructional style which recognizes individual differences in learning achievement.

The hypothesis here is that in classes with a strong emphasis on creating different teaching styles for different pupils the learning results will be better than in classes with less emphasis on differences (see Baumert, Roeder, Sang & Schmitz, 1986; Roeder, 1989).

If all pupils in a class receive adequate instruction according to their abilities, then they will develop their individual talents optimally, which will result in increased differences between the pupils. There is evidence from the literature that the teaching needs of low and high aptitude pupils are different. The extent to which teachers adapt their teaching in heterogeneous classes to these different needs has an effect on pupils' learning results. Low-aptitude pupils seem to benefit most from a systematic, structured instructional approach, whereas high-aptitude pupils are stimulated by a more open style of teaching (De Klerk, Schouten & Van der Sanden, 1988).

This hypothesis can also be seen in the context of the international debate about homogeneous or heterogeneous grouping in education. In a recent review of the literature De Koning (1987) concludes that the question of leveling of achievement in heterogeneous classes plays an important role in this debate. There is also some evidence in the literature that teachers in secondary schools do not always take the considerable differences between their pupils into account. Many teachers are inclined to adapt their teaching to the bulk of the pupils (the steering group) at, or even below, the middle of the ability range: to the detriment of all pupils at the extremes of the curve, but especially of the high achievers (Kerry, 1982; Nijhof, 1978; Bonset & ten Brinke, 1988; Roeder, 1989). We expect that teachers with a differentiated instructional approach have better results than those with a less differentiated approach. This means that differentiation in teaching will result in differences between pupils and that these, in turn, will be reflected in differences between posttest and pretest standard deviations.

In traditional whole-class instruction it seems difficult to attain two possibly conflicting objectives, namely high achievement for each individual learner according to his or her potentialities and at the same time equalizing achievements within the class. Teachers who are especially concerned about "losing" their low aptitude pupils seem to adopt an equalizing approach, whereas teachers who are primarily concerned about creating optimal conditions for high aptitude pupils may adopt a style which may result in maximizing differences between students. There seems to be a trade-off between excellence and equality. High achievers show good results in classes with increased differences between pupils (in which the teacher maximizes differences between pupils from pretest to posttest), whereas low aptitude pupils seem to profit more in classes with a lower emphasis on maximizing differences in achievement (Roeder, 1989).

In theory at least, we expect that this dilemma can be solved by using a curriculum and a form of differentiated teaching in which more than one 'steering group' functions as reference group in the teaching-learning process. This means that the teacher must be able to adapt his or her teaching (or learning environment) to differences between (groups of) pupils. In reality this is probably the case in classes with different aptitude groups or in classes with cooperative groups of four pupils working together (Slavin, 1987). We expect that with the right combination of grouping and instruction a solution of the dilemma can be found. We also expect that a relatively high mean of prior knowledge in a class is a prerequisite for relatively good learning results.

The MTOT1 variable concerns the level of prior knowledge in a class at the start of the implementation of the new curriculum. This variable was determined by the class mean of the pretest (TOT1). The corresponding hypothesis is that learning results correlate positively with prior knowledge.
The background to this hypothesis is contained in the literature about the effects of class composition on learning results for pupils of different aptitude groups (Beckerman & Good 1981; Good & Marshal, 1984; Oakes 1986; Dar & Resh, 1986; Dreeben & Barr, 1987). There is some evidence in the literature that the number of low aptitude pupils has a negative effect on learning results in a class. In general, the mean aptitude score of a class is a determinant of the quality of the teaching-learning processes in that class, which will be reflected in the learning results.

The MPERCIA variable is the mean score of a class on the PERCIA questionnaire. This score indicates the extent to which the curriculum has been implemented in the class (see also the earlier description of LPERCIA).

The relevant hypothesis is that degrees of curriculum implementation correspond with pupils' learning results. In operational terms this implies that MPERCIA scores correlate positively with learning results.

A detailed description of the tests and questionnaires used, as well as their reliability and validity, can be found in the research report (Terwel, Herfs, Dekker & Akkermans, 1988). All research instruments were valid and reliable. Only the reliability of the observation scale (NTAKOB) was not determined.

The Model

The point of departure for our analysis is the regression of the posttest (TOT3) on the pretest (TOT1). A number of hypotheses concerning the effects of the teacher variables and class-room variables on this regression has been formulated in the foregoing section. These cross-level hypotheses are subsequently tested within the random coefficient (RC) model of multilevel analysis. The RC-model consists of two steps. The first step concerns a within-group regression. In the second step the results of the first step are incorporated in a between-group regression (De Leeuw & Kreft, 1986; Aitkin & Longford, 1986; Raudenbush & Bryk, 1986; Goldstein, 1987).

The two steps of the model can be formulated as follows, where classroom equals group level. The within-group regression of the first step is expressed in the following equation:

\[
\text{TOT3}_{ij} = a_j + b_j \text{TOT1}_{ij} + z_{ij}
\]

where

- \( i \): individual pupil (\( i = 1 \ldots I \))
- \( j \): class-room (\( j = 1 \ldots J \))
- \( b_j \): slope of regression of TOT3 on TOT1 in class-room \( j \)
- \( a_j \): intercept of class-room \( j \)
- \( z_{ij} \): error term, with variance \( s^2 \).
- \( \text{TOT3}_{ij} \): mathematics score at the end of the school year.
- \( \text{TOT1}_{ij} \): mathematics score at the beginning of the school year.

In the second step there is a between-group regression of intercept \( a_j \) and slope \( b_j \) on class-room variables \( U_{jm} \) (\( m: 1 \ldots M \)).

\[
a_j = A_0 + A_1 U_{j1} + A_2 U_{j2} + \ldots + A_M U_{jm} + d_j
\]

\[
b_j = B_0 + B_1 U_{j1} + B_2 U_{j2} + \ldots + B_M U_{jm} + e_j
\]

where \( d_j \) and \( e_j \) are error terms belonging to, respectively, \( a_j \) and \( b_j \) with variances \( t^2 \) and \( u^2 \).

\( A_0 \) refers to a general constant, \( A_m \) to the direct effect of a class-room variable \( U_m \), \( B_0 \) with respect to the direct individual effect, and \( B_m \) to the effect of a class-room variable on the influence of \( \text{TOT1}_{ij} \) on \( \text{TOT3}_{ij} \).

The RC-model (De Leeuw & Kreft, 1986) is also known as 'variance decomposition-model' (Winer, 1971; Aitkin & Longford, 1986) and 'hierarchical linear model' (Raudenbush & Bryk, 1986). It resembles the 'separate equation approach' of Boyd and Iversen (1979) and the 'two step
procedure' model (Van den Eeden & Saris, 1984), insofar as the $A_0$, $A_m$, $B_0$, and $B_m$ coefficients can be estimated.

By means of this model it is possible to read the teacher-specific and class-room specific educational chances from the differences in the intercepts and the slopes of a regression in which is expressed the influence of the pupils' TOT1 on their TOT3. Research in a similar vein has been carried out by Aitkin and Longford (1986), Raudenbush and Bryk (1986) and Cuttance and Willms (1986), and in the Netherlands, by, among others, Bosker and Van der Velden (1987), Brandsma and Knuver (1989), Kreft (1988), Koehler and Van den Eeden (1986), and Roeleveld, Koopman, De Jong and Van den Eeden (1988).

Longford's (1988) VARCL-program was used in the analysis. Starting from a full model the variables which offer non-significant coefficients were fixed on 0 in several subsequent 'runs', taking into account the fit between the model and the data.

**Analysis**

On the basis of the distribution of TOT1, in the former study by Terwel, Herfs, Dekker and Akkermans (1988), three aptitude groups Low (1), Medium (2) and High (3) were distinguished. The main object of the analysis is a comparison of these different aptitude groups. In each group the effects were studied of, inter alia, given variables at the general class-room level on pupils' learning results. We may assume that pupils of a given aptitude group receive their instructions in different classrooms and schools, which differ with regard to the degree in which the Freudenthal method is implemented. They also differ in respect of conditions such as average initial knowledge (average TOT1). We can now investigate differences between aptitude groups in sensitiveness to differences in implementation and conditions.

Table 1 contains some characteristics of the distribution of TOT1 and TOT3, as well as the coefficients of the regression of TOT3 on TOT1 per aptitude group. Table 1 shows an increase in the average of TOT1 compared to the average of TOT3. The increase corresponds with the number of good pupils in the aptitude group. It is obtained differently for each aptitude group since the regression coefficients of TOT3 on TOT1 differ between aptitude groups. The regressions of aptitude group 1 and aptitude group 3 correspond most closely. However, aptitude group 2 shows a small percentage of explained variance ($R^2 = .083$), and a low or even negative slope coefficient. Consequently, it is to be expected that the learning process of aptitude group 2 pupils depends on variables in the learning environment other than TOT1 alone.

In our analysis the following five levels are used: pupils, classes, teachers, schools and aptitude groups. Starting with the deepest level the nesting hierarchy is pupils, classes, teachers, schools and aptitude- groups. To a certain extent this hierarchy is arbitrary, but three considerations were decisive in its construction, namely, that aptitude groups hardly play any role at all in the enrollment of pupils by schools, that each school possesses its own team of teachers, and that each teacher usually teaches several classes.

The aptitude group is seen as the highest level. The reasons for this lie in the research problem itself, in its associated theoretical framework, and in the way TOT1 is measured. The research problem is aimed at differences in the effects of TOT1 on TOT3 per aptitude group. TOT1 is assessed in absolute terms, that is, without regard to the division of pupils into classes. The fact that pupils belong to different aptitude groups is already given at the division of pupils over the various classes. At the theoretical level therefore the accent is on the various effects of instruction, without classroom composition being taken into account. If classroom composition were taken into account (as is the case in comparative reference group theory) then aptitude group assessment would be a relative measure, with class average as a yardstick. Consequently, rather than 'aptitude group', the level between classroom and pupil would constitute the top of the hierarchy.
Table 1: Characteristics of the Distributions of TOT1 and TOT3 and the Regression Coefficients of TOT3 and TOT1 by Aptitude Group (Standard error between brackets)

<table>
<thead>
<tr>
<th>Aptitude Group</th>
<th>TOT3</th>
<th>TOT1</th>
<th>Gain</th>
<th>Regression Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>29.54</td>
<td>20.62</td>
<td>8.82</td>
</tr>
<tr>
<td></td>
<td>St. Dev</td>
<td>8.98</td>
<td>4.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>2</td>
<td>4.5</td>
<td>a = 8.74 (2.34)</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>54</td>
<td>26.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cases</td>
<td>197</td>
<td></td>
<td>b = 1.01 (.11)</td>
</tr>
<tr>
<td></td>
<td>Classes</td>
<td>22</td>
<td></td>
<td>R² = .30</td>
</tr>
<tr>
<td></td>
<td>Teachers</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schools</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>Mean</td>
<td>41.62</td>
<td>30.20</td>
<td>11.424</td>
</tr>
<tr>
<td></td>
<td>St. Dev</td>
<td>10.87</td>
<td>2.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>9</td>
<td>27</td>
<td>a = 1.39 (11.19)</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>75</td>
<td>34.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cases</td>
<td>158</td>
<td></td>
<td>b = -.34 (.37)</td>
</tr>
<tr>
<td></td>
<td>Classes</td>
<td>20</td>
<td></td>
<td>R² = .08</td>
</tr>
<tr>
<td></td>
<td>Teachers</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schools</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>Mean</td>
<td>54.93</td>
<td>40.91</td>
<td>14.03</td>
</tr>
<tr>
<td></td>
<td>St. Dev</td>
<td>11.40</td>
<td>5.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>32</td>
<td>35</td>
<td>a = 9.68 (5.88)</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>86</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cases</td>
<td>127</td>
<td></td>
<td>b = 1.11 (.14)</td>
</tr>
<tr>
<td></td>
<td>Classes</td>
<td>18</td>
<td></td>
<td>R² = .33</td>
</tr>
<tr>
<td></td>
<td>Teachers</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schools</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Before we deal with the differences in coefficients we shall look at the extent to which the variables TOT1 and TOT3 for each aptitude group can be explained by the division into classrooms, teachers or schools. This can be read from the eta² coefficient, which expresses what percentage of the total variance constitutes the between-group variance. The values of eta² are shown in Table 2.

Table 2 shows that for all aptitude groups there is, corresponding with height of aptitude group, a decrease in the percentage of TOT1 and TOT3 variance explained by between-class, between-teacher, and between-class differences. This finding supports our division into aptitude groups. Table 2 also confirms that aptitude groups differ from each other in that the differences between classrooms and teachers are considerably greater than for the other aptitude groups. This corresponds with our conjecture that middle-aptitude groups depend more on their learning environment than other aptitude groups. At the two lowest aptitude groups there is an increase in the differences in TOT1 and TOT3 variance explained by between-classroom, between-teacher, and between-school differences.
Table 2: Eta²-values by Aptitude Group

<table>
<thead>
<tr>
<th></th>
<th>aptitude group 1</th>
<th>aptitude group 2</th>
<th>aptitude group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TOT1</td>
<td>TOT3</td>
<td>diff.</td>
</tr>
<tr>
<td>class</td>
<td>.171</td>
<td>.299</td>
<td>.128</td>
</tr>
<tr>
<td>teacher</td>
<td>.147</td>
<td>.213</td>
<td>.066</td>
</tr>
<tr>
<td>school</td>
<td>.027</td>
<td>.078</td>
<td>.051</td>
</tr>
</tbody>
</table>

Table 3: Distribution Characteristics of the Group Variables

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>standard deviation</th>
<th>minimum</th>
<th>maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>teacher LPERCIA</td>
<td>3.332</td>
<td>1.740</td>
<td>2.881</td>
<td>3.559</td>
</tr>
<tr>
<td>class DEVV</td>
<td>3.330</td>
<td>1.855</td>
<td>.408</td>
<td>7.545</td>
</tr>
<tr>
<td></td>
<td>27.804</td>
<td>5.721</td>
<td>18.231</td>
<td>39.032</td>
</tr>
<tr>
<td></td>
<td>3.305</td>
<td>.206</td>
<td>2.881</td>
<td>3.605</td>
</tr>
<tr>
<td></td>
<td>25.597</td>
<td>24.027</td>
<td>.370</td>
<td>74.620</td>
</tr>
</tbody>
</table>

By contrast, there is a small decrease for aptitude group 3 at all levels. On the basis of these results it seems worthwhile to analyze each aptitude group separately and to adopt different levels for the explanation of the TOT3 differences. This is done by performing a multilevel analysis. The results mentioned consist of unstandardized coefficients whose interpretation depends on the coefficients of average, minimum and maximum distribution characteristics. The distribution characteristics of TOT1 and TOT3 have already been shown in Table 1. The group variables are presented below, as used in the explanation of the intercepts (see Table 3). If differences between the slopes are to be explained by group variables, their values can be obtained by multiplication of the individual variables and the group variables concerned.

The School Hypothesis
The school hypothesis holds that there are differences in the intercepts and slopes of the regression of TOT3 on TOT1 between the three schools. The basic equations are the following ones, where the subscript s refers to school (s: 1...S).

\[ TOT3_{is} = a_s + b_s \cdot TOT1_{is} + z_{is} \]
\[ a_s = A_0 + d_s \]
\[ b_s = B_0 + e_s \]

As can be seen, school variables do not feature in these equations.
The result of an analysis based on the above equations is shown in Table 4. (As is the case in subsequent tables, some coefficients equal to zero have been omitted.)

Table 4 shows that the differences in slopes between the various aptitude groups are minimal. Nor do the slopes differ between schools. Again, the pupils in the second aptitude group demand our attention. On the basis of TOT1 a worse prediction is obtained for them than for the other pupils ($R^2$ is .214 in comparison with resp..300 and .326). Moreover, for the middle aptitude group the intercept variance is considerable (18.8 per cent of the total), whereas for the other groups it is totally absent. This points to the importance of school composition for this particular aptitude group.

These differences could be explained post hoc by the school characteristics. The schools do differ in the average level of pupils. In relative terms, school 1 has many low achieving pupils in comparison to both other schools. A second explanation could be that many pupils from ethnic minorities are enrolled in this school. It appears that this school achieves little learning gain in comparison to other schools. The average difference-score between posttest and pretest is relatively small, especially for the middle aptitude group. Probably, enrollment in one school or another is of no importance for the pupils of this aptitude group. A third type of explanation does not concern inflow, but is didactic in nature. The teachers of school 1 have generally had less experience with the new curriculum. This is reflected in the average PERCIA-scores per school: school 1 has the lowest score. This means that this school was the least successful in putting the curriculum into practice.

The "distance" between schools 2 and 3 is slight. In many respects, however, school 2 is the better one, if we take into consideration the average beginning performance (MTOT1), teachers' experience, the degree of implementation of the curriculum (MPERCIA), and the average 'learning gain' of the pupils (MTOT3 minus MTOT1).

In order to explain the large intercept variance of the medium aptitude group we assessed the regressions of TOT3 on TOT1 per school for each aptitude group. It appeared that there is a deviant pattern especially in the medium aptitude group. The intercept differences between the schools are noticeably large for this aptitude group, whereas the explained variances ($R^2$) are relatively small for this aptitude group, especially at school 2. In contrast to the other schools the latter school shows a high intercept value. This school has a slight posttest prediction (TOT3) on the basis of the pretest (TOT1) and so offers a greater chance on a higher posttest score, apart from the pupils' pretest score. At least for the medium aptitude group, school 2 combines, in Cuttance's
terms, high "quality" (high intercept) with a high "compensating capacity" (small slope). The worse prediction of this aptitude group on the basis of the pretest could be interpreted as the fact that other variables than the pretest play an important role in the explanation of the learning results of the pupils concerned, e.g., instruction variables of the pupil-body composition.

The Teacher Hypothesis

The teacher-hypothesis holds that there are differences in the intercepts and in the slopes of the regression of TOT3 on TOT1, and that these differences can be explained by differences in implementation of the curriculum by the teachers (LPERCIA). The equation has the following form, where subscript 1 refers to the teacher (1: 1...L).

\[ TOT3_{1i} = a_1 + b_1 TOT1_{1i} + z_{1i} \]
\[ a_1 = A_0 + A_1 \text{LPERCIA}_1 + A_2 \text{LSD}_1 + d_1 \]
\[ b_1 = B_0 + B_1 \text{LPERCIA}_1 + B_2 \text{LSD}_1 + e_1 \]

Table 5: Effects of the Teacher by Aptitude Group (Standard errors between brackets)

<table>
<thead>
<tr>
<th></th>
<th>aptitude group 1</th>
<th>aptitude group 2</th>
<th>aptitude group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed part</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intercept</td>
<td>( A_1 )</td>
<td>10.839</td>
<td>2.869</td>
</tr>
<tr>
<td>TOT1</td>
<td>( B_1 )</td>
<td>0.918 (.110)</td>
<td>0.000 (.000)</td>
</tr>
<tr>
<td>explaining slope differences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>by LPERCIA</td>
<td>( B_1 )</td>
<td>0.000 (.000)</td>
<td>0.373 (.091)</td>
</tr>
<tr>
<td>random part</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pupil variance</td>
<td>( s^2 )</td>
<td>51.813</td>
<td>91.689</td>
</tr>
<tr>
<td>( s )</td>
<td>7.200</td>
<td>9.642</td>
<td>9.321</td>
</tr>
<tr>
<td>intercept variance</td>
<td>( t^2 )</td>
<td>0.000</td>
<td>9.696</td>
</tr>
<tr>
<td>( t )</td>
<td>0.000 (.000)</td>
<td>3.114 (1.217)</td>
<td>0.000 (.000)</td>
</tr>
</tbody>
</table>

Table 5 confirms the exceptional position of aptitude group 2. There, in contrast to the other aptitude groups, the effect of achievement in a former test (TOT1) on TOT3 is absent. Moreover, the LPERCIA is correlated with the slope, which indicates a pupil's learning process with a given teacher.

Table 5 also reveals that there is a considerable variance between the intercepts per classroom for all aptitude groups. This intercept refers to differences in composition with respect to the average achievement level of the pupils per teacher.
The Classroom Hypothesis

The classroom hypothesis holds that there are differences between the intercepts and slopes, and that these differences can be explained by the following four variables.

1) Time off-task NTAKOB;
2) Classroom composition MTOT1;
3) Delevelling instruction DEVV;
4) Implementation MPERCIA.

Table 6: Effects of the Class by Aptitude Group (Standard errors between brackets)

<table>
<thead>
<tr>
<th>fixed part</th>
<th>aptitude group 1</th>
<th>aptitude group 2</th>
<th>aptitude group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>A₀ 11.310</td>
<td>-6.437</td>
<td>10.581</td>
</tr>
<tr>
<td>TOT1</td>
<td>B₀ .000 (.000)</td>
<td>.000 (.000)</td>
<td>1.117 (.139)</td>
</tr>
</tbody>
</table>

explaining intercept differences

| by MTOT    | A₂ .031 (.003)  | .552 (.165)      | .000 (.000)      |
| NTAKOB     | A₁ .003 (.001)  | .000 (.000)      | .000 (.000)      |

explaining slope differences

| by DEVV    | B₁ .031 (.016)  | .058 (.014)      | .000 (.000)      |
| MTOT       | B₂ .022 (.008)  | .000 (.000)      | .000 (.000)      |
| MPERCIA    | B₃ -1.044 (.494)| .249 (.096)      | .000 (.000)      |
| NTAKOB     | B₄ .000 (.000)  | .000 (.000)      | -.003 (.001)     |

random part

<table>
<thead>
<tr>
<th>pupil variance</th>
<th>s² 48.039</th>
<th>75.973</th>
<th>83.840</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>6.931</td>
<td>8.716</td>
<td>9.156</td>
</tr>
</tbody>
</table>

slope variance

| v²     | .000 (.)  | 1.050  | .000 (.) |
| v      | .000 (.000) | 1.025 (.486) | .000 (.000) |
The relevant equations comprise the classroom variables NTAKOB, MTOT1, DEVV, MPERCIA in order to explain the intercept differences and slope differences. The equations are as follows. The subscript k refers to classroom k (k: 1...K).

\[
\begin{align*}
\text{TOT3}_{ik} &= a_k + b_k \text{TOT1}_{ik} + z_k \\
\text{TOT3}_{ik} &= A_0 + A_1 \text{DEVV}_{ik} + A_2 \text{MTOT}_{ik} + A_3 \text{MPERCIA}_{ik} + \ldots \\
\text{TOT3}_{ik} &= B_0 + B_1 \text{DEVV}_{ik} + B_2 \text{MTOT}_{ik} + B_3 \text{MPERCIA}_{ik} + \ldots
\end{align*}
\]

The results of Table 6 may be summed up as follows. The individual effect of TOT1 on TOT3 is absent for aptitude group 1 and 2, but present for aptitude group 3.

The hypotheses on the effects of MTOT and NTAKOB concern differences in both intercepts and slopes and it turns out that these coefficients differ for different aptitude groups. Only aptitude group 3 appears to be relatively immune. Here, only NTAKOB has a small influence, being negative.

Let us take a closer look at the differences between the slopes. The effect of DEVV is considerably greater for aptitude group 2 than aptitude group 1. MPERCIA has a negative effect on aptitude group 1, but a positive effect on aptitude group 2. Moreover, a classroom composition effect can be detected for aptitude group 1.

Aptitude group 2 show a unique pattern. It appears that advances in achievement for this group depend exclusively on teacher instruction and on classroom composition in comparison with achievement at the beginning of the school year. This constitutes a difference from aptitude groups 1 and 3. It should be noted that aptitude group 2 is unique in showing an as yet unexplained variance in differences between slopes per classroom. A full explanation could only be given via an analysis of the other aptitude groups. For the medium aptitude group other, as yet unknown, factors play a role in the between-class differences of the slopes.

Conclusions

As the above analysis has shown, the effectiveness of the 'Freudenthal model' differs for the three different aptitude groups and the learning results within each of these levels are dependent on different variables.

The learning results for the medium aptitude group are remarkable. The pupils of this group seem to profit most from innovation. Their learning results are totally determined by the teaching-learning situation. The extent to which the new curriculum has been implemented affects learning results, with class composition (mean class level) and increase in standard deviation (from pretest to posttest) playing a facilitating role.

The learning results of the medium aptitude group are relatively independent of prior knowledge. Within this group the hypothesis has been confirmed: the greater the degree of curriculum implementation, the better the results.

Although this hypothesis does not hold for the lows, there is a significant effect of class composition (mean class level). This is in accordance with other research findings in which the low aptitude groups proved to be more sensitive to class composition than the higher levels (Dar & Resh, 1986; Beckerman & Good, 1981; Good & Marshall, 1984). It was concluded in another study that the gains of this aptitude group are relatively low (Terwel, 1988b). Judging by our multilevel analysis this conclusion can be confirmed. The 'Freudenthal model', with an emphasis on comprehension of mathematical problems in stories of real life situations, problem solving and discovery learning, probably assumes a repertoire of concepts and strategies that these pupils do not possess. In addition, the model does not provide remedial help for pupils who fall behind. The
results of our multilevel analysis can be explained by introducing a kind of 'threshold hypothesis': in order to profit from this type of innovation students need a certain level of prior knowledge. Without a minimum of prerequisite skills the gains will be low. Although we were not able to test the validity of this hypothesis, our interpretation is in accordance with the findings of recent studies by Van Streun (1989, in press).

The pupils in the high aptitude group are relatively less affected by their learning environment. These pupils do not respond to differences in the extent to which the curriculum has been implemented. High-level pupils seem to rely more on their own abilities than medium- and low-level pupils. Contrary to our expectations this group proved not to be affected by the increase in the standard deviation. The only effect we find for this group is related to time off-task in the class.

Discussion

In the analysis of this article five levels were available, which were hierarchically nested. It was technically impossible to do an analysis on all those levels at the same time. The numbers of cases at each level were too low and adequate analysis modes were not yet available where the outcomes of the school level could be controlled at the lower levels concerned (the class level and the teacher level). The same holds for the teacher level. However, even in the case of the analysis presented, where three levels were incorporated (the pupil, his/her aptitude and an intermediate level) we were not able to analyse them at the same time but only in a comparative way. An analysis directed to the assessment of differential effects which handle more appropriately each combination of levels, had to rely on an explicit multilevel design, starting from an elaborated theory about the differential effects of the learning environment on the learning process of pupils with low-, medium and high aptitudes.

The application of multilevel analysis in curriculum evaluation enables researchers to take a closer look at the differential effects of the curriculum on learning results. If it is possible to measure the extent to which the curriculum has been implemented, then it is useful to determine the effects of these variations in implementation on learning results by means of a multilevel analysis.

References


Wubbels, Th., et al. (1987). Perceptie van de leraar-leerlingrelatie; constructie en kenmerken van een instrument. Tijdschrift voor Onderwijs research, 12, 3-16.
The Authors

JAN TERWEL is Associate Professor of Education and Instructional Psychology at the University of Amsterdam, Graduate School of Teaching and Learning.

PIETER VAN DEN EEDEN is Associate Professor of Sociology, Methodology and Statistics at Free University Amsterdam, Department of Methodology.