The Hashemite Kingdom of Jordan

National Center for Human Resourses Development

Reform Impact on Math Achievement of Basic Schools in JORDAN

By
Dr. Kapur Ahlawat
&
Hisham Al-Dajeh

Publication Series (47)

March 1997

Preface

The National Center for Human Resources Development (NCHRD) plays a key role in the comprehensive educational reform in Jordan

The Center conducts policy-oriented research to guide educational policy of sustainable improvement and keeps the reform on track through constant monitoring of implementation of the reform plan. To this end, the Center has instituted an operating Education Management Information System (EMIS) and established a sample-based national achievement monitoring system for continuous assessment of educational progress. Student achievement serves as a barometer for the general quality of education and effectiveness of the education system and provides, critical feedback to the concerned parties.

"The national assessment of instructional quality of the basic education in Jordan" project launched by the Center in 1991 is a landmark in the history of national assessment of educational progress in Jordan. While the first sample-based round of student achievement data established baseline benchmark, the second round of the study conducted in 1995, provides vital information about the impact of the educational reform going on since 1990.

This volume compares mathematics achievement of 4th and 8th grade students before the reform with that of those after four years of exposure to reform program. The results provide valuable feedback to all those who are concerned with school education in Jordan. Also, the results open a forum for discussion.

I hope that education community, particularly, curriculum organizers, textbook developers, teacher trainers, supervisors and specifically teachers themselves will take a closer look at the findings and act accordingly.

Victor Billeh
President NCHRD

Acknowledgements

National long-term studies of this magnitude are collaborative and cooperative endeavours. Generally, several institutions and teams of specialists from various disciplines, and a host of field administrators, participated from the beginning to the end of this project.

First and foremost we gratefully thank the thousands of parents and students and hundreds of teachers and school principals who patiently filled in the long questionnaires (cut down to one-third of their size for 1995 administration) in both 1993 and 1995.

The teams of supervisors of regional districts administered the tests in classrooms and those who coordinated the data collection activity from the directorates of education performed a critical task because without their vigilance we could not have collected credible information. Again, we heartily thank all and each one of them.

Professor Victor Billeh, the founding President of the NCERD (now called NCHRD) has been a constant source of inspiration and encouragement. It was his vision to start with, and his guidance all along, that have been realized and embodied in the volumes that present the findings of the Evaluation of the Educational Reform Impact in Jordan.

We gratefully acknowledge the contribution of all the experts who participated in the critical task of achievement tests' construction and we particularly thank the heads of the three teams: Professor Gahzi Hamzeh the head of the Mathematics team; Mr Mohammed Swailem the head of the Science team; and Dr. Abdullah Khabbas, the head of the Arabic team.

The selection of the sample was an arduous task. Eng. Yasser Mohawesh painstakingly drew the stratified random sample and then established its representativeness by using complex statistical procedures. We heartily acknowledge his contribution to this study.

It goes without saying that all the staff of the NCHRD did their best for the successful implementation of this long-range project. We extend our thanks to all of them; particularly to the research assistants viz; Miss Amal Al Kharouf, Miss Hanan Ennabi, and Mr. Ramzi Abu Ghazaleh.

In spite of the wholehearted efforts of all the participants this study, as it is in your hands, would not have seen the light of the day without the enduring patience of Mrs. Hala AL Masoud not only in typing and retyping drafts after drafts but also in designing all the tables and graphs of this study. We express our heartfelt appreciation for Mrs. Hala's relentless efforts to make the text as visually pleasant as it is.

Table of Contents

Table of Contents/ cont..

Table of Contents/ cont..

List of Tables

Part I:

Pre-Reform and Post-Reform Math Achievement of Grade 4 Students in Jordan: Indicators of Reform Impact.

List of Tables/ cont..

Part II:
Value-Added Math Achievement of Grade 8 Students Under Four Years of Educational Reform in Jordan

List of Tables/ cont..

List of Figures

Part I:

Pre-Reform and Post-Reform Math Achievement of Grade 4 Students in Jordan: Indicators of Reform Impact.

Part II:

Value-Added Math Achievement of Grade 8 Students Under Four Years of Educational Reform in Jordan

List of Figures/ cont..

An Overview of the Background

In the wake of the 80s' slump in the regional oil economy, which had given during its boom tangible spillover benefits to Jordan in the form of remittances from Jordanian skilled workforce working in Gulf Cooperation Council (GCC) Countries, Jordanian human-resource-based economy was hard hit.

On the other hand the general education system which had served the country so well in the past, had gradually outlived its utility. This had happened largely due to two major factors. One was that rapid quantitative expansion of the education system undertaken by the MOE to provide free and compulsory basic education (grades 1-9) for all children 6-15 years old, had eroded its quality. Second, the technological revolution and growing use of modern technologies in the industries as well as in other employment sectors including teaching and learning had changed the knowledge and skills requirements of labor markets.

For Jordan it became necessary to upgrade the quality of school graduates in order to meet the changing demands of the domestic labor market and to maintain its skilled workforce advantage in the regionwide labor market.

Under these circumstances Jordan launched upon a comprehensive 10-year-long education reform plan to overhaul the whole general education system.

The overarching objective of the reform plan was to enhance student achievement levels.

After years of thoughtful planning, the actual execution of the Education Reform Plan (ERP) started in 1989. Of the three-phase execulation plan the First Pase (1989-1992) was called the Foundation Phase.

In 1990. the National Center for Educational Research and Development which is now known as the National Centre for Human Resources Development (NCHRD) was created.

Monitoring the implementation of the ERP through continuous assessment and evaluation of the progress of education was one of its major responsibilities. The NCHRD initiated the establishment of an institutionalized system of continuous monitoring and assessment of students' learning achievement in 1990.

A longitudinal study, "The National Assessment of Instructional Quality in Basic Education" was designed to fulfil the following three major objectives.

1. To provide baseline data on the quality of teaching and student achievement levels in basic schools in Jordan.

2. To identify instructional processes most likely to influence the quality of basic education in Jordan.

3. To provide information on the effects of the different interventions initiated under the comprehensive reform on teacher classroom practices and student achievement.

Its sampling design constituted of a stratified two-stage random sample of 245 schools selected from the population of all the schools in Jordan existing in 1992.

Out of all grades (1-10) of basic education cycle, 4th, 5th, and 8th grades; three core subjects, Arabic , Math, and Science were selected on logistic grounds. In the year 1991-92 new curricula and textbooks were introduced to grades (1, 5, 9). Also the teachers of these grades had been trained, through the in service teacher training program, in how to apply new curricula, textbooks and modern instructional methods in the classrooms.

Similar reform elements were introduced in grades (2, 6, 10) in 1992-93; and in grades (3, 7, 11) in 1993-94; and likewise in grades (4, 8, 12) in the scholastic year 1994-95. Therefore, in the year 1992-93 when

the data were collected grades 1 and 4 were untouched by the reform while reform elements had been introduced in the 5th grade for the first time.

Obviously, we selected the 4th and the 8th grades to establish pre-reform bench-marks and the 5th grade to provide formative feedback. In addition to measuring student achievement in the three academic subjects, information was also needed on a host of variables from the parents, principals, teachers, and students, covering the key domains (see Annex I).

Batteries of tests and questionnaires were developed and tried out in 1992, while the data were collected from schools in May 1993. The same batteries of tests and questionnaires were administered again to the 4th and 8th grade classes as well as to parents, teachers and principals related to the same schools, in May 1995. This time the whole system had been fully exposed to educational reform.

The 1993 study was concerned with establishing baseline benchmarks. The results have been described and discussed in several publications of the NCHRD. The main purpose of the 1995 replication of the 1993 study was to investigate about the impact of the comprehensive educational reform in Jordan.

Because of the variety and volume of this rich mine of information, but primarily, for the sake of the convenience of different audiences, we have undertaken to release the findings of the study through a series of topical reports.

The present report entitled "Reform Impact on Mathematics Achievement of the Basic Schools in Jordan" deals with changes in mathematics achievement levels of the 4th and 8th grade students after four years of reform.

The study is presented in two parts. The first part deals with changes in math achievement of the 4th grade students and the second part describes changes in math achievement of the 8th grade students.

Our objective, apparently, was to study the impact of the comprehensive educational reform of the education system in Jordan. But as the evaluators were well aware of the painful fact that impact evaluation of national reform programs, particularly of educational programs, is more like an art of detection than like a scientific proof. In case of the issue before us, we can see if there is a change in any direction or not but we cannot prove beyond a shadow of doubt the cause of change or the cause of stasis.

However like a scientist by eliminating the plausible alternative factors that could cause the observed changes we can establish with reasonable certainty that the changes might have been caused by the planned interventions.

We have compared the baseline achievement of the 4th and 8th grades students as it was in 1993 with their achievement on the same tests in 1995. The 1993 students were schooled in the traditional ways, while the 1995 students were taught under reformed conditions for four years. We have reasoned to attribute any changes in students' test scores to the implementation of educational reform.

Kapur Ahlawat
NCHRD

(Part I)

Value-Added Math Achievement of Grade 4 Students Under Four Years of Educational Reform

Value-Added Math Achievement of Grade 4 Students Under Four Years of Educational Reform

Contextual Background

In 1989 Jordan embarked upon an ambitious educational reform program to restructure and revitalize its basic education system. The main goal of reform focused at enhancing student achievement levels. The key reform elements: reconstructed curricula, newly designed textbooks and instructional materials, and inservice teacher training in classroom applications of innovative instructional methods for using new textbooks and materials were introduced for the first time in 1st, 5th, and 9th grades in the scholastic year 1991-92. Next year they were introduced in 2nd, 6th, and 10th grades. By the year 1994-95 the reform had covered all the grades (1-12).

Recognizing the fact that the ultimate test of the impact of educational reform depends upon enhanced student achievement levels, the National Center for Educational Research and Development (now renamed as the National Center for Human Resources Development (NCHRD)) designed a reform impact evaluation program to assess improvement in instructional quality reflected by student achievement levels. The assessment design entailed a stratified two-stage random national sample of 245 schools. One section each of 4th, and 8th grade classes was randomly selected if schools contained more than one sections in either of these grades. (For detail about sampling characteristics and domains of variables tested, see Ahlawat, 1993; Ahlawat, et al., 1994 a, 1994b).

Students were tested in three subjects, Arabic Language, Mathematics, and Science, at the end of the 1992-93 scholastic year. In addition to achievement in the three subjects, questionnaires were administered to students, parents, teachers and principals to gather various types of information on attitudes, beliefs, and perceptions about schools, classrooms, instruction and management practices. (see Annex I for domains of variables included in the study).

At the first stage information was collected primarily to serve as baseline data to study changes (if any) after about four years of educational reform. The same achievement tests and questionnaires were administered again to the 4th grade students in the same schools. The second testing took place after the students had completed 4 years under the reform program using new curricula and textbooks taught by teachers who had received inservice training in innovative instructional techniques recommended to foster critical thinking, problem solving, self-learning, and higher level cognitive skills in students.

While the whole assessment program covered achievement testing in three major subjects (Arabic, Mathematics, and Science), this study concentrates on Mathematics achievement alone.

The pre and post samples of 4th grade consisted of 4,840, and 5,010 students for the years 1993 and 1995, respectively. In 1993, however, only about one half of the students in each sampled class took the mathematics test. Thus the mathematics test sample in 1993 consisted of 2,428 students. In the year 1995, the mathematics tests were administered to the whole classes. The comparisons of performance on the math tests of 4th grade classes between the years 1993 and 1995 are based on the same school units. It should be clear that schools and grades were the constant units of comparison while the student populations for the two testing occasions were evidently different. The same grades were measured on the same tests in the same schools on two different occasions. On the first occasion 4th grade students had no exposure to the elements of education reform, whereas, on the second occasion the students had been exposed to reform processes continuously for four years.

The following section presents a brief description of the contents and composition of the 4th grade achievement test of mathematics and its psychometric properties on both testing occasions (1993, 1995) based upon the performances of samples.

Content and Structure of the Test

The 1993 mathematics test comprised 46 items, some supply response and some multiple-choice. The test for 1995 was shortened to 36 items, of which 19 were multiple choice and 17 supply - response. While

administering the test it transpired that two of the 19 multiple-choice items had structural faults. The faulty items were dropped from analysis. Table 1 presents the table of specifications for the 34 items which were common in both testings. Six content areas and three cognitive skills were covered by the test.

Table (1)

Table of Specifications for Grade 4 Math Achievement Test

Skills / Content	Conceptual Understanding (CU)	Procedural Knowledge (PK)	Problem Solving (PS)	Total
Numbers (NUM)	1	2	1	4
Operations (OPR)	3	4	3	10
Theory (THY)	3	2	--	5
Fractions (FRA)	2	3	2	7
Decimals (DEC)	1	1	--	2
Geometry (GEO)	2	3	1	6
Total	12	15	7	34

Psychometric Properties of the Test

Reliability Indices

Table (2) presents the alpha coefficients (K-R 20) for the whole test, for each content component and for each skill component of the test on each occasion of testing. The 1995 testing is referred as Post-Reform and 1993 as Pre-Reform in Table (2).

Table (2) clearly shows that for the whole test post- and pre-reform reliability coefficients are nearly the same. In most cases the absolute difference between the two reliability coefficients of a test component does not exceed .03. In three cases namely, Number and Operations

(content areas), and Problem solving (skill) post-reform reliability coefficients show a noticeable increase over pre-reform. The alpha coefficient is an index of homogeneity of a set of items or in other words it represents the degree of consistency of performance over a set of test items.

Table (2)

Math Test Scale and Subscales' α - Reliability's in Pre and Post-Reform Grade 4 Populations (N: Post = 5010, Pre=2428)

Scale/Sub-Scale (No. of Items)	α– Coefficient	
	Post-Reform	Pre-Reform
Total (36)	.89	.88
Number (4)	.56	.46
Operations (10)	.77	.72
Theory (5)	.55	.57
Fractions (7)	.62	.59
Decimals (2)	.34	.33
Geometry (6)	.58	.59
Conceptual Understanding (12)	.68	.66
Procedural Knowledge (15)	.80	.79
Problem Solving (7)	.73	.64

Because the items in a well defined content or skill area are homogeneous by definition, inconsistency in performance should reflect sketchy knowledge or lack of required skills and competencies while on the other hand, consistency of performance will reflect more integrated competency within the particular subdomain represented by a particular set of items. Looking from this perspective, increased reliability indices in the three subscales could be attributed to more levelled understanding of students in these areas, possibly on account of emphasis given to: building solid foundation of basic knowledge and skills such as Numbers and Operations, and enhancing problem solving skills of the students in the education reform program in Jordan. This position is further supported by substantial increase in test performance of the 1995 student population over the 1993 student population on the same sets of test items.

Comparative Difficulty and Discrimination Indices

Difficulty index (Proportion of students answering an item correctly) and discrimination (item-remainder correlation coefficient) were computed for each item in both 1993 (Pre-reform) and 1995 (Post-reform) populations.

Table (3) presents the post-reform and pre-reform difficulty indices for each item, the difference between post and pre-reform indices, and the port-reform and pre-reform discrimination indices of each item.

The difficulty index of an item read without decimal point estimates the percentage of students in the defined student population who gave correct answer to the item.

Table (3)
Item Difficulty and Discrimination Indices of the G4 Math Test in Pre- and Post-Reform Populations (N: Post = 5010, Pre = 2428)

S.		Difficulty (Pi)		Difference	Disc. (Item-Remainter Corr).	
No.	Item	Post	Pre	Post-Pre	Post	Pre
1	M1	.54	.41	.13***	.45	.47
2	M2	.39	.21	.18***	.50	.38
3	M3	.37	.34	.02 NS	.32	.40
4	M4	.49	.38	.11***	.49	.50
5	M5	.27	.22	.05***	.26	.17
6	M6	.43	.37	.05***	.21	.32
7	M7	.69	.52	.17***	.45	.46
8	M9	.65	.46	.19***	.41	.44
9	M10	.66	.65	.01 NS	.41	.46
10	M11	.38	.30	.08***	.41	.39
11	M12	.37	.33	.04***	.37	.36
12	M13	.41	.33	.08***	.43	.28
13	M14	.42	.31	.11***	.40	.32
14	M15	.35	.23	.12***	.52	.49
15	M16	.36	.33	.03**	.31	.28
16	M18	.64	.38	.26***	.48	.41
17	M19	.17	.02	.14***	.53	.27

Table (3) continued

18	M20	.54	.46	.08***	.40	.32
19	M21	.58	.59	-.01 NS	.29	.44
20	M22	.23	.25	-.02*	.32	.28
21	M23	.29	.13	.16***	.52	.45
22	M24	.38	.26	.12***	.33	.27
23	M25	.23	.14	.09***	.47	.40
24	M26	.17	.13	.04***	.48	.42
25	M27	.46	.29	.17***	.47	.44
26	M28	.30	.19	.11***	.57	.53
27	M29	.18	.08	.10***	.47	.41
28	M30	.52	.28	.24***	.33	.36
29	M31	.37	.23	.14***	.57	.52
30	M32	.45	.20	.25***	.48	.38
31	M33	.21	.11	.10***	.50	.43
32	M34	.49	.44	.05***	.32	.29
33	M35	.42	.18	.24***	.52	.41
34	M36	.39	.18	.21***	.30	.30
	Mean (P)	**.41**	**.29**		**.42**	**.38**
	MIN	**.17**	**.02**		**.21**	**.17**
	MAX	**.69**	**.65**		**.57**	**.53**
	VAR	**.02**	**.02**		**.01**	**.01**

Note: * = $p \leq .05$

** = $p \leq .01$

*** = $p \leq .000$

The discrimination index indicates the power of an item to differentiate between the high and low performing students on the test. That is, the extent to which a student can be correctly classified into a high or low performing group of students on the basis of his/her (correct or incorrect) response to this item. Conversely, the lack of discrimination power of an item indicates that the ability required to answer this item is independent of the ability that is measured by the rest of the test, items.

In practice, if an achievement test item is ambiguous such that different students interpret its meanings in different ways and answer accordingly or if an item is so difficult that every student resorts to random

guessing then its discrimination index is bound to be nearly zero; it could be negative if bright students avoid answering the item while others answer it by guessing.

Comparison between the last two columns of Table (3) suggests a slight overall increase in the discrimination power of items from 1993 population to 1995 population of students. The average discrimination index has increased from .38 in 1993 to .42 in 1995; likewise, the minimum index has increased from .17 to .21 and the maximum has increased from .53 to .57. Given the difficulty levels of the items this seems to be a healthy trend.

Going back to increased performance, on 30 items out of 34 there is statistically highly significant difference ($p \leq .001$ in 29 cases and $p \leq .01$ in one case) in favor of the post-reform student population. In general the average proportion correct taken over 34 items has increased from .29 in 1993 to .41 in 1995. This means that while in 1993 the average student answered 29% of the test items correctly in 1995 the average student answered 41% of the test items correctly, a 12% improvement in test performance. The minimum proportion correct index has increased from .02 in 1993 to .17 in 1995 and the maximum has increased from .65 to .69; not necessarily on the same items.

Table (4)

Comparative Frequency Distribution of the Difficulty and Discrimination Indies of the Grade 4 Math Test Items on the Two Testings (Pre-testing (1993) and Post-testing (1995))

(a) **Difficulty Index**

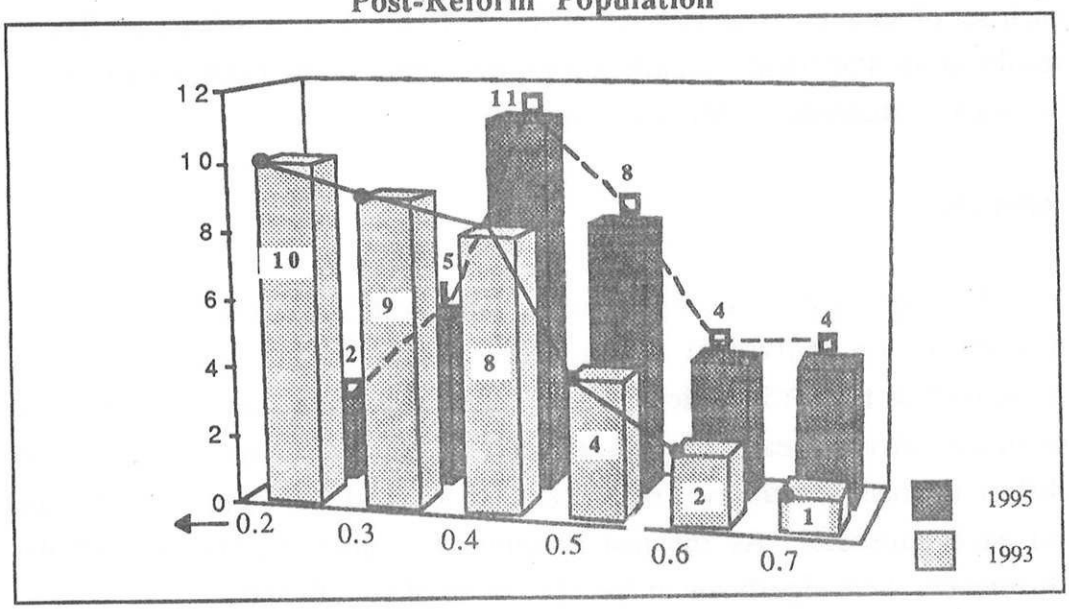
Interval	Post-Reform		Pre-Reform	
	Freq.	% Freq.	Freq.	% Freq.
Lo - .20	2	6	10	29
.21 - .30	5	15	9	26
.31 - .40	11	32	8	24
.41 - .50	8	23	4	12
.51 - .60	4	12	2	6
.61 - Hi	4	12	1	3
Total	34	100	34	100

(b)	Discrimination Index (Item-Remainder Corr.)			
Lo - .30	4	12	8	24
.31 - .40	8	23	9	26
.41 - .50	16	47	15	44
.51 - Hi	6	18	2	6
Total	34	100	34	100

Table (4) presents comparative frequency distributions of: (a) the difficulty indices and (b) the discrimination indices in the post-reform and the pre-reform populations, Figure (1) shows a graphic display of the two distributions of the difficulty index in the post-reform and pre-reform populations of 4th grade students.

It can be seen from Figure (1) that in the pre-reform population the highest frequency was in the lowest category, then in the next, then as the Pi index increased the frequency plummeted and dropped to one in the .61 and above category. The first category on the extreme left accounts for 10 (29%) of the items and the next category accounts for 9 (26%) of the items. This means that in 1993 there were 19 (55%) items which could not be answered correctly by 70% of the students but in 1995 the number of such items dropped down to 7 (21%). On the easier end of the test as the penultimate row (.61 - Hi) of part (a) of Table (4) shows, more than 60% students answered correctly three times more items in 1995 than they did in 1993. Change in these indices is a sign of improvement in mathematics achievement.

Figure (1)
Distribution of the Difficulty Index Pi in the Pre-Reform and Post-Reform Population



In 1995 only 2 (6%) of the test items fell in the 1st category and only 5 (15%) in the 2nd category. That is, in 1995 there were only 21% (compared to 55% in 1993) of the test items that were answered correctly by at most 30% of the students or one can say that 70% of the students failed to answer only 21% items in 1995 as compared to 55% items in 1993.

Looking at the last two categories of Figure (1) and Table (4a) we also note that in 1995, 8 items (23%) were answered correctly by 51% to 69% of the students while in 1993 only 3 items (9%) were answered correctly by 51% to 65% of the students. More items being answered correctly by larger numbers of students is a sign of improvement the education reform program is striving for.

Post-Reform Improvement in Math Achievement of G4 Student Population

From individual items we move up to whole test and subtest mean scores of 1993 and 1995 student populations.

An unequal independent samples t-test was applied to compare the performance of pre-reform and post-reform samples on the whole test.

The six content subtests and the three skill substests comprised two disparate sets of tests. Each set was subjected to Hotellings T^2 with unique sum of squares option in the SPSS MANOVA procedure. The univariate null hypothesis for the whole test and the multivariate null hopthesis of each set of subtests were rejected at ($p < .000$) level of significance. These results show statistically significant improvement in the math achievement of Grade 4 students in 1995 over their 1993 cohorts.

Results

Having rejected the multivariate null hypotheses we examined the differences on individual scales. The results of univariate tests are presented in table (5) which includes the means and standard errors of estimates of the means of post-reform and pre-reform populations, the difference between the two means and its level of statistical significance on each subscale. As the last column of Table (5) shows, all the differences were significant at ($p<.000$) level of significance.

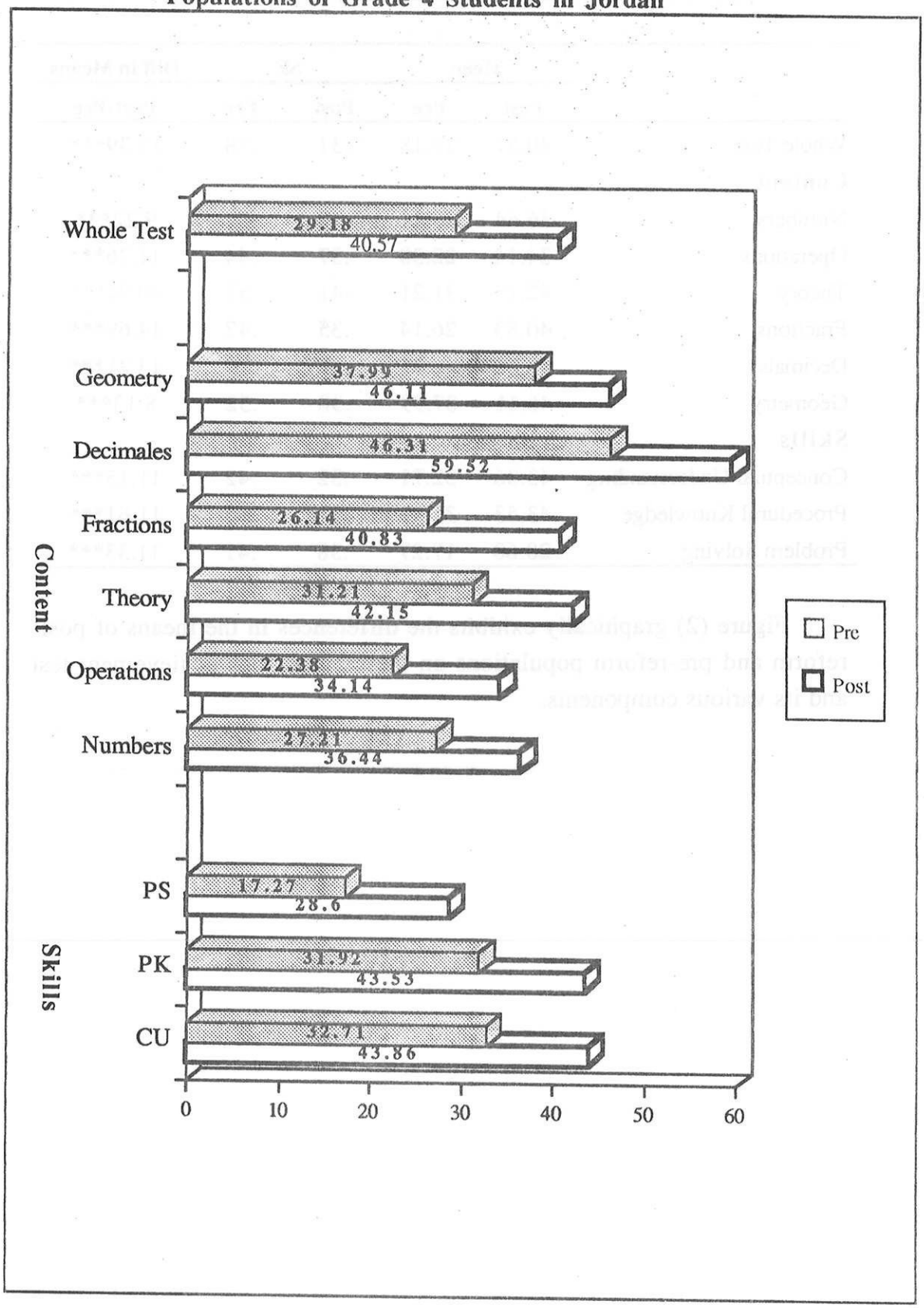
Table (5)

Differential Performance (Percent Correct Score) of Pre-and Post-Reform G4 Populations on the Math Test Components (N: Post = 5010, Pre = 2428)

	Mean		SE		Diff in Means
	Post	Pre	Post	Pre	Post-Pre
Whole Test	40.57	29.18	.31	.38	11.39***
Content					
Numbers	36.44	27.21	.45	.55	9.22***
Operations	34.14	22.38	.37	.44	11.76***
Theory	42.15	31.21	.41	.57	10.94***
Fractions	40.83	26.14	.35	.42	14.69***
Decimals	59.52	46.31	.53	.79	13.21***
Geometry	46.11	37.99	.38	.52	8.13***
Skills					
Conceptual Understanding	43.86	32.71	.32	.42	11.15***
Procedural Knowledge	43.53	31.92	.34	.45	11.61***
Problem Solving	28.60	17.27	.38	.41	11.33***

Figure (2) graphically exhibits the differences in the means of post-reform and pre-reform populations on the mathematics achievement test and its various components.

Figure (2)
Math Test Performance of Post-Reform and Pre-Reform Populations of Grade 4 Students in Jordan



On the whole test there is 11% increase. Among the six content subscales the gain ranges from 8% (on Geometry) to 15% (on Fractions). Along the three skill subscales there is 11% increase in domains of Conceptual Understanding and Problem Solving, and 12% increase in the domain of Procedural knowledge. Variation in improvement across content areas perhaps reflects differential curricular and/or instructional emphasis on different topics in the curriculum and perhaps more so in teachers' instructional practices.

Regional Differences in Reform Impact

Having described the post-reform gains in the mathematics achievement scores of the fourth grade student population at the national level now we investigate the state of improvement within each region (Governorate). Since the time this study was designed to monitor student achievement levels the number of Governorates has been increased from 8 to 12. The following analysis, however, is based upon the old sampling classification of schools in the 8 Governorates. The intention here is to examine whether the improvement in achievement scores is uniform in all regions or it varies considerably from region to region.

To study whether reform gains are uniform across all regions or there are any regions where implementation of educational reform did not raise student math achievement scores, the post-reform and pre-reform math test scores were compared separately within each Governorate by means independent unequal sample t-test.

The results of the t-tests conducted for all the Governorates are presented in Table (6).

Table (6)

Post-Reform Gain in Math Achievement of Grade 4 Students Across Governorates (total score means, difference between post-reform and pre-reform means, t-value, and statistical significance)

Govern.	Post-Reform		Pre-Reform		Post $\bar{x}$ - Pre $\bar{x}$			
	N	$\bar{x}$	N	$\bar{x}$	Diff.	t-value	2-Tail Sig.	DF
Amman	1706	43.74	834	31.06	12.68	15.10	.000	1844.9
Zarqa	749	44.92	379	31.84	13.08	9.55	.000	485.4
Balqa	263	41.96	94	30.62	11.33	5.71	.000	237.1
Irbid	1180	38.21	545	27.94	10.27	9.97	.000	1235.4
Mafraq	380	36.80	203	28.30	8.50	5.40	.000	453.1
Karak	324	36.76	131	17.89	18.78	10.41	.000	366.6
Tafileh	114	27.04	46	26.92	0.12	0.04	.965	106.0
Ma'an	294	33.64	196	27.73	5.91	3.51	.000	428.9

In Table (6) there are eight columns each representing a statistic and eight rows (each representing a Governorate).

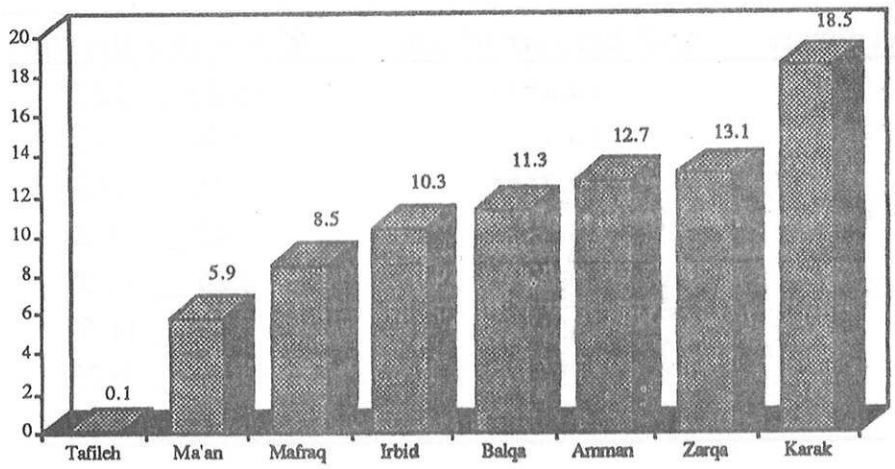
A glance down the 'Sig' column shows statistically significant ($p \leq .000$) post-reform gains in all Governorates except one.

The exception is the Governorate of Tafileh where the value of post-reform test score mean (27.04) and pre-reform test score mean (26.92) is almost identical.

Karak Governorate in contrast has recorded the largest gain (18.8%). Judging from the comparison of observed values of the mean gain score among the Governorates, Karak stands 1st, Zarqa 2nd, Amman 3rd, Balqa 4th, Irbid 5th, Mafraq 6th, Ma'an 7th, and Tafileh the last where there is virtually no gain at all.

Figure (3) presents a bar diagram of the relative gain scores in the eight Governorates arranged in ascending order.

Figure (3)
Post-Reform Gains in the Math Test Scores in the Eight Governorates



The observed differences on the post-reform gain score in different Governorates may paint a misleading picture of the real landscape. In order to examine the significant differences in reform gains across the Governorates we compared the 95% Confidence Intervals for the difference between post- and pre-reform means for different Governorates. This information is presented in Table (7). Nonoverlapping intervals signify significant differences.

Table (7)
Mean Math Gain Scores and Their 95% Confidence Intervals in the Eight Governorates of Jordan

Governorate	Diff. Between Means	95% CI for the Diff.
Karak	18.8***	15.23 __ 22.33
Zarqa	13.1***	10.39 __ 15.77
Amman	12.7***	11.03 __ 14.33
Balqa	11.3***	7.42 __ 15.22
Irbid	10.3***	8.25 __ 12.29
Mafraq	8.5***	5.41 __ 11.59
Ma'an	5.9***	2.60 __ 9.23
Tafileh	0.1***	-5.25 __ 5.49

Note:

a. Karak is significantly better than all other Governorates except Zarqa.

b. Zarqa is significantly better than Ma'an and Tafileh.

c. Amman is significantly better than Ma'an and Tafileh.

d. No significant differences among other Governorates.

e. Significance is established by nonoverlapping 95% CIs.

It is informing to note that Karak Governorate's schools have shown largest gain in 4th grade students mathematics achievement since the implementation of educational reform program. In terms of post-reform math test performance, however, Karak ranks only 5th among the eight Governorates, while at the time of pre-reform testing its rank was eigth.

Although there is no significant difference between the gains of Zarqa and Amman Governorates, both of them have shown significantly better gain than the Governorates of Ma'an and Tafileh.

There are no significant differences on the post-reform gain score among other Governorates.

Differential Impact Among Education Authorities

Post-reform and Pre-reform math test performance of Grade 4 students was compared separately for the four education authorities in Jordan.

The data on number of students taking the test and their mean test score for each occasion of testing, the mean difference score, the unequal independent sample t-value, unequal sample Degrees of Freedom and 2-Tail significance of the difference is presented in Table (8).

Table (8)

Post-Reform Gain in Math Achievement of Grade 4 Students Across Education Authorities (Total score means, difference between post-Reform and Pre-Reform means, t-value, and statistical significance)

Authority	Post-Reform		Pre-Reform		Post $\bar{x}$ - Pre $\bar{x}$				95% CI
	N	$\bar{x}$	N	$\bar{x}$	Diff.	t-value	DF	2-Tail sig	
MOE	3889	39.53	1867	26.99	12.55	23.29	4328.5	.000	11.49--13.61
MOD	125	33.01	81	41.39	-8.38	-2.95	167.9	.004	-13.97-- -2.78
UNRWA	449	36.11	212	30.37	5.74	3.44	455.6	.001	2.46 -- 9.02
Private	547	53.34	268	39.86	13.48	8.88	613.4	.000	10.50 -- 16.46

Note:

MOE = Ministry of Education.

MOD = Ministry of Defence.

UNRWA = United Nations Relief and Works Agency.

The 2-Tail statistical significance column in Table (8) clearly shows that the change in test performance of students in all the education authorities is highly significante.

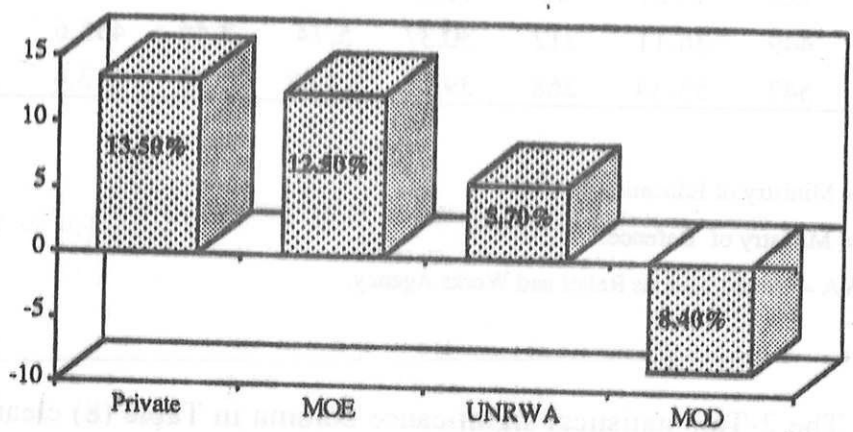
Also in three education authorities namely, MOE. UNRWA, and Private, the change is, as anticipated, in the positive direction. In the performance of both Private and MOE schools there is about 13% improvement. Nearly 6% improvement in the math test performance of UNRWA schools, however, is only nominal.

On the other hand, contrary to expectations, the significant change in the MOD schools is in the negative direction. The (Grade 4) students' math test performance in the MOD schools has decreased over 8%. While there is no transparent explanation, two alternative hypotheses can be

postulated. One relates to administration of the Grade 4 mathematics test in the MOD schools; the other refers to real decline in instructional standards. Neither savors pleasant.

Regarding relative gain among the four education authorities, Table (8) presents the observed gain scores and their 95% Confidence Intervals. The observed mean gain scores for the four authorities are displayed by Figure (4). On the basis of the observed gain values, one can see that private schools with 13.50% average gain rank first, then, MOE, neck and neck, with over 12.5% gain comes second. UNRWA with a modest gain of nearly six percent takes the third place. MOD, as stated earlier, has shown decrement of over eight percent.

Figure (4)
Percent Gain in Different Education Authorities



Statistically speaking, however, there is no significant difference between the gain of MOE and Private schools; whereas, both Private and MOE schools' gain scores are significantly better than the mean gain score of UNRWA schools. MOD schools, in contrast, have suffered a significant loss in terms of depleted performance on the (Grade 4) math test.

Comparative Gains of Male and Female Students

Table (9) presents the pre-and post-reform math test means, respective sample size, mean gain score, unequal sample t-value, unequal

sample Degrees of Freedom and 2-Tail probability for male and female groups separately.

Table (9)

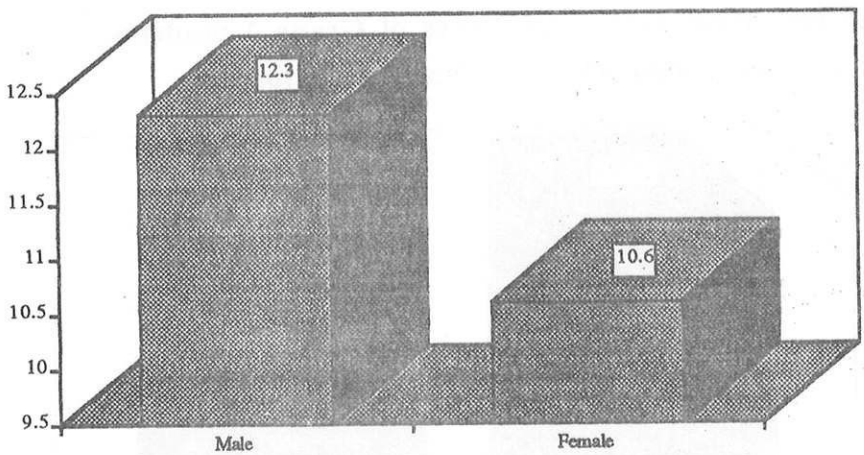
Post-Reform Gain in Math Achievement of Grade 4 Students Across Student Gender (Total score means, difference between post-reform and pre-reform means, t-value, and statistical significance)

Student	Post-Reform		Pre-Reform		Post $\bar{x}$ — Pre $\bar{x}$				95%
Gender	N	$\bar{x}$	N	$\bar{x}$	Diff.	t-value	DF	2-Tail Sig	CI
Male	2581	40.31	1130	28.02	12.29	17.14	2452.6	.000	10.89 --13.70
Female	2426	40.82	1298	30.20	10.63	15.64	3065.86	.000	9.30 -- 11.96

Both groups have shown significant ($p < .000$) improvement in their math achievement after four years of reform.

Figure (5) presents the observed mean gain scores of the Male and Female subpopulation's of Grade 4 students.

Figure (5)
Mean Gain Scores of Grade 4 Male and Female Groups of Students on the Math Test



Despite slightly higher mean gain score of the male students, there is no statistically significant difference between the average improvement of male and female students. The 95% CIs for the mean gains of male and female students are highly overlapping.

Comparative Gains in Urban and Rural Schools

For the country as a whole it was investigated whether reform effects vary in urban and rural schools. The gain scores separately analysed for urban and rural school students are presented in Table (10).

Table (10)

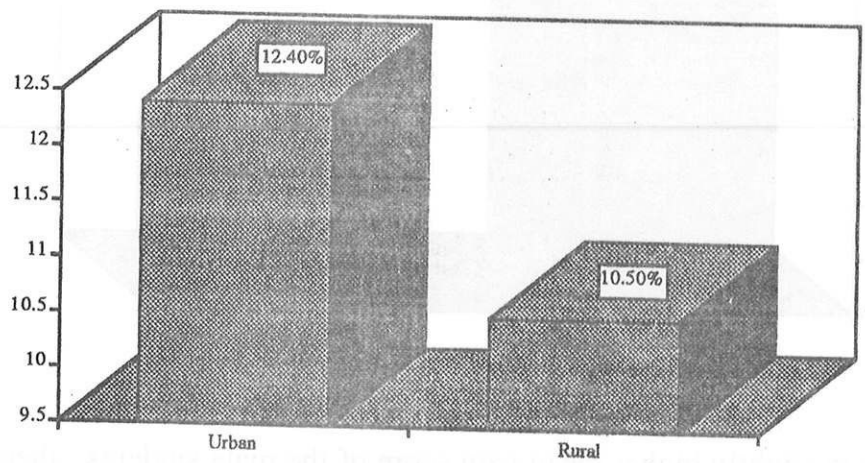
Post-Reform Gain in Math Achievement of Grade 4 Students Across Urban and Rural Schools (Total score means, difference between post-reform and pre-reform means, t-value,and statistical significance)

Location	Post-Reform		Pre-Reform		Post $\bar{X}$ - Pre $\bar{X}$				95%
	N	$\bar{X}$	N	$\bar{X}$	Diff.	t-value	DF	2-Tail Sig	CI
Urban	2310	43.80	1110	31.41	12.39	16.77	2481.97	.000	10.94 -- 13.84
Rural	2700	37.81	1318	27.31	10.50	16.10	3031.37	.000	9.22 -- 11.78

It is clear from Table (10) that math test performance of Grade 4 students has significantly improved in both urban and rural schools. The urban schools have shown 12.4% improvement and the rural schools 10.5% improvement. Comparative gain of urban and rural schools is displayed in Figure (6).

Figure (6)

Mean Math Test Gain Scores of Grade 4 Students in Urban VS Rural Schools of Jordan



Looking from the observed mean gain scores, urban schools seem to have an edge over rural schools. Comparison of the 95% CIs for the two means, however, reveals that observed difference in the two means is not statistically significant at ($\alpha = .05$) level of significance. But this does not mean that there are no differences in the achievement levels of urban and rural schools.

Also most high quality private schools are located in big cities and they further tilt the balance in favour of urban schools. The issue addressed here is of the value added due to reform elements, that is, the gains over their performance in the base year (1993). In that respect, the data here show that urban and rural area schools have benefited almost equally by the introduction of educational reform in Jordan. This, however, does not imply that within different regions there are no specific differences in the gain scores of different groups of schools.

Summary of Results

1. At the national level there is significant improvement in the math achievement of Grade 4 post-reform students. This improvement covers all components of the test, the six content areas and the three cognitive skills. While gain across the three cognitive skills is almost uniform, the gain across content areas varies from 8% in Geometry to about 15% in Fractions.

2. The reform gains vary across the eight Governorates. All Governorates with the only exception of Tafileh have recorded positive gains but to a varying degree. Besides Tafileh with zero gain, the lowest significant gain (6%) is in Ma'an while the highest (19%) is in Karak.

3. With respect to education authority, all education authorities have reported significant positive gain expect the MOD where there is significant decline of over 8%.

   Among the three positively gaining authorities, both MOE and Private have gained significantly better than UNRWA, while there is no significant difference between the average gain score of Private and MOE schools.

4. The reform gain is almost uniformly reflected in the achievement of both male and female student populations.

5. While both urban and rural schools have registered significant gain, there is no significant difference between the average gain of urban and rural schools.

CONCLUSIONS AND DISCUSSION

This report, focused upon tracking students' mathematics achievement from pre-reform status to post-reform conditions. Apparently, there is evidence of improvement in the math achievement of Grade 4 students whether measured interms of individual items, content or cognitive skill subscales or the total test score; all indicators consistently provide evidence of improvement. Post-reform population performance is significantly better than the performance of pre-reform population in all areas of mathematics achievement included in this survey.

Having concluded about definite evidence of improvement of student achievement levels in mathematics we ponder over the thorny issue of attributing the improvement to educational reform. To study differences in the performance of two different populations at two points of time is one thing but to account for the observed change and relate it to legitimate causes is quite a different ball of game. On one hand, it is inconceivable to assert that nothing in Jordan has changed from 1992 to 1995 except the introduction of educational reform in the schools; while, on the other hand, no significant changes have occurred in peoples' way of living or thinking or in their social and cultural values. Theoretically, however, there can be infinite rival hypotheses each claiming for the cause of improvement in student achievement levels. In the absence of controls which are impossible to be exerted without upsetting the natural conditions of national education system, it is extremely difficult to refute the rival hypotheses. Nevertheless, researchers who study social environments in their natural settings have developed a line of argument which runs as follows.

We observed the phenomenon (students' math achievement) as it prevailed in 1993 when Grade 4 students had spent 4 years in schools under normal conditions before the introduction of reform program. In the

same schools we measured students' math achievement on the same math test again in 1995. This time the students had spent 4 years in the same schools but under conditions of reform. No cataclysmic events have occurred during this period, nor any significant social, cultural, economic changes have taken place to alter the living conditions, value system, attitudes or outlook of people and communities in Jordan. Therefore, one can safely assume that schools and their feeding communities are essentially the same in 1995 as they were in 1993 except the implementation of certain reform elements in all schools.

New curricula and textbooks have been introduced and teachers have attended the inservice training program in how to apply modern effective instructional techniques and instructional materials in classrooms.

Apart from introduction of these reform components no other substantive changes have taken place during this period of time. The only conceivable differences in the Grade 4 population of 1993 and that of 1995 may lie in the fact that 1993 students completed their primary education before the introduction of reform program but 1995 Grade 4 students completed their four years of schooling under reform conditions.

If we can accept this assumption then we can see that the two populations are comparable in all respects except the reform treatment. Pre-reform population serves as control whereas 1995 population has gone through treatment conditions. This directly leads to attributing the improvement of student achievement level in 1995 to the reform elements that had been implemented by 1995, mainly: reconstructed curricula, new textbooks and teaching materials, and inservice training of teachers and principals, and all other cognitive and affective factors that imperceptibly develop with the introduction of reform elements.

In essence it is the integrated conglomerate reform environment, the whole reform context, rather than isolated elements of reform that impact improvement. Following the preceding line of thinking we are inclined to assert that, reform program as a whole, has started showing signs of positive impact in some areas of student achievement and educational progress, which is encouraging for all those who have any stake in the educational reform.

(Part II)

Value - Added Math Achievement of Grade 8 Students Under Four Years of Educational Reform

Value - Added Math Achievement of Grade 8 Students Under Four Years of Educational Reform

The Context

In 1989 Jordan embarked upon an ambitious educational reform program to restructure and revitalize its basic and secondary education system. The main goal of reform focused at enhancing student achievement levels. The key reform elements: reconstructed curricula, newly designed textbooks and instructional materials, and inservice teacher training in classroom applications of innovative instructional methods for using new textbooks and materials were introduced for the first time in 1st, 5th, and 9th grades in the scholastic year 1991-92. Next year they were introduced in 2nd, 6th, and 10th grades. By the year 1994-95, educatinal reform had covered nearly all the school grades (1-12).

Recognizing the fact that the ultimate test of the impact of educational reform depends upon enhanced student achievement levels, the National Center for Educational Research and Development (now renamed as the National Center for Human Resources Development (NCHRD)) designed a reform impact evaluation program to assess improvement in instructional quality reflected by student achievement levels. The assessment design entailed a stratified two-stage random national sample of 245 schools. One section was randomly selected if schools contained more than one sections in the 8th grades. (For detail about sampling characteristics and domains of variables tested, see Ahlawat, 1993; Ahlawat, et al., 1994 a, 1994b).

Students were tested in three subjects, Arabic Language, Mathematics, and Science, at the end of the 1992-93 scholastic year. In addition to achievement in the three subjects, questionnaires were administered to students, parents, teachers and principals to gather various types of information on attitudes, beliefs, and perceptions about schools, classrooms, instruction and management practices. (see Annex I for domains of variables included in the study).

At the first stage information was collected primarily to serve as baseline data to study changes (if any) after about four years of educational reform. The same achievement tests and questionnaires were administered again to the 8th grade students in the same schools. The second testing took place after the students had completed 4 years under the reform program using new curricula and textbooks taught by teachers who had received inservice training in innovative instructional techniques recommended to foster critical thinking, problem solving, self-learning, and higher level cognitive skills in students.

While the whole assessment program covered achievement testing in three major subjects (Arabic, Mathematics, and Science), this study concentrates on Mathematics achievement alone.

The samples of 8th grade consisted of 2,484, and 3,747 students, respectively, for the years 1993 and 1995. In 1993, however, only about one half of the students in each sampled class took the mathematics test. Thus the mathematics test sample in 1993 consisted of 1750 8th grade students. In the year 1995, the mathematics tests were administered to the whole classes. The comparisons of performance on the math test of 8th grade classes between the years 1993 and 1995 were based on the same school units. It should be clear that schools and grades were the constant units of comparison while the student populations on the two testing occasions were evidently different. Thus, the same grades were measured on the same tests in the same schools on two different occasions. On the first occasion students of 8th grade had no exposure to the elements of education reform, whereas, on the second occasion the students of 8th grade had been exposed to reform processes continuously for four years.

The following section presents a brief description of the contents and composition of the 8th grade achievement test of mathematics and its psychometric properties on both testing occasions (1993, 1995) based upon the performances of samples.

Composition of the 8th Grade Math Achievement Test

The 8th grade mathematics achievement test consisted of 46 items of which 30 were multiple-choice and 16 supply-response. The test was designed to cover both new and old curricular contents and objectives. Generally, the old and new curricula shared the same contents but a few new topics namely, Probability and Trigonometry were added to the newly constructed math curriculum while a few topics were shifted to the 7th grade textbook.

The test was designed to study improvement in student achievement levels (if any) after the introduction of new curricula, textbooks and other changes stipulated by the comprehensive basic .and secondary education reform in Jordan. The test covered three broad cognitive skills (Conceptual Understanding, Procedural Knowledge, and Problem Solving) and eight content topics (Basic Concepts, Algebra, Geometry, Measurement, Ratios, Numbers, Probability, and Trigonometry). The last two topics i.e., Probability and Trigonometry, to which five test items were devoted, were not covered by the old (pre-reform) curricula. When the test was reproduced for the 1995 testing, faults crept in two of the items, so they were discarded. The three Probability items and two Trigonometry items were analysed separately.

Therefore, math achievement comparison of the pre-reform and post-reform 8th grade students' populations were based on test and sub-test scores computed from various combinations of the 39 items which were covered by both old and new curriculums.

Table (1) presents the distribution of the 39 test items across various cells of a table of specifications.

Table (1)

Table of Specifications for the 39 Math Test Items Common to Both Old and New Curricula for the Eighth Grade

Content / Skills	Conceptual Understanding (CU)	Procedural Knowledge (PK)	Problem Solving (PS)	Total
Basic Concepts (BAS)	3	3	--	6 15%
Algebra (ALG)	4	6	--	10 26%
Geometry (GEO)	5	2	--	7 18%
Measurement (MEA)	--	7	--	7 18%
Ratios (RAT)	2	1	1	4 10%
Numbers (NUM)	2	3	--	5 13%
Total	**16** **41%**	**22** **56%**	**1** **3%**	**39** **100%**

As can be seen from Table (1) there is relative preponderance of Algebra items (26%) as compared to only 10 percent items from the topic of ratios. Among the three cognitive skill areas, 'Problem Solving' has only a nominal representation. There is only one item (3%) in the whole test that falls into this category. This however is a different issue. At the moment it should suffice to note that in this study the post-reform and pre-reform student populations' math achievement has been compared on the basis of the same 39 items.

In the following section we compare the classical psychometric properties of the test between the post-reform and pre-reform samples of 8th grade students. We remind the reader that in the post-reform testing all the students in a sampled class took the math test whereas in the pre-reform testing only one half of the students in a sampled class were administered the math test, the other half took the science test.

Psychometric Properties of the Grade 8 Math Test in the Pre-Reform and Post-Reform Populations

Reliability Indices of the Whole Test and Subscales in the Post and Pre-Reform Populations.

Table 2 presents the alpha coefficients of the whole test and of all the subscales computed from the two populations.

Table (2)
Math Test Scale's and Subscales' α - Reliability Coefficients for Pre and Post-Reform Grade 8 Populations (N: Post = 3747, Pre=1750)

Scale/Sub-Scale (No. of Items)	α– Coefficient	
	Post-Reform	Pre-Reform
Whole Test (39)	.74	.70
Basic Concepts (6)	.18	.31
Algebra (10)	.36	.38
Geometry (7)	.31	.13
Measurement (7)	.62	.52
Ratios (4)	.17	.29
Numbers (5)	.24	.27
Conceptual Understanding (16)	.46	.45
Procedural Knowledge (22)	.67	.60

Cranbach's alpha reliability coefficients were computed for the whole test, for every one of the six content subscales and the two cognitive skill subscales (the third, skill scale 'Problem Solving' had only one item) for each of the two populations. The magnitude of α -coefficient reflects the consistency in the sample's performance over all items defined for a scale or subscale. Higher values of the coefficient indicate homogeneity of performance across items and lower values indicate the reverse, lack of homogeneity of student performance across items. The homogeneity coefficients range from moderate to low. For a 39-item math test, covering six divergent content areas and three different skills, the homogeneity reliability coefficients of .74 and .70 for the post-reform and pre-reform samples, respectively, are reasonably good.

Among the six items, measuring basic concepts, the homogeneity of student performance has decreased by 13%. For the pre-reform sample it was .31 but for the post-reform sample it fell down to .18. For the 10-item Algebra subscale, the degree of homogeneity is about the same. Also for Numbers subscale (5 items) it is about the same. For the seven Geometry items it has increased from .13 in the pre-reform sample to .31 in the post-reform sample, and for the seven Measurement items it has increased from .52 to .62. The 4-item Ratios subscale has shown a 12 points' decrease from .29 in the pre-reform sample to .17 in the post-reform sample.

The homogeneity of performance over the 16 items tapping Conceptual Understanding has not changed. It was .45 for the pre-reform sample and .46 for the post-reform sample. Over the 22-item Procedural Knowledge subscale it has increased from .60 in the pre-reform sample to .67 in the post-reform sample.

Given the same set of items increase in homogeneity of students' performance may be interpreted as better grasp over the specifics and consolidation of knowledge and understanding of the various aspects connected with a defined area of knowledge and content; the decrease may be interpreted as its converse.

The Indices of Difficulty and Discrimination Power of the Items.

Difficulty and Discrimination indices of the items for both post-reform and pre-reform samples are presented in Table (3). Difficulty index of an item is actually an index of easiness of the item. In fact, it is the proportion of students answering the item correctly. The larger the value of the difficulty index the easier the item.

Table (3)
Comparison of Difficulty and Discrimination Indices of Items Between the Post -Reform and Pre-Reform Samples (N: Post-Reform = 3747, Pre-Reform = 1750)

		Difficulty (Pi)		Difference	Disc. Item-Remainder Corr.	
S. No.	Item	Post	Pre	Post-Pre	Post	Pre
1	M1	.55	.28	.27***	.16	.23
2	M2	.18	.27	-.09***	.14	.31
3	M3	.43	.47	-.04*	.27	.25
4	M4	.23	.19	.04***	.00	.12
5	M5	.62	.41	.21***	.19	.21
6	M6	.61	.49	.12***	.42	.41
7	M7	.45	.33	.12***	.48	.41
8	M8	.68	.47	.21***	.38	.38
9	M9	.39	.21	.18***	.23	.14
10	M10	.58	.44	.15***	.28	.15
11	M11	.31	.32	-.01NS	.17	.17
12	M12	.42	.37	.06***	.10	.08
13	M13	.45	.57	-.11***	.26	.32
14	M14	.35	.07	.28***	.40	.11
15	M15	.36	.28	.08***	.41	.36
16	M16	.57	.47	.01***	.13	.15
17	M17	.02	.02	.00NS	.26	.19
18	M18	.06	.04	.02**	.28	.29
19	M19	.37	.42	-.06***	.12	.19
20	M20	.29	.23	.06***	.14	.01
21	M23	.49	.42	.07***	.05	-.04
22	M24	.51	.43	.08***	.43	.38
23	M25	.21	.14	.07***	.22	.18
24	M26	.33	.22	.11***	.33	.08
25	M27	.28	.17	.11***	.04	.05
26	M29	.30	.25	.05***	.27	.13
27	M30	.23	.18	.05***	.12	.04
28	M31	.39	.20	.19***	.45	.34
29	M32	.39	.17	.21***	.43	.32
30	M33	.28	.43	-.16***	.17	.26
31	M35	.28	.21	.07***	.25	.15
32	M36	.30	.22	.08***	.21	.01
33	M37	.26	.27	-.01NS	.09	.11

Table (3) continued

34	M38	.24	.29	-.05***	.01	.18
35	M39	.04	.06	-.02***	.33	.32
36	M40	.004	.04	-.04***	.14	.20
37	M41-A	.51	.39	.12***	.11	.32
38	M42	.14	.07	.07***	.41	.32
39	M43	.41	.30	.11***	.19	.04
	Mean ($\bar{p}$)	.35	.28	.07	.23	.21
	MIN	.0035	.02	-.16	.00	.01
	MAX	.68	.57	.28	.48	.41
	VAR	.20	.18	.01	.02	.01

Note:

* = p≤.05

** = p≤.01

*** = p≤.000

In Table (3), the column titled "Difficulty Pi" is divided into two columns. The "post" column contains the item difficulty index (proportion of students who answered the item correctly) for the post-reform population. Likewise, the column headed "Pre" contains item difficulty index for the pre-reform population of students.

The "Difference" column contains the value of difference between the post-reform and pre-reform difficulty indices of an item. It is generally expected that in the post-reform sample the proportion of correct responses will increase. If this happens then the difference will be positive. If no change has occurred then the difference will be zero. If post-reform students' performance on the item has decreased then the difference will be negative.

We tested the statistical significance of these differences by unequal-independent-sample two-tailed t-test. If the difference is not significant at the .05 alpha level, it is marked NS (Not significant). Whether positive or negative, if the difference is significant it is marked by appropriate number of asterisks. A single asterisk represents statistical significance at less than or equal to (p≤.05) five percent level but greater than one percent level. Two asterisks indicate the statistical significance at the level ranging from

one percent to greater than one per thousand. Three asterisks indicate the significance level of one per thousand or less.

The column headed "Disc. Item-Remainder Corr." contains the coefficient of correlation between the item and the total score computed over the rest of the items in the test.

This item-remainder correlation coefficient is called a discrimination index of the item. The value of discrimination index indicates the power of the item to differentiate between high and low performing groups of students when performance is measured by a student's total test score computed without the score on the item whose discrimination power is being estimated.

If students' ability to answer the item in question is related to their abilities to answer the remaining items in the test then the item's discrimination index represents the strength of this relationship.

The direction of relationship (if the relationship exists) can be positive or negative. On account of the logic of interpretation of the test score, items with negative discrimination are recommended to be discarded from achievement tests. An item will show a near zero discrimination index if: (i) it is ambiguous and therefore interpreted by different students in different ways irrespective of their achievement level in the subject; (ii) the knowledge and skills required to answer this item share nothing in common with ability and skills tapped by other items in the test; (iii) it is so easy that all examinees answer it correctly; (iv) it is too difficult to be answered even by the very bright students; (v) it has been miskeyed. An achievement test item may have a negative discrimination if, due to some reasons, low achieving students answer the item correctly and high achieving students answer it wrong.

Bearing of the Difficulty Index on Change in Performance.

When an item is scored (0/1), its difficulty index, the proportion of students who answered the item correctly is actually the sample mean score of the item. Thus, for each item in the test, we have average performance of the post-reform sample, as well as, that of the pre-reform sample. Given this, we can directly see the difference in the performance

of pre-reform and post-reform populations of the 8th grade students in Jordan on each individual item in the test.

Examining the entries in "Difference" column in Table (3) we see that:

1. Three out of 39, nearly 8%, of the test items register no difference whatsoever between the performances of pre-reform and post-reform populations of the 8th grade students.

2. On 8/39, about 20% items, post-reform students' performance has deteriorated significantly.

3. On the brighter side, on 28 out of 39 (72%) items post-reform students have outperformed their pre-reform cohorts at a high level of statistical significance ($p \leq .000$).

The three items on which there is no significant post-reform improvement belong to rather abstract theoretical and analytical topics in the math curriculum. One item belongs to irrational numbers, one to factorization and one to algebraic analysis. One item is supply response and two are multiple choice. In both testings only 2% students answered the supply-response item correctly. The performance on both multiple - choice items is just above chance level and discrimination power is very low. It seems that teachers are not skilled enough to present these concepts in a way that students could have a good grasp of them. Although full chapters are devoted to these topics in the textbook, inservice training under reform does not seem to have impacted the instructional skills of teachers in these topics. These areas were weak before the reform and these are still weak after four years of reform.

The 8 items on which performance has decreased can be classified into two groups: (1) items which were taught in the 7th grade, and (2) items that related to analysis. Two items belong to the topic which has been shifted down to Grade 7 and one belongs to the topic which was covered in the beginning of the year. The performance on these items has deteriorated perhaps due to forgetting effect. Students, everywhere, try to remember things to pass the exam. A substantial portion of learning that was driven by examination is rapidly lost after the exam is over, particularly if the learned material has no chance for later application.

In the pre-reform testing the topic of sets was taught in the eighth grade, while in the new curriculum it has been shifted down to 7th grade. Two factors are at work here. First, the way the teachers introduce new concepts of mathematics do not help pupils understand them clearly. Second, students who learnt the concepts to pass the 7th grade exam had no motivation to load their memory with the knowledge that had no application in their real life after the exam is over. A combination of such factors has resulted in lowering the performance of post-reform students on the contents which were taught in the 7th grade.

In spite of the fact that the reform program overly emphasizes the analytical skills and critical thinking, on almost all the items of mathematical analysis post-reform students' performance has shown significant decline from the baseline performance standard of the pre-reform students. Although, student performance in this area has always been poor. For instance, one of the four items was answered correctly by only 4% of the pre-reform and 0.4% of the post-reform students. On another item from the same group the performance was 6% and 4% respectively for pre-reform and post-reform students. On the face value, the items look simple, straightforward, and ordinary. This raises the question, how the whole area of analysis is handled in the 8th grade classrooms.

On a pleasanter note, there is statistically significant improvement on 72% of the test items, even though the gains are low to moderate, ranging from 2% to 28%. The average gain on these 72% items is .11 with a standard deviation of .07. The comparison of distributions of item difficulty index in the two samples can provide information about the differences in their performance. The frequency distributions of the difficulty and discrimination indices of the two samples are presented in Table (4).

Table (4)

Comparative Frequency Distributions of the Difficulty and Discrimination Indies of the Grade 8 Math Test Items From the Two Samples (Pre-Reform (1993) and Post-Reform (1995))

(a)	Difficulty Index			
Interval	Post-Reform		Pre-Reform	
	Freq.	% Freq.	Freq.	% Freq.
Lo - .20	6	15	11	28
.21 - .30	11	28	13	33
.31 - .40	8	21	4	10
.41 - .50	6	15	10	26
.51 - .60	5	13	1	3
.61 - Hi	3	8	--	--
Total	39	100	39	100
(b)	Discrimination Index (Item-Remainder Corr.)			
Lo - .30	27	69	27	69
.31 - .40	5	13	10	26
.41 - .50	7	18	2	5
Total	39	100	39	100

From the distribution of the difficulty index in the post-reform and pre-reform populations one can see that in the lowest range of difficulty index (.2 or less) the percentage of items has decreased from 28% in the pre-reform to 15% in the post-reform sample. We restate that low values of difficulty index actually represent items with high difficulty levels.

Here the first row of Table (4) designated (Lo-.20) contains those items which were answered correctly by 20% or less than 20% of students. So, this is the category of most difficult items. The percentage of items in this category has fallen from 28 to 15.

On the easier end of the test, the last two rows of Table (4) record items that were answered correctly by more than 50% of the students in each sample. In the pre-reform sample only 3% of the test items were

answered correctly by more than 50% of the students, while in the post-reform sample the percentage of items falling in this category increased from 3 to 21.

This is an indication of relatively better performance of the post-reform students.

Figure (1)
Grouped Frequency Distribution of Difficulty Index in Post-Reform and Pre-Reform Samples

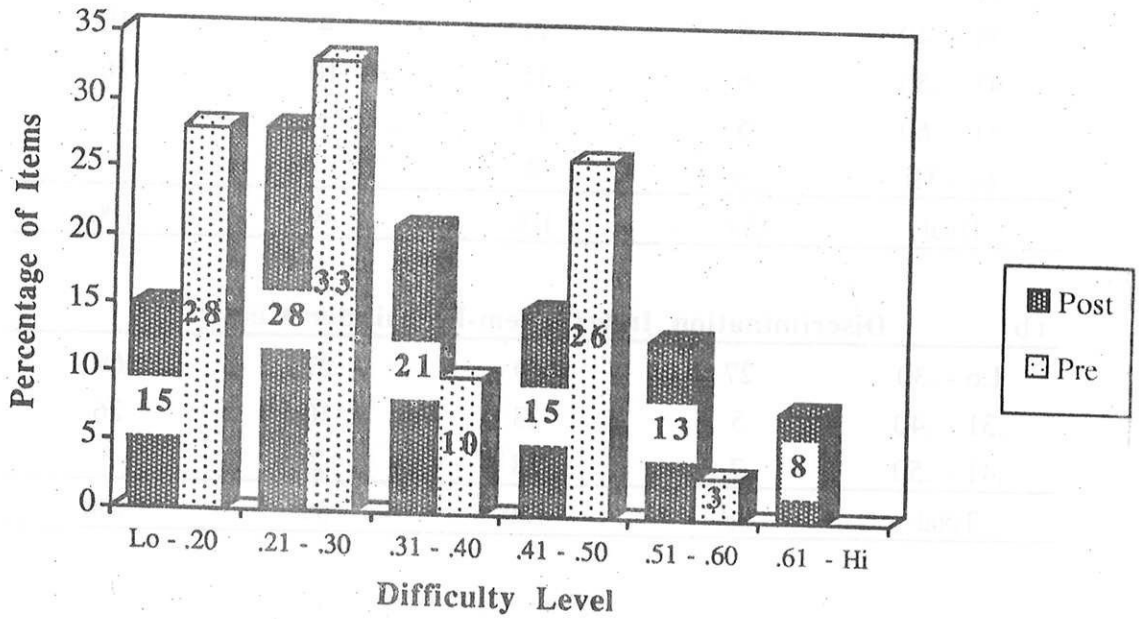


Figure (1) presents a graphic display of the two distributions of item difficulty index.

Item Discrimination Levels in the Two Populations

Regarding discrimination power of the test items in the two samples, the items were categorized into three groups; low, medium, and relatively high. As can be seen from part (b) of Table (4), in both samples the percentage of items in the low group remains the same (69%). In the moderate group the percentage of items has decreased from 26% in the pre-reform sample to 13% in the post-reform sample. In the medium discrimination power group the number of items has increased from 5% in

the pre-reform sample to 18% in the post-reform sample. Given the same test, increase in discrimination power reflects clearer conceptualization of what is measured by the test in the minds of the groups of examinees.

Post-Reform Gains in Grade 8 Math Achievement

Having compared the psychometric properties of the test in the post-reform and pre-reform samples now we examine the reform gains with respect to the total test score and individual subscale scores of the six content and three cognitive skill subscales. The percentage correct scores on each subscale and the whole test were analyzed. Unequal independent sample t-test was used to test the statistical significance of the difference between post-reform and pre-reform means on each subscale and the whole test. The results are presented in Table (5).

Table (5)

Differential Performance of Pre-and Post-Reform G8 Populations on the Math Test Components (N: Post = 3747, Pre = 1750)

Component	Mean		SE		Diff. in Means	95% CI
	Post	Pre	Post	Pre	Post-Pre	
Totscore	34.7	27.8	.22	.29	6.9***	6.2__7.9
Basic Concepts	37.5	31.1	.33	.51	6.3***	5.1__7.5
Algebra	21.4	19.1	.23	.34	2.3***	1.5__3.1
Geometry	39.1	26.9	.34	.41	12.1***	11.1__13.2
Measurement	45.9	31.4	.44	.55	14.5***	13.2__15.9
Ratios	34.6	30.6	.40	.59	4.0***	2.6__5.4
Numbers	36.1	35.1	.39	.54	1.1^{NS}	-.23__2.4
Conceptual Understanding	39.9	34.4	.26	.37	5.6***	4.7__6.4
Procedural Knowledge	31.5	23.6	.25	.31	7.9***	7.1__8.6
Problem Solving	20.8	13.7	.66	.82	7.1***	5.0__9.2

On the whole test designated by 'Totscore' row there is a gain of about 7%; post-reform students have scored significantly ($p \leq .000$) higher than their pre-reform cohorts.

The overall gain, however is composed of different components. Content-wise, it is comprised of the six content scales with varying degrees of gain. Among the six content areas the gain ranges from 14.5% in "Measurement" to no gain in "Numbers". "Geometry" with 12% gain takes the second place. "Basic Concepts" with 6% gain comes third, "Ratios" with 4% gain comes fourth and "Algebra" with a nominal 2% gain comes fifth.

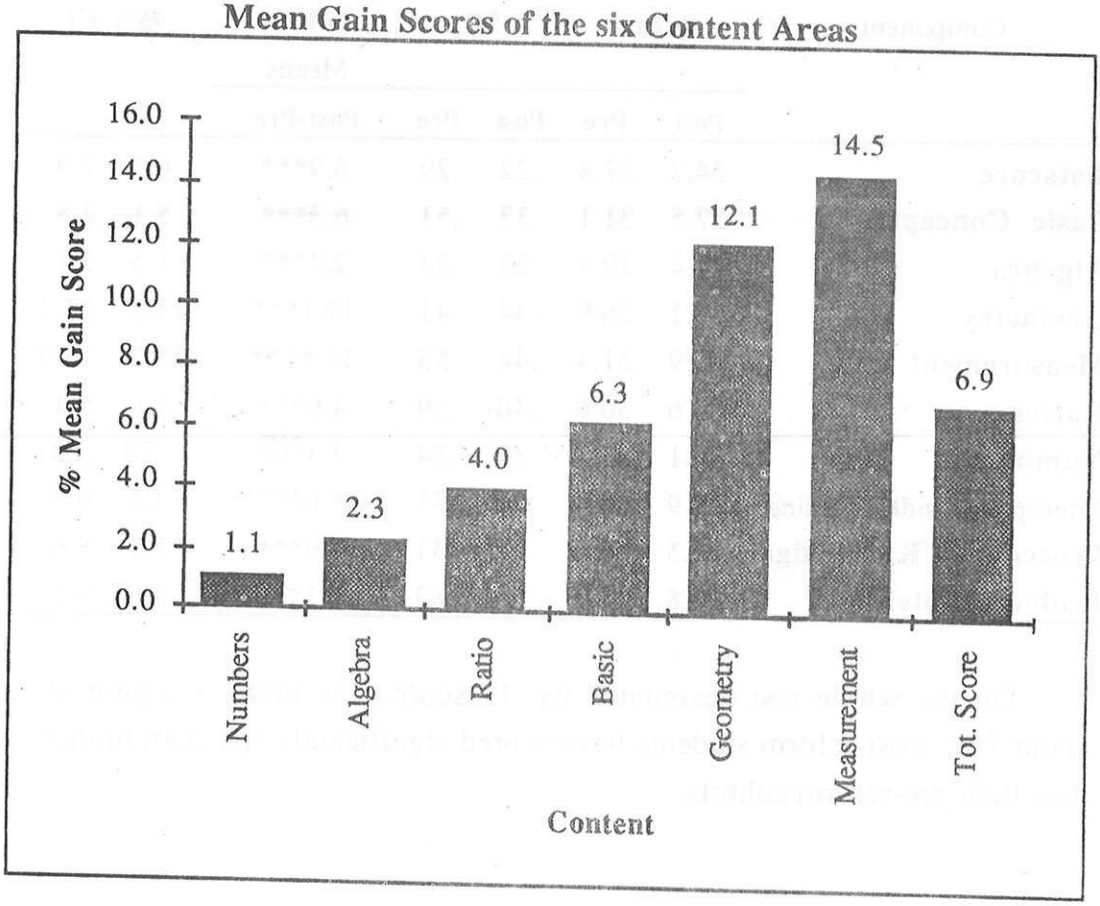
Along the cognitive skills dimension, the gain in each skill area is nearly the same, the observed mean gain score varies from 7.9 (Procedural Knowledge) through 7.1 (Problem Solving) to 5.6 (Conceptual Understanding).

Differential Gain Over Different Areas of Content

Areas of differential gain are clear from Figure (2) which displays a Bar-Graph of mean gain scores in the six content areas.

Figure (2)

Mean Gain Scores of the six Content Areas



In terms of statistical significance (α =.05) of difference in mean gains, judging from the nonoverlapping 95% Confidence Intervals for gain score means, gain in the area of measurement is significantly better than all other content areas.

Geometry gain is again significantly higher than the gain in Algebra, Ratios, and Basic Concepts. There is no significant difference between the gain means of Basic Concepts and Ratios, but gain in Basic Concepts is significantly higher than that in Algebra. We already stated that there was no significant gain in Numbers, so we left it out of comparisons.

Numbers, Algebra, and Ratios turn out to be the major areas of concern. In fact, on the 'Analysis ' items, the performance of both samples was very poor, and the items on which gains were generally negative belong to these three content areas.

Differential Gain Across the Three Cognitive Skills

Among three areas of cognitive skills, gain is significantly higher in 'Procedural Knowledge' than in both 'Conceptual Understanding' and 'Problem Solving', while there is no significant difference between the latter two skills.

We remind the reader that enhancing critical thinking and problem solving skills was an explicitly emphasized objective of the Educational Reform Program (ERP). Enhancing critical thinking and problem solving skills especially after generation-old deep-rooted habits of rote learning and practice drill is easier said than done.

It is however, encouraging to note that there is some progress in this area after the implementation of the reform.

Reform Gains In Different Education Authorities

Having described the impact of educational reform on the math achievement of the eighth grade students for the country as a whole, in this section we examine whether the reform gain is uniformly distributed among the four education authorities or it varies from one authority to another. It should be recalled that all schools, irrespective of their administering authorities, are obliged by the education law to follow the national curriculum and use the same textbooks. In order to study the reform impact in each education authority, post-reform and pre-reform means in each education authority were compared using independent unequal sample two-tailed t-test. The sample size and the mean math test scores of the 8th grade students for both (post-reform and pre-reform) testings; the difference between the two means; its t-value, Degrees of Freedom, two-tailed statistical significance, and 95% Confidence Interval, separately for each education authority are presented in Table (6).

Table (6)
Post-Reform Gain in Math Achievement of Grade 8 Students Across Education Authorities (Total score means, difference between post-reform and pre-reform means, t-value, and statistical significance)

Authority	Post-Reform		Pre-Reform		Post $\bar{x}$ - Pre $\bar{x}$				95%
	N	$\bar{x}$	N	$\bar{x}$	Diff.	t-value	DF	2-Tail sig.	CI
MOE	3035	33.3	1349	26.6	6.6	17.5	2928.1	.000	5.9--7.4
MOD	87	30.8	61	28.5	2.3	1.0	126.4	.313	-2.2-- 6.8
UNRWA	348	39.0	200	29.9	9.1	7.2	417.0	.000	6.6 -- 11.6
Private	277	45.7	140	35.5	10.2	7.0	291.2	.000	7.3 -- 13.1

Note:

MOE = Ministry of Education.

MOD = Ministry of Defence.

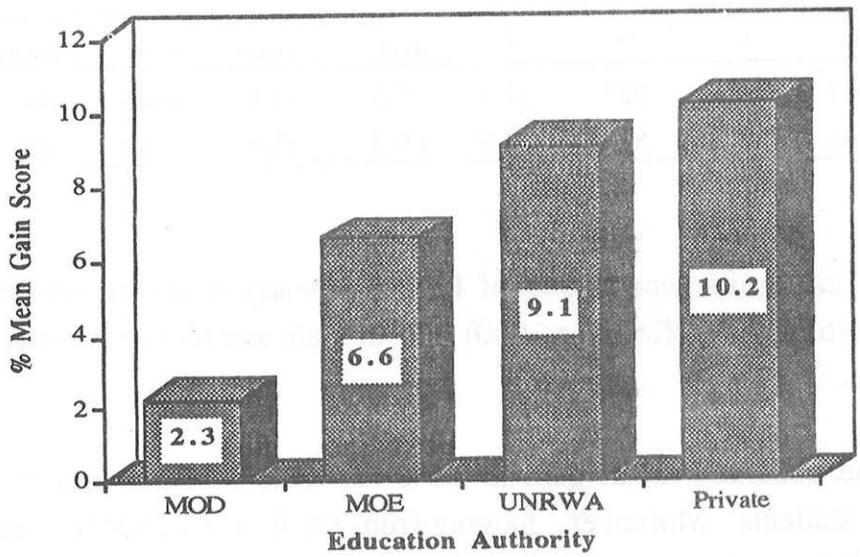
UNRWA = United Nations Relief and Works Agency.

The Column of 2-Tail significance in Table (6) clearly shows that reform gain in the eighth grade students' math test performance is statistically significant ($p \leq .000$) in three education authorities viz., MOE, UNRWA, and Private. In the MOD schools, unfortunately, the

improvement was not at all significant. The observed gain scores of the four education authorities are presented in Figure (3).

Figure (3)

Mean Gain In 8th Grade Students Math Achievement In the Four Education Authorities



Among the three authorities with statistically significant gain there was, however, no significant difference. Thus, apart from astonishingly poor performance of the MOD schools, the schools in the other three education authorities registered positive change in math achievement, but there was no statistical basis to distinguish among their relative gains.

Reform Impact Across Male/Female Populations of Grade Eight Students

Table (7) presents the sample size, post-reform and pre-reform math test score means, mean gain, t-value, Degrees of Freedom, two-tailed significance and 95% Confidence Interval for the mean gain separately for male and female samples of 8th grade students.

Table (7)

Post-Reform Gain in Math Achievement of Grade 8 Students Across Student Gender (Total score means, difference between post-reform and pre-reform means, t-value, and statistical significance)

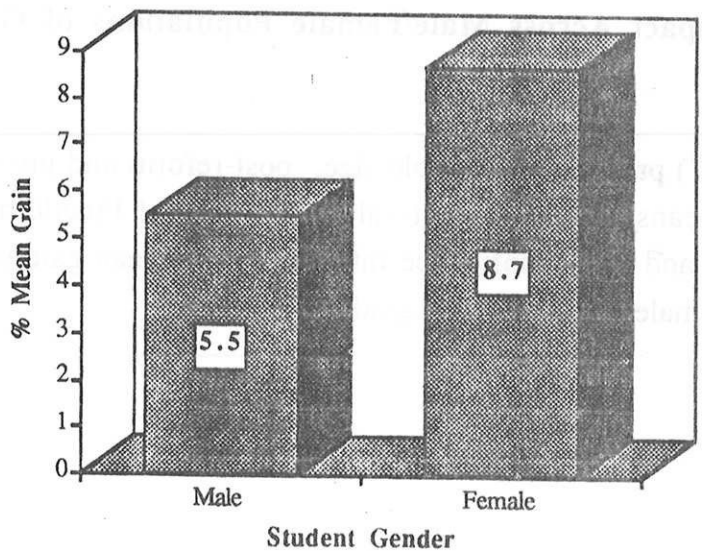
Student	Post-Reform		Pre-Reform		Post $\bar{x}$ - Pre $\bar{x}$				95%
Gender	N	$\bar{x}$	N	$\bar{x}$	Diff.	t-value	DF	2-Tail Sig	CI
Male	2049	32.6	927	27.1	5.5	11.4	1850.8	.000	4.5 -- 6.4
Female	1698	37.2	823	28.5	8.7	15.9	1874.3	.000	7.6 -- 9.7

2-Tail-significance column of Table (7) clearly shows the reform gains are statistically significant ($p \leq .000$) in both male and female populations of students.

The observed mean gain is 5.5% for male students and 8.7% for female students. Moreover, judging from the 95% Confidence Intervals, this difference between the degree of male and female improvement is statistically significant. This means that female students have derived more benefit from the education reform than the male students have done. Figure (4) displays the mean gain for male and female student samples.

Figure (4)

Varying Reform Impact on Male and Female Students



Reform Impact Across Urban and Rural Schools

The statistics needed to compare the reform gains between urban and rural area schools are given in Table (8). The post-reform and pre-reform difference column in Table (8) shows the mean gain scores of urban and rural schools. In both cases the reform gain is statistically significant ($p \leq .000$).

On the average urban schools have shown 8.1% gain against 5.5% mean gain of the rural schools in their 8th graders math achievement .

Table (8)

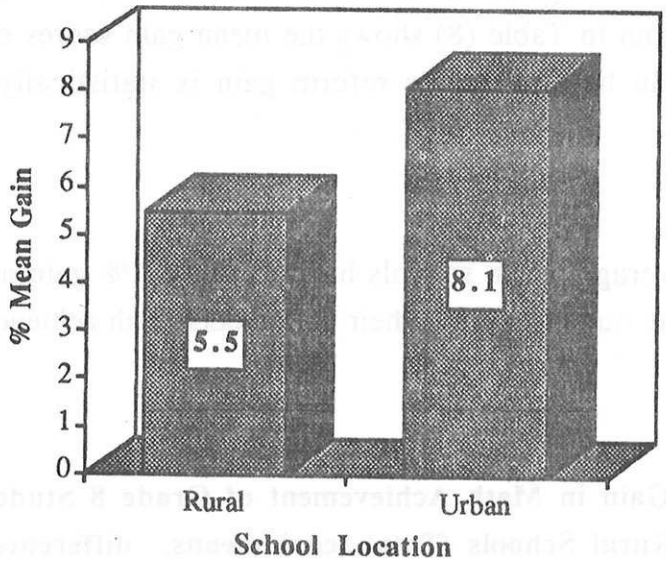
Post-Reform Gain in Math Achievement of Grade 8 Students Across Urban and Rural Schools (Total score means, difference between post-reform and pre-reform means, t-value, and statistical significance)

Location	Post-Reform		Pre-Reform		Post $\bar{X}$ - Pre $\bar{X}$				95%
	N	$\bar{X}$	N	$\bar{X}$	Diff.	t-value	DF	2-Tail Sig.	CI
Urban	1964	36.8	870	28.7	8.1	15.1	1852.0	.000	7.0 -- 9.1
Rural	1783	32.3	864	26.8	5.5	11.4	1808.9	.000	4.6 -- 6.5

This shows that urban schools have reaped significantly more benefits from the reform elements than the rural schools have done. Figure (5) gives a graphic display of the differential reform gain of urban and rural schools.

Figure (5)

Differential Reform Impact In Urban and Rural Schools



Regional Differences in Reform Gains on 8th Graders' Math Performance

Regional differences on critical indicators are of interest because they imply issues related to regional equity. When we collected baseline pre-reform data in 1993 there were only eight Governorates in Jordan but by the time we collected post-reform data in 1995 the number of Governorates had been raised to 12.

Because our sampling design was based upon schools from the eight Governorates, we have adhered to the same classification of schools for the purposes of studying regional differences. The relevant information about post-reform and pre-reform math test performance of the 8th grade students in each of the eight Governorates is presented in Table (9). In Table (9) the rows representing Governorates have been arranged according to descending order of the observed mean gain score.

Table (9)

Post-Reform Gain in Math Achievement of Grade 8 Students Across Governorates (total score means, difference between post-reform and pre-reform means, t-value, and statistical significance)

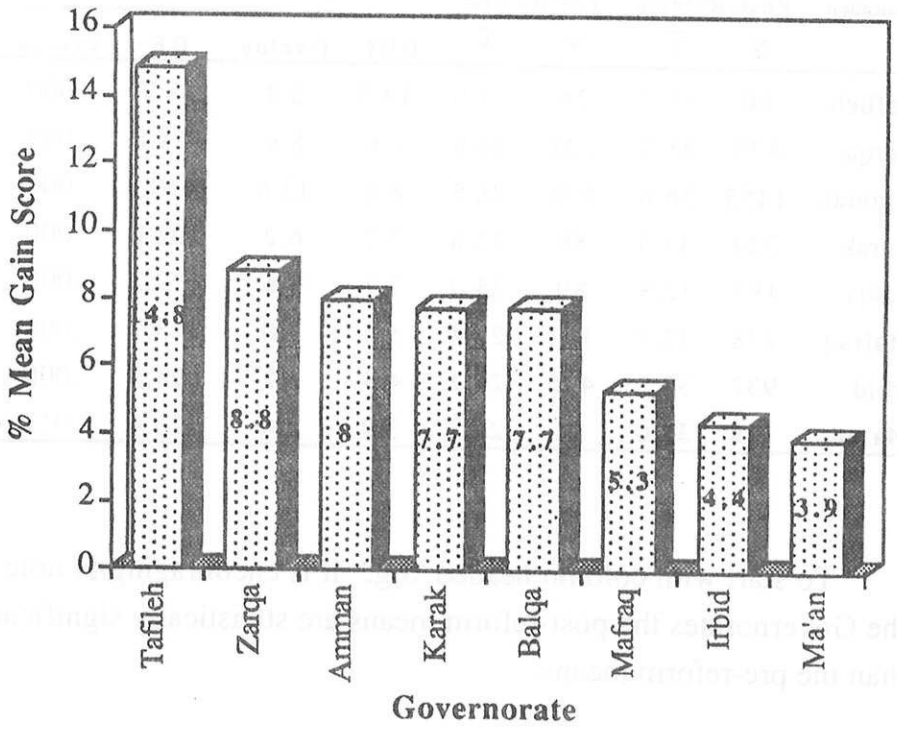
Location	Post-Reform		Pre-Reform		Post $\bar{X}$ - Pre $\bar{X}$				95% CI
	N	$\bar{X}$	N	$\bar{X}$	Diff.	t-value	DF	2-Tail Sig.	
Tafileh	60	35.8	26	21.0	14.8	5.4	68.7	.000	9.3 __ 20.2
Zarqa	475	35.7	228	26.9	8.8	8.9	528.8	.000	6.9 __ 10.8
Amman	1455	36.6	675	28.5	8.0	12.8	1459.6	.000	6.8 __ 9.3
Karak	224	33.3	89	25.6	7.7	6.2	202.9	.000	5.2 __ 10.2
Balqa	195	32.8	89	25.0	7.7	5.6	179.2	.000	5.0 __ 10.5
Mafraq	238	32.7	111	27.4	5.3	4.0	323.7	.000	2.7 __ 7.9
Irbid	932	33.5	435	29.1	4.4	6.2	824.5	.000	3.0 __ 5.9
Ma'an	168	38.6	81	24.7	3.9	2.6	123.6	.012	0.9 __ 6.9

To start with column headed 'Sig.' it is encouraging to note that in all the Governorates the post-reform means are statistically significantly better than the pre-reform means.

The level of statistical significance is very high, ($p \leq .000$) in all the Governorates except in Ma'an where the p-value is. 012.

Figure (6) displays the observed mean gain score in each Governorate.

Figure (6)
Observed Mean Gain Score on the Grade 8 Math Test Across Governorates



From Figure (6) it is quite clear that the improvement in Grade 8 students' math test performance with 14.8% mean gain score is by far the largest in Tafileh, of all the Governorates.

Then, in four Governorates namely, Zarqa, Amman, Karak, and Balqa, with their mean gain scores of 8.8, 8.0, 7.7, and 7.7 respectively, there is moderate level of improvement.

In the remaining three Governorates of Mafraq, Irbid, and Ma'an which have meagre mean gain scores of 5.3, 4.4, and 3.9 respectively, the reform impact is only modest. The wide variation in the gain is crystal clear, it ranges from 3.9% the lowest in Ma'an Governorate to 14.8% the highest in Tafileh.

Regarding the statistical significance of the reform-gain between pairs of Governorates we note that Tafileh has gained from reform statistically significantly better than Amman, Marfaq, Irbid, and Ma'an; whereas, there are no dintiguishable differences among Amman, Zarqa, Karak, and Balqa.

On the other hand, Tafileh, Zarqa, and Amman have gained significantly better than Irbid, while Zarqa also has gained significantly more than Ma'an. The preceding conclusions have ensued from comparing pairs of nonoverlapping 95% Confidences Intervals for the mean gain in each Governorate.

DISCUSSION

A national sample of Grade 8 student population was administered a math achievement test in 1993 in order for setting up pre-reform baseline achievement levels in mathematics. The same test was administered again, following the same procedures, to the 1995 eighth grade students' sample from the same schools which were included in 1993. It is apparent that we are talking about two distinct populations of eighth grade students (one of 1993, and the other of 1995). These two populations are assumed to be similar in all respects except one which is that the 1993 students were schooled under pre-reform conditions while the 1995 students have attended classes 5th through 8th under reform conditions which included new curricula, textbooks, instructional materials, and instructional methods.

In addition to reform's technical inputs the reform environment engendered, a new mood, a new awareness, and a new outlook towards the quality of education not only in the education community but also in the public at large, nationwide.

From the above-stated premises we can adduce that any real changes in 8th grade students' math test performance could be reasonably attributed to reform conditions. It is from this perspective that we talk of reform impact, gain scores, value added or reform gains.

The results of data analysis, presented in the preceding sections of this paper reveal the following facts.

1. On the whole, in the field of mathematics, the educational reform has made a positive impact on 8th grade students' achievement. The post-reform students outperformed their pre-reform cohorts on the math test by a margin of 7%.

2. The improvement, however, is uneven over different content areas. The observed average improvement ranges from 1% (not statistically significant) in 'Numbers' to 15% in 'Measurement'. Algebra and Rations are the areas of minimal improvement. From the analysis of individual items we know that in the topics related to abstract theoretical concepts such as indices, roots, irrational numbers, set functions, and in the whole area of analysis there is significant deterioration in performance. When these topics have direct relevance to analytical thinking (a proclaimed objective of reform), why is there no improvement? This is a pressing question.

   It seems that reform activities exert a mixed impact, both positive and negative, depending upon the right and wrong applications of instructional techniques and materials in the classrooms. After informal discussions with some math supervisors it transpired that (i) in a very short inservice training some teachers do not fully grasp some modern instructional techniques and therefore they use them in the classrooms in a wrong way; (ii) some other techniques cannot be properly applied in overcrowded classrooms but teachers feel obligated to apply them; (iii) the math curriculum is too extensive to be properly covered in time allocated to math instruction, but teachers feel under pressure to cover all the topics in the textbook; (iv) some topics are rushed through; (v) teachers and students do not have enough time needed for effective teaching and learning of new concepts and materials; (vi) some topics have been shifted down to 7th grade syllabus; (vii) there seems to be a general weakness in the mastery of basic foundation skills and concepts. All these factors have deleterious efect on students' learning and achievement.

3. The improvement along the three cognitive skills is almost uniform, although the area of 'Conceptual Understanding' is generally the weakest. This further supports the above alluded weaknesses in the foundation skills and concepts.

4. Improvement in math performance varies across education authorities. While there is no improvement in the MOD schools; in the Private, UNRWA, and MOE schools it is 10%, 9%, and 7% respectively. However, statistically there is no significant difference among mean gain of the latter three education authorities.

5. Female students have shown significantly higher improvement (9%) than the male students (6%). This seems to lend support to common belief that apart from early maturity of females at this stage both female teachers and female students take their respective duties of teaching and learning more earnestly than their male counterparts.

6. Reform impact is significant in both urban and rural community schools. Improvement in the urban schools (8%), however, is statistically significantly better than that in rural schools (5%).

   Due to compounded influence of numerous social, cultural, economic, demographic and environmental factors, in general, urban community schools' students perform better on academic achievement tests than their rural counterparts. But here we are investigating the gains due to educational reform, which also seem to have exceeded in urban schools over the gains in rural schools.

7. Reform gains in math achievement vary across regions (Governorates). The gain (15%) is highest in the Governorate of Tafileh and lowest in Ma'an, Irbid, and Mafraq (4%, 4%, and 5%), respectively). In Zarqa the gain is 9% while in the Governorates of Amman, Karak, and Balqa it is 8%.

   The reasons for variation of reform impact on math achievement of students across Governorates are not clear. The case of Tafileh is still more puzzling because there was absolutely no gain in the math performance of Grade 4 students whereas there is highest gain in the math performance of Grade 8 students.

Ma'an Governorate schools have shown minimum gain (4%). We should like to point out that Ma'an included the MOD schools in which there was no virtual difference between the pre-reform and post-reform means. While analyzing the per-reform and post-reform performance of the fourth graders we had noted that MOD students performance had decreased by 8%. This does not seem to make sense, what has happened in the MOD schools needs to be properly investigated.

Looking from the brighter perspective it is somewhat encouraging to note that, on the whole, their is statistically significant improvement in 8th grade students' math test performance after four years of reform. We are inclined to attribute the gain to the educational reform. On the other hand, it is hard to say what part of this gain is on account of changes in curriculum, textbooks and instructional practices of the teachers, and what part is due to some sort of Hawthorn effect. Irrespective of the substantive reform inputs, implementation of educational reform had infused the educational atmosphere with a spirit of awareness that something is going on to raise students' achievement levels.

Should this raised consciousness be attributed to reform or not is a matter of philosophical argument. But clearly it should not be confused with such reform inputs as new textbook or new method of teaching. Nevertheless, it is quite conceivable that injection of something like this in the national consciousness could have had a salubrious effect on educational achievement of students.

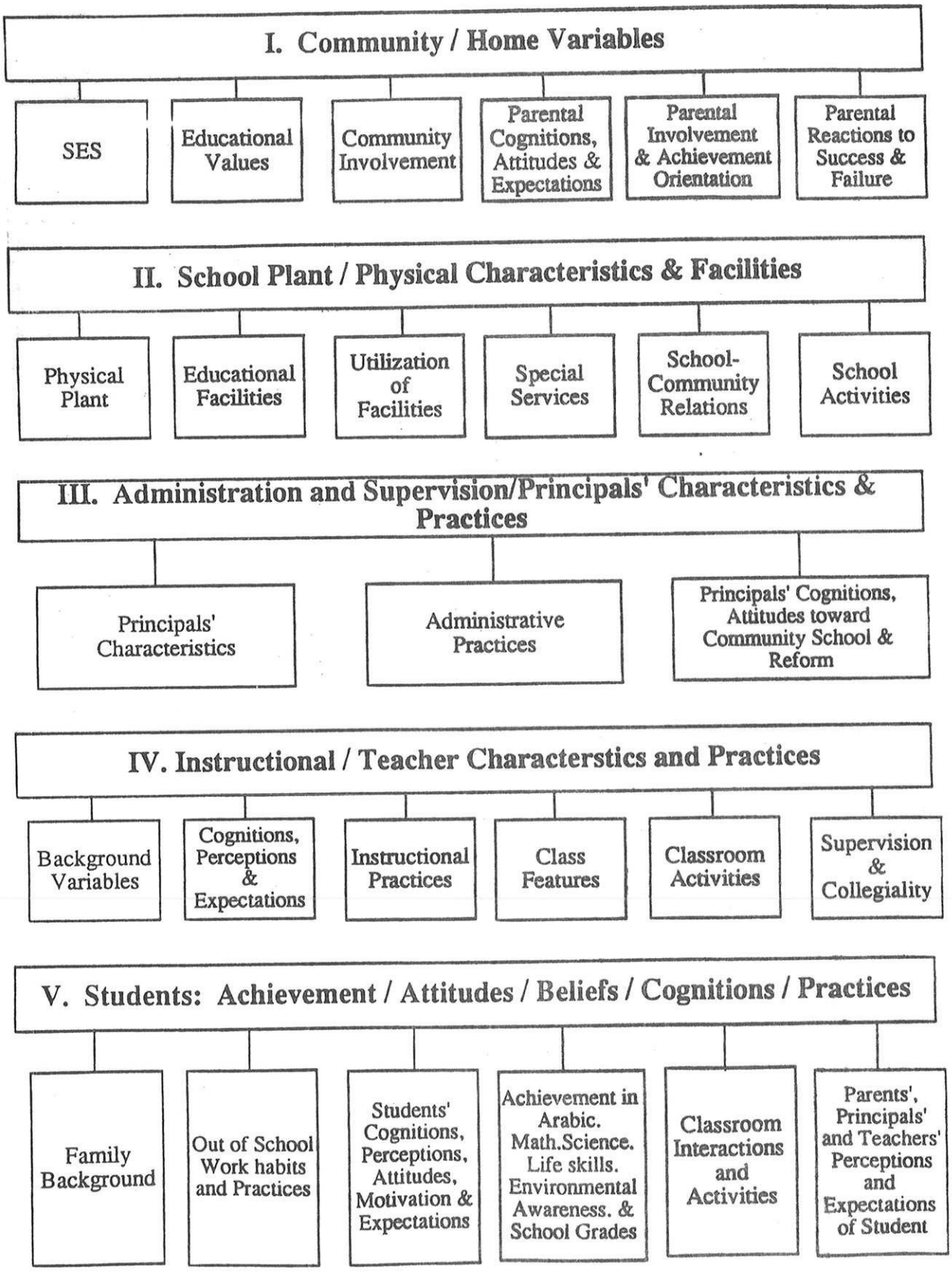
On the other hand, the reform does not seem to have affected students' analytical skills and other higher cognitive abilities which were specifically emphazied by the educational reform. Judging from the eighth grade students' performance in this area, one apprehends a trace of decline in the development of these critical skills.

Despite reform efforts that include new curricula, new textbooks, teacher guides, and inservice short-term teacher training, changing deep rooted attitudes and practices of teachers and students (not to speak of parents and community) is an uphill task, it requires strong determination and sustained effort. Moreover, as pointed out earlier, to master new

techniques of teaching and to apply them successfully in the classrooms requires dedicated effort on the part of the teachers which few of them are inclined to make without proper incentives.

Annex I

DOMAINS OF VARIABLES OF THE COMPREHENSIVE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS



References

Ahlawat, Kapur (1993). National assessment of instructional quality in basic education: An approach to address the critical issues at the tertiary stage of the baseline data phase of the study. **NCERD Document.**

Ahlawat, K., Billeh, V., Akasheh, T., and Shahbaz, M. (1994a). National survey of environmental knowledge and awareness of 8th and 10th grade students in Jordan. **Amman: NCERD Publication Series No.20.**

Ahlawat, K., Al-Nahar, T., and Billeh, V., (1994b). Assessment of learning achievement of grade 4 students in Jordan: A preliminary report. **Amman: NCERD Publication Series No.20.**