

**C.M.Goethe Middle School:  
An Evidence-Based\* Middle School Model  
Evaluation for First Year of Implementation  
November 1998**

The Evidence-Based Model at C.M. Goethe Middle School uses interventions that have proven effective. Evidence-based\* approaches are not experimental. Through close partnership with the University of Oregon and the Sacramento County Office of Education, the Goethe staff have had access to the latest research-based knowledge for reaching their goals.

Overview:

[Goethe Middle School Demographics and Background](#)

[The Instructional Treatment](#)

[The Measures](#)

[Results](#)

**C.M. Goethe Middle School Demographics and Background**

Table 1 displays the demographic statistics for C.M. Goethe Middle School. Goethe had the reputation of being one of the poorest performing schools in the district.

Table 1. C.M.Goethe Middle School Demographics

Economic Level Indicators	Percentage of students living with families receiving Aid for Dependent Children	58%
	Percentage of students receiving free or reduced lunch	61%
Cultural groups	Asian—Hmong	24%
	Asian—Other	16%
	African-American	27%
	Hispanic	23%
	Caucasian—English Language Learners	2%
	Caucasian—English-speaking	9%
English-language proficiency	Emergent English	8%
	Primary	16%
	Advanced	14%
	Nearly Fluent	9%
	Native Speaker	53%

---

\*See webpage of the National Center to Improve the Tools of Educators for Handbook for Creating a “Smart” School, a school that uses evidence to make decisions.

<http://darkwing.uoregon.edu/~ncite/smart.htm>

**The Instructional Treatment**

Three major components of the intervention were (a) curricular materials that included detailed lesson plans, (b) in-class coaching, and (c) monthly progress monitoring. Direct Instruction curricular materials were chosen because of their extensive research base. The programs used in Year1 are listed in Table 2 below.

Table 2. Curricular materials used in the Goethe Model during the 1997-98 school year.

Class Title	Programs Used	Completion	Size of Group	Grouping
Reading Level A	Corrective Reading—Decoding Level A*	To Lesson 30 out of 65 lessons	5 students	All 7 <sup>th</sup> and 8 <sup>th</sup> grade students grouped together according to placement test performance.
Reading Level B1	Corrective Reading—Decoding Level B1*	To end (lesson 60)	12 students	
Reading Level B2	Corrective Reading—Decoding Level B2*	To end (lesson 65)	15-18 students	
Reading Level C	Corrective Reading—Decoding Level C*	To Lesson 45-70 out of 125 lessons	25-28 students	
Reading Level D	Expressive Writing II*	To end (lesson 45)	30-33 students	
Reading Level E	Reasoning & Writing E*	To end (lesson 80)	33-35 students	
Reading Level F	Reasoning & Writing Level F*	To end (lesson 80)	33-35 students	
History	Understanding U.S. History**	Chapters 1-7	20 (low level ELL classes) -35 students	8 <sup>th</sup> grade students only
Mathematics Level C	Problem Solving with Addition and Subtraction; with Multiplication and Division (videodisc program)***	To lesson 37 in Multiplication and Division in one class; to end (lesson 50) in other	25-30 students	All 7 <sup>th</sup> and 8 <sup>th</sup> grade students grouped by grade, by ELL or English-Speaking, and by placement test level.
Mathematics Level D	Connecting Math Concepts Level D*	To lesson 50 or to end (lesson 120)	33-35 students	
Mathematics Level E	Connecting Math Concepts Level E*	To lesson 105 or to end (lesson 125)	33-35 students	
Mathematics Level F	Connecting Math Concepts Bridge to F and Level F*	To end	30 students	

\*published by Science Research Associates

\*\*published by University of Oregon

\*\*\*published by BFA Phoenix Film

The school electives were dropped and reading became a required subject for all students. All teachers in the school taught reading during one period of the day, the last period. The language arts teachers taught various levels of *Corrective Reading*. The math and science teachers taught the *Reasoning and Writing* program. The other programs listed in Table 2 were taught during the relevant subject area period.

**In-class coaching.** The in-class coaching followed 2 days of workshop training in each of the programs. Each teacher received an average of 4 to 5 coaching sessions over the course of the year. In most cases coaching began with a demonstration lesson. During the second session the coach monitored class behavior

allowing the teacher to focus his/her efforts on becoming fluent with the teaching procedures suggested for use with the instructional program. In a third session, the coach would observe the teacher carrying on both functions—teaching the lesson and monitoring student behavior.

Teachers reported monthly each student's performance on in-program tests. Summary reports were developed that guided the coaching and additional interventions for individual student problems, such as attendance problems.

### Measures

Three standardized, norm-referenced tests were administered to students of C.M. Goethe Middle School during the first year of implementation (see Table 3). Two tests were multi-level tests, meaning that students were compared with several norm groups of different ages through the administration of one test. One of these multi-level tests, Woodcock-Johnson Word Attack Subscale, required the examiner to listen to each individual student read a list of syllables. This test was administered only to the lowest performing students. A third standardized test, the SAT-9, was first administered only in the spring as part of California's statewide assessment program.

Table 3. Standardized norm-referenced tests administered at Goethe MS.

Name of Test	Type of Norms	Pretest	Posttest	Sample size
Multi-Level Academic Survey Test in reading comprehension and math	Multi-level	First week of September, 1997	Last week of May, 1998	Whole school, Number taking both pre and post=528.
Woodcock-Johnson Word Attack Subscale in word reading	Multi-level	First week of October, 1997	Third week of May, 1998	Readers scoring in the 1 <sup>st</sup> to 20 <sup>th</sup> percentile range. N= 184
SAT-9 in all subjects	Single level	No pretest scores	April, 1998	Whole school, posttest only, N=711

### Results

According to the Multi-Level Academic Survey Test, the middle student at Goethe improved by 2 grade levels during the first year both in reading comprehension and in mathematics. The middle student in reading moved from 4<sup>th</sup> grade level to 6<sup>th</sup> grade level, and the middle student in math moved from 5<sup>th</sup> grade level to 7<sup>th</sup> grade level. No differences were found between 7<sup>th</sup> and 8<sup>th</sup> grade students; therefore, scores across both grades were aggregated for the analysis. Figure 1 shows how students at Goethe moved across the grade levels in reading and Figure 2 shows how students moved in mathematics.

<b>Reading</b>	<b>Pretest</b>	<b>Posttest</b>
<b>Mathematics</b>	Begin 4 <sup>th</sup> grade level	Begin 6 <sup>th</sup> grade level
	Begin 5 <sup>th</sup> grade level	Begin 7 <sup>th</sup> grade level

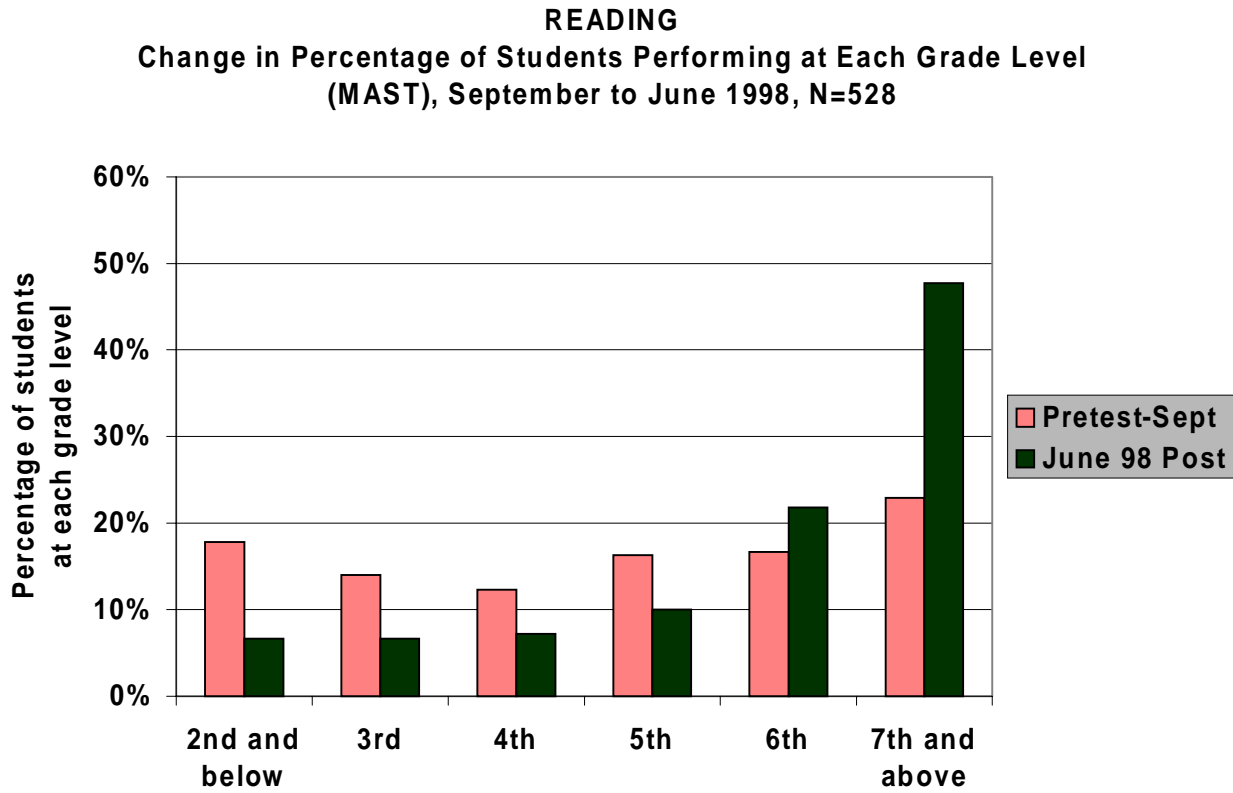


Figure 1. Change in Reading Performance as indicated by the MAST.

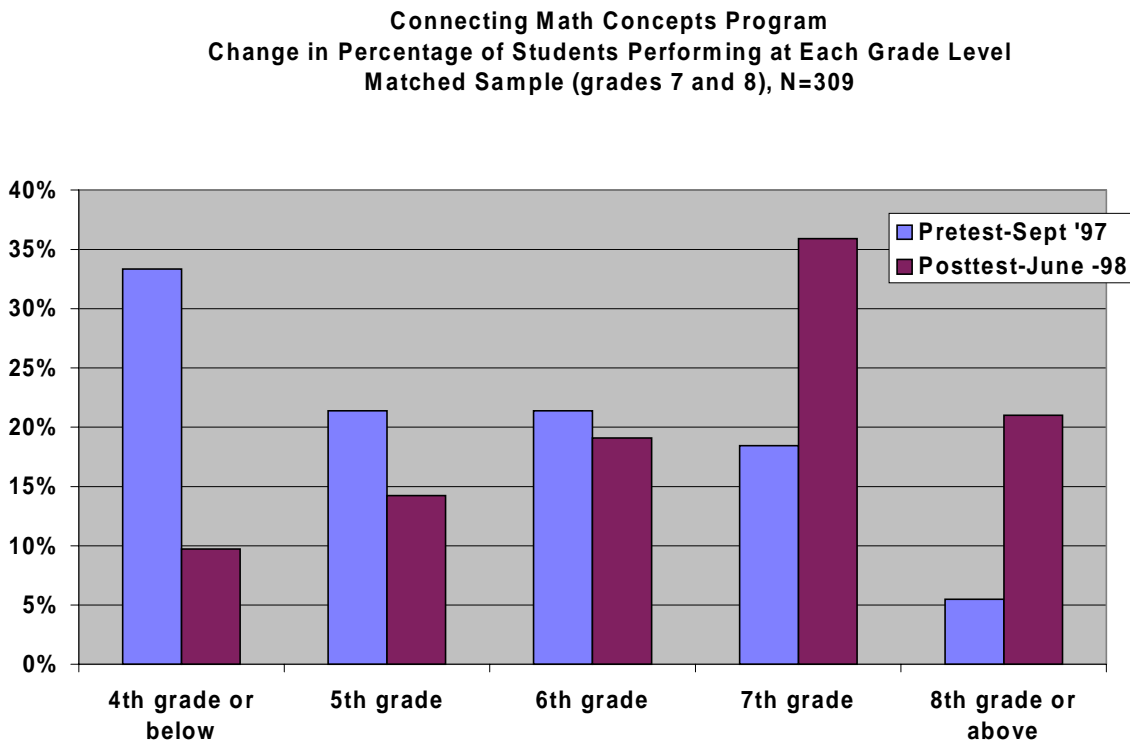
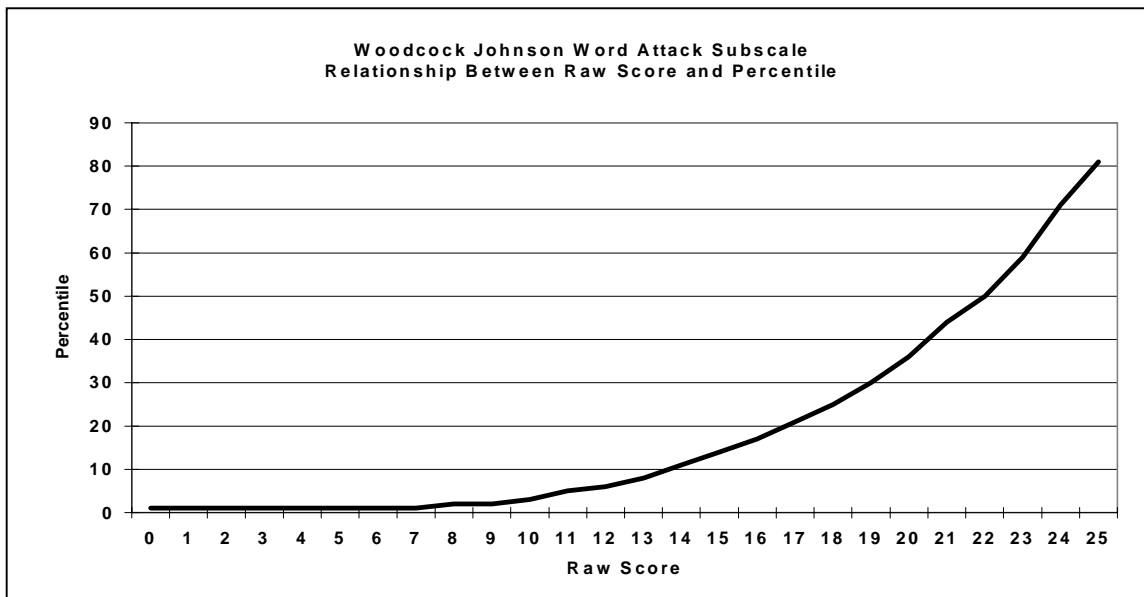


Figure 2. Change in mathematics performance as indicated by the MAST.

A change in percentile scores cannot be used as an indicator of growth. This is illustrated in Figure 3. Figure 3 shows the relationship between each raw score and the corresponding percentile score for the Woodcock-Johnson Word Attack Subscale. A low achiever that gets a raw score of 2 will score in the 1<sup>st</sup> percentile and a higher achiever that gets a raw score of 18 will score at the 25<sup>th</sup> percentile. Let’s say that after a year both students learn a lot and on the posttest, both students increase their raw score by 6. The low achiever can read 8 words now, a 400% gain in raw score. This reader moved essentially from being a nonreader to a reader. The high achiever also reads 6 more words, making only a 33% gain in raw score. However, the low achiever’s percentile score is still only in the 1<sup>st</sup> percentile, but the higher achiever moved from the 25<sup>th</sup> percentile to now the 70<sup>th</sup> percentile. The change in percentile, especially for older students, does not represent the amount of growth in performance.

Percentile scores indicate how a student compares to the norm: A score at the 50<sup>th</sup> percentile means that the student performs exactly in the middle of the norm group; a score at the 20<sup>th</sup> percentile means 20 percent of the group of students the same age performed below this score; a score at the 1<sup>st</sup> percentile means that only 1 out of 100 students from the norm group performed below this score. A multilevel norm-referenced test includes items that range from very easy to very difficult and allows comparisons with groups of students at different ages, as displayed in Figures 1 and 2 above. A single level norm-referenced test includes items of the same level of difficulty and allows comparisons only with students the same age (the normal level of performance for students that age, or “the norm”).

Figure 3. Relationship between raw score and percentile score on a multi-level test.



An important difference between a multi-level test and a single level test is that the multi-level test can be used as a measure of growth (as in Figures 1 and 2), by comparing a student’s performance with different age groups whereas a single-level test cannot. Both can be used to show how the student compares to his or her age-appropriate norm group.

Another way to use a multi-level test to show progress for a low-achieving school is to pick a lower standard to see to what extent student performance rose to meet that standard. We chose end of 6th grade level performance as a lower standard that is meaningful because it represents accomplishment in basic academic foundation skills necessary for later academic success. Sixth-grade level is often used to designate functional literacy. Figure 4 shows how student performance in relation to this standard changed from the pretest to the posttest for reading. Figure 5 shows the change from pretest to posttest in mathematics. In both cases we can see that illiteracy and innumeracy (represented by the portion showing “below basic standard”) have been cut in half in our first year of implementation.

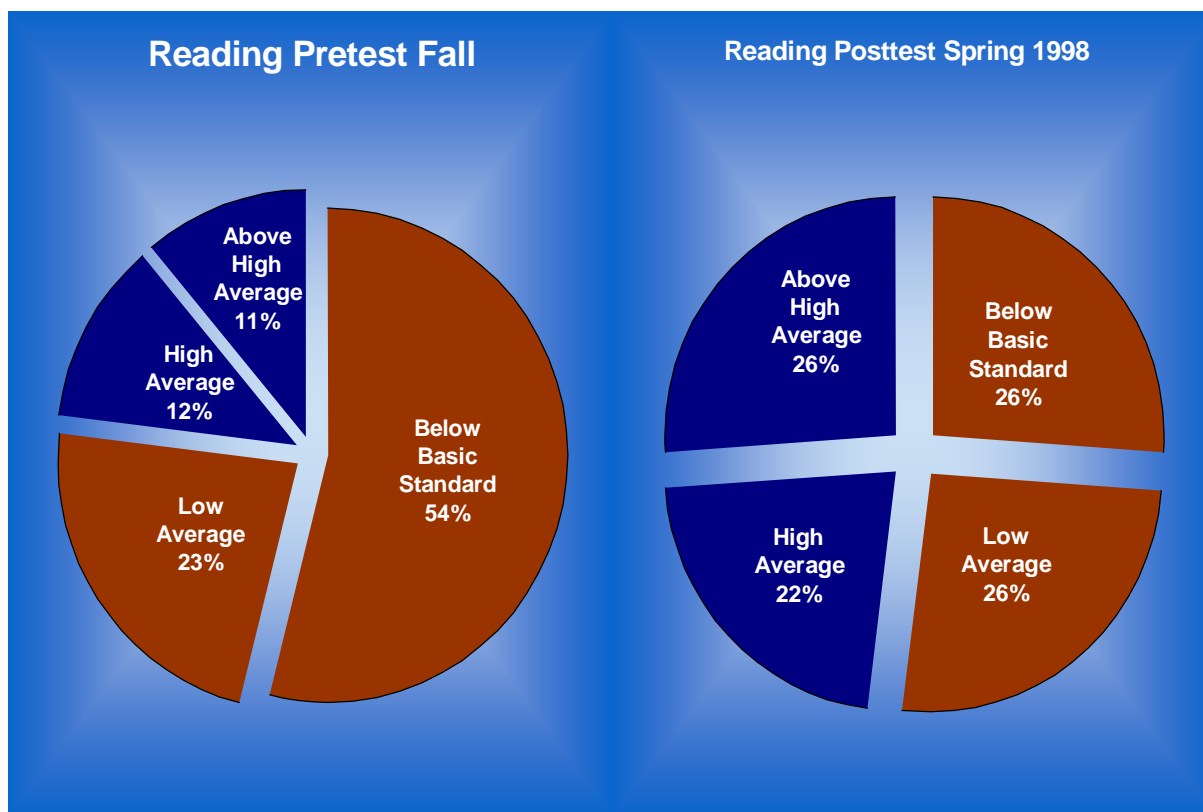


Figure 4. Change in reading performance from pretest to posttest compared to a standard of sixth-grade performance. (Each category represents a quartile in the distribution using 6<sup>th</sup> grade norms.)

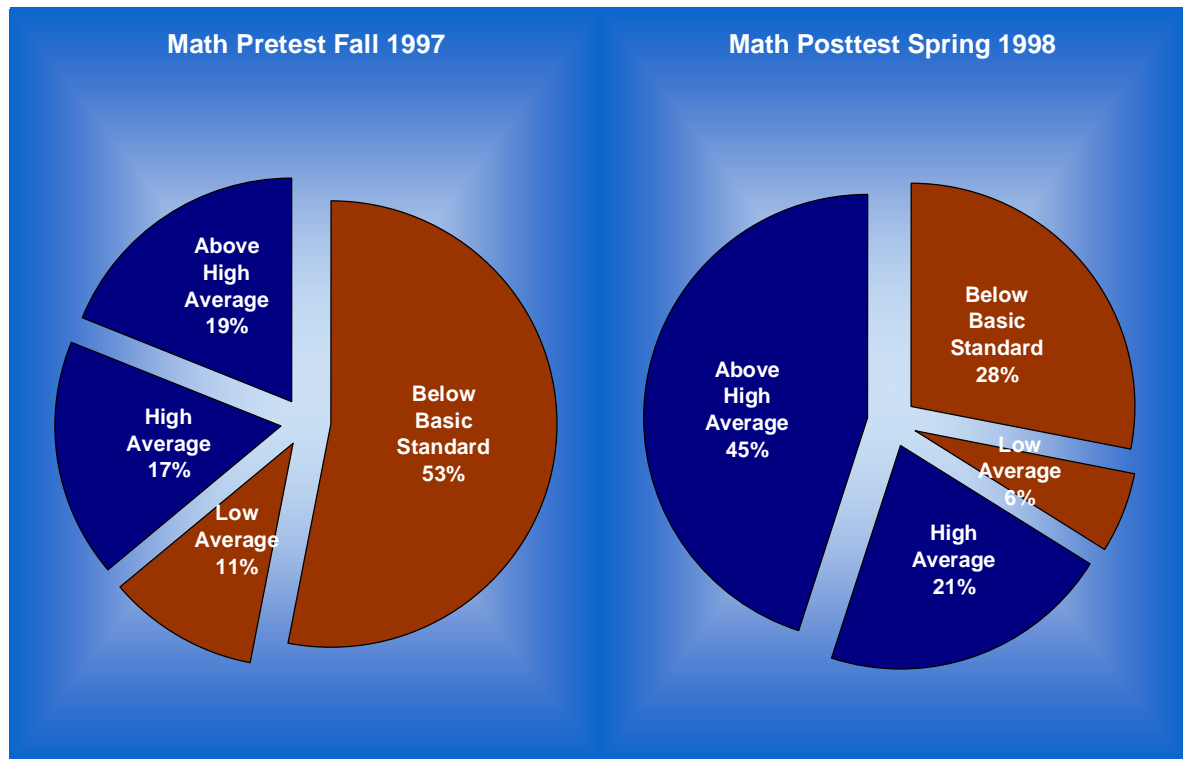


Figure 5. Change in mathematics performance from pretest to posttest compared to a 6<sup>th</sup> grade standard. (Each category represents a quartile in the distribution using 6<sup>th</sup> grade norms.)

### Growth of the Lowest Performing Students

All the tests except the Woodcock Johnson (WJ) were administered in a group setting with a multiple-choice response form. The WJ is administered one-on-one, the student reading words and syllables to the examiner. The same examiner administered the pre- and posttest to all the students, increasing the reliability of this test. Because very low-achieving children often do not do well on these group-administered tests, we administered the WJ to the lowest students to identify students performing in the 1<sup>st</sup> to 20<sup>th</sup> percentile range on the pretest. 30% of the students at Goethe scored in this low range where only 20% of the norm group scored. Much more surprising though was to find that 18% of the students at Goethe scored in the first percentile, where only 1% of the norm group would score. The first percentile represents performance below a second grade level. To move out of the first percentile a student would have to read 8 words correctly on this test. For the many students at Goethe who were virtually non-readers, to move from reading zero, one, or two words on the pretest to reading more than 8 words correctly was change of great magnitude. These students moved from being non-readers to being readers. Yet the percentile score based on the normal performance of students their age leaves them with a percentile score of only 1.2. Figure 6 shows that of the 39 students who read zero to three syllables correctly on the pretest, only 11 students remained in this very low range.

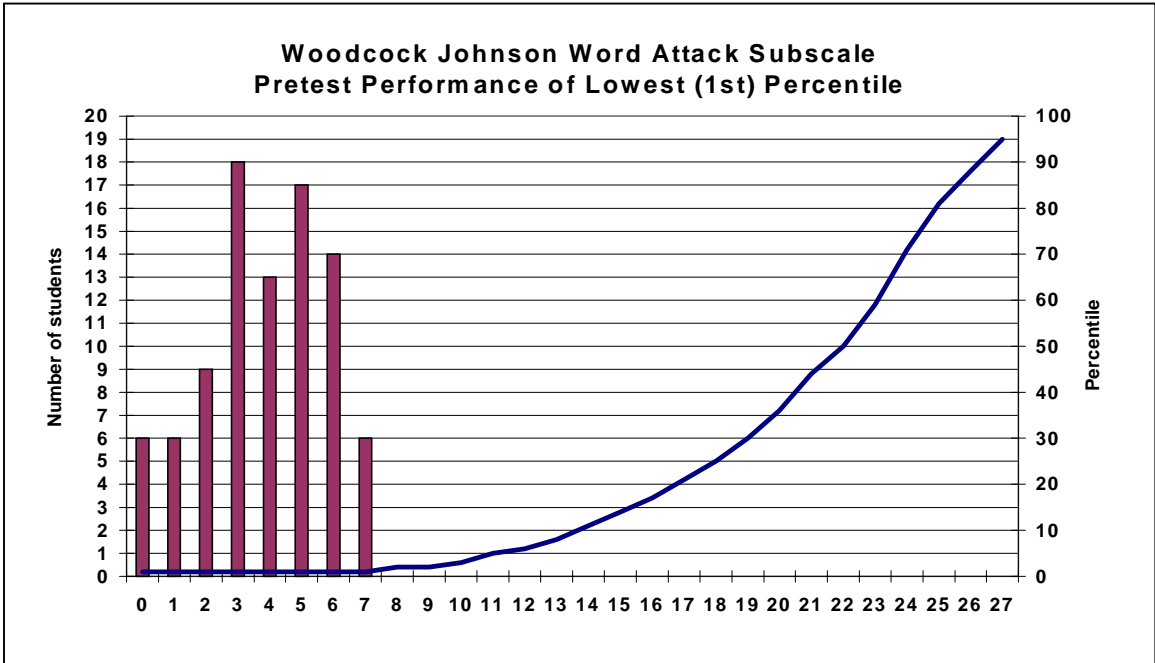


Figure 6. Number of students achieving each raw score point within the first percentile range of the Woodcock-Johnson Word Attack Subscale.

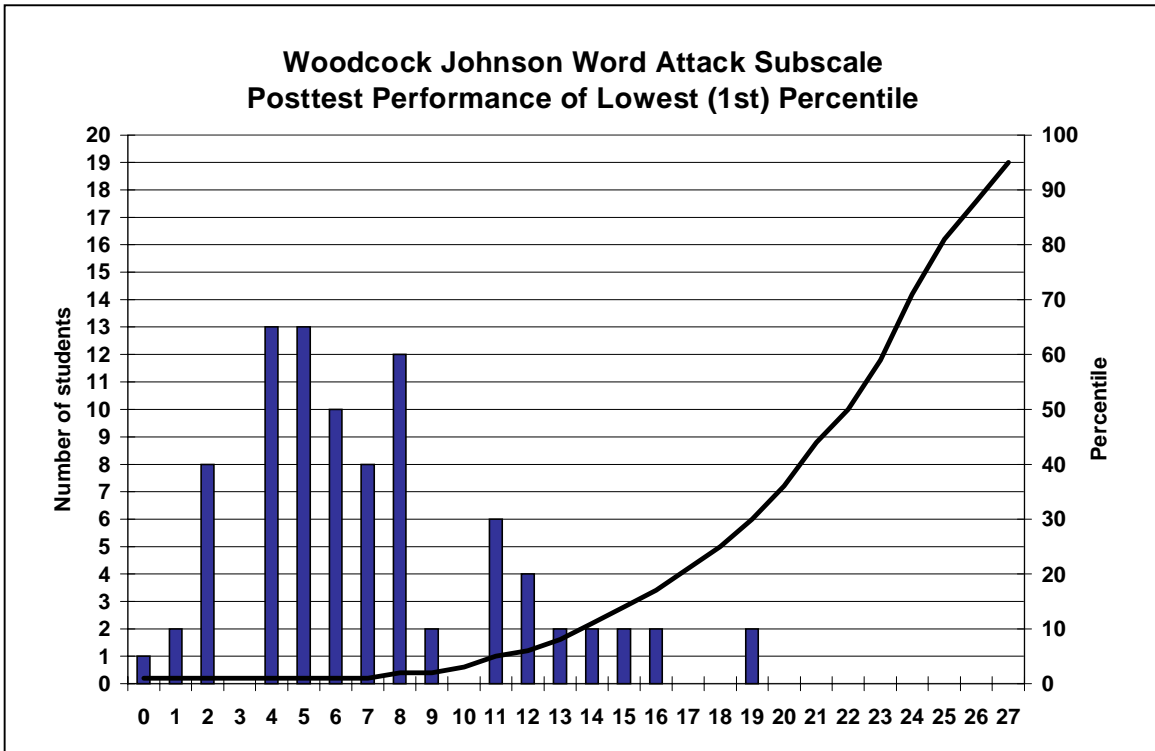


Figure 7. Posttest performance of students performing within the first percentile.



The drop from 39 to 11 students reading 0 to 3 syllables indicates that student performance was lifted up off the bottom. Of those students scoring in the 1<sup>st</sup> percentile on the pretest, Figure 8 below shows that the growth of students identified with disabilities was comparable to that of nonidentified students.

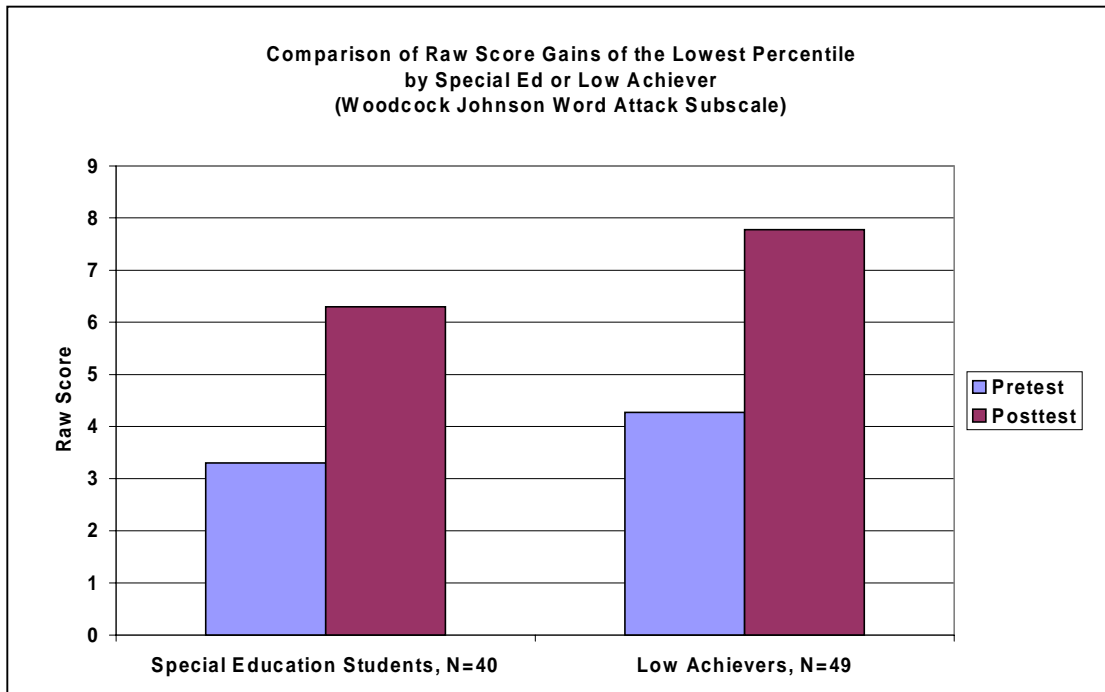


Figure 8. Comparison of growth of students with disabilities and low achievers, all performing within the 1<sup>st</sup> percentile.

### English Language Learners

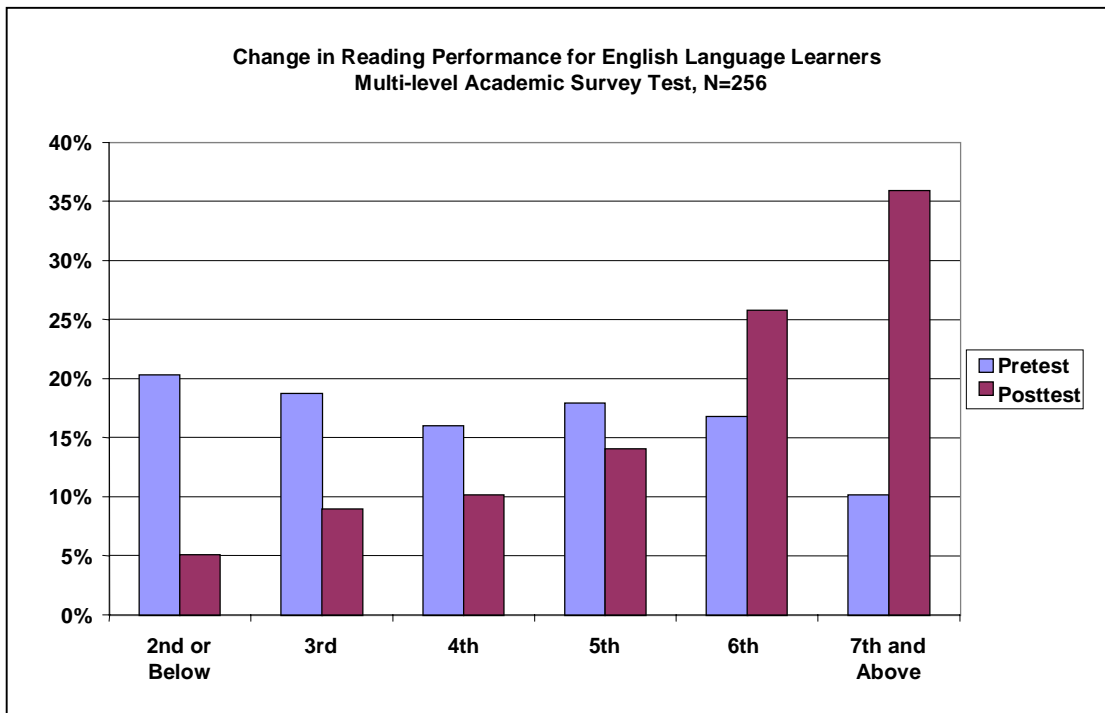


Figure 9. Percentage of ELL students scoring in each grade level range on the pre- and posttest.

### Other Results From the Multi-Level Academic Survey Test

#### Gifted and Talented

Reading		
	Mean Pretest Score	Mean Posttest Score
7 <sup>th</sup> Grade	35% ile	86% ile
8 <sup>th</sup> Grade	36% ile	85% ile

#### By Ethnic Groups

Mean Grade Equivalent Scores in Reading				
	N	Pre	Post Extended	Change
Asian	210	4.6	5.9	+1.3
Black	142	4.6	5.8	+1.2
Latino	120	3.9	5.5	+1.6
White	48	5.2	7.3	+2.1
Native American	4	3.2	4.9	+1.7

Mean Grade Equivalent Scores in Math				
	N	Pre	PostExtended	Change
Asian	111	5.7	7.4	+1.7
Black	88	5.3	6.1	+8
Latino	79	5.1	6.3	+1.2
White	21	5.3	6.0	+7

In reading, all ethnic groups made substantial progress. The mathematics instruction had more differential effects by ethnic group. Perhaps the different grouping arrangements used for math and reading can explain these differences. For reading instruction students were mixed ethnically and grouped according to specific reading needs. For mathematics instruction, students were grouped according to their English language level, with native speakers being grouped separately from English language learners who were taught in smaller groups (20). In effect, this meant the African-American and white students were grouped separately from most of the Asian and Latino students and taught in larger groups of 30-35. For year 2, students were grouped for math according to their skill needs regardless of their language level or disability.

#### C.M. Goethe Teachers Accelerated Growth at All Levels

##### Change in Mean Grade Level Equivalent in Math Computation and Problem Solving

	N	Pretest	Posttest	Change in Grade Level
Level D	104	5.6	7.0	+1.4
Level E	88	5.6	7.4	+1.8
Level F	25	7.0	9.0	+2.0

## Change in Mean Grade Level Equivalent in Reading Comprehension

	N	Pretest	Posttest	Change in Grade Level
Level B2	120	2.7	4.2	+1.5
Level C	150	4.0	5.7	+1.7
Level D	80	4.8	5.8	+1.0
Level E	230	5.5	7.0	+1.5
Level F	85	7.0	9.0	+2.0

Students in all the different skill levels, as identified by their placement in the different levels of instructional programs, generally made significantly better progress than normal (i.e., more than 1 year's growth). The "level D" group in reading made only average growth. However, average growth is an increase over these students' previous rate of learning, which was only about ½ the average rate of growth.

The students in level D were students who could decode well enough not to place in the decoding programs below level D, but did not comprehend well enough to place in *Reasoning and Writing* levels E or F. The level D instructional groups used *Expressive Writing II*, which may not have aligned as closely with the reading comprehension skills measured by the test as did the other programs.

**SAT-9 Scores**

California's statewide assessment, as of Spring 1998, is the SAT-9. The SAT-9 makes possible comparisons with other schools on standardized measures. On this measure, Goethe's 8<sup>th</sup> graders' mean percentile score in reading (24) was in the bottom quartile along with that of two other middle schools with mean scores of 22 and 23. In math Goethe's 8<sup>th</sup> grade students ranked second from the bottom with a mean score of 25.

Unfortunately, with only one testing occasion, it is impossible to know if these other schools provide a fair comparison. In other words, it is impossible to know how the performance of Goethe's students compared with other schools at the beginning of the year, prior to the intervention. The communities these schools serve differ greatly in economic level and in the achievement levels of the students leaving the feeder elementary schools. Goethe has had the reputation of serving the lowest achieving population in the district.

Another comparison group is provided by the entering 7<sup>th</sup> grade cohort. If we assume that their performance is representative of student performance before coming to Goethe, we can compare their scores with those of students completing one year at Goethe. The district's evaluation department reported these scores as shown below. By definition, students making 1 year's growth in achievement in one year's time would show a zero increase in percentile score from one year to the next. The higher percentile scores for the students who received one year of instruction at Goethe would indicate that Goethe's students grew at a faster than normal rate during the first year of implementation. (The district-reported math scores are not included because schoolwide data included the students who remained in the former program.)

## Percent of students at or above grade level

	Reading	Language
Goethe 7 <sup>th</sup> grade's 1998 scores (before 1 year at Goethe)	15%	22%
Goethe's 8 <sup>th</sup> grade's 1998 scores (after 1 year at Goethe)	21%	30%

**Mathematics**

The new mathematics program (*Connecting Math Concepts, SRA*) was implemented only with students who were not placed in Prealgebra or Algebra. To evaluate each program and each level of each program, we analyzed the data by program group, using the MAST for a pretest score.

1997-1998 C.M. Goethe Math Scores

Level	N	Pretest MAST Mean Percentile Score	Posttest SAT-9 Mean Percentile Score
Level C (Videodisc Math)	62 (9%)	3	25
Level D CMC	234 (36%)	8	28
Level E CMC	174 (27%)	22	36
Level F CMC	16 (2%)	50	36.5
Prealgebra	130 (20%)	50	36
Algebra	53 (8%)	75	28

**Other Questions**

**Which better prepares students for Algebra—Level F CMC or the district’s Prealgebra course.** The measure used to evaluate student preparation for algebra was the algebra placement test, a test that had been used for this purpose for several years and was not designed to align particularly with the Level F CMC program. Part 1 of the test is test of prerequisite skills for algebra. Part 2 is a test of very basic algebraic concepts. Students in Level F CMC scored significantly higher on Part 1 of Algebra Placement test ( $t,52=2.2$ ); differences on Part 2 were not significant ( $t,52=1.32$ ). Students were better prepared for Algebra by taking Level F CMC than by taking the district’s Prealgebra course.

	Pretest		Posttests	
	N	Mean (SD)	Prerequisites for Algebra Mean (SD)	Algebra Mean (SD)
Level F CMC	19	15.1 (3.2)	19 (5.5)*	7.1 (4.2)
Prealgebra Course	34	14.6 (2.6)	15.1 (6.6)*	8.5 (3.3)

\*Difference was significant at  $p<.05$  level on 2-tailed test.

**Attitude Surveys**

**Results of Student Attitude Survey June 98—Reading**

- 1. How many times each week did your teacher give you grades for your written work (including in-class work, independent work, and tests)? 2.5 times per week
- 2. How long did you usually wait for your papers to be graded? 2.4  
(where 2 = one day and 3=one week)
- 3. Do you think you have learned a lot in this class? 91% yes
- 4. Do you feel your reading or writing improved more this year than in previous years? 85% yes
- 5. Do you enjoy reading or writing more than you did last year? 55% yes
- 6. Do you feel you have been a better student this year than you were last year? 66% yes
- 7. Are you looking forward to learning a lot more next year? 80% yes

**Results of Student Attitude Survey June 98—Mathematics**

- |  |                    |
|--|--------------------|
| 1. How many times each week did your teacher give you grades for your written work (including in-class work, independent work, and tests)? | 2.6 times per week |
| 2. How long did you usually wait for your papers to be graded?<br>(where 2 = one day and 3=one week)                                       | 2.4                |
| 3. Do you think you have learned a lot in this class?  | 90% yes            |
| 4. Do you feel you can solve math problems better this year than in previous years?  | 92% yes            |
| 5. Do you enjoy your math class more than you did last year?   | 59% yes            |
| 6. Do you feel you have been a better student this year than you were last year?   | 75% yes            |
| 7. Would you like to take more math classes in high school?  | 58% yes            |

**Some responses to Questions 8 and 9 for Reading**

<b>Reading Q.8: What did you like most about your reading or writing class?</b>	<b>Reading Q.9: What did you like the least about your reading or writing class?</b>
---	--

**Some responses from students who did NOT find that they learned more this year:**

Easy	Paragraphs, took more than a day for each lessons
Read stories	Checkout
The pace, class, teacher	Stories, questions, work attack, boring lectures
Learning, comprehending, and reading	Vocabulary

**Some responses from students who DID find that they learned more this year:**

Improved my work a lot	Don't know
Making jokes out of some of the work	Homework
Learning, reading	Writing answers
I learned more than I did last year	Writing and reading
Nothing	Reading
Small class	Some of the students
Reading good stories	Reading aloud
The way the teacher taught me	Work everyday
Learned a lot of things, made new friends	Nothing
Learned a lot, teacher	Homework
Fun, learning	Nothing
Team activities	Work
Not a lot of tests	Work everyday
Teacher, sometimes fun	Work was hard sometimes
Fun	Read too little
Word attack, workbook	Story reading
Learned proper English, had fun	Book was long
Enjoyed the work, work wasn't too hard	Too much writing
Workbook, reading words in the box	Stories
Learning and the way the teacher explains	Nothing
Mustard Jar story	Homework
Road work	Workbook
Teacher, the way we were taught	Substitutes, different info about reading