Does Schooling Mean Progress? It Depends!

Anat Ben-Simon & Yoav Cohen

National Institute for Testing & Evaluation, Israel

The underlying premise of all educational systems is that schooling leads to learning. A unique opportunity to test this premise from an unusual angle, presented itself, when the Israeli Ministry of Education decided to retest the 8th grade cohort, which participated in the Israeli NAEP at the end of the 9th grade. The primary motivation for this study was to develop a scale for measuring progress which would serve as a "national yardstick" and later be used to assess the effectiveness of various educational interventions.

In this study the same tests (assessing proficiency in Hebrew, Arabic and mathematics) were administered twice, one year apart, to the same schools and classes. This design was adopted in order to minimize sampling error as much as possible and to allow for the calculation of individual gain scores, in addition to class gain scores.

The results of this study indicate that several factors affect progress, as measured by proficiency gain scores: schooling, test subject, school type and school's SES. In addition, it was found that previous exposure to test items also had a differential effect on gain scores.

The effects revealed in this study shed some light on the contribution of schooling to educational progress. The results also indicate that no single yardstick can be used to measure such progress, but that many factors should be taken into account while setting standards for expected learning progress.

Introduction

In 1994 the Israeli Ministry of Education decided to renew the administration of the National Assessment of Educational Progress (I-NAEP), but in a new format. The objectives of the new assessment, as defined by the Ministry, were to gather information on a national level on student achievement and on curriculum coverage and to monitor changes across time. The first administration of the I-NAEP in its new format was conducted in June 1996 and focused on the assessment of achievements of 4th and 8th grade students in mathematics and language.

Soon after the 1996 administration the Ministry decided to re-administer the mathematics and language tests in 1997 to the same sample that took the test in 1996. The objectives of the re-administration, as defined by the Pedagogical Secretariat in the Ministry that promoted this initiative were: (1) to evaluate the extent to which a further year of studies improves achievement in various school subjects

taught in 8th grade; and (2) to establish a normative measure of progress to aid policy makers and researchers in examining the effect and efficacy of various intervention programs implemented in the educational system.

Research sample and methodological concerns

Pre- and post-research models in general tend to be problematic when applied within the framework of the educational system. The main source of difficulty stems from the constant change in the composition of classes due to geographical and educational mobility of students and other educational or organizational considerations that call for reorganization of students in different classes in a given school. This frequent change in class composition poses a challenge to any study attempting to monitor the effect of schooling on achievements. The current study is no exception in this sense.

Seven thousand and thirty three students were tested in the Mathematics Test in 1996. The corresponding number of examinees in the Language Test was 7,081. The total number of classes that participated in the administration of each test was 256. Only one class was sampled from each school.

Since the objective of the study was to monitor learning progress for classes as well as for individuals, the 1997 re-administration of the Mathematics and Language Tests was limited to classes that had not undergone systematic changes in their student composition.

Given this constraint, only intermediate schools populated by 7th-9th grade students could participate in the re-administration. Elementary schools having 7tht-8th grade students had to be excluded from the study, since their graduates apply to various other schools for further studies. This exclusion resulted in shrinkage of the original 1996 sample by approximately a third, leaving only 154 and 155 classes qualifying for the re-administration of the Mathematics and Language tests correspondingly.

An additional 14 classes of the 154 classes tested on the 1996 Mathematics Test, and 20 classes of the 155 classes tested on the 1996 Language Test were excluded due to major changes in the composition of their students. Such changes resulted mostly from redistribution of students among existing classes or from the closing of a few schools. These classes were also removed from the 1997 sample.

Table 1 presents the number of intermediate school students and classes that were tested in the I-NAEP Mathematics and Language tests in 1996 and in 1997. The number of students and classes is reported by sectors: the Hebrew secular sector, the Hebrew religious sector and the Arabic sector. Since the curriculum in each of these sectors is slightly different, a separate analysis was performed for each of them. It should be noted that the figures reported in this table and throughout the whole report refer to "regular" (mainstream) students only. This category excludes new immigrants and special education student. (I-NAEP results are routinely reported separately for the above three populations).

It should also be noted that while the 1996 sample was a representative sample of students in each of the above educational sectors, the significant reduction in the number of classes in the 1997 re-administration resulted in a certain bias in this representation, as the classes remaining in each sector were no longer a representative sample of the schools belonging to each sector. Thus the generalizability from this sample to the general population of students in each sector is somewhat limited.

Description of the tests

All students participating in the re-administration received the same test forms that were administered in the 1996 administration, including the Student's Questionnaire. In addition, all of the teachers who taught Mathematics or Language subjects that year to the classes that were retested were requested to fill out a Teacher's Questionnaire. This questionnaire contained all of the questions appearing in the Teacher Questionnaire in the previous year, as well as some additional questions pertaining to teacher training.

<u>The Mathematics Test</u>

The Mathematics Test consisted of 112 items altogether that assessed achievements in 10 topics. About half of the items were of multiple-choice type while the remaining items were open-ended. Three parallel forms were administered. The three mutually exclusive forms were identical for all three sectors that participated in the administration.

The Language Test

The Language Test assessed three language skills: reading comprehension, writing and grammar. Each test included 10 texts (67-72 items) in the reading comprehension section, 11-13 tasks (prompts) in the writing section and about 10 grammar items. Six test forms were administered in the Hebrew sectors and seven test forms were administered in the Arabic sector. The test forms in each language (Hebrew/Arabic) were mutually exclusive and different for both languages.

Table 2 presents the structure of the mathematics and language tests that were administered.

Only one test (Mathematics or Language) was administered in each class. The various forms of each test were distributed randomly in the class. Thus, 33% of the students who were tested in mathematics in 1997 had an *a priori* chance of receiving the same test form as in 1996. The corresponding percentages for Language tests were 17% in the Hebrew sector and 14% in the Arabic sector. The tests were administered by external proctors, and the open-ended items were assessed by professional raters.

Calculation of progress score

Two progress score indices were defined:

Individual progress score - the difference between the achievement score that was obtained by a given student in the 1997 administration and the achievement score obtained by him or her in the 1996 administration. This score was naturally calculated only for identified repeaters, i.e., students who were tested in both administrations.

Class progress score - the difference between the mean achievement score that was obtained in a given class in the 1997 administration and in the 1996 administration. The class' mean achievement score was calculated for all students who took the test in a given administration. Accordingly, the class progress score is based on students who were tested in a given class in the first administration only, in the second administration only, or in both administrations.

In both cases a positive progress score indicates that the scores obtained in the 1997 administration were higher than the scores obtained in the 1996 administration.

In order to identify the students who participated in both administrations and construct an individual progress index for them, all students participating in both test administrations were matched by class and name. The first stage of the matching was performed by a computerized algorithm, which was corrected and updated manually in the second stage. Students identified by this process are referred to as 'identified repeaters'.

The numbers of identified repeaters in both tests and all sectors appear in Table 1.

The rate of identified repeaters, as can be seen in Table 1 is relatively small. The following factors account for this low rate:

1. Absence from school on the day of the test administration – on average only 85% of the students were present in the class at each administration (in 1996 and in 1997). Thus the expected overlap rate, taking the absence rate into account (on the assumption that each year different students were absent from class on the day of the test) was 72%. In other words, a maximum re-identification rate of 72% was expected. It should be noted that the absence rate tends to increase towards the end of the school year, when both administrations took place.
2. Change in class composition - each year a number of students leave their class and/or school for various reasons and new students join that class.
3. Poor or partial recording of the student's name - some students wrote fictitious names in one of the test administrations. Other students wrote first name or family name only. This partial recording made it difficult to definitively identify and match these students. Some classes had several students with the same name (first name and family name), which prevented definitive matching of the students in the two administrations. This phenomenon was particularly common in the Arabic sector and is responsible, inter alia, for the low re-identification rate of students in this sector. Another problem in identifying repeaters that was especially characteristic to the Arabic sector was that students wrote their names in different ways.

   It should be noted that many of the identification problems could have been

avoided if the re-administration of the test had been planned in advance, prior to the 1996 administration.

4. Change in the student's status. As mentioned above, all of the analyses in the I-NAEP were performed on 'regular' students only – students who were not new immigrants and who were special education students. A small number of students changed their status between the two administrations and therefore had to be removed from the sample of the identified repeaters.

Calibration of test form

To correct for the differences in the difficulty level of the various forms of a given test, a linear equating based on an equal populations design was applied to all forms of a given test. Accordingly, a standard score was calculated for each student in the 1996 sample. This standard score is the relative score (in terms of standard deviation), which that student obtained on the test, compared to all of the students who took that test form in the full sample (1996). A standard score was also calculated for each student in the 1997 sample using the mean and standard deviation obtained in the full sample (1996), on the form which the student took in the re-administration.

For purposes of convenience all subsequent results will be reported using the normal standard score scale.

As stated above, the objectives of the re-administration were:
(1) to evaluate the extent to which a further year of studies improves achievement in various school subjects taught in 8^{th} grade; and (2) to establish a normative measure of progress to aid policy makers and researchers in examining the effect and efficacy of various intervention programs implemented in the educational system.

The current study was designed to assess the effect of schooling on achievement and to investigate the feasibility of establishing a standardized measure of progress, as well as to identify the constraints associated with the use of such a standardized measure. Given the fact that it was a field study and not a control study, the research was limited to the following questions only:

- Does schooling mean progress?

- Is the progress score invariant to the subjects and topics taught?
- To what extent is student achievement consistent across time?
- What is the effect of a re-administration of identical test forms, as opposed to different test forms, on the observed progress score?
- Who benefits more from schooling, boys or girls?
- Do students from high SES background benefit more from schooling than do those from low SES background?

Results

Does schooling mean progress?

Mathematics progress score

Table 3 present the means and standard deviations of the standardized achievement scores obtained in the 1996 sample and the 1997 sample of the I-NAEP tests in Mathematics for identified repeaters. Also presented in this table are the means and standard deviations of the individual progress score and of the class progress score as defined above.

As can be seen from Table 3 the distribution of achievement scores in mathematics obtained for the identified repeaters in the 1996 administration was somewhat different from that of the original sample. In addition, the mean achievement score in mathematics obtained for this group was consistently higher than the mean achievement score obtained for all the students tested in the same class sample. This discrepancy was particularly great in the 1997 sample, in which the mean achievement score for the identified repeaters was 3.2-5.6 points higher across both subjects and all sectors than the mean achievement score for all the students that were tested in the same class sample. This discrepancy may be due to at least two causes: (1) the absence of weaker students from class on the day of the test in both administrations, particularly in the second administration; (2) a repeated testing effect – the identified repeaters were tested twice on the I-NAEP Mathematics Test, and the testing experience itself could have contributed to their performance on the test.

With regard to the individual progress score in mathematics, an improvement of .27-.31 points (in terms of normal standardized scores) was observed for the two Hebrew sectors and 0.8 points for the Arabic sector. The corresponding effect sizes for these progress scores were .27-.33 and .08. It is interesting to note that a somewhat different picture was observed regarding the class progress score, where a relatively large difference was found between the Hebrew sectors and a negative score was observed for the Arabic sector.

In order to examine the degree of progress in the different study topics assessed in the Mathematics Test, progress scores were calculated for each of these topics separately. These scores were derived only for identified repeaters. Again progress scores in each topic were calculated in terms of using standardized scores. The individual achievement scores in each topic were standardized for each topic separately in accordance with the distribution of scores obtained for that topic on a given test in the original (complete) 1996 sample. The mean individual progress scores observed for each Mathematics topic appear in Table 4. The topics detailed in this table are ranked by the educational level (grade) at which they are supposed to be taught according to the national curriculum.

The mean individual progress scores found for all topics, except 'elementary school topics', were positive. On average, students improved their achievement in these topics by .11-.76 standard deviations in the Hebrew sector and 0-.30 standard deviations in the Arabic sector. Significantly greater progress was made in topics that are taught in the 8^{th} and 9^{th} grades relatively to those taught at elementary school and in the 7^{th} grade. This trend can be attributed to the fact that the latter topics were taught in eighth grade only at a basic level, whereas in the ninth grade, they were taught in greater depth. Students improved their raw scores for the study topics by 8 to 28 points. This trend was found mostly for the Hebrew sector. Only minor improvement was observed in most topics in the Arabic sector, with the exception of 'introduction to functions,' which showed relatively higher improvement.

Language progress score

Table 5 present the means and standard deviations of the standardized achievement scores obtained in the 1996 sample and the 1997 sample of the I-NAEP tests in

language for identified repeaters. Also presented in this table are the means and standard deviations of the individual progress score and of the class progress score as defined above.

The mean achievement scores in language obtained for the identified repeaters in the 1996 administration were markedly higher, especially in the Hebrew sector, than those of the original 1996 sample, and slightly higher than those obtained for all the students that were tested in the same classes in 1996.

With regard to the individual progress score in language, an improvement of .05, .14 and .16 points (in terms of normal standardized scores) was observed for the Hebrew secular, Hebrew religious and Arabic sectors respectively. The corresponding effect sizes for these progress scores were .05, .18 and .23. Similar results were observed for the class progress score, yielding somewhat larger effect sizes for the Hebrew religious and Arabic sectors.

Relatively low correlations (.40-.51) were found between the language achievement scores on the two administrations. These low coefficients may indicate that measurement error is far greater than indicated by the reliability measures obtained for this test (.83-.84). This conclusion is supported by the fact that the “test-retest” correlations were far higher (.76-.85) for students who received identical test forms than for students who received different test forms (.38-.53). Content sampling error may be the reason for the above phenomenon. In light of the instability of language achievement scores across time, no analysis of the various language skills (i.e., reading comprehension, writing and grammar) was performed.

Are individual and class achievement scores consistent across time?

In order to examine the stability of individual achievement scores, a Pearson correlation was calculated between the individual achievement scores that were obtained in both 1996 and 1997 administrations (see Table 6). The correlations were calculated for both tests and the three sectors. Fairly high correlations (.73-.81) were found for mathematics. These correlations were slightly lower than the median reliability coefficients of the various test forms. Far lower correlations (.40-.58) were found for language. These correlations were markedly lower than the median reliability coefficients that were calculated for the Language Test forms.

A similar analysis was applied to the mean class scores. The correlations found for mean class score ranged from .68 to .86 for mathematics and .33-.78 for language. Somewhat higher correlations were found for mean class achievement scores when they were calculated for identified repeaters only. These correlations ranged from .67 to .94 for mathematics and .63-.84 for language.

To what extent is progress affected by re-administration of the same test form versus a different test form

Three test forms of the Mathematics Test and six to seven forms of the Language Test (six in the Hebrew sectors and seven in the Arabic) were administered in 1996 and 1997. Test forms were distributed randomly in the class; thus, 33% of students tested in mathematics in 1997 had an *a priori* chance of receiving the same form that they had received in 1996. The parallel percentages in the Language Tests were 17% for the Hebrew sectors and 14% for the Arabic sector.

In order to examine the marginal effect of receiving an identical form as opposed to a different form on the individual progress score, these scores were calculated separately for students who were tested on identical test forms and students who were tested on different test forms in the two administrations (see Tables 7 and 8).

The mean individual progress scores in mathematics for students who received identical test forms were only .02-.06 SDs higher than those obtained for students who received different test forms. The mean individual progress score in language for students who received identical test forms were .07 SDs higher for the Hebrew sector than those obtained for students who received a different test form, yet much higher (.28 SDs) for the Arabic sector.

In addition to the above comparisons, Pearson correlations were calculated for each of the above testing conditions (identical versus different forms) in order to examine the effect of this condition on the consistency of the achievement scores. Relatively high correlations (.76-.85) were found for both subjects and for both sectors between achievement scores in the two administrations for students who took identical test forms. Lower correlations were found between achievement scores in the two administrations for students who took different test forms (.38-.79). The later correlations were particularly low (.38-.53) in the Language Test. These correlations

point to the relatively low reliability of this test in assessing individual achievement. However, it should be noted that this correlation increased (.63-.75) when calculated for mean class scores based on identical students.

Who benefits more from schooling, boys or girls?

Individual progress scores were calculated for each gender separately, for both tests and three sectors. The results of this analysis as presented in Tables 9 and 10 show no gender effect except for a slight increase in mean achievement scores obtained by boys from the Hebrew religious sector in comparison to their female counterparts. Also no gender differences were revealed in the correlation between the 1996 scores and the 1997 scores except for slight yet inconsistent differences within the various sectors.

Do students from high SES background benefit more from schooling than those from low SES background?

In order to examine the differential effect of instruction on students from different SES backgrounds all classes participating in the re-administration were divided into three SES levels based on the school SES index. The school SES index is based on five variables: (1) family income; (2) parents' education; (3) family size; (4) percentage of new immigrants in the student body; and (5) proximity to the closest city. Next, a comparison was made of the individual progress scores obtained in classes belonging to the three different SES levels. The mean and standard deviation of the progress scores obtained in the Mathematics and Language Tests in all three groups appear in Tables 9 and 10.

A comparison of class and individual progress scores in mathematics by the three SES levels showed a small and inconsistent effect for SES on individual progress scores, both in the Hebrew and Arabic sectors. Similar results were found for class progress scores in the Arabic sector. A marked increase in class progress score was found in the Hebrew sector for classes of medium and high SES.

With regard to progress score in language, results indicate no SES effect on both individual and class progress scores obtained in the Hebrew sector, moderate advantage for the low and medium SES groups in the individual progress scores and an opposite trend for the class progress scores.

Summary and Discussion

The objectives of this study, as stated above, were to examine the effect of schooling and to explore the possibility of developing a national yardstick that may be used to assess the effectiveness of educational interventions. The results indicate that schooling does indeed yield progress but many factors affect the extent of this progress, and thus no single yardstick can be used to judge it, unless all these factors are taken into account in the process of setting standards for expected learning progress.

The effect of several factors on learning progress was examined in this study: schooling, school subject and topics, school type, gender and school SES index. These factors are but a few among many other factors that may affect learning outcomes, yet they do shed some light on the effect of schooling.

With regard to school subject and topics the results of this study showed relatively higher gains in achievement in mathematics than in language. In mathematics greater progress was observed for topics that were taught in the period between the two administrations (in the 9th grade) than for topics that were taught prior to the first administration.

A possible explanation for the difference between the progress scores observed in mathematics and in language may be the fact that mathematics lessons involve the teaching of knowledge (e.g., concepts, rules, algorithms) while language lessons focus on practicing skills. It may be the case that far greater effort is required to improve one's skills than to teach him or her new topics. This conclusion is supported by the relatively high progress score that was found in the Hebrew religious sector as compared to the secular sector. Male students from the Hebrew religious sector engage daily for 2-4 hours in reading and discussing religious texts; accordingly the greatest progress score in language was observed for this group. A significantly greater improvement in achievement was also observed in the Arabic sector than in the Hebrew secular sector. This may be due to the fact that the literary Arabic taught in school is in fact a second language for these students (whose first language is spoken Arabic), and each additional year of studies helps improve their proficiency in this language.

The fact that far greater progress was observed for topics that were taught in the period between the two administrations (in the 9th grade) as opposed to topics that were taught prior to the first administration, may indicate a rather limited transfer between the various topics in mathematics.

As to the gender and SES factors, results indicate no effect for gender and only a minor and overall inconsistent effect for SES. The latter finding is interesting in the light of two different phenomena which are expected to have opposing effects on gain scores: (1) while the mean achievements of students from a low SES background are markedly lower than those of students from a high SES background, due to the "regression towards the mean" effect the difference between their achievements on the second administration is expected to be smaller than the difference between their achievements in the first administration; (2) many studies investigating the effect of instruction on students with various ability levels suggest that better students tend to benefit more from instruction than do weaker students. It seems that the findings of the current study cannot resolve these conflicting trends.

Lastly, it should be emphasized that policy makers often focus on class or school progress scores and use these estimates as an indication of changes in the quality of instruction. As can be seen from the results of the current study, changes in the composition of a class can greatly affect class progress scores. This in turn suggests that whereas mean progress scores of classes or schools may increase or decrease across time, such changes do not necessarily indicate changes in the quality of instruction but may merely result from changes in class or school composition. Thus studies designed to examine the effectiveness of instruction should be categorically limited only to identified repeaters.

Table 1: Number of students and classes that participated in the 1996 & 1997 administrations of the I-NAEP in Mathematics and Language*

		Mathematics			Language		
		Hebrew Secular	Hebrew Religious	Arabic	Hebrew Secular	Hebrew Religious	Arabic
1996 Adm. Full sample	N students	2,274	1,956	1,943	2,140	2,034	2,074
	N classes	88	90	78	88	90	78
1996 Adm. Intermediate schools only	N students	1,720	1,343	1,519	1,727	1,398	1,570
	N classes	55	50	49	56	50	49
1997 Adm.	N students	1,495	1,050	1,342	1,282	1,075	1,320
	N classes	51	42	47	47	43	45
Identified Repeaters		766	406	544	613	424	226

Table 2: The structure of the Mathematics and Language Tests and reliability estimates obtained for the different test versions

	Mathematics		Language	
	Hebrew Secular & Religious	Arabic	Hebrew Secular & Religious	Arabic
Number of test versions	3	3	6	7
Number of items	112 Ques.	112 Ques.	R - 72 Ques. W - 13 Prompts	R - 67 Ques. W - 11 prompts
Reliability range	.87 - .91	.90 - .91	.77 - .87	.81 - .88

Table 3: Mean and standard deviation of standardized scores obtained in the 1996 and 1997 administrations of the Mathematics Test and of individual and class progress scores

		Hebrew Secular	Hebrew Religious	Arabic
1996 Adm.	Mean	-.11	-.05	.13
	Std	.98	.93	.90
1997 Adm.	Mean	.16	.26	.21
	Std	1.06	.97	1.00
Individual Progress Score	Mean	.27	.31	.08
	Std	.64	.59	.70
	Effect size	.27	.33	.08
Class Progress Score	Mean	.24	.10	-.08
	Std	.25	.58	.43
	Effect size	.48	.15	-.14

Table 4: Progress scores observed in the Mathematics Test by topics

Topics	Grades taught	No. of items	**Hebrew Sector**		**Arabic Sector.**	
			Progress score	Effect size	Progress score	Effect size
Arithmetic	4^{th} – 6^{th}	34	-.05	-.05	-.09	-.09
Statistics	7^{th}	6	.17	.17	.12	.12
Algebra – 1 variable	7^{th} – 8^{th}	24	.16	.17	.09	.09
Geometry: basic concepts	8^{th}	3	.11	.11	.01	.01
Probability	8^{th}	6	.24	.24	.15	.15
Algebra – 2 variables	8^{th} – 9^{th}	9	.23	.23	-.01	-.01
Geometry	8^{th} – 9^{th}	12	.29	.30	.13	.13
Linear functions	8^{th} – 9^{th}	3	.39	.39	.11	.11
Algebraic technique	8^{th} – 9^{th}	9	.45	.44	.07	.07
Introduction to functions	8^{th} – 9^{th}	3	.76	.71	.31	.30

Table 5: Mean and standard deviation of standardized scores obtained in the 1996 and 1997 administrations of the Language Test and of individual and class progress scores

		Hebrew Secular	Hebrew Religious	Arabic
1996 Adm.	Mean	.58	.53	.25
	Std	.88	.92	.82
1997 Adm.	Mean	.63	.69	.45
	Std	.89	.89	.94
Individual Progress Score	Mean	.05	.16	.20
	Std	.97	.91	.83
	Effect size	.05	.18	.23
Class Progress Score	Mean	-.02	.14	.16
	Std	.42	.38	.49
	Effect size	-.04	.27	.38

Table 6: Test-retest correlations for individual and mean class score as obtained in the 1996 and 1997 administrations of the Mathematics Test

	Mathematics			**Language**		
	Hebrew Secular	Hebrew Religious	Arabic	Hebrew Secular	Hebrew Religious	Arabic
Correlations between individual scores						
	.81	.81	.73	.40	.48	.58
Correlations between average class scores						
Across all students	.86	.68	.69	.60	.78	.33
Across repeaters	.92	.94	.67	.63	.84	.63
Median reliability	.90		.91	.83		.84

Table 7: Mean and standard deviation of individual progress score in Mathematics obtained by students who received identical versus different test forms

	Hebrew Secular & Religious		Arabic	
	Same form	Different form	Same form	Different form
Mean	.30	.28	.11	.05
Std	.58	.65	.67	.72
Corr.	.84	.79	.78	.70
Median Reliability	.90		.91	
N	405	767	184	360

Table 8: Mean and standard deviation of individual progress score in Language obtained by students who received identical versus different test forms

	Hebrew Secular & Religious		Arabic	
	Same form	Different form	Same form	Different form
Mean	.15	.08	.43	.15
Std	.60	1.00	.53	.90
Corr.	.76	.38	.85	.53
Median Reliability	.83		.84	
N	158	879	37	189

Table 9: Mean and standard deviation of Mathematics achievement obtained in the 1996 & 1997 administrations and of individual progress score by gender

		Hebrew Secular		Hebrew Religious		Arabic	
		M	F	M	F	M	F
1996 Adm.	Mean	-.02	-.19	-.08	-.05	.24	.07
	Std	1.0	.94	1.1	.86	.94	.86
1997 Adm.	Mean	.26	.09	.22	.28	.29	.15
	Std	1.1	.98	1.1	.92	1.0	.97
Individual Progress Score	Mean	.28	.28	.30	.33	.05	.08
	Std	.68	.58	.56	.56	.71	.70
Test-retest correlation		.80	.82	.87	.80	.75	.72
	N	346	419	121	285	219	325

Table 10: Mean and standard deviation in Language achievement obtained in the 1996 & 1997 administrations and of Individual progress score by gender

		Hebrew Secular		Hebrew Religious		Arabic	
		M	F	M	F	M	F
1996 Adm.	Mean	.38	.75	.31	.63	.08	.37
	Std	.95	.78	1.1	.83	.80	.82
1997 Adm.	Mean	.44	.79	.57	.75	.28	.57
	Std	.92	.84	1.1	.81	1.0	.87
Individual Progress Score	Mean	.06	.04	.26	.12	.20	.20
	Std	1.1	.89	1.0	.87	.86	.77
Test-retest correlation		.35	.39	.53	.44	.56	.58
	N	289	321	132	292	93	133

Table 11: Mean and standard deviation of Mathematics achievement and mean and effect size of progress scores by SES

		Hebrew Sectors			Arabic Sector		
	SES	Low	Medium	High	Low	Medium	High
Individual scores for identified repeaters							
1996 Adm.	Mean	-.42	-.02	.11	-.10	.07	.24
	Std.	.98	.91	.95	.91	.88	.89
1997 Adm.	Mean	-.16	.28	.41	-.15	.19	.31
	Std	1.04	.99	.98	.93	.95	1.03
Progress Score	Mean	.26	.30	.30	-.05	.12	.07
	Effect size	.26	.32	.31	-.06	.13	.07
N		328	417	423	71	192	281
Average class scores							
Progress Score	Mean	-.01	.25	.29	-.11	-.06	-.03
	Effect size	-.02	.57	.70	-.22	-.13	-.05
N classes		33	30	.29	21	17	9

Table 12: Mean and standard deviation of Language achievement and mean and effect size of progress scores by SES

		Hebrew Sector			Arabic Sector		
	SES	Low	Medium	High	Low	Medium	High
Individual scores of identified repeaters							
1996 Adm.	Mean	.38	.52	.72	.09	.14	.38
	Std.	.93	.94	.79	.98	.81	.78
1997 Adm.	Mean	.50	.63	.80	.36	.41	.51
	Std.	.97	.90	.80	.83	1.00	.91
Progress score	Mean	.12	.11	.08	.27	.27	.13
	Effect size	.12	.12	.09	.30	.29	.15
N		292	356	389	26	90	110
Average class scores							
Progress score	Mean	.08	.05	.05	.13	.08	.24
	Effect size	.13	.13	.13	.28	.17	.64
N classes		33	26	31	8	16	21