„Accelerated Mathematics“ in Grades 4 through 6

-

Evaluation of an experimental program
in 15 schools in Northrhine-Westphalia

Prof. Dr. Dr. Rainer H. Lehmann
Dr. Susan Seeber

edukatione
Agentur für Beratung im Bildungsbereich

Humboldt Universität zu Berlin

Contents

1 Pretest sample and test instruments

1.1 Achieved pretest sample

The sample design for the evaluation of „Accelerated Mathematics“ in NRW included 5 primary schools, Grade 4, with one experimental and one control group each (i.e., a total of 10 classrooms), as well as 10 secondary schools with one experimental and one control group from Grades 5 and 6 each (i.e., a total of 40 classrooms). Secondary schools were stratified according to school type: 3 Basic (*Hauptschulen*), 3 Intermediate (*Realschulen*), 2 Academic (*Gymnasien*), and 2 Comprehensive Schools (*Gesamtschulen*). In one school (the Intermediate Carl-Benz-Realschule, Düsseldorf), no Grade 5 class mathematics teacher could be convinced to participate. Instead, an additional Grade 6 class for the group control could be recruited here. Thus, the achieved secondary school sample is constituted by 19 experimental and 20 control classes (for response rates, cf. Annex 1). The following Table 1 summarizes the achieved sample:

Table 1: **Achieved pretest sample: number of classrooms by grade, school type, and participation status**

School type	**Treatment group**		**Control group**	
	classes	students	classes	students
Primary schools (Grade 4)	5	118	5	129
Secondary schools (Grade 5)				
• *Hauptschule* (Basic)	3	60	3	62
• *Realschule* (Intermediate)	2	61	2	62
• *Gymnasium* (Academic)	2	62	2	62
• *Gesamtschule* (Comprehesive)	2	56	2	55
Secondary schools (Grade 6)				
• *Hauptschule* (Basic)	3	67	3	68
• *Realschule* (Intermediate)	3	80	4	111
• *Gymnasium* (Academic)	2	64	2	63
• *Gesamtschule* (Comprehesive)	2	57	2	56
Σ	**24**	**681**	**25**	**668**

1.2 Initial data collection, response rates, and data entry

With the exception of the Comprehensive Käthe-Kollwitz-Gesamtschule, Recklinghausen, initial data collection took place between February 9th and 20th, 2004. Due to sick leave on the part of the responsible school vice principal, the Recklinghausen school was obliged to delay testing until March 5th to March 12th. With this exception, all testing material was returned to the project coordinating center at Humboldt University, Berlin, until February 27th, 2004. All data were entered and cleaned between March 15th and April 15th.

Class-level response rates lay between 80 and 100 percent within the treatment group and between 69 and 100 percent within the control group, respectively. Lower response rates tend to have occurred primarily in the basic *Hauptschule* track, which is in accordance with expectations. As a rule, experienced test administrators (thoroughly trained in the PISA exercise) were, in cooperation with the schools, able to guarantee the required testing conditions. In most schools, the students worked on the tests at high levels of effort and motivation. The only exception reported pertains to the *Hauptschule* Erftstadt, where some students were reported to have quit their work on the test before the end of the testing time allowed, perhaps, as waqs speculated, in reaction to inadequate testing conditions with unacceptably high noise levels outside the building during the testing session. The effects of such irregularities were investigated by way of appropriate data analyses, searching for high missing rates or unexpectedly low levels of performance. When no conspicuous irregulaities were found, it was concluded that the data collected in Erftstadt could safely be used for subsequent investigations.

1.3 Test instruments and questionnaires

The instruments used in the initial data collection include curricularly valid mathematics tests and the German adaptation of CATTELL'S non-verbal „Culture Fair Intelligence Test" (CFT 20, short form). Also, a Student Questionnaire was used in order to obtain information on the students' socio-biographic backgrounds and their general attitudes towards mathematics. Using IRT technology, the mathematics test results have been successfully projected onto scales pertaining to comparable data sets from large scale assessments in Hamburg and Berlin, Germany.

Mathematics test

In the subsamples from Grades 4 and 5, the students' mathematical competencies were assessed using the *Hamburger Schulleistungstest für 4. und 5. Klassen – HST 4/5* (MIETZEL & WILLENBERG 2000); in Grade 6, the *Hamburger Schulleistungstest für 6. und 7. Klassen – HST 5/6* was used. Both tests have been shown to function satisfactorily in the context of the longitudinal school achievement census in the City of Hamburg „*Aspekte der Lernausgangslage von Hamburger Schüler und Schülerinnen*" (LAU; cf. Lehmann *et al*. 1997; Lehmann *et al*. 1999).

The HST 4/5 reflects a relatively wide array of mathematical competencies the mastery of which is expected by the end of Grade 4. It reflects the syllabus for the 4-Grade Primary School. In the domain labelled "*concepts of numbers*", the mastery of numerical and digital values, of the relationship between verbal and non-verbal representations of numbers, of basic arithmetic operations, and of histograms are assessed. In the domain „*Measurement*", knowledge of units of measurement is tested, and by definition, the domain „*Computation*" includes items which require the mastery of arithmetic / computational skills (vgl. Lehmann, Peek & Gänsfuß, 1997, 32).

(1) HST 4/5:

Below, two examples taken from the mathematics test for Grades 4/5 are given, one representing a relatively a low proficiency level and the other a relatively high one.

Example for a mathematics item from the HST 4/5, lower proficiency level:

You have the digits 3 7 8 9 4 at your disposal. Which is the smallest number you can form using these digits?

A 37 894

B 98743

C **34 789**

D 49 873

Example for a mathematics item from the HST 4/5, upper proficiency level:

A football stadium has 37 000 seats. 18 400 tickets were sold during early reservation. How many tickets can still be sold?

A 17 600

B **18 600**

C 17 400

D none of these

HST 6/7:

The mathematics test for Grade 6 comprised 35 items, five from the domain of geometry, 13 from arithmetic, and 17 from algebra; they can be considered a valid representation of bthe curricular content of Grades 5 and 6.

Below, there are two examples which again represent the lower and the upper profiency level.

Example for a mathematics item from the HST 6/7, lower proficiency level:

Three childrfen want to share 72 cards showing pictures of baseball players. How many does each child get?

A 12

B 14

C **24**

D none of these

Example for a mathematics item from the HST 6/7, upper proficiency level:

1 000 youngsters have a party in a disco. $^{2}/_{5}$ of them are girls. How many boys are there?

A 400 boys

B 450 boys

C **600 boys**

D none of these

All items are constructed as multiple-choice items with four response categories. Both tests required 45 minutes of testing time. Each was used in two pseudo-parallel forms whiich differed only in the order of items and response categories. As post-tests, the students received the parallel form they had not already worked on during the pretest.

Culture Fair Intelligence Test

In order to control for inter individual differences in learning ability, the mathematics test was complemented by CATTELL'S non-verbal *Culture Fair Intelligence Test – CFT 20* in an adapted short version (Weiss 1998). This test aims at measuring the so-called "fluid intelligence" which includes aspects of reasoning abilities, *i.e.,* very general learning prerequisites. The items are organized into four sub-tests: *Continuation of Sequences,* *Classifications, Matrices, Topological Reasoning.* They are based on graphical representations constructed to allow a multiple-choice format. Within each sub-test, they are ordered according to difficulty. Due to the use of figural representations, this standardized test is virtually independent of the command of the language of instruction, as well as of substantive subject-matter knowledge. Thus, it allows the identification of discrepancies between a student's cognitive resources and his or her achievement in terms of subject matter. The administration of the short version of the CFT 20 (46 items) requires 40 minutes of testing time including instruction. All items are constructed as multiple-choice items with four response categories (cf. Weiss, 1998, 10).

The CFT 20 can be used for children and adolescents between 8.5 and 18 years of age (roughly Grades 3 to 12 in the German school system) and for adults with moderate amounts of education. There are standardized norms for these groups available. Moreover, there are reference values for Grades 3 and 4 in Primary School and for Grades 5 through 10 in Secondary School. If so required, raw scores can be transformed into IQ scores, standardized T-scores, and percents rank scores.

Test items are based upon 4 response categories each. Including a short introduction, the administration of the test requires 45 minutes.

Like the mathematics test, the CFT 20 has two pseudoparallel forms, A and B. Once again, these differ only by the order of the items and the position of the response categories.

Student Questionnaire

The application of a Student Questionnaire facilitates the measurement of attitudes and beliefs (*e.g.,* subject-related self-efficacy, interest in mathematics) as well as the collection of relevant background information. It may be mentioned here that in the second data collection wave, this component of the evaluation has been substantially broadened.

1.4 Psychometric adequacy of test instruments

The summary below presents the statistical properties of instruments as used in the initial data collection. Conjoint IRT scaling has been conducted for the pretest results on the HST 4/5 from Grades 4 and 5. The scaled vaues are the basis for the subsequent analyses (*cf.* Table 2).

Table 2: **Statistical characteristics of responses from tests and questionnaires, initial data collection**

Scale	test instrument	number of tasks	alpha	number of students
Mathematics, Grade 4	Hamburger Schulleistungstest für 4. und 5. Klassen (HST 4/5)	35	0,88	234
Mathematics, Grade 5	Hamburger Schulleistungstest für 4. und 5. Klassen (HST 4/5)	35	0,86	457
Mathematics, Grade 6	Hamburger Schulleistungstest für 6. und 7. Klassen (HST 6/7)	35	0,88	530
General learning prerequisites, "fluid intelligence"	Culture fair Intelligence test (CFT 20); Grade 4/5 Grade 6	46 46	0,83 0,78	641 530
Subject-related selroup. f-efficacy	Students Questionnaire, Grade 4/5 Grade 6	11 11	0,83 0,86	621 521

2 Posttest sample and test-instruments

2.1 Posttest sample

Due to some individual schools' decisions as to participation in the experimental program, there have been slight changes in the structure of the sample. Thus, the elementary Lenningskamp-Grundschule in Schwerte discontinued cooperation shortly after the beginning of the experiment; obviously this meant also that it failed to participate in the final data collection, and the alredy obtained initial data were to be excluded from later analyses. Therefore, only 4 primary schools with two classroms each – one experimental and one control – participated both in the initial and the final data collection. These eight classrooms form the basis of subsequent analyses.

The ten secondary schools which had participated in the initial data collection also participated in the final exercise. However, one teacher from the experimental class in Grade 6, Hauptschule in Wermelskirchen, had to discontinue his work with the Renaissance Learning approach because of technical difficulties, early in the experiment. This group was therefore redefined as an control class and then kept in the data set for analysis, because there were data on both entry and exit characteristics.

Thus, the sample of the final data collection consisted of 22 experimental and 25 control classes (for response rates, see Appendix 1). Table 3 gives an overview for the sample structure in the final data collection for the evaluation of the experimental implementation of "Accelerated Mathematics". Data analses are based on this sample structure.

***Table 3*: Achieved posttest sample: Number of classrooms by grade, school type, and participation status**

School type	Treatment group		Control group	
	classes	students	classes	Students
Primary schools (Grade 4)	4	92	4	105
Secondary schools (Grade 5)				
• *Hauptschule* (Basic)	3	60	3	62
• *Realschule* (intermediate)	2	61	2	62
• *Gymnasium* (academic)	2	62	2	62
• *Gesamtschule* (comprehesive)	2	56	2	55
Secondary schools (Grade 6)				
• *Hauptschule* (Basic)	2	45	4	90
• *Realschule* (intermediate)	3	80	4	111
• *Gymnasium* (academic)	2	64	2	63
• *Gesamtschule* (comprehesive)	2	57	2	56
Σ	**22**	**577**	**25**	**666**

2.2 Final data collection, response rates and data entry

The final data collection took place between June 21st and July 16th, 2004. 42 of the 49 participating classes were tested between June 21st and July 2nd. The school in Recklinghausen was allowed to postpone the posttest by two weeks, due to ist late start into the experiment. Similarly, the four classes of the *Hauptschule* in Ahlen were one week behind most others in their test scheme; this was owed to the schedule of one test administrator who had personally been requested by the school, but was available only after the end of the first week in July. As will be shown below, the postponement of testing at this school by a margin of one to two weeks was not accompanied by noticeable effects in either the treatment or the control group.

The test administrators returned the completed tests and questionnaires to the evaluation group at Humboldt University until the end of July. There, the data were entered manually and verified electronically. BY the end of August, the data were ready for scaling and analysis.

For the final data collection, response rates lay between 82 and 100 percent of all students enrolled in the treatment groups and between 76 and 100 percent in the control groups. Lower response rates were typical for *Hauptschulen* and one class at the Primary level (*cf.* Appendix 1).

Log entries and feedback from test administrators show that the final data collection was planned and conducted very carefully in most schools. The students worked their way through the test at high levels of motivation and engagement. In Grade 4, the relatively lengthy questionnaire pertaining to the characteristics of their schools and instructional perceptions was somewhat problematic. As a reaction to this, help from test administrators and teachers was required so that less familiar intruction-related concepts could be explained. Also, in the case of this instrument, concentration and perseverance were , in some cases, somewhat lessened by the demands which the preceding mathematics tests had placed, especially among the younger students. Thus, the questionnairer were not always completed and returned in complete fashion and with evidence of a thoughtful response mode.

2.3 Psychometric adequacy of posttest instruments

The instruments for the final data collection consisted of two parts: the mathematics tests (HST 4/5; HST 6/7) which have already been described in some detail in Section 1.3 and a student questionnaire. The latter addressed general perceptions of and attitudes towards school in general and mathematics instruction in particular. In addition, those students who had participated in “Accelerated Masthematics” as the experimental treatment were asked to articulate their judgements on “accelerated Mathematics”. A number of items and scales were used that had proved their merits in the above mentioned achievement census in the City of Hamburg and in the Study „*Qualität von Schule und Unterricht*“ („Quality of School and Instruction“) which is being conducted at Ludwig-Maximilian-Universität, Munich (*cf.* Ditton, 2001).

Moreover, a teacher questionnaire was used that illustrates characteristics of mathematics instruction from the perspective of the respective teacher. Teachers’ general attitudes towards school and instruction as well as aspects of their satisfaction with work and professional career choices and their perceived stress and pressure on the job were also measured (*cf.* Table 4).

Table 4: **Statistical characteristics of responses from tests and questionnaires, final data collection**

scale	test instrument	Num. of items	alpha	Stu-dents
Mathematics, Grade 4	HST 4/5	35	0,88	234
Mathematics, Grade 5	HST 4/5	35	0,86	457
Mathematics, Grade 6	HST 6/7	35	0,88	530
Subject-related self-efficacy	Students Questionnaire, Grade 4/5	11	0,77	648
	Students Questionnaire, Grade 6/7	11	0,86	536
Appropriateness: difficulty level	Students Questionnaire, Grade 4/5	3	0,70	639
	Students Questionnaire, Grade 6/7	3	0,77	530
Appropriateness: level of encouragement	Students Questionnaire, Grade 4/5	6	0,69	644

scale	test instrument	Num. of items	alpha	Stu-dents
	Students Questionnaire, Grade 6/7	6	0,72	538
Formal-cognitive structure of lessons	Students Questionnaire, Grade 4/5	5	0,60	645
	Students Questionnaire, Grade 6/7	5	0,68	538
	Teachers Questionnaire	5	0,74	38
Emotional support	Students Questionnaire, Grade 4/5	9	0,77	648
	Students Questionnaire, Grade 6/7	9	0,78	540
Teacher's diagnostic competencies, individualized feedback	Students Questionnaire, Grade 4/5	9	0,84	648
	Students Questionnaire, Grade 6/7	9	0,89	539
Motivational potential	Students Questionnaire, Grade 4/5	7	0,80	645
	Students Questionnaire, Grade 6/7	7	0,77	537
	Teachers Questionnaire	5	0,78	35
Classroom management	Students Questionnaire, Grade 4/4	9	0,71	643
	Students Questionnaire, Grade 6/7	9	0,76	537
	Teachers Questionnaire	4	0,84	39
Anxiety	Students Questionnaire, Grade 4/5	2	0,68	637
	Students Questionnaire, Grade 6/7	2	0,67	533
Evaluation of teacher's behaviour	Students Questionnaire, Grade 4/5	8	0,88	637
	Students Questionnaire, Grade 6/7	8	0,90	535
Classroom climate	Students Questionnaire, Grade 4/5	3	0,71	631
	Students Questionnaire, Grade 6/7	3	0,72	530
	Teachers Questionnaire	4	0,86	38
Attitude towards school	Students Questionnaire, Grade 4/5	6	0,70	637
	Students Questionnaire, Grade 6/7	6	0,60	537
High student potential	Teachers Questionnaire	5	0,78	37
Appropriateness: individualization	Teachers Questionnaire	5	0,87	36
General teacher's attitudes towards school: curricular compliance	Teachers Questionnaire	5	0,73	35
Teacher's attitudes to school: achievement orientation	Teachers Questionnaire	3	0,75	37
Teacher's attitudes towards school: individual support and promotion	Teachers Questionnaire	2	0,94	38
Job satisfaction	Teachers Questionnaire	6	0,91	35
School-induced dissatisfaction	Teachers Questionnaire	3	0,76	38
pressure of work, stress	Teachers Questionnaire	3	0,78	36
Positive experience with Accelerated Mathematics	Students Questionnaire	7	0,73	301
		7	0,71	336
Critical attitudes towards Accelerated Mathematics	Students Questionnaire, Grade 4/5	5	0,63	300
	Students Questionnaire, Grade 6/7	5	0,63	334
	Teachers Questionnaire	8	0,84	19
Improvement of diagnostics and personalizing instruction	Teachers Questionnaire	7	0,85	19
Inappropriate performance expectations in Acc. Mathematics	Teachers Questionnaire	5	0,77	20
Self-regulated learning by Acc. Math.	Teachers Questionnaire	3	0,83	20

3 Initial situation and growth in mathematics, Grade 4

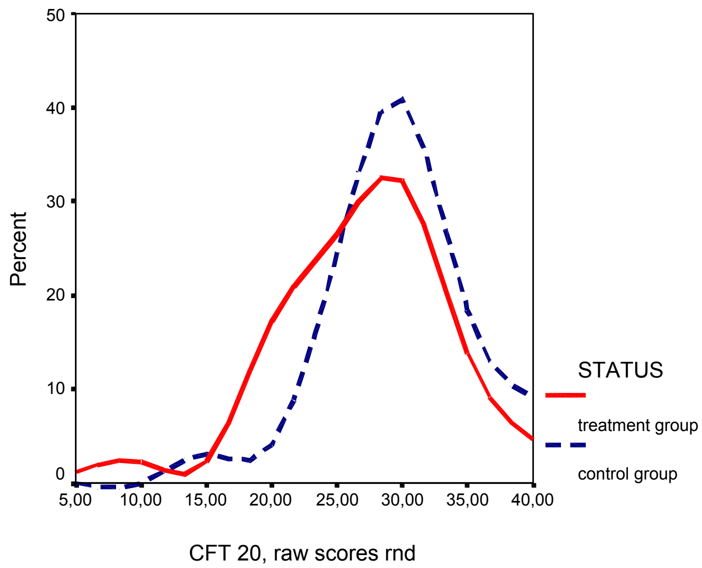
3.1 Initial situation, Grade 4

3.1.1 Cognitive disposition towards learning and social context, Grade 4

Cognitive diposition towards learning

The CFT 20 measures reasoning abilities by way of nonverbal tasks / items. Figure 1 below illustrates that in Grade 4, both the treatment and the control groups are characterized by a distribution that is biased towards high ability levels. With an average raw score of 28.50 points (SD=6,30), the tested 4^{th}-graders have obtained, in the initial data collection in the middle of the school year, a value which is almost three points or a little less than one half standard deviation below the reference point derived from the calibration sample of the year 1977 (*cf.* Weiss, 1998, 53) and still by about two points above the mean of the reference sample from the Hamburg longitudinal study, tested at the beginning of Grade 5 (*cf.* Lehmann, Peek & Gaensfuss, 1997; see also Weiss 1998, 52).

Figure 1: **Distribution of nonverbal reasoning ability (CFT 20), by treatment status, Grade 4**



The experimental groups display a mean score lying almost one forth of a standard deviation below the average of all 4^{th} grade students tested while the control groups are about one fifth of a standard deviation above that mean (*cf.* also Table 5). In other words: the treatment groups and even more so the control groups represent a sample that is very clearly positive selection with respect to measured intelligence. Since the participation of schools and teachers in the experiment was voluntary, one may assume that primary schools with favorable patterns of motivation, greater interest in innovative practices, and higher levels mathematics achievement were more likely to participate than others and that there was a tendency within these schools for the higher achieving classrooms to join the sample.

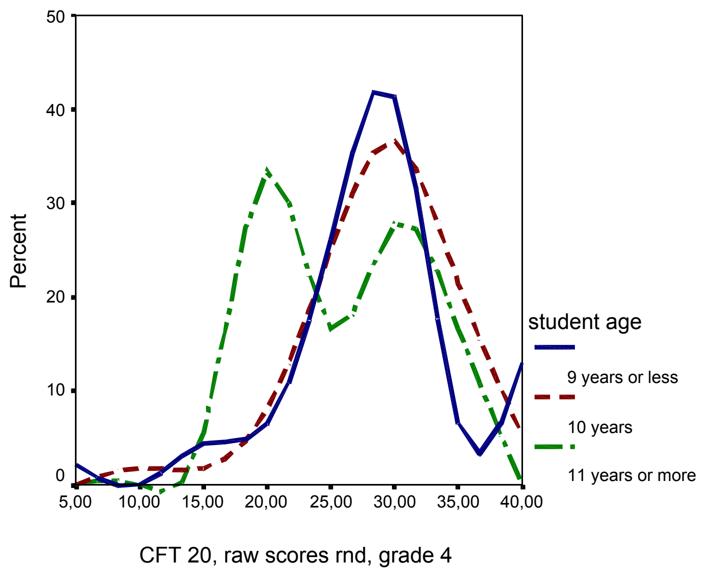
Table 5: **Nonverbal intelligence (CFT 20 raw scores), Grade 4: means and standard deviations by experimental status**

Status	Class	Mean	Number of students	Standard deviation
treatment	12	26,18	22	6,87
	21	27,00	26	7,74
	31	28,16	19	4,79
	51	26,85	20	7,21
	Total	*27,01*	*87*	*6,76*
control	11	28,89	28	6,38
	22	29,66	29	5,97
	32	29,81	16	4,76
	52	31,08	25	4,60
	Total	*29,83*	*98*	*5,57*
Treatment and control	*Total*	*28,50*	*185*	*6,30*

The age of the students from Grade 4 varied, according to the self-reports, between 8 and 14 years, with 10-year-olds (64 percent) and 9-year-olds (26 percent) representing the two largest groups. About 10 percent of the sampled students are 11 years and olderf (*cf.* Appendix 2, Table 2).

Since this level of heterogeneity within classes with respect to age implies biased ability estimates, if grade-related reference norms are applied, test results will have to be interpreted by taking age norms into account as well. The following Figure 2 shows the distribution of CFT 20 raw scores within Grade 4 broken down by age.

Figure 2: **Distribution of nonverbal reasoning ability (CFT 20), by age group, Grade 4**



While reasoning abilities among the 9-year-olds and among the 10-year-olds are distributed in a slightly skewed fashion which illustrates once again the positive selection of students even within age groups, the older students (11 or more years of age) show a bimodal distribution. This suggests that this group is constituted as a result of rather diverse processes. Students with reasoning abilities below average (those grouped around the lower maximum) will in many cases still attend a 4th grade class because of belated school entry or grade repetition, both of which are often related to unfavorable cognitive prerequisites for learning. The reasons for those with a normally developed learning potential but a higher age than normal to be still enrolled in a 4th grade classroom are not known. There may be motivational factors behind this phenomenon (*e.g.*, "underachievement"), or biographical reasons (*e.g.*, migration from another country and subsequent enrolment in a relatively low grade which matched the current command of the language of instruction, but not necessarily a sound prognosis of future development). This second group is more frequent in the control classes, while the first – that of older students with limited cognitive resources – is more common in experimental classes. It should be noted, however, that the number of students in this rather unusual and problematic age group is small.

A comparison of test results with the respective age norms also demonstrates the cognitive superiority of the groups from Grade 4, involved in the present study. The 9-year-olds have achieved, on average, about 5 raw points more than the adjusted calibration mean for the respective age group. Similarly, the 10-year-olds who comprise almost two thirds of the 4th grade sample have reached an average which is above the respective age norm by approximately 4 raw points. As opposed to this, students of 11 years and older from experimental classes have reached test results below the average for 11 year old students in the calibration sample, while the corresponding group from the control classes performed at a higher level than the reference norm for this age.

Social context of achievement

It is well known from international studies of educational achievement (*cf.,* for example, Deutsches PISA-Konsortium, 2002) and also regional assessments that the relationship between cultural capital in the home and school achievement is reinforced rather than mitigated against by German schools (Lehmann *et al*., 2001, 139ff.). In order to take this effect properly into account in the analysis and interpretation of test results, some general socio-demographic background information was collected by way of the student questionnaire.

The students' unusually high abilities in terms of cognitive prerequisites to learning (as compared with the norms for grade and age derived from the calibration sample of the CFT 20 and the referential Hamburg study) had already suggested that the present Grade 4 sample not be representative with respect to its social composition, but rather a positively selected group.

Studies relevant to the investigation of such relationships often refer to parents' education, namely the parents' years of schooling or their highest leaving certificate as operationalized in the *International Standard Classification of Education* (ISCED). Also, the number of books in the home, the availability of an encyclopedia or dictionary, and the students' access to a room of their own are frequently used as indicators for the existence of educationally relevant resources, implying higher degrees of cultural experience and better individual opportunities for learning (for Germany, *cf.*, for instance, Lehmann, Peek & Gaensfuss, 1997; Deutsches PISA-Konsortium, 2001).

Unfortunately, the respective information collected in the course of the present study for the students of Grade 4 allows only limited insights because of missing data. Thus, only 30 percent of the 4th-graders were able to indicate their parents' vocational education and training, to give just one example. With respect to the current employment situation, student responses appear to be somewhat more informative. According to the students, about 80 percent of the fathers and 50 percent of the mothers were employed at the time of the initial data collection. About three percent of each group were indicated to undergo pre-employment training or vocational retraining. About six percent of the mothers and four percent of the fathers were indicated to be unemployed. These are figures which are remarkably plausible and in accordance with official labor-market statistics, if one is prepared to accept the supposition that the sample is, culturally, socially, and economically speaking, upwardly biased.

About 80 percent of the students in Grade 4 have a room of their own; about 90 percent have indicated to possess their own desk. About two thirds of the respondents have access to a dictionary and/or an encyclopedia. If judged by the students' indications as to the language(s) spoken in the home, 17 percent of them come from immigrant families, a ratio which is rather low as compared with official statewide statistics. Within the group of immigrant students, about one half speaks regularly („always“ or „almost always“) German at home.

To summarize, it must be noted that the Grade 4 classes investigated here clearly represent a positively selected group, as far as their cognitive prerequisites for learning and their extrascholastic background are concerned. The students are characterized by a level of nonverbal intelligence that is far above average by any standard of comparison and by a cultural and social environment which also appears to be unusually conducive to learning. As these selection effects are stronger with respect to the control group as compared with the treatment group, the effectiveness of the experimentally implemented *Accelerated Math* approach is likely to be underestimated in the analyses below.

3.1.2 Initial proficiency in mathematics, Grade 4

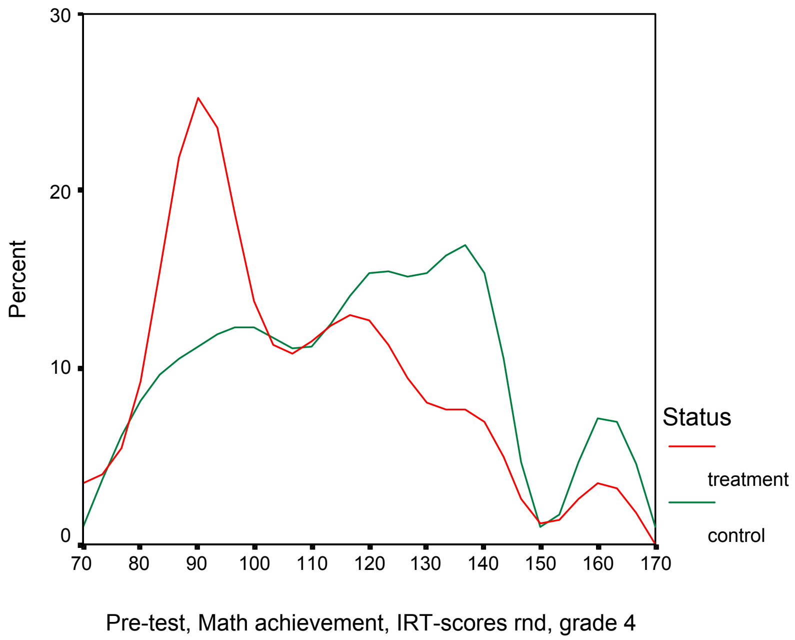
Out of the total of 35 items, the Grade 4 students have, on average, solved 22.5 items correctly 5 (SD=6,9). This is equivalent to an average 64 percent correct. By comparison, the students in Hamburg who took a slightly shorter version of the test (30 items) at the beginning of Grade 5 or half a school year later attained an average percent correct of 58 percent (mean raw score = 17.6; SD = 6.2; *cf*. Lehmann, Peek & Gaensfuss, 1997, 32).

As already described in Section 1.3, the raw scores were subjected to IRT analysis and the corresponding nonlinear, probabilistic transformations and thus were made comparable to statistics obtained from two regional studies for which data collections took place in 2003: the so-called KESS Achievement Census in the City of Hamburg („*Kompetenzen und Einstellungen von Schülerinnen und Schülern“*; $N \approx 13.000$; Bos et al. 2004) and the ELEMENT Study in the City of Berlin („*Erhebung zum Lese- und Mathematikverständnis, Entwicklungen in den Jahrgangsstufen 4 bis 6*“; $N \approx 2.700$; Lehmann & Nikolova, 2004). The joint mean for these two studies was set to be 100 and the joint standard deviation 25. The transformed test scores will henceforth be referred to as „proficiency scores“ (for the general approach, see Lehmann, Gaensfuss & Peek, 1998, 87ff.). The correlation between raw scores and the IRT-based proficiency scores thus defined is $r = 0.98$. This means that the information obtained from raw and proficiency scores is essdentially very similar, except that IRT scores have a number of advantages such as taking differential item difficulties into account and allowing for test equating in situations where respondents have worked on different sets of items. IRT-based proficiency scores and their corresponding levels are particularly suited to comparisons between initial and final achievement, *i.e.,* the analysis of growth over a period of time.

The students from Grade 4 in Northrhine-Westphalia (NRW), covered by the present study, have attained an average proficiency score of approx. 110 points in the pretest (mean = 109,9; SD = 25,0). Thus, these students have obtained test results that are clearly above the average test performance in Berlin and Hamburg, 2003, displayed also in the second half of Grade 4, *i.e.,* at a comparable point in time. The fact that the difference is equivalent to not much less than a year's learning toeards the end of primary school, underscores once more that the students from Grade 4 in the experiment represent a highly select group.

The following Figure 3 shows the initial distributions of mathematics proficiency among the students from the treatment and the control groups. Similar to the superiority of the control classes in terms of reasoning / nonverbal intelligence, there is a tendency for the control group to have shown higher levels of mathematics performance at the beginning of the experiment. Here, the proportion of low-achieving students is very noticeably smaller and the proportion of high-achieving students clearly higher than in the treatment groups.

Figure 3: **Initial proficiency in mathematics by experimental status, Grade 4**



On average, the treatment classes attained 103 initial proficiency score points, while the control classes reached 116 points in the initial testing (*cf.* Table 6). Thus, the latter were, at the outset of the experiment, slightly more than one half standard deviation above the treatment group.

As the comparison with the raw scores obtained by the students in the longitudinal Hamburg study (LAU) had already shown, the Grade 4 classes from NRW, especially those from the control group, were by any measure unusually advanced in their mathematics proficiency. Nevertheless, there were also quite remarkable differences at the classroom level in NRW. Thus, the classroom with the lowest mean proficiency score (class 21; see Table 6) achieved about one half of a standard deviation below the grand mean for all classes oof Grade 4, while the class with the highest mean performed at a mean level of almost two thirds above the grand mean (class 52; see Table 6). According to what is known from other studies, the latter value is equivalent to more than one year of mathematics learning.

Table 6: **Pretest results by experimental status and class, Grade 4**

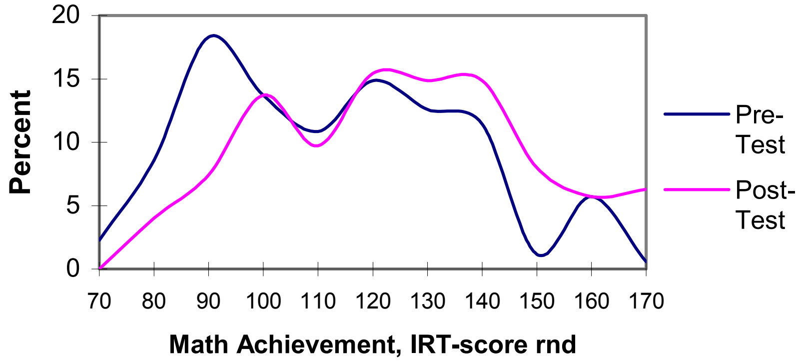
Status	**Class code**	**Number of students**	**Mean**	**Standard deviation**	**Minimum**	**Maximum**
Treatment	12	22	103,87	16,27	82,37	141,28
	21	26	97,49	18,58	62,51	130,78
	31	19	106,79	27,68	28,02	141,28
	51	20	104,83	35,69	28,02	156,64
	Total	*87*	102,82	24,90	28,02	156,64
Control	11	28	112,12	28,96	49,15	169,16
	22	29	111,36	19,78	73,07	156,64
	32	16	115,86	19,30	91,09	156,64
	52	25	126,27	20,91	79,36	156,64
	Total	*98*	116,12	23,46	49,15	169,16
Treatment and control	*Total*	*185*	*109,86*	24,98	28,02	169,16

It may be noted here, that students both from the treatment and the control groups have expressed relatively positive concepts of their mathematical abilities. On a four-category response scheme ranging from 1 = „disagree“ to 4 = „agree“, the respective class means lay between 2.77 and 3.14. In other words: the students were convinced to be capable of solving mathematical tasks and problems with a fair degree of certainty, and, according to their own responses, they enjoyed dealing with mathematics.

3.2 Growth of mathematics proficiency during the experimental treatment, Grade 4

Figure 4 below shows the shift of the distribution of mathematics proficiency in Grade 4 between the initial data collection (February 2004 or middle of the school year) and the final data collection (June / July 2004, towards the end of the school year).

Figure 4: **Distribution of mathematics proficiency at the beginning and end of the experiment, Grade 4**



It is noteworthy that the bimodal distribution which existed initially persisted during the course of the experiment, albeit with a clear shift towards higher proficiency levels.

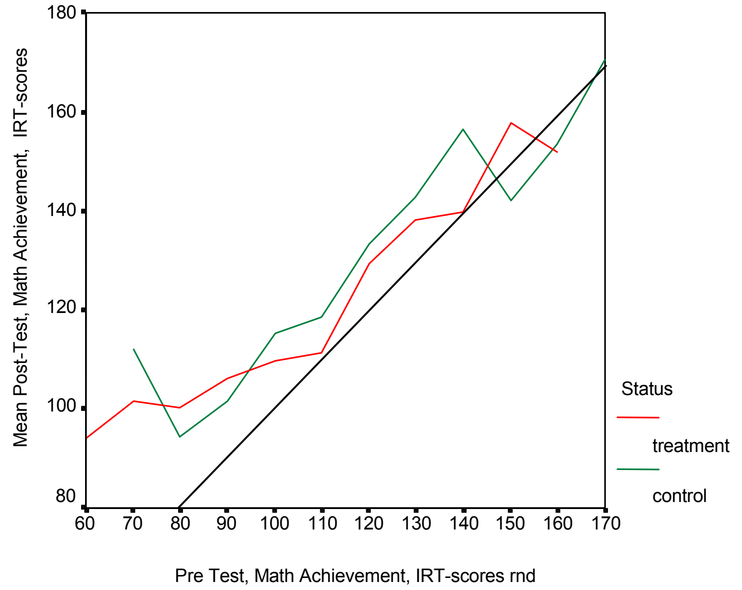
If one compares the average initial proficiency (mean = 109,9, SD = 25,0) with that observed in the final data collection (mean = 122,4, SD = 24,9; *cf.* Table 7), one can see that there has been an increase for all classes participating in the experiment, including both the treatment and control classes, by approximately half a standard deviation. Thus, the average growth in these classes, measured over an interval of four months, is more than the reported mean annual increment in Grades 5 and 6 in the longitudinal Hamburg study (*cf.* Lehmann, Gaensfuss and Peek, 1999, 114) and even more favorable as compared with the German annual growth rates for mathematics, Grades 7 and 8, observed in the IEA Third International Mathematics and Science Study (TIMSS; *cf.* Baumert *et al*., 1997, 144ff.).

> In Hamburg, average proficiency gains in mathematics between the beginning of Grade 5 and the end of Grade 6 were 0.68 standard deviations or 0.34 standard deviations *per annum.* Thus, the increment of 0.50 standard deviations observed in the present study for an interval of four months only appears to be extraordinary, even if some deceleration in the rates of learning between primary and secondary grades is suggested by other studies in this field.

It has to be noted, however, that the Hamburg study is based on an achievement census and TIMSS on a large nationwide representative study, while in the present context the existence of positive selection effects had to be acknowledged (*cf.* Section 3.1 above). As a consequence, proficiency gains were to be expected which were somewhat above what was observed in other studies.

As Figure 5 illustrates, there has been a remarkable growth both in the treatment and the control classes virtually across the whole spectrum of initial ability (the black line has been added as a reference to indicate zero growth). The almost parallel shape of the two lines for treatment and control classes is noteworthy, with only a slight advantage fort he control group in the middle range of initial ability. This indicates that the average as well as the differential effectiveness of instruction is very similar in the two groups.

***Figure 5*:** **Proficiency growth between the initial and the final data collection, by initial achievement and experimental status, Grade 4**



At the class level, however, there are quite substantial differences in proficiency gains, if classes of equal or similar initial achievement are compared (see Figure 6).

Figure 6: **Mathematics proficiency in the initial and the final data collection, by class and experimental status, Grade 4**

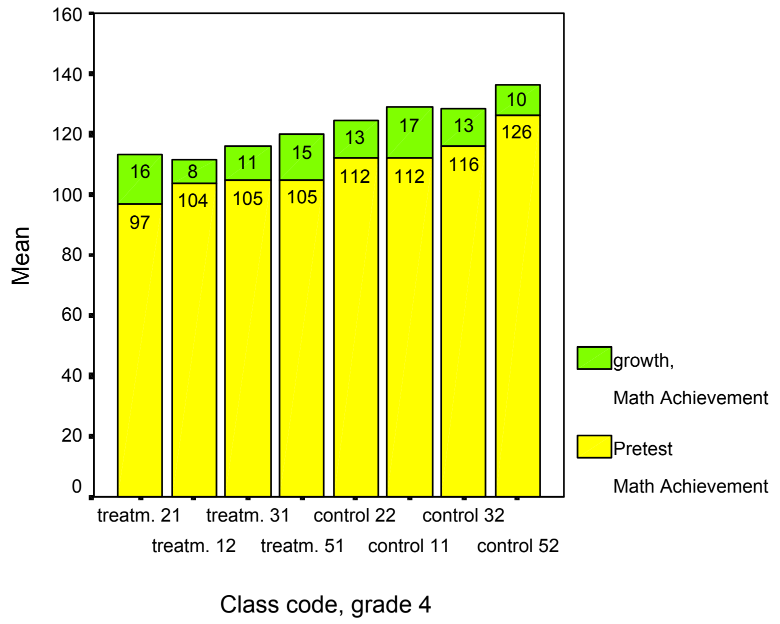


Figure 6 depicts the mean increments in mathematics proficiency between the initial and the final data collection by class – implying also the distinction by experimental status. This shows that there are quite significant differences in growth between individual classrooms. Thus, the treatment class no. 21, for instance, which had obtained the lowest scores in the initial data collection (*cf.* Table 7), had the highest gains. However, a significant increase in mathematics proficiency can also be reported for control classes no. 22, 32, and 11.

Table 7: **Mean mathematics proficiency at the beginning and end of the experiment, by experimental status and class, Grade 4**

Status	Class code	Mean Pretest (IRT-scores)	SD	Mean Posttest (IRT-scores)	SD	Diff. Pre-post-test	Effect size	Students (pre and post)
Treatment	12	103,87	16,27	109,86	21,25	5,99	0,37	22
	21	97,49	18,58	113,32	18,49	15,83	0,85	25
	31	106,79	27,68	117,27	20,19	10,48	0,38	17
	51	104,83	35,69	120,80	28,36	15,97	0,45	20
	Total	*102,82*	*24,90*	*115,00*	*22,24*	*12,16*	*0,49*	*84*
Control	11	112,12	28,96	128,92	28,27	16,80	0,58	28
	22	111,36	19,78	124,66	22,25	13,30	0,67	28
	32	115,86	19,30	128,11	22,41	12,25	0,63	16
	52	126,27	20,91	134,76	27,38	8,49	0,41	25
	Total	*116,12*	*23,46*	*129,07*	*25,35*	*12,95*	*0,55*	*97*
Treatment and control	*Total*	*109,86*	*24,98*	*122,44*	*24,89*	*12,58*	*0,50*	*181*

At first sight, there does not seem to be much of a difference between treatment and control classes with respect to mathematics proficiency: there is an increment in the order of half a standard deviation both in traditionally instructed classes and in those taught using the „*Accelerated Math*“ approach which furnishes targeted and detailed information on the students' currently attained level and thus supports individualized learning.

As was already mentioned, the most pronounced gains were achieved in class no. 21, namely 85 percent of a standard deviation. At the same time, this was the class in which „*Accelerated Math*“ was used most intensively.

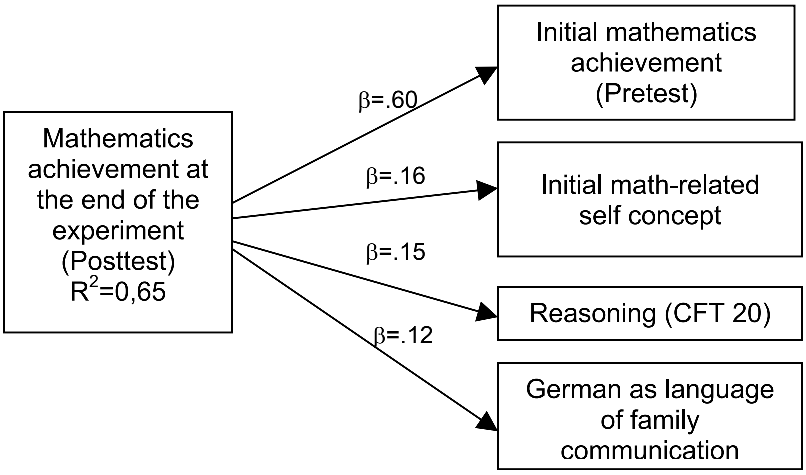
There were marked differences between classrooms in terms of utilizing the program „*Accelerated Math*“, a circumstance which impedes the experimental identification of treatment effects. In one of the four treatment classes, the program was used only on a very limited scale (user group2; *cf.* Appendix2, Table 1). In another two classes, the program was implemented only at a „regular“ level (user group 3). In one of the classes, namely that with the least favorable average initial mathematics proficiency and, at he same time, the highest average gains, a very intensive utilization was recorded.

Finally, it was ascertained by way of a multiple regression analysis which characteristics are most important in predicting the posttest proficiency scores. While these were set to bet he dependent variable, pretest achievement, nonverbal reasoning ability, social background factors, and instructional characteristics were entered as independent variables.

In order to measure the intensity of program utilization, log data on individual, student-level usage were aggregated to the class level. User groups were defined according to the average number of problems attempted in any given class. These groups can be taken as representing various degrees of implementing the „*Accelerated Math*“ approach. Four groups were distinguished, namely those with „minimal“, „partial“, „regular“, and „intensive“ utilization. These groups could then be represented by three dichotomous „dummy variables“. Empirically, only the distinction between intensive usage against all other forms proved to be significant.

Figure 7 below shows which factors could be demonstrably associated with an independent influence on mathematics proficiency at the end of the evaluation. Minor influences, irrespective of their statistical significance, may be omitted from the graph below.

Figure 7: **Relative importance of several predictors of posttest mathematics achievement (Beta-weights derived from a multiple regression analysis), Grade 4**



The pretest performance has turned out to be the best unique predictor of mathematics achievement at the end of the experiment (posttest). The initial math-related self-concept, nonverbal reasoning ability, and the frequency with which German is used as the language of communication in the family are other important predictors.

In this mode of analysis and in the case of Grade 4, variables which represent student perceptions of math instruction and the intensity of program utilization do not render an independent contribution to the explanation of interindividual posttest achievement differences[1]. This will be primarily due to the fact that the lack of comparability between treatment and control classes which was described above has masked the respective effects of the experimental approach. Moreover, a so-called "John Henry Effect" cannot be ruled out, i.e., a tendency among control group students and teachers to consciously try and outperform the students in the treatment group. An finally, it can be shown that 4th-grade students do not have highly differentiated perceptions required for providing valid assessments of instructional quality. This is illustrated by the relatively low reliabilities of the respective scales for students of Grades 4 and 5 as documented in Appendix 3, while the reliabilities are substantially higher for the same items / scales as applied to the Grade 6 sample (*cf.* Appendix 4). In view of the attenuated reliability and, consequently, reduced validity, a clear demonstration of the effectiveness of mathematics instruction based on students' assessments of *Accelerated Math* could not really be expected at this point.

[1] Normally, a multi-level analysis would have been the preferable mode of statistical analysis to ascertain the impact of such classroom-level variables. Given the small number of classrooms included in this evaluation, however, such approach did not seem advisable here.

4 Initial situation and growth in mathematics, Grade 5

4.1 Initial situation, Grade 5

4.1.1 Cognitive disposition towards learning and social context, Grade 5

Cognitive disposition towards learning

The Grade 5 students have achieved an average CFT 20 raw score of 28.5 which is again unusual and clearly above the norms for nonverbal reasoning (*cf.* Table 8). The treatment classes have attained an average value of 29.6 on the CFT 20 (SD=6.5, N=226), *i.e.*, approximately two raw score points above the adjusted calibration weight for this grade (*cf.* Weiss, 1998, 49).

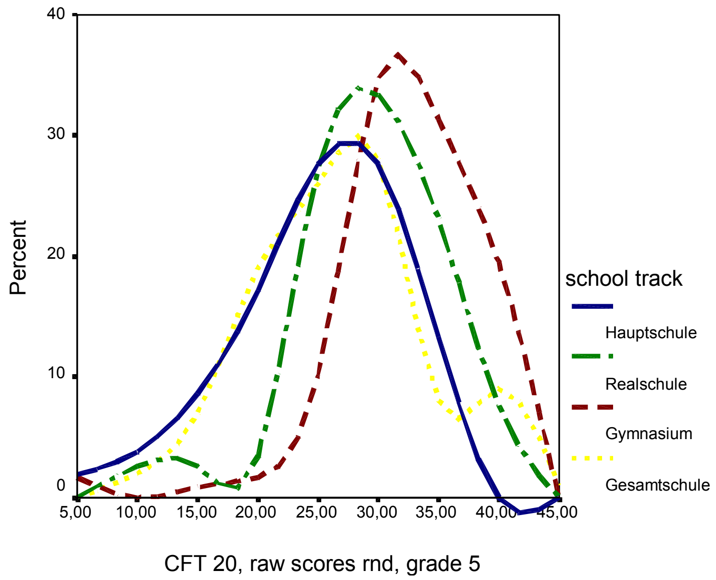
Compared with the Grade 5 data from the longitudinal achievement study in Hamburg which was repeatedly mentioned above the difference is even greater: about 3 raw points above the average of 25.6 points which had been attained in Hamburg. The control classes have nearly the same mean scores as the calibration sample (for the grade norms, see Weiss, 1998, 49; for the Hamburg reference sample, *cf.* Lehmann, Peek & Gaensfuss, 1997; Weiss, 1998, 53f). The standard deviation of 7.0 raw points differs noticeably from the calibration value (SD=5,8); it is also slightly higher than that encountered in the Hamburg data (SD=6,56; *cf.* Weiss, 1998, 49. 53f.). This could be indicative of a rather special social context of the sample, a point that will be investigated and discussed below. Considering the overrepresentation of students from *Hauptschule* and *Gesamtschule* in the sample, the observed bias towards intellectually more able and socially privileged students is very likely to underestimate the true magnitude of autoselectivity in the process of volunteering for participation in the experiment.

Table 8: **Mean nonverbal reasoning ability (CFT 20, raw scores) by experimental status, school type, and class, Grade 5**

Status	School type	Class	Mean	Number of students	Standard deviation
Treatment	*Hauptschule*	62	23,45	11	6,76
		72	26,24	21	7,46
		81	28,22	18	5,29
	Realschule	101	29,81	31	5,54
		111	26,97	29	5,63
	Gymnasium	122	32,77	31	5,28
		131	34,55	31	4,20
	Gesamtschule	142	26,04	24	4,60
		151	31,83	30	6,78
		Total	*29,60*	*226*	*6,48*
Control	*Hauptschule*	61	25,57	14	6,72
		71	24,67	24	5,72
		82	22,24	21	7,46
	Realschule	102	30,77	31	7,53
		112	28,96	28	5,29
	Gymnasium	121	29,97	29	7,03
		132	31,68	31	6,71
	Gesamtschule	141	23,74	23	6,59
		152	24,59	29	6,69
		Total	*27,36*	*230*	*7,35*
Treatment and control		*Total*	*28,47*	*456*	*7,01*

The differences between school types and classes within school types are evident (*cf.* Table 8; Figure 8). As was to be expected, the four classes of *Hauptschule* (class nos. 61, 62, 71, 72, Table 8) as well as the two control classes of *Gesamtschule* (class nos. 141 and 152) hadd raw score means that were quite noticeably below the adjusted grade norm (calibration sample, Grade 5: mean = 27,6; SD = 5.8). The distributions for these two school types present evidence for a very broad range of fundamental learning abilities. This is especially true for the bimodal distribution of intelligence test scores in *Gesamtschule* where the secondary maximum is based on the unusually good reasoning abilities apparent in one experimental class (Table 8, experimental class no. 151). Classes from *Realschule* and *Gymnasium* are characterized – with the exception of control class no 82 in *Realschule* – by above-avergage mean test results in comparison with both the age norm and the reference values provided by the Hamburg longitudinal study of the year 1996 (*cf*. Lehmann, Peek & Gaensfuss, 1997, 105).

Figure 8: **Distribution of nonverbal reasoning ability (CFT 20) by school type, Grade 5**

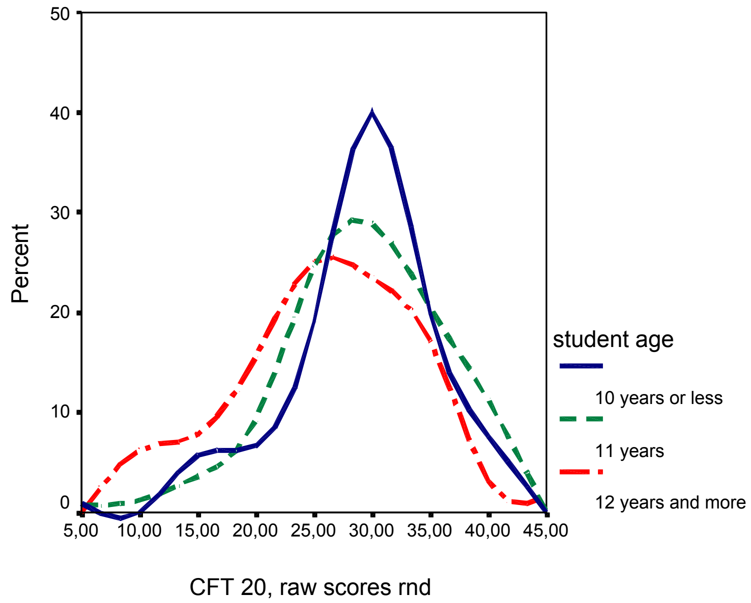


As in the case of Grade 4, there are significant and substantial differences in terms of reasoning ability, depending on age (*cf.* Figure 9). The age of the students in Grade 5 as indicated by self-reports varies between 8 and 13 years. The highest proportion in this grade pertained to the 11-year-olds (52 percent); about one forth of the participating students was 10 years old during the initial data collection (*i.e.*, at the middle of the school year) and about 13 percent were 12 years old. The remaining students were either older than 12 or younger than 10 years old (see Appendix 2, Table 3). In order to present the finding referring to nonverbal reasoning abilities, three age groups were defined: students of 10 years or less, students of 11 years, and students of 12 years or older.

The distributions displayed in Figure 9 show that the heterogeneity of reasoning ability increases with age. While the averages for the youngest and the middle age groups differ only marginally, the heterogeneity of abilities is greater for the 11-year-olds, and there is an even greater variance and also a rather low mean for those who are 12 years or older (mean = 26,11, SD = 7,58).

Theoretically, the age range within a grade should not exceed one year, if one considers the legal requirements for the first school enrolment. In reality, however, the differences are much greater, due to belated school entry, grade repetition, and mismatch between biological age and competency level as determined by the school authorities in the late entries among immigrant students. These phenomena also give rise to differences between school types: *Gymnasium* students had the lowest mean age as against students from *Hauptschule* and *Gesamtschule* with the highest.

***Figure 9*:** **Distribution of nonverbal reasoning ability (CFT 20) by age group, Grade 5**



Social context of achievement

The data pertaining to the social background of the students from Grade 5 are incomplete, much like those in the case of Grade 4, especially with respect to the parents' educational attainment. Less than half of the students were able to indicate their parents' occupational training; among the students from *Hauptschule*, the missing rate was close to 80 percent. Thus, the respective data could not be included in the subsequent analyses. For the remaining variables which were to cover the students' socio-cultural and family background, the response rate reached 80 percent and more so that the respective information will be discussed below.

Among the parents of the 5th-graders, the employment situation appeared to be relatively favourable: about 51 percent of the mothers and approximately 80 percent of the fathers were (self-)employed during the initial data collection; about 5 percent of both mothers and fathers were being trained or re-trained vocationally or enrolled in higher education. The percentage of women who were predominantly engaged in family-related activities and for this reason not on the labor market was close to 30 percent and thus below the average for Germany. Four percent of the mothers and six percent of the fathers were, according to the students' responses, unemployed; these are ratios which are also below the respective national averages. It should be noted, however, that only 56 percent of the students from *Hauptschule* provided some information as to their parents' employment situation, while the corresponding ratios fort he other school types were all above 75 percent (*Realschule* 78 percent, *Gesamtschule* 85 percent, *Gymnasium* 99 percent). Thus the present overall estimates for the students' and their families' social background are likely to be positively biased, especially as far as the *Hauptschule* is concerned. More elaborate analyses would have to take this into account.

The second approach to explaining the privileged vs. Underprivileged position of certain student groups pertains to the existence – or lack – of educationally relevant resources in the parents' homes. This involves the availablity of a place for study in the family home, access to dictionaries / encyclopedias and other books, the possibility of using a computers and other modern media. About 86 percent of the 5th-grade students questioned have a room of their own, and 89 percent can use their own desk. About three quarters of the students confirmed that they can use a dictionary and / or encyclopedia at home, and more than half have a computer of their own. This percentage of computer-owners is 61 percent in *Gymnasium* and 52 percent in *Hauptschule.* Howewver, the percentage of students who have a TV-set of their own is higher in *Hauptschule* than it is in *Gymnasium*.

About 19 percent of the students come from immigrant families, as judged by their responses as to their mother tongue. Of these 19 percent, roughly one third has indicated that German is used as the predominant language in the family: in about one half of the cases this is true only occasionally, and for about 14 percent of the students from immigrant families (or 12 students, respectively), German is almost never the language for communicating at home.

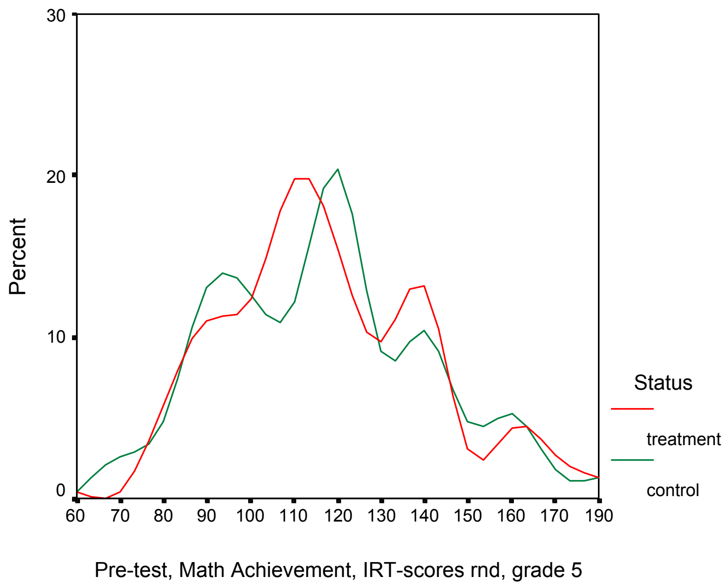
4.1.2 Initial proficiency in mathematics, Grade 5

As the standardized mathematics test HST 4/5 covers essential parts of the curriculum for Grades 4 *and* 5, classes from Grade 5 were subjected to the same test as those from Grade 4. On the average, the Grade 5 students solved 24.1 out of the 35 items correctly (SD = 6,4). This is equivalent to 69 percent correct. Comparing these figures with the corresponding figures derived from the longitudinal achievement census in Hamburg, both the treatment and the control classes in NRW have reached clearly higher proficiency levels than the students in the reference group, even if the additional five months of instruction are taken into account (for a discussion of this difference, see section 3.1.2). Especially in view of the high scores for nonverbal reasoning, as compared with the Hamburg reference group, such favorable outcomes as to mathematical proficiency were to be expected.

As in the case of the Grade 4 sample, the raw scores were transformed according to an IRT-based algorithm so that the test results could be allocated to certain competency or proficiency levels. The transformation rules were the same as the ones used for the Grade 4 sample; in fact, the respective scaling was conducted in a single, conjoint routine. In Grade 5, the initial mean proficiency score was found to be 116.0 (SD = 25.1; N = 457). The difference between the initial mean scores for Grades 4 and 5 was in the order of a quarter of a standard deviation, *i.e.,* somewhat less than what one might have expected, given that in these grades an average annual growth of a third of a standard deviation has been reported in other studies, as has been mentioned already. The present comparison between 4^{th} and 5^{th} grade is not a longitudinal one, however, and much of the reduced difference will be due to the unusually high initial mathematics performance in Grade 4. Thus, the extraordinary selection effects which were observed for Grade 4 are underscored once again.

The distribution of initial mathematics proficiency was almost identical fort he treatment and the control classes, including a distinction between multiple modes. The fact that four different school types were covered by the investigation implied that a rather broad range of mathematics abilities was present in the data and that school-type differences should ideally become visible in the form of a multimodal distribution.

Figure 10: **Initial proficiency in mathematics by experimental status, Grade 5**



As was to be expected, the lowest mathermatics skills were observed among students from *Hauptschule* whose mathematics proficiency scores were more than two thirds of a standard deviation below the grand mean for all Grade 5 students (*cf.* Table 9). Moreover, the *Hauptschule* students had a very heterogeneous achievement profile with extraordinary differences at the level of individual classes. Whereas treatment class no. 72 had a mean proficiency score of more than a full standard deviation above the overall mean for all *Hauptschule* classes and thus almost reached the level of *Realschule*, the students of the control class no. 82 appeared to have attained a very low proficiency level, indeed, with 45 percent of a standard deviation below the grand mean for *Hauptschule*, Grade 5. On the whole, the students of treatment classes have started the experiment from a somewhat more favorable position.

As opposed to this finding, the two treatment classses from *Realschule* showed significantly lower initial mathematics proficiencies than the two control classes belonging to this school type.

Treatment classes from the participating *Gesamtschulen* had test scores which were slightly below the grand mean for Grade 5, even if there was a marked difference between the two classes from *Gesamtschule,* with class no. 45 having displayed a mean proficiency score of about three quarters of a standard deviation below the grand mean for Grade 5, while class no. 151 was, at the beginning of the experiment, above the overall average.

Gymnasium classes, both in the treatment and the control segment, have shown a very clear advantage in terms of mathematics achievement; they were nearly 90 percdent of a standard deviation above the grand mean for initial mathematics proficiency in Garde 5.

Table 9: **Pretest results by experimental status, school-type, and class, Grade 5**

Status	School track	Class code	Mean	Number of students	Standard deviation
Treatment	Hauptschule	62	96,06	11	15,93
		72	107,45	22	18,21
		81	99,30	18	14,46
	Realschule	101	116,92	31	13,75
		111	106,19	29	23,84
	Gymnasium	122	131,44	31	21,37
		131	143,71	31	22,62
	Gesamtschule	142	97,08	24	16,46
		151	123,49	30	18,28
		Total	*116,64*	*227*	*24,42*
Control	Hauptschule	61	99,89	14	13,58
		71	96,74	24	16,87
		82	90,81	21	18,18
	Realschule	102	123,93	31	29,49
		112	124,07	28	17,63
	Gymnasium	121	131,72	29	22,60
		132	140,20	31	17,84
	Gesamtschule	141	96,79	23	15,08
		152	110,36	29	17,83
		Total	*115,45*	*230*	*25,77*
Treatment and control	Hauptschule		98,52	109	17,08
	Realschule		117,81	119	22,97
	Gymnasium		136,85	122	21,59
	Gesamtschule		108,13	106	20,22
		Total	*116,00*	*457*	*25,08*

Like the students from Grade 4, those from Grade 5 had a rather positive self-esteem with respect to their mathematics abilities. The mean for Grade 5 – on the scale which was already mentioned to be based on for categories ranging from 1 = „disagree“ to 4 = „agree“ was found to be 3.0 (SD = 0,55) for all students from Grade 5, with a range of 2.6 to 3.3 at the aggregate classroom level.

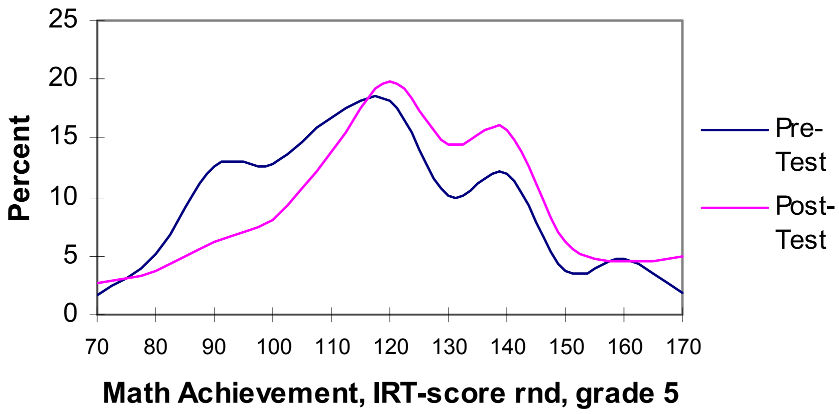
The proportion of variance of mathematical self-concept which can be associated with school type did not exceed 3 percent and thus was rather small. At the class level, it amounted to nine percent. The self-esteem in terms of mathematics abilities was by a third of a standard deviation lower in *Hauptschule* than in the other school types; here, the range for class-level aggregates was given as 2.6 to 3.0. There were no significant differences in the respect of mathematical self-efficacy between *Gymnasium, Realschule,* and *Gesamtschule.* It may be noted, however, that che classroom means in *Gesamtschulen* where the proficiency scores were clearly below the overall average*,* the self-esteem with respect to mathematics ability was found to be above the average.

Relevant studies have shown that the correlation between subject-related self-esteem and test achievement is often surprisingly low. This observation which is based on a number of in-depth analyses may be indicative of a tendency among the students to evaluate their abilities within a frame of reference constituted by their class / learning group and, by implication, by the school type they belong to (*cf.* Horstkemper, 1987, Lehmann, Gaensfuss & Peek, 1997).

4.2 Growth of mathematics proficiency during the experimental treatment, Grade 5

Figure 11 demonstrates the changes in the distribution of mathematics proficiency which have take place between the beginning (mid-school year, Grade 5, February 2004) and the end of the experiment (end of Grade 5, June / July 2004). The comparison illustrates that very noticeable achievement increments have occurred not only in Grade 4, but in Grade 5 as well, albeit not quite as uniformly as in the former case.

***Figure 11*: Distribution of mathematics proficiency at the beginning and end of the experiment, Grade 5**



It can be seen from Table 10 substantial mathematics proficiency gains have been observed both in treatment and control classes, but these gains are about twice the size in the former as compared with the latter. Whereas those classes in which *Accelerated Math* was utilized obtained an increment of more than one third of a standard deviation – an amazing growth, given the short duration of the experiment – the corresponding figure for the control classes is 19 percent of a standard deviation, *i.e.,* a still respectable, but significantly lower rate of progress. To express these findings from yet another angle: Over a period of half a school year, an average growth was observed in the treatment classes which was equivalent to the typical *annual* growth reported in comparable studies or twice the growth observed in relevant reference groups (for the Hamburg achievement census, see Lehmann, Gaensfuss & Peek, 1999, 112ff.; for TIMSS, see Baumert *et al.*, 1997). This is quite remarkable, even if one takes into account that some of these reference figures apply to grades 6, 7, and 8, *i.e.,* school years which may have slightly slower growth rates, as some evidence suggests.

***Table 10*:** **Mean mathematics proficiency at the beginning and end of the experiment, by experimental status, school type, and class, Grade 5**

Status	School track	Class code	Pretest Mean	Pretest Standard deviation	Posttest Mean	Posttest Stand. deviation	post-pre-difference	Effect-size	Num. Stud. (pre and post)
Treatment	Hauptschule	62	96,06	15,93	99,52	17,54	3,46	0,22	10
		72	107,45	18,21	120,95	19,93	13,50	0,74	22
		81	99,30	14,46	108,91	19,67	9,61	0,66	18
	Realschule	101	116,92	13,75	122,66	19,86	5,74	0,42	30
		111	106,19	23,84	122,63	17,45	16,44	0,69	28
	Gymnasium	122	131,44	21,37	144,32	18,10	12,88	0,60	30
		131	143,71	22,62	157,14	21,82	13,43	0,59	30
	Gesamtschule	142	97,08	16,46	102,65	20,24	5,57	0,34	23
		151	123,49	18,28	126,31	21,71	2,82	0,15	28
		Total	*116,64*	*24,42*	125,43	26,15	*8,79*	*0,36*	*219*
Control	Hauptschule	61	99,89	13,58	104,32	15,91	4,43	0,33	14
		71	96,74	16,87	107,46	19,13	10,72	0,64	22
		82	90,81	18,18	93,66	19,47	2,85	0,16	20
	Realschule	102	123,93	29,49	132,04	20,14	8,11	0,28	31
		112	124,07	17,63	120,99	17,97	-3,08	-0,17	28
	Gymnasium	121	131,72	22,60	138,63	18,93	6,91	0,31	28
		132	140,20	17,84	142,76	17,45	2,56	0,14	30
	Gesamtschule	141	96,79	15,08	102,44	19,01	5,65	0,37	20
		152	110,36	17,83	114,76	22,08	4,40	0,25	29
		Total	*115,45*	*25,77*	120,18	24,75	*4,73*	*0,19*	*222*
Treatment and control	Hauptschule		98,52	17,08	106,37	20,51	7,85	0,46	106
	Realschule		117,81	22,97	124,63	19,18	6,82	0,30	117
	Gymnasium		136,85	21,59	145,75	20,12	8,90	0,41	118
	Gesamtschule		108,13	20,22	112,47	22,88	4,34	0,21	100
		Total	*116,00*	*25,08*	*122,79*	*25,56*	*6,79*	*0,27*	*441*

Figure 12 shows the differential growth rates for initial proficiency groups, broken down by experimental status. The two curves are almost parallel: In (almost) all initial proficiency groups, there are proficiency gains, and these are (almost) constantly higher in the treatment group as compared with the control group. As in Grade 4, the increments are largest for students with relatively low initial proficiencies, with the black line marking zero-growth as a reference. There is also some indication that regression towards the mean is somewhat less effective in the treatment group, at the high end of the achievement spectrum.

***Figure 12*:** **Proficiency growth between the initial and the final data collection, by initial achievement and experimental status, Grade 5**

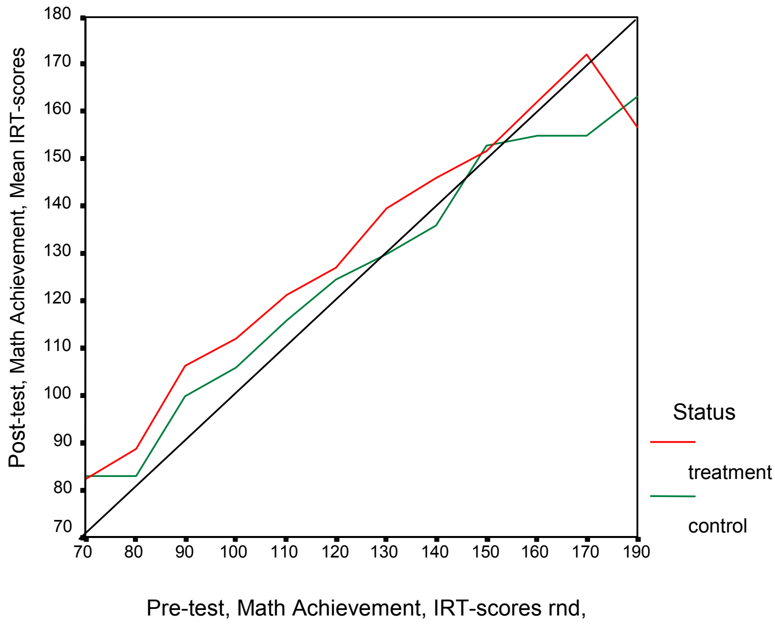
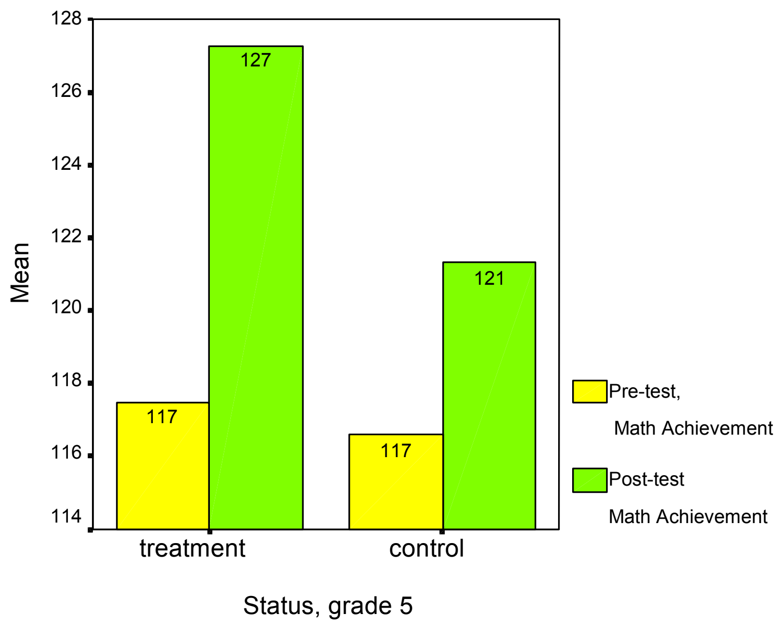


Figure 13 presents the difference between the treatment and the control group in terms of average proficiency gain. Based on nearly equivalent initial proficiency levels, the group taught under the premises of *Accelerated-Mathematics* has attained, by the time of the final data collection, on the average a clearly higher achievement level than the control group.

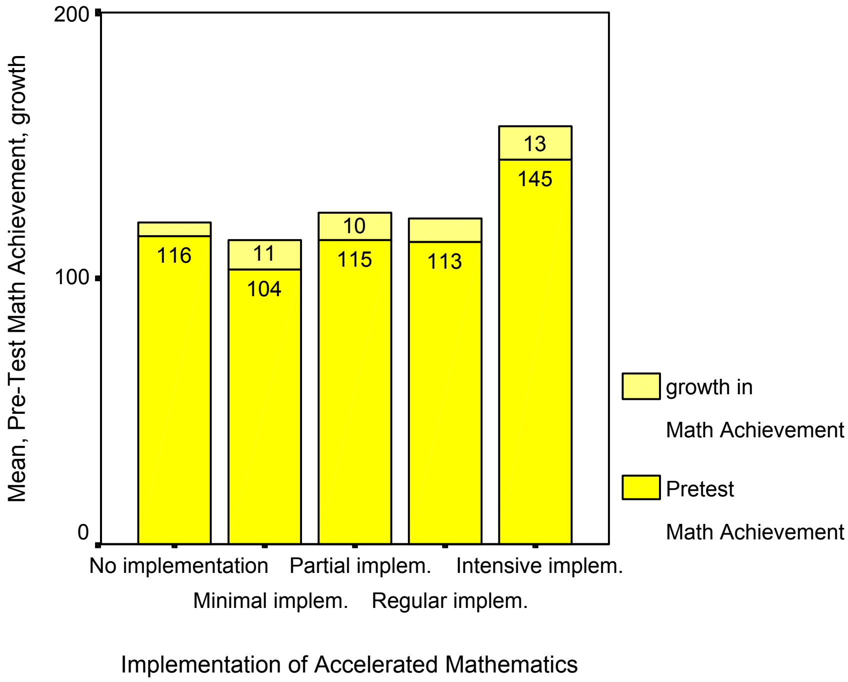
***Figure 13*:** **Mathematics proficiency in the initial and the final data collection, by experimental status, Grade 5**



As in Grade 4, there was substantial variance between classes in terms of the intensity with which *Accelerated Math* was implemented in the individual classroom. Thus, two of the nine treatment classes have made only minimal use of the program and two more have utilized it only partially. In other words: these were classrooms with a predominantly traditional instruction and only occasional enrichment furnished by *Accelerated Math.* Four classrooms can be described as having been regularly exposed to *Accelerated Math* and only one as a setting of intensive utilization.

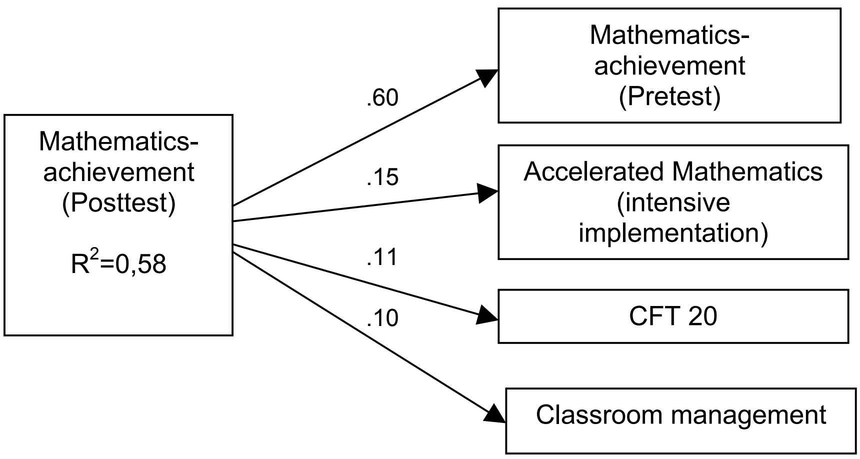
If average proficiency gains are considered in their relationship to the intensity of program utilization, the superior proficiency gains in groups with intensive exposure become apparent (swee Figure 14). If the non-random nature of this difference were to be tested statistically, greater numbers of intensively exposed classrooms would obviously be required. Unfortunately, under the constraints of the present experiment, this aim could not be attained in any of the three grades investigated.

***Figure 14*:** **Mathematics proficiency in the initial and the final data collection, by experimental status, depending on the intensity of utilizing *Accelerated Math*, Grade 5**



It remains to investigate the degree to which the new approach to mathematics instruction has rendered a contribution to the emergence of mathematics proficiency during the second half of Grade 5 which was independent of other determinants of mathematics learning. A multiple regression analysis was conducted, entering pretest performance, nonverbal reasoning, social background factors, math-related self-esteem, student assessments of classroom management during math instruction, and the intensity of program utilization (four user groups according to Appendix 2, Table 1) as independent variables and considering posttest achievement as dependent. Minor predictors, however statistical significant, may be omitted from the graph below.

Figure 15: **Relative importance of several predictors of posttest mathematics achievement (Beta-weights derived from a multiple regression analysis), Grade 5**



It was to be expected that pretest achievement came out as the strongest predictor of the posttest scores, as in most pre-posttest designs. Other independent contributions are derived from intensive utilization of *Accelerated Math* (user group 4; Appendix 2, Table 1) with the second highest beta weight ($\beta = 0.15$), followed by nonverbal reasoning (CFT 20; $\beta = 0.11$). Instructional characteristics subsumed under “classroom management” which mirror a well controlled, effective way of teaching have a beta weight of $\beta = 0.10$, still significant at the 5-percent level, as against the 1-percent level for the other predictors (see Figure 14).

5 Initial situation and growth in mathematics, Grade 6

5.1 Initial situation, Grade 6

5.1.1 Cognitive disposition towards learning and social context, Grade

Cognitive disposition towards learning

Contrary to what was observed in Grades 4 and 5, the average nonverbal reasoning abilities in Grade 6 (mean raw score = 29.4) are very close to the adjusted age norm in the calibration sample (mean = 29.2). The standard deviation observed in the NRW sample (SD = 6.75) is larger than the one computed for the reference sample collected fort he calibration, however (see Weiss, 1998, 49).

***Table11*: Mean nonverbal reasoning ability (CFT 20, raw scores) by experimental status, school type, and class, Grade 5**

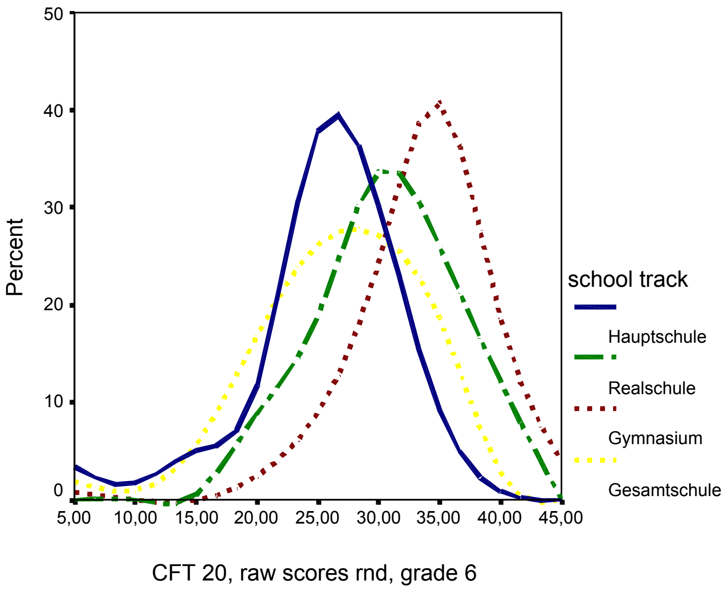
Status	School track	Class	Mean	Number of students	Standard deviation
Treatment	Hauptschule	64	25,69	16	5,16
		83	23,20	20	7,47
	Realschule	94	31,00	25	4,67
		104	32,73	26	3,79
		113	29,58	24	5,33
	Gymnasium	123	32,77	31	6,34
		133	35,33	30	3,85
	Gesamtschule	144	23,50	26	6,13
		153	28,39	28	5,40
		Total	*29,62*	*226*	*6,68*
Control	Hauptschule	63	25,12	17	7,97
		73	24,77	22	5,80
		74	26,57	23	6,05
		84	28,52	21	4,66
	Realschule	92	29,41	27	6,53
		93	29,19	27	6,93
		103	33,74	27	4,07
		114	27,64	25	4,92
	Gymnasium	124	32,91	32	6,96
		134	33,73	30	4,72
	Gesamtschule	143	24,80	25	7,30
		154	29,32	28	6,74
		Total	*29,18*	*304*	*6,80*
Treatment and control	Hauptschule		25,69	119	6,35
	Realschule		30,50	181	5,59
	Gymnasium		33,67	123	5,66
	Gesamtschule		26,61	107	6,77
			29,37	*530*	*6,75*

If the treatment and control groups are considered in their totalities, the average scores are almost identical. As is to be desired for an experimental design, neither of the two groups appears to be superior to the other in terms of this control variable. There are, however, marked differences between the school types. Thus, five classes from *Hauptschule* and two classes from one of the participating *Gesamtschulen* have produced test results which were by a wide margin below the grand mean (*cf.* Table 6 classes no. 63, 64, 73, 74, 83, 143, 144).

The differences in nonverbal reasoning ability according to school type attended are illustrated in the diagram below (Figure 16). Similar to the situation in Grade 5 there is a concentration of low test results in the group attending *Hauptschule*, arelatively broad range of results in *Gesamtschule,* and a rather privileged position of students attending *Gymnasium* mit a distribution tending towards favorable results. Overlapping areas between school types are also apparent.

In Grade 6, the proportion of variance in the students' nonverbal reasoning ability associated with school type attendance is 21 percent.

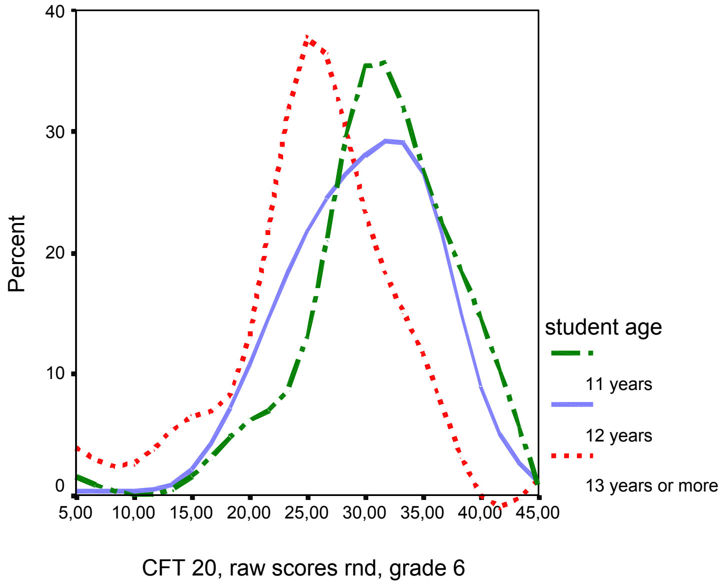
Figure 16: Distribution of nonverbal reasoning ability (CFT 20), by school type, Grade 6



At the time of the initial data collection, more than half (58 percent) of the students from Grade 6 were 12 years old, about one quarter was 11, and 16 percent 13 years and older. As was to be expected – presumably a consequence of belated school entry and grade repetition – , the older students (13 years and higher) have shown a somewhat less favorable position with regard to their average reasoning abilities (see Figure 17). This group has attained a mean CFT 20 score of 25.1 raw points (SD=7,66) which is almost two third of a standard deviation below the respective grade mean (for the age distribution in Grade 6, see Appendix 2, Table 4).

Well in accordance with this, the proportions of ‚older' students in *Hauptschule* (30 percent) and *Gesamtschule* are substantially higher. By way of contrast, the respective proportion of students of 13 years and higher is 10 percent in *Realschule* and only 4 percent in *Gymnasium.* It may be noted at this point that a change of school type attended (*e.g.,* from *Gymnasium* to *Realschule*) is frequently recommended or even required as an alternative to an otherwise inevitable grade repetition.

Figure 17: Distribution of nonverbal reasoning ability (CFT 20) by age group, Grade 6



Social context of achievement

Although they were included in the questionnaire, some important indicators for the support functions of the home such as parents' educational attainment could not be taken into account in the respective analyses because of missing data, not unlike the case of the other two grades investigated. Only about half of the students reported on their parents' highest educational and / or vocational training certificate, such that the required representativeness – also across school types – could not be assumed to be given

About 84 percent of the male parents and approximately 60 percent of the female ones were employed at the time of the data collection; 3 percent of the males and 5 percent of the females were indicated as enrolled in continuing or higher education, and about 5 percent of the fathers as against 4 percent of the mothers were, according to the students, registered as unemployed.

The number of books in the home which has appeared in a broad body of research as a valid indicator for the socio-cultural conditions in the home and one of the best predictors of school achievement (*cf.* Lehmann, Peek & Gaensfuss, 1997, 68) can be characterized as follows: About half of the respondents has indicated that there are more than 100 books in the home. About one quarter has indicated less than 50 books and the remaining quarter between 50 and 100 books. As in other studies, the correlation between number of books in the home and test achievement was highly significant. In Grade 6, the proportion of variance in test achievement which is associated with the number of books in the home was found to be $Eta^2 = 0.12$ for the pretest and $Eta^2 = 0.09$ for the posttest.

Comparable studies have also demonstrated a relationship between test results and the number of siblings, with students from small families (one or two children) being at an advantage. Students from families with three or more children had, on average, less favorable results. It has to be noted, however, that this variable is confounded with other social characteristics, above all immigrant status (see Lehmann, Peek & Gaensfuss, 1997, 66). About half of the students in the present Grade 6 sample grow up without a sibling in the family, and another third with only one. These values are typical for the situation in middle-class families.

In their majority, the students have access to a wide array of educationally relevant material resources such as a room of their own (85 percent), a desk of their own (91 percent), dictionaries / encyclopedias of their own (79 percent), or a computer of their own (65 percent). There are characteristic differences between school types in terms of investment patterns: Whereas no differences were apparent in Grade 6 with respect to domestic study conditions and access to computers, such differences were all the more visible as to the existence of reference sources (dictionaries etc.) in the homes. More than 85 percent of the students from *Realschule* and *Gymnasium* have indicated ownership of and access to dictionaries, while only about one half of the students from *Hauptschule* have such resources available in their homes.

According to the students, about 20 percent of them come from families with a history of immigration. About half of these speak regularly German in their families. Only 7 percent have indicated that German is almost never used in their homes.

5.1.2 Initial proficiency in mathematics, Grade 6

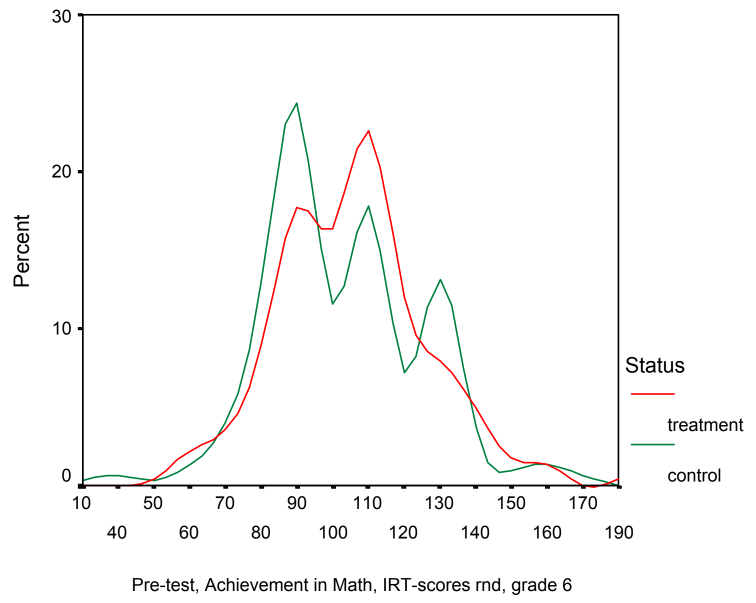
The subtest mathematics within the HST 6/7 comprises five geometry items, 13 arithmetic items, and 17 algebra items, all in a multiple choice format. For the NRW context, curricular validity for Grades 5 and 6 may be assumed. The test may be applied in Grade 6 or at the beginning of Grade 7. Ten of the 35 items serve as bridge items, in order to link the test to the HST 4/5 in the case of longitudinal designs. It will be left for future analyses to invesitigate such relationships across grades.

Out of the 35 items, the students from Grade 6 have reached an average 18.3 correct (SD=7,2); this corresponds to a mean of 52 percent correct.

> In order to take initial differences between treatment and control groups properly into account, the data from *Hauptschule* and *Realschule* were appropriately weighted, because the externally initiated changes in the design had caused some imbalance in the respective proportions: In *Hauptschule*, two treatment classes were matched by four control classes, and in *Realschule*, there were three treatment classes as against four control classes. Due to the relatively large variance components associated with school type (for mathematics achievement: Eta^2_{Pre}=0,28 and Eta^2_{Post}=0,31), these imbalances could not be ignored. Therefore, the respective weights are used where school type provides the frame of reference in subsequent computations. For computations at the class level (which are unaffected by weighting anyway) the real numbers of cases apply.

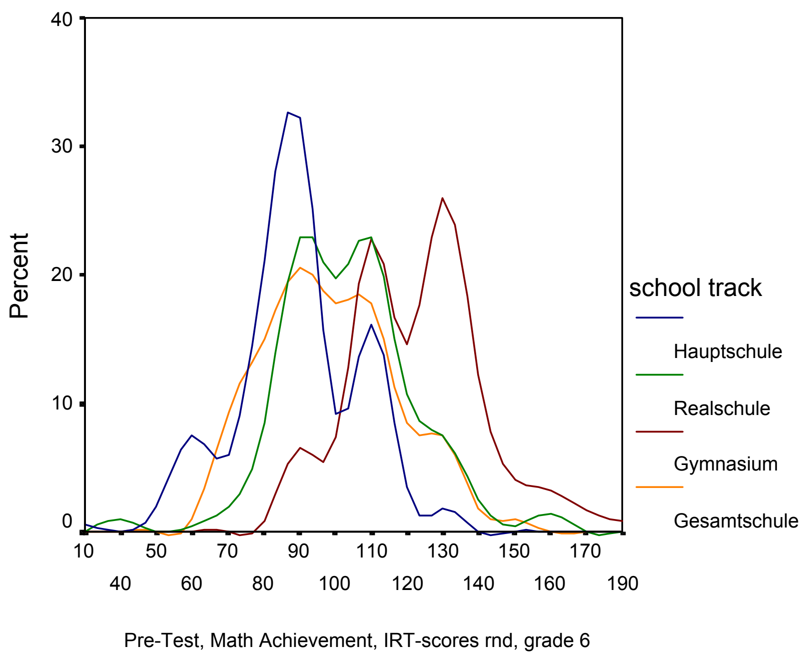
Once again, raw scores were transformed according to the Item Response Theory, one of the advantages being that test results can be associated with definable levels of competency and the respective probabilities of success. Here, the average proficiency score was set to be 103.9 (SD=22,1; N=530). During the initial data collection, there was no notable difference in mathematics achievement between treatment and control classes. However, the multimodal distributions (suggesting rather heterogeneous samples) are remarkable, especially in the control group (see Figure 18).

Figure 18: Initial proficiency in mathematics by experimental status, Grade 6



In accordance with expectations, there were sizeable mathematics achievement differences between school types. Students from *Hauptschule* have attained an average which is about two thirds of a standard deviation below the grand mean for Grade 6; students from *Gesamtschule* have reached a mean which is about one forth of a standard deviation below the grand mean, as opposed to the students from *Gymnasium* whose average was 88 percent of a standard deviation above the grand mean (see Table 12; Figure 19).

Figure 19: Initial proficiency in mathematics by school type, Grade 6



As becomes apparent from Figure 19, all school types are characterized by multimodal proficiency distributions, and this pattern is repeated in nearly all participating classes. Moreover, the large overlapping areas between school types in terms of achievement will be noted.

Table 12: **Pretest results by experimental status, school-type, and class, Grade 6**

status	school track	Class code	Math Achievement Mean	Number of students	Standard deviation
treatment	Hauptschule	64	85,63	16	17,58
		83	90,35	20	18,54
	Realschule	94	101,79	25	14,53
		104	104,45	26	6,80
		113	104,13	24	14,74
	Gymnasium	123	119,35	31	17,11
		133	130,75	30	20,45
	Gesamtschule	144	87,07	26	15,62
		153	108,08	28	17,38
		Total	*105,53*	*226*	*21,39*
control	Hauptschule	63	88,08	17	11,29
		73	87,36	22	12,27
		74	93,37	23	16,74
		84	90,75	21	24,85
	Realschule	92	96,47	27	20,03
		93	87,47	27	15,39
		103	127,58	27	16,14
		114	103,36	25	16,58
	Gymnasium	124	119,26	32	16,76
		134	123,38	30	21,65
	Gesamtschule	143	85,83	25	11,17
		154	110,70	28	15,85
		Total	*102,73*	*304*	*22,55*
Treatment and control	Hauptschule	Total	89,49	119	17,38
	Realschule	Total	103,62	181	19,03
	Gymnasium	Total	123,09	123	19,39
	Gesamtschule	Total	98,46	107	18,98
		Total	*103,92*	*530*	*22,09*

In accordance with expectations, substantial differences are to be noted between classes in terms of their mathematics proficiencies. Thus, one of the classes from *Realschule* has attained an average which is above the mean of three out of the four *Gymnasium* classes (see Table 12, class no. 103). At the same time, class no. 93 (also *Realschule*) has an average which is even below the general average for *Hauptschule*. Some of the differences within *Gesamtschule* are also very noteworthy: Whereas students from school no. 14 have shown a proficiency level which is typical of the *Hauptschule* track, both the treatment and the control classes in the *Gesamtschule* no. 15 have demonstrated achievement levels which were clearly above the *Realschule* average.

Students' self-appraisals based on such statements as „I will find the solution even for difficult math problems." or „When I do math problems, it can happen that forget how time is passing." are condensed into two scales labelled „mathematics self-concept" (see Appendix 2, Tables 1 and 2), the only difference between the two being that one refers to the initial and the other to the final data collection.

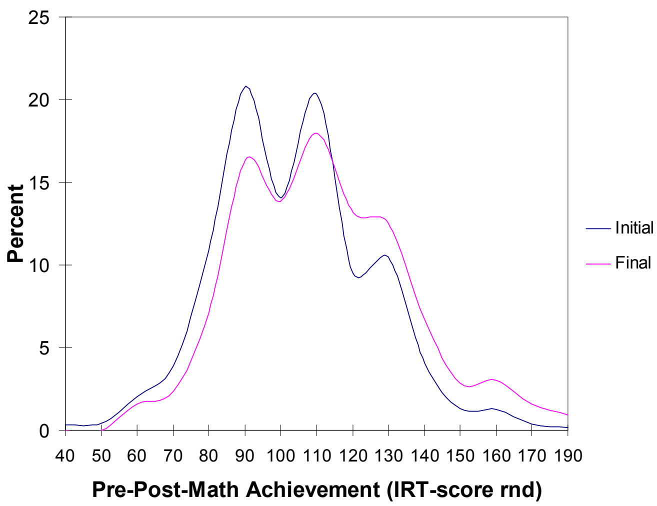
While there were no substantial differences in terms of these self-appraisals in Grades 4 and 5, rather drastic differences are apparent in Grade 6. On the four-point metric ranging from 1 = "disagree completely" to 4 = "agree completely", class means varied between 2.4 (*i.e.,* a level of self-esteem even lower than the scale mid-point) and 3.1 (*i.e.,* almost unmitigated trust in the students' own ability to cope with math problems). The two classes with the lowest and highest means in this dimension are located in *Hauptschulen*. Among *Realschule* and *Gymnasium* classes, class means are generally positive, towards agreement with respect to the statements mentioned above. The class means from *Gesamtschule* were in the range between 2.4 and 2.9; similar to the situation in *Hauptschule*, this school type is characterized by a remarkable level of heterogeneity in this respect.

Correlations between math-related self-concept and pretest achievement were in the rather modest order of r = 0.22 ($p < 0.001$). If differential effects according to school type are explored, it is noteworthy that – with the exception of *Realschule – the* respective within-school type correlations are higher; they range from 0.29 for *Gymnasium* and 0.30 for *Hauptschule* to 0.41 for *Gesamtschule.*

5.2 Growth of mathematics proficiency during the experimental treatment, Grade 6

Figure 20 shows the changes in mathematics proficiency which have occurred between the initial and the final data collection; they correspond to the general increment in terms of mathematical knowledge and ability during the experiment.

***Figure 20*: Distribution of mathematics proficiency at the beginning and end of the experiment, Grade 6**



The increase of mathematics proficiency within four months in the order of 36 percent of a standard deviation is clearly higher than what has been reported in comparable achievement studies (for TIMSS, see Baumert et al., 1997, 144ff.; for the Hamburg longitudinal census, see Lehmann, Gaensfuss & Peek, 1999, 113ff.). In Hamburg, the mean increment begin the beginning of Grade 5 and the beginning of Grade 7 was 0.68 standard deviations or 0.34 standard deviations *per annum.* As opposed to these figures, an increment which was even slightly larger was attained in a considerably shorter period of time. Between school types, slightly divergent tendencies are to be observed. Whereas in *Gymnasium* and *Realschule* an increment of about half a standard deviation could be attained, the respective growth rate in *Hauptschule* was still a remarkable 38 percent of a standard deviation. The gains in *Gesamtschule*, however, were in the order of 17 percent and thus did not exceed the reference findings from other studies cited above, if one takes the duration of the experiment into account.

Since the present study is based on a relatively small sample only, growth should be investigated at the class level. As Table 13 shows, there are large between-class differences, even within identical school types.

Table 13: **Mean mathematics proficiency at the beginning and end of the experiment, by experimental status, school type, and class, Grade 6**

status	school track	Class code	Pretest, Math Ach., IRT-scores	Pre-Stand. Dev.	Posttest, Math Ach., IRT-scores	Post Stand. Dev.	Post-pre-diffe-rence	Effect size	Num. of students (pre and post)
Treatment	Hauptschule	64	85,63	17,34	94,80	15,68	9,17	0,53	13
		83	90,35	18,34	92,66	18,14	2,31	0,13	18
	Realschule	94	101,79	14,48	113,65	12,66	11,86	0,82	23
		104	104,45	6,78	115,97	15,68	11,52	1,70	25
		113	104,13	14,69	109,15	14,22	5,02	0,34	24
	Gymnasium	123	119,35	17,12	133,40	21,59	14,05	0,82	31
		133	130,75	20,45	142,71	21,39	11,96	0,58	30
	Gesamtschule	144	87,07	15,62	87,35	15,76	0,28	0,02	24
		153	108,08	17,38	111,09	15,81	3,01	0,19	27
	Total		*103,75*	*21,27*	*111,48*	*24,11*	*7,73*	*0,36*	*215*
Control	Hauptschule	63	88,08	11,44	93,91	13,47	5,83	0,51	14
		73	87,36	12,39	89,59	14,39	2,23	0,18	20
		74	93,37	16,90	103,82	15,77	10,45	0,62	21
		84	90,75	25,11	103,72	19,05	12,97	0,52	20
	Realschule	92	96,47	20,10	106,96	20,47	10,49	0,52	27
		93	87,47	15,44	102,23	17,16	14,76	0,96	27
		103	127,58	16,20	137,94	17,41	10,36	0,64	27
		114	103,36	16,64	106,00	14,40	2,64	0,16	25
	Gymnasium	124	119,26	16,76	125,63	24,48	6,37	0,38	32
		134	123,38	21,65	133,28	25,39	9,90	0,46	29
	Gesamtschule	143	85,83	11,17	88,60	12,79	2,77	0,25	24
		154	110,70	15,85	116,15	20,90	5,45	0,34	28
	Total		*103,85*	*22,64*	*111,75*	*24,56*	*7,90*	*0,35*	*294*
Treatment and control	Hauptschule	Total	89,12	17,53	95,66	16,96	6,54	0,37	106
	Realschule	Total	103,59	18,21	113,02	18,81	9,43	0,52	178
	Gymnasium	Total	123,09	19,39	133,56	23,78	10,47	0,54	122
	Gesamtschule	Total	98,46	18,98	101,88	20,99	3,42	0,18	103
	Total		*103,80*	*21,94*	*111,61*	*24,31*	*7,81*	*0,36*	*509*

The mean increments, both for treatment and control classes, are in the order of a third of a standard deviation. This is to say that the average instructional effectiveness was equivalent in the two groups. Figure 21 illustrates the parallel growth patterns in treatment and control classes.

***Figure 21*: Proficiency growth between the initial and the final data collection, by initial achievement and experimental status, Grade 6**

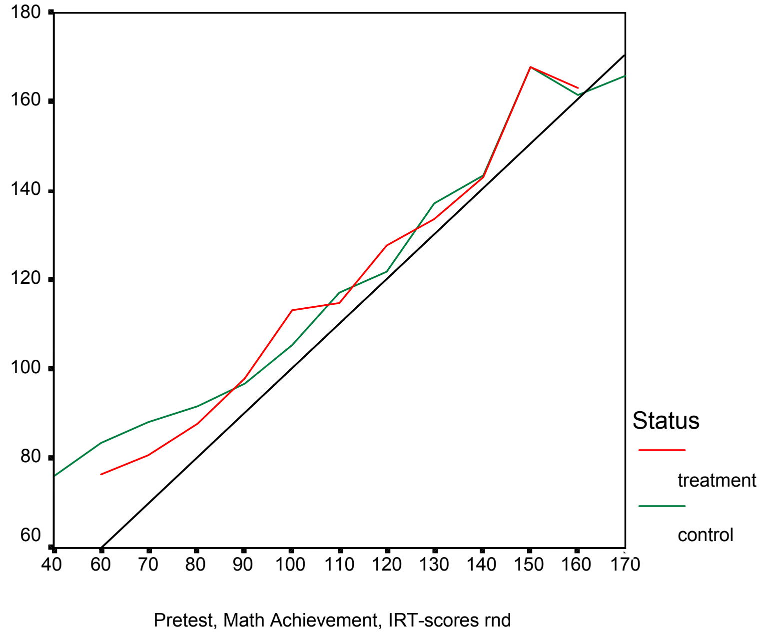
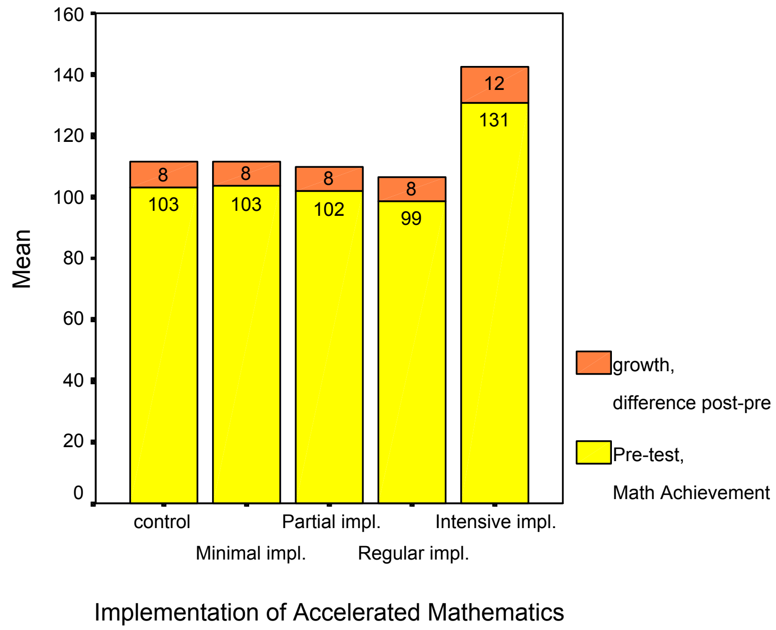


Figure 22 below demonstrates the average proficiency growth in treatment and control classes which are divided into four groups according to the degree to which *Accelerated Math* was effectively implemented (for this distinction of user groups, see Appendix 1 Table 1). This graph which simply represents mean differences can be taken to refer to the average effectiveness of teaching and learning in the respective groups.

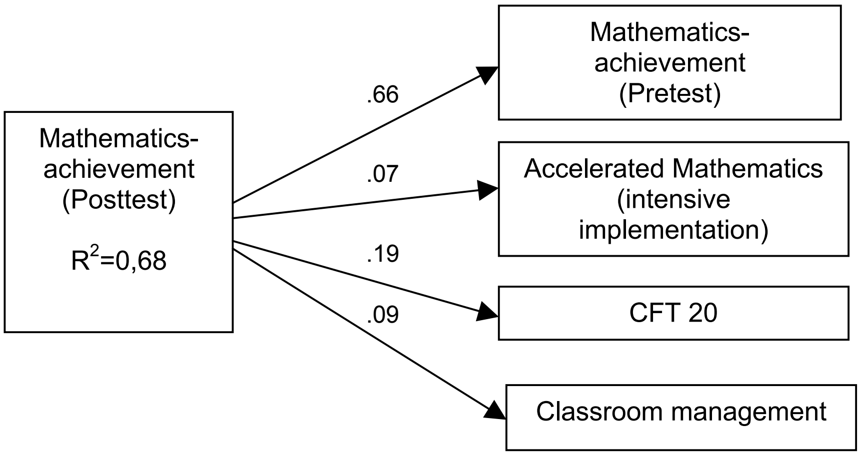
Figure 22: Mathematics proficiency in the initial and the final data collection, by experimental status, depending on the intensity of utilizing *Accelerated Math*, Grade 6



While the control group and classes where *Accelerated Math* was used only superficially expefrienced an average gain of eight proficiency score points, there were greater increments in the group of intensive users. Admittedly, this group of intensive users encompassed no more than one class in either Grade 4 or Grade 5 also. This observation, however, underscores the point that the effectiveness of *Accelerated Math* can be convincingly demonstrated only on the basis of a sufficiently large number of classes where the program is appropriately implemented such that its effects are likely to unfold.

As in the case of Grades 4 and 5, a multiple regression analysis was conducted also for Grade 6, in order to ascertain which variables (predictors) have had any sizeable independent impact on posttest achievement (Figure 23). The independent variables included mathematics proficiency as measured in the pretest, nonverbal reasoning ability (CFT 20 scores), social background factors, math-related self-concept, student perceptions of mathematics instruction, and the intensity of using *Accelerated Math* in class (as an aggregate measure). In analogy to the respective analyses for Grades 4 and 5, minor predictoers are omitted in the graph.

***Figure 23*: Relative importance of several predictors of posttest mathematics achievement (Beta-weights derived from a multiple regression analysis), Grade 6**



As usual, the pretest performance as measured at the beginning of the experiment proved to be by far the strongest predictor, as indicated by a beta weight of 0.68. Nonverbal reasoning abilities constitute the second most important influence. Classroom management (as perceived by the students) and an intensive application of *Accelerated Math* are also predictors that are statistically significant at the 1-percent level.

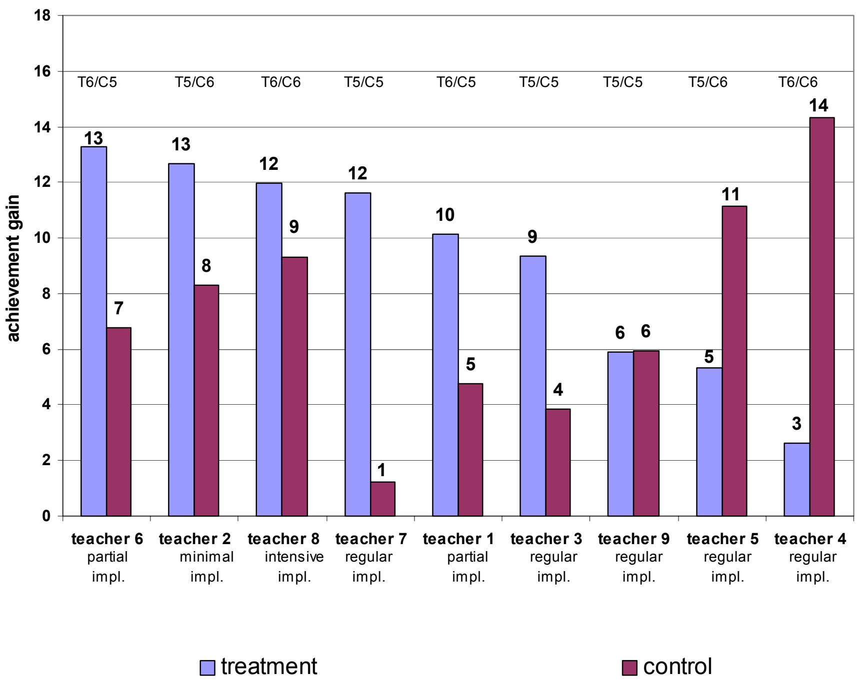
6 Selected case studies: teachers who taught in both experimental and control classrooms

Pertinent studies show that cognitive gains are highly dependent on the learning context and the learning opportunities as arranged by the individual teacher (*cf*. Ditton, 1997). Thus, it may be assumed that in the case of *Accelerated Math* instructional characteristics and teacher preferences are equally associated with decisive impacts on the success of learning mathematics. Since the original experimental design which had called for identical teachers operating at the same class level in *both* a treatment *and* a control class in order to control for teacher personality and didactic style could not be implemented under field conditions, those cases where the respective condition was, indeed, met received special attention. The following case studies are meant to investigate this special case.

There was a total of nine cases where teachers have taught, in separate classrooms, both in accordance with *Accelerated Mathematics* and following their traditional modes of teaching. Four of these nine teachers taught both groups at the same grade level (treatment and control class either both Grade 5 or both Grade 6). The remaining five have taught one of the groups in Grade 5 and the other in Grade 6.

The respective analysis conveys the following results: *Six of the nine teachers have attained higher gains in their treatment class*, in two cases, the control classes appear to have been more successful, and in one case the difference between treatment and control group in terms of mathematics gains is practically nil.

***Figure 24*:** **Differential gains in treatment and control groups with identical teacher**



If, moreover, teacher assessments of *Accelerated Math* are taken into account, it becomes apparent that in six of the nine cases a positive judgement clearly prevails, especially with regard to the diagnostic potential of this program and its inherent opportunities for individualizing instruction. Only one of the nine teachers considered here has expressed a markedly critical and even skeptical position; it is all the more surpriising that this person has used the program rather extensively.

In summary, then, the majority of teachers involved has expressed rather supportive views with respect to using *Accelerated Math* in class. One of the strong points emphasized refers to the fact that students receive a multitude of application problems which provides opportunities for practice and reinforcement. Most teachers held the belief that this approach offers a number of possibilities to implement individualized support fort he students of mathematics. Moreover, *Accelerated Math* appears to imply options for monitoring, at the student level, individual progress. They were also convinced that students displayed a higher level of math-related interest in those classes which were taught according to *Accelerated Math.*

Critical opinions pertained to the supposition that the development of a deeper mathematical understanding be underemphasized within the *Accelerated Math* approach and that the inclusion of more demanding tasks would be desirable. However, opinions are far from homogeneous as to this point which suggests that the teachers judge this point very much from the perspective of their own work context (school type, average proficiency level in their class).

If asked whether or not these teachers would prefer to return to their old teaching methods, their responses are very clearly in favor of *Accelerated Math*. A large majority of 13 has expressed unambiguous support fort his new method, while only seven would rather abandon such fresh approach and return to conventional teaching techniques. It may be interesting to note that teachers who work with relatively slow learners are tendentially more open towards *Accelerated Math* and perceive its advantages more acutely than teachers in stronger groups of learners who hold a slightly more critical view.

7 Executive Summary

The *Accelerated Math* program, developed by Renaissance Learning Inc., was implemented, on an experimental basis, early in 2004 in 14 schools in the German State of Northrhine-Westphalia. The experiment involved 5 classrooms from Grade 4 as well as 9 classrooms each from Grades 5 and 6 where four different school types were to be distinguished. The experimental classrooms were matched by an approximately equal number of control classrooms.

A careful evaluation of the instruction based on the *Accelerated Math* program demonstrated the effectiveness of this approach in terms of growth over a period of approximately four months.

The respective study involved, among others, the following measures:

- Pre- and post-test mathematics scores (IRT scores), based on standardized achievement tests;
- nonverbal intelligence (Cattell's CFT 20);
- students' ratings of classroom instruction;
- a measure on the intensity of using *Accelerated Math* in class.

Analyses included

- comparisons of pre- and post-test scores in experimental and control classrooms, distinguished by grade and school type;
- comparisons of achievement levels and growth to reference scores from regional assessments conducted in the Cities of Berlin and Hamburg;
- comparisons of pre- and post-test scores in experimental and control classrooms, distinguished by grade and school type;
- analyses of gains as a function of the intensity of program use;
- multiple regression analyses, controlling post-test achievement for pretest scores, nonverbal intelligence, students' perceptions of classroom instruction, and program implementation.

Key findings of the present study are:

1. Schools in which the *Accelerated Math* program was accepted as an experimental approach had a positively select group of students. Their average levels of nonverbal intelligence and initial mathematics achievement compared very favorably with reference values from other sources. This was particularly true for the control group in Grade 4.
2. Achievement gains were unexpectedly high *both in experimental and control classrooms*. Apart from the unusually favorable conditions in terms of learning potential (intelligence) and initial achievement level, it is possible that the existence of an innovative approach to instruction is conducive to the emergence of increased levels of achievement orientation among teachers.
3. In all grades investigated, classrooms could be identified in which the *Accelerated Math* program was used intensively, accompanied by rather unusual gains in mathematics achievement. By way of contrast, gains accompanying inadequate levels of implementing *Accelerated Math* were not systematically superior to those encountered in control classes.
4. Originally, it was envisaged to compare program-based mathematics instruction to regular teaching without support by the *Accelerated Math* program, inviting the same teachers to serve in the experimental and the control group. While this design proved

impracticable on the whole, eight teachers finally did teach both in a control and an experimental group. In six out of these eight cases, where – by fortuitous circumstance – the ‘personal factor’ could be held constant, achievement gains in the experimental class were higher than those in the control group.

5. Generally speaking, teachers and students accept *Accelerated Math* as an interesting new approach to teaching and learning mathematics, more or less independent of grade. In terms of school type, acceptance is highest at the levels of *Gymnasium* and *Realschule* where a general achievement orientation may be assumed to be best developed. It may be interesting to note that the most acclaimed advantage is seen to be that of *Accelerated Math* supporting self-regulated, highly individual learning.

6. Irrespective of these commendable traits of the evaluated program, teachers feel that improvements are desirable in terms of better adapting the program libraries to the German educational context, most notably to the currently valid curricula.

References

Baumert, J., Köller, O., Lehrke, M. & Brockmann, J. (2000). Anlage und Druchführung der Dritten Internationalen Mathematik- und Naturwissenschaftsstudie zur Sekundarstufe II (TIMSS/III) – Technische Grundlagen. In Baumert, J., Bos, W. & Lehmann, R. H. (Hrsg.), *Dritte Internationalen Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Opladen: Leske und Budrich, 31-84.

Bos, W. & Pietsch, M. (2004). Erste Ergebnisse aus KESS 4 – Kurzbericht. http://www.erzwiss.uni-hanburg.de/Kess/kurzbericht.pdf.

Cattel, R. B. (1960). Culture Fair Intelligence Test, Scale 2 (Handbuch). 3. Auflage, III., IPAT.

Fischer, G.H. & Molenaar, I. W. (1995). *Rasch models – Foundations, recent developments, and applications.* New York: Springer.

Ditton, H. (1997). Wirkung und Wirksamkeit der Einzelschule – Analysen der Bedeutsamkeit der verschiedenen Ebenen des Schulsystems. In Lehmann, R. H., Venter, G., Buer, J. van, Seeber, S. & Peek, R. (Hrsg.), Erweiterte Autonomie für Schule – Bildungscontrolling und Evaluation. Studien zur Wirtschafts- und Erwachsenenpädagogik aus der Humboldt-Universität zu Berlin. Band 13.2. Abschlussband zur gleichnamigen Sommerakademie vom 31. August bis 6. September 1997 in Ungarn. Nyȋregyháza-Berlin.

Ditton, H. (2001). Skalendokumentationen zum DFG-Projekt „Bildungsqualität von Schule und Unterricht“. www.quassu.net/seite4.html.

Lehmann, R. H., Peek, R. & Gaensfuss, R. (1997). *Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen.* Behörde für Schule, Jugend und Berufsbildung, Amt für Schule, Hamburg.

Lehmann, R. H., Gaensfuss, R. & Peek, R. (1999). *Aspekte der Lernausgangslage und der Lernentwicklung. Ergebnisse der Längsschnittstudie für die Klassenstufe 7*. Behörde für Schule, Jugend und Berufsbildung, Amt für Schule, Hamburg.

Lehmann, R. H., Peek, R., Gaensfuss, R. & Hussfeld, V. (2001). *Aspekte der Lernausgangslage und der Lernentwicklung. Ergebnisse der Längsschnittstudie für die Klassenstufe 9*. Behörde für Schule, Jugend und Berufsbildung, Amt für Schule, Hamburg.

Lehmann, R. H., & Nikolova, R. (2004). Erhebung zum Lese- und Mathematikverständnis, Entwicklungen in den Jahrgangsstufen 4 bis 6 in Berlin. http://www. Senbjs.berlin.de/ bildung/qualitaetssicherung/element_ersterbericht.pdf.

Weiss, R. H. (1998). Grundintelligenztest Skala 2, CFT 20. Handanweisung. Göttingen: Hogrefe.