

**The Hashemite Kingdom of Jordan**



**National Center for Human Resources Development  
(NCHRD)**

**Continuous Monitoring And  
Assessment Of Learning  
Achievement In Primary Basic  
Education In Jordan :  
Progress Form 1998 To 2004**

**By**

**Kapur Ahlawat**

**Publications Series No.  
2005**

**114**



## *Table Of Contents*

	<u>Page</u>
Table of Contents	i
List of Tables	vi
List of Figures	x
Preface	1
Acknowledgements	2
Executive Summary	4
<b>PART I. BACKGROUND AND INTRODUCTION</b>	<b>13</b>
I.1 Historical Synopsis	13
I.2 Goal and Objectives	14
I.3 Objectives of the Fourth Study	14
I.4 Sampling Strategy	14
I.5 Instruments of the Study	16
I.6 Content and Format of the Achievement Tests	16
I.7 Structure of the Math Test	17
I.8 Structure of the Arabic Language Test	18
I.9 Data Collection	20
I.10 Coding Scoring and Data Entry	20
I.11 Data Analysis	20
<b>PART II RESULTS</b>	<b>22</b>
II.1 MATH TESTS	22
II.1.1 Psychometric Properties of the Math Tests	22
II.1.2 Students' Performance on Math Tests	23
II.1.2.1 Test1: General Competency of Understanding and Application of Procedures: 1998 VS 2004	23
II.1.2.1.1 Competency Levels on Math Test 1: 1998 VS 2004	23
II.1.2.1.2 Comparative Performance on Math Test 1 in 1998 and 2004 in Each Education Authority	26

## *Table of Contents*

	<u>Page</u>
II.1.2.1.2.1 Comparative Performance of MOE G4 Students in 1998 and 2004 on the Math Test1 (Understanding and Application of Procedures)	26
II.1.2.1.2.2 Comparative Performance of MOD G4 Students in 1998 and 2004 on the Math Test1 (Understanding and Application of Procedures)	28
II.1.2.1.2.3 Comparative Performance of UNRWA G4 Students in 1998 and 2004 on the Math Test 1 (Understanding and Application of Procedures)	29
II.1.2.1.2.4 Comparative Performance of Private Schools Grade 4 Students on Math Test 1 General and Sub-competencies in 1998 and 2004	29
II.1.2.1.3 Differential Progress in Urban and Rural Schools in Math Test1 (General Competency in Understanding and Application of Procedures)	30
II.1.2.1.4 Differential Progress of Male and Female 4 <sup>th</sup> Graders on Math Test 1 (Understanding and Application of (Procedures) from 1998 to 2004	32
II.1.2.1.4.1 Male 4 <sup>th</sup> Graders Performance on Math Test 1 in 1998 and 2004	32
II.1.2.1.4.2 Female 4 <sup>th</sup> Graders Performance on Math Test 1 in 1998 and 2004	33
II.1.2.2 Math Test2: General Competency In Mathematical Thinking And Communication	34
II.1.2.2.1 National Competency Levels of Grade 4 Students on Math Test2 (Mathematical Thinking and Communication) in 1998 and 2004	34
II.1.2.2.2 Fourth Grade Students' Achievement on Math Test2 (General Competency in Mathematical Thinking and Communication) in 1998 and 2004 in Each Education Authority	36
II.1.2.2.2.1 MOE 4 <sup>th</sup> Graders' Achievement on the General Competency in Mathematical Thinking and Communication in 1998 and 2004	37
II.1.2.2.2.2 UNRWA Grade 4 Students' Progress in Math Test2 (General Competency in Mathematical Thinking and Communication) from 1998 2004	38



## *Table of Contents*

	<u>Page</u>
II.1.2.2.2.3 Comparative Achievement of the MOD 4 <sup>th</sup> Graders on the Math Test2 (General Competency in Mathematical Thinking and Communication) in 1998 and 2004	39
II.1.2.2.2.4 Comparative Achievement of the Private Grade 4 Students on the Math Test2 (General Competency in Mathematical Thinking and Communication) between 1998 and 2004	41
II.1.2.2.3 Urban / Rural School Differences in Gain Scores (2004-1998) on Math Test2 (General Competency in Mathematical Thinking and Communication)	42
II.1.2.2.3.1 Urban Fourth Graders' Performance on Math Test 2 (General Competency in Mathematical Thinking and Communication in 1998 and 2004	42
II.1.2.2.3.2 Gain Scores of Rural School 4 <sup>th</sup> Graders on Math Test 2 (General Competency in Mathematical Thinking and Communication) from 1998-2004	43
II.1.2.2.4 Differential Achievement Gains of Male and Female 4 <sup>th</sup> Graders	44
II.1.2.2.4.1 Male 4 <sup>th</sup> Graders' Progress in Math Test2 Achievement from 1998 to 2004	45
II.1.2.2.4.2 Female 4 <sup>th</sup> Graders' Progress in Math Test2 Achievement From 1998 to 2004	46
<b>II.1.2.3 MATH TEST 3: PROBLEM SOLVING</b>	47
II.1.2.3.1 Comparative Competency Levels of Grade 4 Students on Math Test3 (Problem Solving)	47
II.1.2.3.2 National Progress in Math Problem Solving Competency of the 4 <sup>th</sup> Grade Students from 1998 to 2004	49
II.1.2.3.3 Differential Change in Performance of Male and Female 4 <sup>th</sup> Graders on Math Test 3 (Mathematical Problem Solving) from 1998 to 2004	50
II.1.2.3.3.1 Male 4 <sup>th</sup> Graders' Progress in Math Problem Solving Competency from 1998 to 2004	50
II.1.2.3.3.2 Female 4 <sup>th</sup> Graders' Progress in Math Problem Solving Competency From 1998 to 2004	51



## *Table of Contents*

	<i>Page</i>
II.1.2.3.4 Urban and Rural School 4 <sup>th</sup> Graders' Progress in Math Problem Solving Competency from 1998 to 2004	52
II.1.2.3.4.1 Urban School 4 <sup>th</sup> Graders' Progress in Math Problem Solving Competency From 1998 to 2004	52
II.1.2.3.4.2 Rural School Fourth Grade Students' Progress in Math Problem Solving Competency from 1998 to 2004	53
II.1.2.3.5 Grade 4 Students' Achievement on Math Test 3 (General Competency in Math Problem Solving) in Each Education Authority	54
II.1.2.3.5.1 MOE 4 <sup>th</sup> Graders' Performance on Math Test3 (Problem Solving) in 1998 and 2004	54
II.1.2.3.5.2 UNRWA 4 <sup>th</sup> Graders' Performance on Math Test3 (Problem Solving) in 1998 and 2004	55
II.1.2.3.5.3 Private School 4 <sup>th</sup> Graders' Performance on Math Test3 (Problem Solving) in 1998 and 2004	56
II.1.2.3.5.4 MOD School 4 <sup>th</sup> Graders' Performance on Math Test 3 (Problem Solving) in 1998 and 2004	57
<b>II.2 ARABIC LANGUAGE TEST</b>	<b>59</b>
II.2.1 Psychometric Properties	59
II.2.2 Arabic Language Competency Levels of Grade 4 Students in 1998 and 2004	59
II.2.2.1 Arabic Language Comprehension Competency Levels of 4 <sup>th</sup> Grade Students in 1998 and 2004	60
II.2.2.2 Arabic Writing Competency Levels of Grade 4 Students in 1998 and 2004	60
II.2.2.3 Arabic Grammar Competency Levels of Grade 4 Students in 1998 and 2004	61
II.2.2.4 Grade 4 Students' Competency Levels on the Use of Dictionary in 1998 and 2004	62
II.2.3 Grade 4 Students' Arabic Language Competencies in 1998 and 2004	63
II.2.4 4 <sup>th</sup> Graders Progress in Arabic Language Within Each Education Authority: From 1998 to 2004	64

## *Table of Contents*

	<u>Page</u>
II.2.4.1 MOE 4 <sup>th</sup> Graders Progress in Arabic Language From 1998 to 2004	64
II.2.4.2 UNRWA Schools 4 <sup>th</sup> Graders' Progress in Arabic Language from 1998 to 2004	65
II.2.4.3 Private Schools 4 <sup>th</sup> Graders' Progress in Arabic Language From 1998 to 2004	66
II.2.4.4 MOD Schools Grade 4 Students' Progress in Arabic Language From 1998 to 2004	67
II.2.5 URBAN and RURAL Area Schools' Grade 4 Students' Progress on Arabic Language Competencies from 1998 to 2004	68
II.2.5.1 Urban Schools Grade 4 Students' Progress in Arabic Language from 1998 to 2004	68
II.2.5.2 Rural Schools' Grade 4 Students' Progress in Arabic Language from 1998 to 2004	69
II.2.6 Male/Female 4 <sup>th</sup> Graders' Progress in Arabic Language from 1998 to 2004	70
II.2.6.1 Male 4 <sup>th</sup> Graders Progress in Arabic Language from 1998 to 2004	70
II.2.6.2 Female 4 <sup>th</sup> Graders' Progress in Arabic Language from 1998 to 2004	71
<b>PART. III. GENERAL CONCLUSION</b>	<b>73</b>
<b>PART.IV. RECOMMENDATIONS</b>	<b>74</b>
EXHIBIT – 1 MATH1	75
EXHIBIT – 2 MATH2	76
EXHIBIT – 3 MATH3	77
EXHIBIT – 4 ARABIC LANGUAGE	87
الملخص التنفيذي للدراسة	1

## *List of Tables*

Table (1):	Sampling Distribution	<u>Page</u> 15
Table(2):	Table of Specifications for Math Test Form 1: Understanding And Application of Procedures	2
Table (3):	Table of Specifications for Math Test Form 2: Mathematical Thinking and Communication	3
Table (4):	Table of Specifications for Math Test Form 3: Problem Solving .	18
Table (5):	Table of Specifications for the Arabic Language Test	19
Table (6):	Comparative Achievement of G4 Students on Math Test1 (Understanding and Application of Procedures) in 1998 and 2004 (National Level)	25
Table (7):	Comparative Achievement of G4 MOE Students in 1998 and 2004 on Math Test1 General Competency (Understanding and Application of Procedures) and its Sub-competencies	27
Table (8):	Comparative Achievement of G4 MOD Students in Math Test 1 (Understanding and Application of Procedures) in 1998 and 2004	28
Table (9):	Comparative Performance of UNRWA 4 <sup>th</sup> Graders on Math Test 1 (Understanding and Application of Procedures) in 1998 and 2004	29
Table (10):	Comparative Performance of Private School 4 <sup>th</sup> Graders on Math Test1 (Understanding and Application of Procedures) in 1998 and 2004	30
Table (11):	Differences on Math Test 1 (Understanding and Application of Procedures) Achievement in 1998 and 2004 and Their Statistical Significance in Urban and Rural Schools	31
Table (12):	Male 4 <sup>th</sup> Graders' Differential Achievement on Math Test1 (Understanding and Application of Procedures) in 1998 and 2004	32
Table (13):	Female 4 <sup>th</sup> Graders' Differential Achievement on Math Test1 (Understanding and Application of Procedures) in 1998 and 2004	33
Table (14):	Mean, Mean Difference, df, t-value and 2-Tailed Significance of the Mean Difference Between 1998 and 2004 Performance of Grade 4 Students on Math Test 2 in Jordan	36
Table (15):	Means, Mean Difference, Degrees of Freedom, t-value and Their 2-tailed Significance, on the Math Test 2 Scores of the 4 <sup>th</sup> Grade MOE Students in 1998 and 2004	37

## *List of Tables*

	<i>Page</i>
Table (16): Mean, SD, Mean Difference, df, t-Value, Std. Err. Of the Difference and 2-Tailed Significance of the Mean Difference for UNRWA 4 <sup>th</sup> Graders between 1998 and 2004	38
Table (17): Mean, SD, Std. Error of the Mean, Mean Difference, df, t-Value and 2-Tailed Sig. of the Mean Difference (2004-1998) on Math Test2 (General Competency in Mathematical Thinking and Communication) of the MOD Grade 4 Students	40
Table (18): Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value and 2-Tailed Significance of the Mean Difference (2004-1998) on Math Test2 (General Competency in Mathematical Thinking and Communication) of the Fourth Grade Private School Students	41
Table (19): Mean, SD, Mean Difference, Std. Error of Mean Difference, t-Value, df and Two-tailed Significance of the Difference in Urban 4 <sup>th</sup> Grade Students' Gain Scores on Mathematical Thinking and Communication Competency and its Sub-competencies from 1998 to 2004	43
Table (20): Mean, SD, Mean Difference, Std. Error of Mean Difference, t-Value, df and Two-tailed Significance of the Difference in Rural 4 <sup>th</sup> Grade Students' Gain Scores on Mathematical Thinking and Communication Competency and its Sub-competencies from 1998 to 2004	44
Table (21): Mean, SD, Mean Difference, Degrees of Freedom, t-Value, Std Error of the Difference of Male Students Achievement in Math Test 2 (General Competency in Mathematical Thinking Communication) from 1998 to 2002	45
Table (22): Mean, SD, Mean Difference, Degrees of Freedom, t-Value, Std. Error of the Mean Difference and 2-Tailed Significance Comparing Female 4 <sup>th</sup> Graders Performance in (General Competency in Mathematical Thinking and Communication) in 1998 and 2004.	46
Table (23): Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. Of the Difference (2004-1998) on Math Test 3 (General Competency in Math Problem Solving) at the National Level	49
Table (24): Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) for Male Fourth Graders in 1998 and 2004	50
Table (25): Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) for Female Fourth Graders in 1998 and 2004	51

## *List of Tables*

	<i>Page</i>
Table (26): Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) between 1998 and 2004 Testings of the Urban School Grade 4 Students	52
Table (27): Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) between 1998 and 2004 Testings of the Rural School Grade 4 Students	53
Table (28): Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) Scores of the MOE Grade 4 Students in 1998 and 2004	54
Table (29): Mean, SD, std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) Scores of the UNRWA Grade 4 Students in 1998 and 2004	55
Table (30): Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test3 (Problem Solving) Scores of the Private Grade 4 Students in 1998 and 2004	57
Table (31): Mean, SD, Std. Err. of the Mean. Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) Scores of the MOD Grade 4 Students in 1998 and 2004	58
Table (32): Mean, SD, Std. Err. of the Mean. t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference of the 4 <sup>th</sup> Grade Students' Performance on the Arabic Test in 1998 and 2004	64
Table (33): Mean, SD, Std. Err. of the Mean. t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between the 1998 and 2004 Arabic Test Scores of the MOE 4 <sup>th</sup> Grade Students.	65
Table (34): Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between the 1998 and 2004 Arabic Language Test Scores of the UNRWA Schools 4 <sup>th</sup> Grade Students	66
Table (35): Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the Private Schools Grade 4 Students	67

### *List of Tables*

Table (36):	Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between the 1998 and 2004 MOD Schools Grade 4 Student Scores on the Arabic Language Test	<u>Page</u> 68
Table (37):	Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the Grade 4 Students in Urban Schools	69
Table (38):	Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the 4 <sup>th</sup> Grade Rural Schools Students	70
Table (39):	Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the 4 <sup>th</sup> Grade Male Students	71
Table (40):	Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the 4 <sup>th</sup> Grade Female Students	72



### *List of Figures*

Figure (1):	National Competency Levels in Knowledge and Application of Mathematical Procedures in 1998 and 2004	<u>Page</u> 24
Figure (2):	Percentage of G 4 Students Reaching Each Competency Level on the General Competency in Mathematical Thinking and Communication in 1998 and 2004	35
Figure (3):	Comparative National Competency Levels of Grade 4 Students in Mathematical Problem Solving in 1998 and 2004	48
Figure (4):	Grade 4 Students Reaching Each Competency Level in Arabic Language Comprehension in 1998 and 2004	60
Figure (5):	Grade 4 Students Reaching Each Competency Level in Arabic Writing in 1998 and 2004	61
Figure(6):	Percentage of 4 <sup>th</sup> Grade Students at Each Competency Level of Arabic Grammar in 1998 and 2004	62
Figure (7):	Percentage of 4 <sup>th</sup> Grade Students at Each Level of Competency in the Use of Arabic Dictionary in 1998 and 2004	63



## *Preface*

This study is the fourth in the series of the National Center for Human Resources Developments' Monitoring of Learning Achievement studies which aim to provide timely feedback for continuous improvement needed to reach a wide spectrum of goals and objectives of education. Improved performance of students is a cause for rejoicing, though the learning process is a long and arduous one.

The general trend in Math Tests of this study is defined as positive concerning MOE students, where they have shown increased performance on five of the eight sub-competencies in Test 1 and ten of the thirteen sub-competencies in Test2. UNRWA students' scores have shown a slight decrease on all subtests. Private schools students' scores have witnessed a positive trend on five of the eight subtests. Students of schools run by the Army have shown an improvement on all the subtests in Math1 and Math2.

As for Arabic Language competencies, MOE students have registered an overall progress. UNRWA students' performance has declined in two of the four competencies. Private schools have suffered a decline on all the four competencies, while Army schools have gained on all the competencies

This mixed performance is quite expected due to many factors including the benchmarks used for comparisons, the instructional differences, the management issues, among other variables.

I hope this study will be of benefit for decision makers, educators, and researchers.

**Dr. Munther Masri**  
**President**  
**NCHRD**

---

## Acknowledgments

Research studies produced as books, bulletins or in any other form usually have a bit of history associated with them. This report, comparing between the test performance of 4<sup>th</sup> grade students in 1998 and 2004, has utilized data collected by the 3<sup>rd</sup> and 4<sup>th</sup> studies conducted under the on-going NCHRD series of National Monitoring Learning Achievement studies. In the process of conducting both studies numerous people participated in their diverse capacities in various ways. I consider it a privilege and an honor to extend my sincere thanks to all of them, individually and collectively. MLA studies have been a cooperative activity between the NCHRD and the MOE. The General Directorate of Examinations at the MOE was particularly helpful in drawing the sample: their valuable support is acknowledged with appreciation.

H.E. Dr. Munther Masri, President NCHRD, from the beginning of this exercise, has been constantly interested in the progress of the report and provided all the support and facilities to create a conducive environment for writing the report. I express my profound thanks and deep gratitude for his kindly support.

In particular, this report owes its comparative theme to H.E. Dr. Tayseer Nhar, Secretary General for Technical Affairs, Ministry of Education, who rightly apprehended the importance of comparing the achievement of the 2004 Primary Basic students with that of their cohorts in 1998 at this juncture of time. It gives me special pleasure to extend my profound thanks for his initiative and guidance.

This report could not have been written without the data for 1998 and 2004 studies. I acknowledge the SPSS output files of the t-test analyses, first as hard copies and subsequently, on request, as e-files supplied by Dr. Khattab Abulebdeh. Besides supplying data, he was kind enough to translate the Executive Summary into Arabic. I express my deep gratitude to him for his support.

Since SPSS output files were not directly usable, the relevant information had to be culled from different places and reorganized in proper tables for reporting display. After wasting a good deal of time to construct the tables manually, I requested Mrs. Shrouq Zawati of NCHRD to help compile the remaining tables electronically. Despite her being overly busy with her work, she helped me by compiling some of the tables. I thank her for all the support.

---

My acknowledgments will remain incomplete without my wholehearted expression of deep appreciation for Mr. Azmi Jarrars' valuable contribution in terms of his editorial suggestions.

Finally, having all said and done, typing a report that is replete with complex tables that were cluttered with cells full of numbers with five decimal places which required to be rounded to two decimal places was a redoubtable toil. Without Mrs. Hala Masoud's unflagging dedication to high quality work delivered on time, this report could not have been in the shape as it is, nor could it have been delivered at the right time. I express my profound thanks for her relentless effort.

**Dr. Kapur Ahlawat**

---

## EXECUTIVE SUMMARY

### *Background*

This is the fourth Monitoring Learning Achievement and Instructional Quality study conducted since 1993 to track the progress and quality of primary education in Jordan.

### *Objective*

In the tradition of monitoring learning achievement, the primary objective of this report is to focus on the changes that have occurred in Math and Arabic Language achievement of the 4th grade student population and its various well-defined subpopulations, over the past six years or so from 1998 (the 3<sup>rd</sup> study) to 2004 (the present, 4<sup>th</sup> study).

### *Sampling Strategy*

The 4<sup>th</sup> study (2004) followed the same sampling procedures, employing two-stage stratified random sampling technique, as were used in the previous studies in this series. The size of the sample in this study, however, was substantially reduced to 2000. Of which 1056 were male and 944 female students. According to school location 1,363 students were from Urban schools and 637 from Rural schools. Across the Education Authority dimension, 1,514 students were from the MOE; 240 from UNRWA, 193 from Private and 53 from the MOD.

### *Instruments of the Study*

The instruments used in the 2004 study were basically the same as used in the 1998 study, Math (three Forms) and Arabic (two Forms) with the exception of Social Studies subject which was dropped from the 2004 study.

### *Findings*

Competency Level reached by the national sample as a whole in 2004 and 1998 in each subject i.e., three Math tests and four major competency areas in Arabic Language, were compared with intent to study change in student achievement in these subjects.

On the basis of percent correct score obtained on the test each student is assigned to one of the three successive competency levels. Students scoring 30% or less are put in the level 1 (L1) category; those scoring

more than 30% but less than 70% are classified in Level2 (L2) category and those who score 70% or more are put in the Level3 (L3) category which is the Mastery level.

### **Comparative Competency Levels in Math (1998-2000).**

In Math Test1, which measured the general competency in Knowledge and Application of Mathematical Procedures, there was a drop of 1 and 2 percentage points, respectively, at L1 (Unacceptable Level) and L2 (Modest or Partial Mastery) from 1998 to 2004 while, at the Mastery Level (L3) there was increase of 3 percentage points.

Decrease of students at the lower competency levels and increase at the highest mastery level, though nominal is a desirable trend.

In Math Test 2 which measured the General Competency in Mathematical Thinking and Communication, 12 percentage points decrease at L1; 6% increase in L2 and 6% increase in L3 categories, taken together indicate a healthy trend of progress in this area of Math from 1998 to 2004.

In general competency area of Problem Solving measured by Math Test 3 there seems little improvement from 1998 to 2004. There is 5% increase at the lowest competency level (L1); 5% decrease at the L2 and no change at the L3 (Mastery level).

### **National Competency Levels in Arabic Language in 1998 and 2004.**

In all the four general competencies tested in 1998 and now in 2004 the levels reached by students on both testings in each area of competency were compared in terms of observed percentages.

In Comprehension there is absolutely no change from 1998 to 2004. The percentage of students reaching levels L1, L2 and L3 are 23%, 72% and 5%, respectively, in both 1998 and 2004.

Regarding national competency levels in Writing, there is 13% points decrease (from 46% in 1998 to 33% in 2004) at L1 (the lowest and unacceptable level) which is desirable trend which is further supported by 11%-age points increase at L2 and 2%-age points increase at L3 (the Mastery level).

In the general competency area of Grammar again the trend (1998 vs. 2004) is in the right direction, though the change is minimal; 3%-age

points decrease at L1; 1% point decrease at L2 and 3% points increase at L3 (the Mastery level).

On the general competency in Dictionary Use, there is 4% points decrease (from 31% to 27%) at L1; 1% point decrease at L2 and 4% point increase at L3, the Mastery level. Again the trend is in the right direction; the change however is in frugal measures.

### **Math Test 1: Understanding and Application of Procedures**

At the national level, on three out of eight sub-competencies viz., Subtraction, Multiplication and Fractions there is statistically significant increase whereas, on Geometry and Measurement subscale there is highly significant decrease from 1998 to 2004.

### **The type and magnitude of change varies across different Education Authorities.**

The general trend defined as positive, negative or mixed varies. MOD has clearly a positive trend with positive gains on all the eight subtests, on six of them gains are statistically significant.

UNRWA has a negative trend showing decrease on all the subtest scores, though all of them statistically nonsignificant. Both MOE and Private have mixed trends. MOE has, on five of the eight subtests, increased performance; on two of them statistically significant (Subtraction and Multiplication) while on three subtests there is decline, on one of them statistically significant (Geometry and Meas). Private schools also showed a mixed trend. On five of them +ve including one statistically significant while on 3 subtests there is loss, on Geometry and Meas it is statistically significant ( $\alpha \leq .05$ ).

### **Location-wise Progress**

Progress on the achievement in the general competency in Understanding and Application of Procedures is rather positive in the Urban schools and all-round negative in the Rural Schools.

In Urban schools there is increased mean score on six of the eight sub-competencies, on three of them the increase is statistically significant (Subtraction, Multiplication and Fractions) whereas, on two subtests there is decrease, on one of them (Geometry and Meas.) the decline is statistically significant.



---

In **Rural** schools, there is lack of progress across-the-board clearly signaled by decrease in the mean percentage correct scores on all the 8 subtests, although it is statistically nonsignificant.

### **Male vs Female 4<sup>th</sup> Graders' Progress**

**Male** students' mean percent correct scores decreased in 2004 on all the subtests except one (Multiplication); none of the differences, however was statistically significant.

**Female** students' mean percent correct scores, in contrast, showed positive gains on all the subtests except one (Geometry and Measurement) and the whole test. Moreover, on three of the subtests and the whole test the increase was statistically significant.

### **Math Test2: (General Competency in Mathematical Thinking and Communication).**

The classification of students into three successive competency levels indicated a modest degree of progress as shown by 12 percentage points decrease at the unacceptable level (L1) with corresponding increase of 6% points at (L2) and also 6% points at (L3), the Mastery level.

Regarding G4 student's progress at the national level on each sub-competency subsumed under the general competency in Mathematical Thinking and Communication, there is all-round increment in the mean percent correct scores on all the 13 sub-competencies and the general competency. In spite of the consistent gains across-the-board on all the subtests, and the whole test, none of them is statistically significant at ( $\alpha \leq .05$ ) level of significance.

### **Progress on Math2 at the Education Authority Level**

MOE students gained on 10 subtests out of 13 and on the whole test and scored loss on three subtests, not even a single difference (gain or loss) was, albeit, statistically significant.

UNRWA schools registered decrement on eight of the 13 subtests and on the whole test; on two of them the decrease was statistically significant. On the remaining five subtests there was slight increase but not statistically significant.

**Private** schools fared no better than the UNRWA schools. On ten counts, including the whole test, there was decline, statistically significant in two cases, while the marginal gains on 4 subtests were not statistically significant.

---

**MOD** schools, in contrast, have shown spectacular increase constantly on all the subtests and the whole test; on top of that, all the gains are statistically significant at ( $\alpha \leq .01$ ) or better levels of significance.

### **Location-wise Progress on Math Test2 (General Competency in Mathematical Thinking and Communication and its Sub-Competencies) of the 4<sup>th</sup> Grade Students from 1998 to 2004**

The **Urban** Schools, on all the 14 counts (13 subtests and the whole test) have shown consistent gains with no exception. None of them is statistically significant, though.

**Rural** schools, on the other hand, have recorded constant decline in all the 14 cases but none of them reached statistical significance.

### **Gain Scores for Male and Female Sub-Samples**

**Male** 4<sup>th</sup> grade students in Jordanian schools have scored less than their cohorts did in 1998 on 12 of the 13 subtests; on one of them the decrease is also statistically significant.

**Female** students' constantly positive pattern, indicating increased performance on all the subtests and the whole test, is a welcome contrast with their male cohorts' largely negative pattern. Over and above the gains on all the 14 counts, seven of them are statistically significant at ( $\alpha \leq .01$ ) level.

### **Math Test3 (General Competency in Problem Solving)**

#### **Progress in National Competency Levels.**

In respect of the progress made on competency levels on problem solving from 1998 to 2004, the situation is rather depreciatory. 5% points increase at the Unacceptable level (L1) followed by another 5% decrease at the L2 and no change at L3 (the Mastery Level) rings an alarm bell.

#### **National Progress on Sub-competencies in Problem Solving Area**

On four of the eight sub-competencies the 4<sup>th</sup> graders have gained a little with none of them being statistically significant, and on the other four they have lost, the loss being statistically significant, at least, on one of the sub-competencies (1 Step Geometry Problems).

## **National Progress on Math Test3 (Problem Solving) in Each Education Authority**

The MOE students have shown decreased performance on seven of the eight subtests as well as on the whole test, on three of the subtests the decrease in scores is statistically significant. UNRWA schools in comparison with the MOE schools have recorded gains on seven of the eight subtests in addition to the whole test; on four of them including the whole test, the improvement is statistically significant. Private Schools, contrary to expectations, have shown decreased performance on five of the eight subtests; none of them is, however, statistically significant. On the other hand, on two out of three subtests the improvement in test performance is statistically significant.

MOD schools students too, in comparison to their noticeable performance on Math Test1 and Math Test2, did not fare so well on the Problem Solving test. Although they have registered gain on five of the eight subtests in addition to the whole test, none of them is statistically significant, nor is any of the losses for that matter.

### **Location-wise Performance on the Math Test3 (General Competency in Problem Solving)**

**Urban** school 4<sup>th</sup> graders have registered a fully balanced pattern of gains and losses, increase on four and decrease on the other four of the eight subtests, while none of the eight is statistically significant.

**Rural** school students test scores display a continuous array of negative signs broken by one positive. Their performance has decreased on seven of the eight subtests as well as on the whole test. None of the changes is statistically significant.

### **Gender-wise Progress on the Math Test3 (General Competency in Problem Solving).**

**Male** 4<sup>th</sup> graders' performance has declined on five of the eight subtests in addition to the whole test; none of the six positives and three negatives is statistically significant, though.

**Female** 4<sup>th</sup> graders, in line with their male cohorts, have registered gains on four and losses on the other four of the eight subtests, none of them is statistically significant.

---

## **Arabic Language Test General Competencies in Grammar, Writing, Comprehension and Dictionary Use**

### **National Progress on the Competency Levels Reached by 4<sup>th</sup> Grade Students on Each of the Four Competencies in 1998 and 2004**

In the field of Arabic language Comprehension judged from the competency levels reached by percentages of students in the two testings the progress seems to be stagnant. Apart from the 1% increase at L2 (the modest level) there is absolutely no change in classification of students in the three competency levels.

On the Writing Competency, decrease of 13% points at L1 and increase of 11% points and 2% points, respectively at levels L2 and L3 is a sign of improvement in student achievement in this field.

On 4<sup>th</sup> graders' competency in Grammar, 3% points decrease at L1, 1% point decrease at L2 and 3% points increase at L3 indicate a borderline case of rather dubious progress.

Regarding progress on Dictionary Use competency of 4<sup>th</sup> grade students, 4% point decrease at L1, 1% point decrease at L2 and 4% points increase at L3, indicate marginal progress but a trend in the right direction.

### **Progress in Arabic Language Competencies in Each Education Authority from 1998 through 2004.**

In the whole national sample, there is general trend of improvement on all the four competencies. Moreover, on three of them viz, Grammar, Writing and Dictionary Use the improvement is statistically significant ( $p < .01$ ).

### **Progress in the Arabic Language General Competencies in Each Education Authority.**

In MOE like the whole sample there is an overall progress which is statistically significant on the whole test and the same three competencies as is the whole population.

UNRWA schools have shown decline on two of the four competencies, and on the whole test and nominal increase on Writing and Dictionary Use. No change, increase or decrease, on any measure is statistically significant.

---

Private schools show an unexpected decline on all the measures of competencies which is statistically significant on two of the four competencies as well as on the whole test.

MOD schools, maintaining the progress trend, have registered statistically significant ( $p < .01$ ) gains on all the subtests and the whole test.

### **Progress in Rural and Urban Schools on the Arabic Language Competencies.**

In the Urban schools while the trend of improvement is consistent over all the subtests and the whole test, the increase is statistically significant only on two of them, Grammar and Dictionary Use.

Rural schools show mixed performance, decline on two counts and nominal increase on three with no statistical significance on any of them.

### **Progress of Male and Female 4<sup>th</sup> Graders in Arabic Language Competencies.**

The Male 4th graders show a consistent pattern of progress across-the-board; on three of the four subtests the improvement in test scores is statistically significant. With respect to Female Grade 4 students, there are four pluses and one minus, none of them is statistically significant.

### *General Conclusion*

In general, on most tests and subtests in Math and Arabic Language, we found no clear-cut, indications of definite improvement in 4<sup>th</sup> grade students' test scores from 1998 to 2004. There are, all the same, encouraging trends of progress marred only by a few instances in specific subpopulations in some areas of achievement where there are signs of declining trend. On the whole, however, such decline is not statistically significant and therefore, statistically speaking we can only conclude that there is no significant progress in those cases.

There are differences in progress in achievement among various subpopulations on different tests and subtests. The causes of such differences and of the lack of improvement, in general, need to be investigated systematically on their own right but such an investigation falls outside the scope of this report.

---

Progress at a glance in the three Math Tests and the Arabic Language Test is presented, respectively in Exhibits 1, 2, 3 and 4.

### *Recommendations*

The causes of lack of progress in Math and Arabic Language at the Primary level are diverse and manifold. Even the results could have been marred by lack of rigor in the study itself. Since there is generally a tradeoff between cost and quality of research, it is paramount that validity of results and the conclusions drawn from them should not be compromised. Based on the cumulative experience of the four studies the following suggestions are offered for serious consideration by the concerned authorities.

1. First and foremost, the authorities should have a firm conviction about the real importance and need for national studies like Monitoring and Assessment of Learning Achievement, and the objectives of such studies should be crystallized and prioritized in order to guide the design and allocation of adequate resources.
2. The choice should be between conducting a rigorous study and relinquishing it altogether because conducting a sloppy study is a sheer waste of resources.
3. If learning achievement is monitored to study the educational realities on the ground and to measure student achievement for guiding improvements where they are needed, then the authorities should openly accept the realities revealed by the valid findings of the study. Only the sincere realization and acceptance of the shortfalls can lead to earnest effort for improvement.
4. The causes of low performance and lack of progress of different segments of primary cycle populations on different tests and subtests should be properly investigated and pin pointed.
5. Having identified the causes of low performance appropriate measures and actions should be proposed and properly implemented to improve the situation.

---

# **Continuous Monitoring and Assessment of Learning Achievement in Primary Basic Education in Jordan: Progress from 1998 to 2004**

## **PART.I BACKGROUND AND INTRODUCTION**

### **I.1 Historical Synopsis**

The launching of Basic Education reform in 1989, gave rise to a realization of need for continuous monitoring and evaluation of the educational progress. Keeping track of the learning achievement of the student, important in its own right, became critical for monitoring the effects of the multifarious costly inputs, tangible and intangible, in the education system, and to provide necessary formative feedback to the education reform process.

Need for reliable and valid baseline data to gauge short-term and long-term outputs and outcomes of the education reform provided further impetus to establish an effective and efficient learning achievement monitoring system at the National Center for Human Resources Development (NCHRD). About the same time, in 1991, IAEP 2<sup>nd</sup> study conducted in 1989, in which Jordan was one of the 24 participants, released its results. The country that gave priority to improving the quality and standards of education was chagrined by the revelation that it came second from the bottom. The authorities and educational leadership found it hard to believe. A special committee was established to thoroughly investigate this vexing problem, while NCHRD, replicated the study on smaller but representative sample and came up with almost identical results. All these factors provided further impetus to designing and conducting the first national study on monitoring the learning achievement and instructional quality in the basic education system.

While the NCHRD was establishing a national system for continuous monitoring and assessment of the progress of basic education, in September 1992, UNICEF and UNESCO launched a global project, Monitoring Learning Achievement (MLA) to assess the midterm achievement of the goals of Education for All. The NCHRD assessment project fitted nicely with the goals of MLA and Jordan became one of the five countries in the world that participated at the initial stage of the UNESCO/UNICEF sponsored MLA project. Since 1992, UNICEF-Amman has provided financial assistance for all the four monitoring

---

learning achievement and instructional quality studies conducted by NCHRD in 1993, 1995, 1998 and 2004.

## I.2 Goai and Objectives

This is the fourth study in the series of Monitoring Learning Achievement and Instructional Quality at the primary level of the basic education. The third study was conducted nearly at the end of the 10-year general education reform program and the present study has taken place in the beginning of implementation of the current education reform project, "The Education Reform for Knowledge Economy (ERfKE)". ERfKE is an ambitious multidonar supported reform which aims to transform the whole education system to comprehensively address Jordan's need for globally competitive knowledge economy and, in the long run, to change Jordan into a knowledge society.

## I.3 Objectives of the Fourth Study

The primary objective of the 1993 study was to collect baseline data to serve as bench marks for later comparisons. The 1995 study focused upon assessment of reform impact. The 1998 study, in addition to measuring learning achievement and tracking educational progress, underlined the assessment of 4<sup>th</sup> grade students' general and specific competency levels in predefined content areas.

While the general objectives shared with the previous studies remain the same (See Ahlawat and Al-Nhar, 1999, pp. 2-5) Specific objective of this study is to compare Grade 4 student achievement in 1998 with that in 2004 in order to find any progress or lack of it. In addition to this, another objective is to establish baseline bench marks against which the impact of newly developed ICT-based e-learning supported curricula and pedagogy introduced under the knowledge economy focused reform could be measured and evaluated in due course of time.

## I.4 Sampling Strategy

Conforming to the definition of the population and sampling design of the previous studies, the two stages stratified random sample entailed selection of school from all the schools that had Grade 4 classes and then one class at random from all the Grade4 classes in each sampled school. The stratification dimensions used for school selection were Education Authority (Ministry of Education (MOE), Private, United Nations Relief Works Agency (UNRWA), and Ministry of Defence (MOD)) and School



Location (Urban and Rural). There was, however, a critical change introduced in the size of the sample due to resource constraints. Number of schools was reduced to 58 from 216 in the third study and the number of students was reduced from 5,929 to 2000. The sampling procedures and distribution of the 1998 sample are described in the NCHRD publication (Ahlawat and Al-Nhar, 1999 pp. 6-8) while the sampling distribution of this study is given in the following Table (1).

**Table (1)**  
**Sampling Distribution**

Sex	Location		Authority				Total
			MOE	UNRWA	Private	MOD	
Male	Urban	Count	471	55	98	10	634
		Row %	74.3	8.7	15.5	1.6	100
		Column %	57.4	48.2	88.3	100	60
		Total %	44.6	5.2	9.3	0.9	60
	Rural	Count	350	59	13		422
		Row %	82.9	14.0	3.1		100
		Column %	42.6	51.8	11.7		40
		Total %	33.1	5.6	1.2		40
	Total	Count	821	114	111	10	1056
		Row %	77.7	10.8	10.5	0.9	100
		Column %	100	100	100	100	100
		Total %	77.7	10.8	10.5	0.9	100
Female	Urban	Count	512	112	62	43	729
		Row %	70.2	15.4	8.5	5.9	100
		Column %	73.9	88.9	75.6	100	77.2
		Total %	54.2	11.9	6.6	4.6	77.2
	Rural	Count	181	14	20		215
		Row %	84.2	6.5	9.3		100
		Column %	26.1	11.1	24.4		22.8
		Total %	19.2	1.5	2.1		22.8
	Total	Count	693	126	82	43	944
		Row %	73.4	13.3	8.7	4.6	100
		Column %	100	100	100	100	100
		Total %	73.4	13.3	8.7	4.6	100

---

## 1.5 Instruments of the Study

Being a continual replication of the earlier studies executed under this program, the study used the same instruments as used in the 1998 study with one notable exception of dropping out the Social Studies from the three subjects tested in 1998. As a consequence, only the competency-based achievement tests of Arabic Language and Mathematics were applied in this study.

In addition to the achievement tests, student, teacher, and principal questionnaires were used to gather information on some key contextual factors deemed to influence student learning and achievement.

## 1.6 Content and Format of the Achievement Tests

In 1998 the 10-year-long education reform had nearly been completed and reformulated objectives, reconstructed curricula, redesigned textbooks and instructional materials and appropriately modified pedagogy were fully operational. This required achievement tests to be reconstructed to align them with the new educational realities on the ground. The reconstructed educational objectives were in line with the Education for All (EFA) objectives. Likewise, innovative instructional strategies applying appropriate available information and communication technologies were also in accord with the EFA Action Plan recommendations.

The new strategy adopted to measure the achievement of new learning outcomes entailed briefly the following:

Development of several independent forms of a test to ensure adequate coverage of knowledge, skills and competencies as well as content areas without increasing the length of the tests.

Different forms of a test were independent in the sense that each form incorporated exclusive sets of competencies from different content areas. The test forms thus, were complementary and NOT parallel or equivalent in the sense of the classical test theory.

In all, three new forms were developed for the Math test and two for the Arabic test.

The new forms measured the skills and competencies emphasized by the reform while the old Math and Arabic tests were retained as test forms to compare student achievement in these subjects in 1998 with that of their

cohorts in the years 1995 and 1993. In this way, the Math test had four forms and the Arabic test three.

## I.7 Structure of the Math Test

The three newly developed forms of the Math test tested three distinct broad competencies, each form covering a distinct broad competency, its constituent general competencies and their underlying specific competencies. The three broad competencies, each encompassing a separate form included: Understanding and Application of Procedures (Form 1); Mathematical Thinking and Communication (Form 2); and Problem Solving (Form 3).

The test specifications of the Math test forms are presented in the following tables, 2, 3 and 4, respectively for Form 1, Form 2 and Form 3.

**Table (2)**  
**Table of Specifications for Math Test Form 1: Understanding**  
**And Application of Procedures**

Content	Competency	Items
Natural Numbers	Properties	5
	Addition	2
	Subtraction	4
	Multiplication	4
	Division	2
Fractions		5
Decimals		5
Meas. & Geometry		3
<b>Total</b>		<b>30</b>

**Table (3)**  
**Table of Specifications for Math Test Form 2:**  
**Mathematical Thinking and Communication**

Content	Competency			Total
	Mathematical Thinking	Thinking & Communication	Communication	
Natural Numbers	2	5	2	9
Fractions	1	1	2	4
Meas. & Geometry	3	8	-	11
<b>Total</b>	<b>6</b>	<b>14</b>	<b>4</b>	<b>24</b>

Table (4)  
Table of Specifications for Math Test Form 3: Problem Solving

Content	Competency		
	One Step (One Operation)	Two or More Steps (Operations)	Total
Numbers	6	8	14
Meas. & Geometry	2	4	6
<b>Total</b>	<b>8</b>	<b>12</b>	<b>20</b>

## I.8 Structure of the Arabic Language Test

The Arabic Language Test was designed to measure student achievement levels in four major linguistic skills: Reading Comprehension, Grammar, Writing and Using Dictionary. Each major competency was subdivided into a number of minor or specific subcompetencies while each subcompetency was measured by one or more test questions. There were 52 test questions distributed over 2 test forms. Form 1 comprised 28 items and Form 2 comprised 24 items. The table of specifications for the Arabic Language Test is given in Table (5).

**Table (5)**  
**Table of Specifications for the Arabic Language Test**  
*(Form 1 and Form 2)*

Competency	Sub-competency	Form 1	Form 2	Number of Items
<b>Reading Comprehension</b>	* Give meanings of new words.	Q2: a, b, c		16
	* Identify main ideas.	Q1: 1, 2		
	* Identify implicit ideas.	Q1: 3, 4, 5		
	* Justify events.	Q6		
	* Give real life situations related to the text.	Q4: a, b, c		
	* Determine the expected end.	Q7		
	* Ask questions.	Q5: a, b		
<b>Grammar</b>	* Follow events in order.	Q3		17
	* Use language structures in a proper way.	Q10: a, b, c, d	Q5: a, b, c, d, e, f	
	* Apply language patterns.	Q11: a, b, c	Q6: a, b, c, d	
	* Fix word endings using grammatical rules.	--	Q7: a, b, c, d	
	* Relate structure to meaning.	--	Q9: a, b	
		--	Q8: a, b, c, d	
		Q9: a, b, c, d	--	
<b>Writing</b>	* Use punctuation marks.	--	Q3	8
	* Write a story about a picture or a set of pictures.	--	Q4	
	* Write a letter.	--	Q1	
	* Write a composition on a subject using a set of given questions.	--	Q2	
	* Use dialogue in writing a narration.	--	--	
<b>Using Dictionary</b>	* Order a set of words according to the first letter.	Q8		1
<b>Total</b>		<b>28</b>	<b>24</b>	<b>52</b>

---

## I.9 Data Collection

The study being a collaborative endeavor of the MOE and the NCHRD, MOE supervisors and NCHRD staff shared various responsibilities of the data collection process. While NCHRD staff handled the administration and management responsibilities, the MOE staff was in charge of the field operation.

A team of MOE supervisors was trained at the NCHRD in the procedures of test administration in school classrooms and supervision and monitoring of the whole data collection operation in the field.

The data were collected from students in sampled 4<sup>th</sup> grade classes in May 2004.

## I.10 Coding Scoring and Data Entry

When the tests reached back to NCHRD, they were checked for the accuracy of school and student IDs. Any discrepancies were cross checked and resolved.

Coding and scoring of each item in each form of the achievement tests required a specialized operation. Achievement tests were coded and scored in such a manner that student responses on each item could be classified into meaningful conceptual categories where each category stood for a well-defined type of error or misconception.

Having accomplished the tasks of coding and scoring, the data entry screens were created and debugged, and then data were entered and cleaned.

## I.11 Data Analysis

In accord with requirements of the intended objectives of this study a data analysis strategy was laid out. To start with, the psychometric properties of the instruments were determined, and then various statistical analyses appropriate to produce information required to answer the following generic questions of the study were conducted.

- ① What are the current levels of primary pupils' (Grad 4) general achievement in Arabic and Mathematics?

- 
- ② What are the national levels of primary pupils' general and specific competencies in different content and skill areas of Arabic and Mathematics?
  - ③ What is the progress made by the Grade4 students on the competency levels in three broad competencies in Math from 1998 to 2004?
  - ④ What is the progress made by the Grade 4 students on the competency levels in four broad competencies in Arabic Language from 1998 to 2004.
  - ⑤ What are the achievement gains made by different groups of 4<sup>th</sup> grade students on each test and subtest of Math and Arabic Language from 1998-2004 where student groups are defined by Education Authority, Student Sex, and School Location?

---

## **PART.II RESULTS**

This section of the report presents the results of the achievement tests in terms of students' performance at the national level on various general and specific competencies in each subject; then comparisons are made between performance levels in 1998 and 2004 for the national sample, for each Authority, for males, females, rural and urban schools measuring general and specific competencies in each subject.

### **II.1 MATH TESTS**

#### **II.1.1 Psychometric Properties of the Math Tests**

The math test had three new forms. Since each form measured performance on a different general competency and its constituent specific competencies, instead of using the terms Form 1, Form 2, Form 3 we have used the terms Test 1, Test 2, Test 3, respectively, in the following description.

Math Test 1 comprised of 30 items measuring the general competency of Understanding and Application of Mathematical Procedures in the content areas of Natural Numbers (17 items), Fractions (5 items), Decimals (5 items) and Measurement & Geometry (3 items). Test 1, was administered to 1472 students in 1998 and to 507 students in 2004.

The homogeneity or internal consistency reliability index (Coefficient Alpha) of this 30-item test was 0.89 on the occasions, 1998 and 2004, which is considered very good. The item difficulty indices ranged from 0.08 to 0.77 with a mean value of 0.38 in 1998 and 0.12 to 0.77 with a mean value of 0.39 in 2004. The discrimination power of all the items was very good. The Corrected Item Total Correlation Coefficients ranged from 0.30 to 0.58 in 1998 and from 0.17 to 0.61 in 2004.

Math Test 2 consisted of 24 items measuring the general competency of Mathematical Communication and Thinking in the content areas of Natural Numbers (9 items), Fractions (4 items), and Measurement and Geometry (11 items). Three subcompetencies were measured: Mathematical Thinking (6 items), Communication and Mathematical Thinking (14 items) and Communication (4 items). 1435 students took this test in 1998 and 604 in 2004.

The internal consistency reliability was reasonably high (Alpha = .88 in 1998 and .87 in 2004).



Difficulty indices ranged from 0.10 to 0.81 in 1998 and from 0.10 to 0.59 in 2004. Discrimination index values ranged from 0.09 to .65, in 1998; all of them were above 0.31 except one (0.09). In 2004 discrimination indices varied from 0.08 to 0.59. The discrimination index of 0.30 is considered reasonably good for such items.

Math Test 3 contained 20 questions designed to measure the general competency of Problem Solving in the content areas of Numbers (14 items) and Measurement and Geometry (6 items). 8 problems required only one step operation while 12 required two or more steps or operations to solve a problem. Alpha coefficient of this 20-item test was 0.93 in 1998 and 0.82 in 2004. Item mean score varied from 0.04 to 0.77 in 1998 and from .000 to .72 in 2004 while discrimination indices ranged from 0.10 to 0.59 in 1998 and from .00 to .57 in 2004.

### **II.1.2 Students' Performance on Math Tests**

#### **II.1.2.1 Test1: General Competency of Understanding and Application of Procedures: 1998 VS 2004**

In 1998 the national average percent correct score on procedures' knowledge and applications was only 37.9 with a minimum score of zero% and maximum of 100% correct. This means that an average 4<sup>th</sup> grade student in Jordan would obtain 38 marks out of 100 on this test, or the average student would be able to answer 38% of the items correctly. In other words, if we define level of attainment of this competency in terms of percentage of correct answers to test items then this would show that the national competency level is 38%, which is very low by every standard.

In 2004, the mean percent correct score on the same test increased by 2 percentage points, it is 40. The increase, however, is not statistically significant at  $\alpha=.05$  level of significance.

##### **II. I.2.1.1 Competency Levels on Math Test 1: 1998 VS 2004**

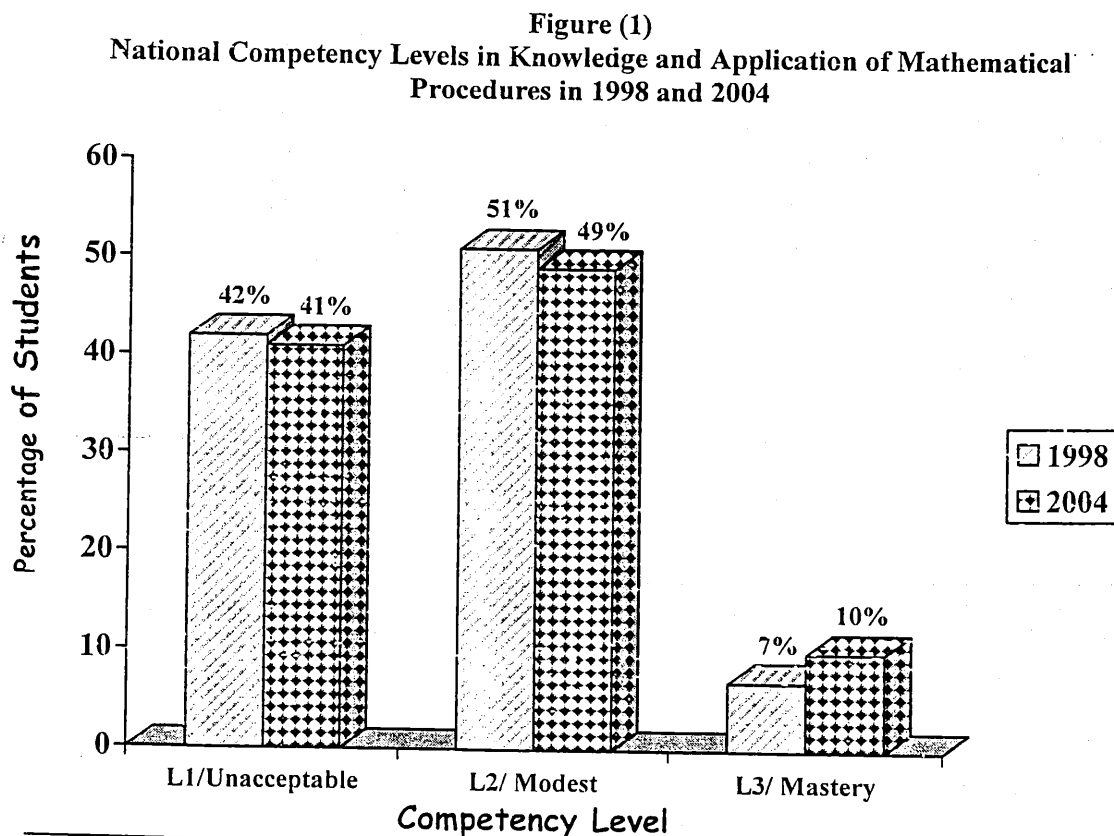
According to test performance, students were classified into three levels on the competency continuum as follows:

Level 1: Very low competency or unacceptable mastery: 30% or less;

Level 2: Modest competency or partial mastery: more than 30% but less than 70%;

Level 3: Mastery level: 70% or above.

Figure (1) shows the comparative percentages of students at each competency level in 1998 and 2004



The picture presented by Figure 1 is by no means flattering even when we set mastery level at 70% in 1998 only 7% of students had reached or crossed the mastery boundary line leaving the vast majority of 93% below the mastery cut-score of 70% correct. 51% of the Grade 4 students had scored more than 30% but less than 70%; we demurely call it a partial mastery or modest level. On the lower end of the scale it is disappointing to know that as much as 42% of the 4<sup>th</sup> grade students in 1998 acquired little competency in the knowledge and application of simple arithmetical procedures.

In 2004 the situation has hardly improved. As clearly shown by Figure 1, percentage of students scoring less than 30% marks has decreased by 10%, percentage of students in the modest mastery category (30%-70%) has decreased by 2 percentage points and the percentage of students attaining mastery (Level 3) has increased by 3% points (from 7% to 10%). The trend is certainly in the right direction where numbers at the lower-levels are on the decline and at the higher level on the increase.

National achievement levels indicated by mean percent correct scores on the general competency (Understanding and Application of Procedures) measured by TOTMP (total percent correct score on Math Test 1) and on the sub-competencies measured in terms of understanding and application of procedures to specific contents and operations such as properties of Natural Numbers (NUMBERP); Addition (NUMADP), Subtraction (NUMSUBP), Multiplication (NUMMULP) and Division (NUMDIVP) of natural numbers; Fractions (FRACTIOP); Decimals (DECIMALP); and Measurement and Geometry (GEUETVP) in the years 1998 and 2004 are presented in Table (6).

**Table (6)**  
**Comparative Achievement of G4 Students on Math Test 1 (Understanding and Application of Procedures) in 1998 and 2004 (National Level)**

Competency	Year	N	Mean	Std. Deviation	Std. Error Mean	Mean Diff.	df	Sig 2-tailed	t
TOTMP	2004	506	38.92	22.22	.99	.99	1976	.36	.89
	1998	1472	37.93	21.56	.56				
NUMBERP	2004	506	31.15	28.65	1.27	-2.30	1976	.12	-1.54
	1998	1472	33.45	29.09	.7				
NUMADP	2004	506	71.74	38.56	1.71	.31	1976	.86	.16
	1998	1472	71.43	37.39	.97				
NUMSUBP	2004	506	48.81	33.94	1.51	.39	1976	.02	2.30
	1998	1472	44.87	33.03	.86				
NUMMULP	2004	506	23.51	29.64	1.32	3.65	1976	.01	2.66
	1998	1472	19.87	25.53	.67				
NUMDIVP	2004	506	22.23	32.62	1.45	.87	1976	.60	.53
	1998	1472	21.36	31.83	.83				
FRACTIOP	2004	506	50.19	31.24	1.34	3.55	838*	.03	2.24
	1998	1472	46.64	29.62	.77				
GEUETVP	2004	506	38.14	30.12	1.34	-5.06	884*	.001	-3.25
	1998	1472	43.20	30.44	.79				
DECIMALP	2004	506	33.83	29.24	1.30	1.21	1976	.40	.84
	1998	1472	32.62	27.59	.72				

\* Degrees of freedom and corresponding statistics for unequal variance.

Now, if we glance over the Mean Difference column in Table (6), we see a positive trend. All the differences (2004-1998) are positive except on two sub-competencies viz., Natural Numbers (NUMBERP) and Measurement and Geometry (GEUETVP). On three of the eight sub-competencies, namely, Subtraction of Natural Numbers (NUMSUBP), Multiplication of Natural Numbers (NUMMULP) and Fractions (FRACTIOP) the increase in 2004 is statistically significant at ( $p < .02$ ), ( $p < .01$ ) and ( $p < .03$ ) respectively. Out of the two sub-competencies with decreased scores the decrease is statistically significant ( $p < .001$ ) on only one, Geometry and Measurement. On the whole, the performance level is rather low. The highest level (71.74% correct) in 2004 and (71.43% correct) in 1998 was reached on Addition Sub-competency. The second

best achievement is way down (50.20%) and (46.64%) in 2004 and 1998, respectively on Fractions Sub-competency. The lowest achievement in 1998 was (19.87%) on Multiplication while in 2004 it was on Division (22.23%).

National performance level in the general competency areas of Understanding and Application of Procedures has not really changed over the past six years (1998 to 2004). The total test score mean was 37.93% in 1998 and 38.92% in 2004.

#### **II.I.2.1.2 Comparative Performance on Math Test 1 in 1998 and 2004 in Each Education Authority**

Education Authority was a major dimension of stratification used in sampling. There are four Education Authorities: Ministry of Education (MOE), Ministry of Defence (MOD), United Nations Relief and Works Agency (UNRWA) and Private Schools (Private). By and large, schools under all authorities are mandatorily required to follow the national curricula prescribed by the MOE. Some schools, mainly private, may, however, teach additional subjects and prepare students for such examinations as SAT of the USA and GCE of the UK. Besides, all schools are under the supervision of the MOE. With respect to educational facilities and school inputs, substantial diversity may exist among schools under different educational jurisdictions (Authorities).

In view of this situation, it is logical to study educational progress under different Education Authorities. In the following we present comparative data on student achievement in the general competency of Understanding and Application of Procedures (Math Test 1) in 1998 and 2004.

##### **II.I.2.1.2.1 Comparative Performance of MOE G4 Students in 1998 and 2004 on the Math Test1 (Understanding and Application of Procedures)**

The relevant statistics including Mean, Standard Deviation (SD), Std. Error of the Mean Difference, degrees of freedom (df), two-tailed statistical significance obtained using independent sample, equal or unequal variance t-test analysis are given in Table (7).

Table (7)  
Comparative Achievement of G4 MOE Students in 1998 and 2004 on Math Test 1  
General Competency (Understanding and Application of Procedures)  
and its Sub-competencies.

Competency	Year	N	Mean	Std. Deviation	Std. Error Mean	Mean Diff.	df	Sig 2-tailed	t
TOTMP	2004	385	37.14	22.04	1.12	.82	1552	.51	66
	1998	1169	36.31	20.92	.61				
NUMBERP	2004	385	29.77	28.25	1.44	-1.99	1552	.24	-1.19
	1998	1169	31.75	28.54	.83				
NUMADP	2004	385	69.09	39.93	20.3	-.63	1552	.78	-28
	1998	1169	69.72	38.24	1.12				
NUMSUBP	2004	385	46.43	34.07	1.74	3.90	1552	.05	2.00
	1998	1169	42.54	32.79	.96				
NUMMULP	2004	385	22.14	28.73	1.46	3.90	572*	.02	240
	1998	1169	18.24	24.07	.70				
NUMDIVP	2004	385	20.39	30.72	1.57	1.14	1552	.52	640
	1998	1169	19.25	30.29	.89				
FRACTIOP	2004	385	47.79	31.94	1.63	2.99	605*	.10	1.62
	1998	1169	44.81	28.96	0.85				
GEUETVP	2004	385	37.58	30.75	1.57	-4.71	1552	.01	2.63
	1998	1169	42.29	30.40	.83				
CECIMALP	2004	385	32.05	28.32	1.44	.32	1552	.84	-1.92
	1998	1169	31.74	26.90	.79				

\* df for unequal variance t-test.

The mean performance of the MOE G4 students on the general competency of Understanding and Application of Procedures has increased by 1% point over the last five years from 1998 to 2004. Also it was quite low just like the national average (only one point less than the national mean on this competency) in both 1998 and 2004 assessments. The MOE mean scores on this competency were 37.13% and 36.31%, respectively in 2004 and 1998 as compared to those at the national level (the whole sample) of 38.92% and 37.93%, in the same order. There is a mixed trend here. Under the Mean Difference column in Table (7), there are three negative signs indicating a decrease in performance on Number, Addition, and Geometry and Measurement sub-competencies; the decline is statistically significant only on one of them, Geometry and Measurement ( $p \leq .01$ ).

On the positive side, on five of the eight sub-competencies there is a sign of increase; on two of them the increase is statistically significant, Subtraction ( $p < .05$ ) and Multiplication ( $p < .02$ ). On these two sub-competencies (Subtraction and Multiplication) increase was statistically significant at the national level also.

### II.1.2.1.2.2 Comparative Performance of MOD G4 Students in 1998 and 2004 on the Math Test1 (Understanding and Application of Procedures)

The results of statistical analysis using unequal independent sample t-test are provided in Table (8).

Table (8)  
Comparative Achievement of G4 MOD Students in Math Test 1 (Understanding and Application of Procedures) in 1998 and 2004

Competency	Year	N	Mean	Std. Deviation	Std. Error Mean	Mean Diff.	df	Sig 2-tailed	t
TOTMP	2004	14	54.52	22.21	5.94	23.61	52	.00	3.80
	1998	40	30.92	19.19	3.03				
NUMBERP	2004	14	45.71	36.31	9.71	21.21	16*	.05	2.63
	1998	40	24.50	21.48	3.40				
NUMADP	2004	14	78.57	37.80	10.10	8.57	52	.45	.74
	1998	40	70.00	37.21	5.88				
NUMSUBP	2004	14	69.64	20.00	5.36	31.52	40	.00	3.19
	1998	40	38.12	34.89	5.52				
NUMMULP	2004	14	32.14	35.93	9.60	19.02	16*	.08	2.39
	1998	40	13.12	21.17	3.35				
NUMDIVP	2004	14	28.57	37.80	10.10	12.32	52	.21	1.27
	1998	40	16.25	28.62	4.52				
FRACTIOP	2004	14	68.57	24.45	6.53	29.07	52	.001	3.51
	1998	40	39.50	27.36	4.32				
GEUETVP	2004	14	54.76	24.83	6.64	19.76	52	.04	2.15
	1998	40	35.00	31.08	4.91				
CECIMALP	2004	14	55.71	33.45	8.94	30.71	52	.001	3.61
	1998	40	25.00	25.12	3.97				

\* df for unequal variance t-test.

It is evident from Mean Difference column in Table (8) that in contrast with the gains made at the national level and by MOE students, the MOD's achievement is somewhat impressive. There are substantial gains on the total score as well as on all the sub-competencies. Moreover, the increase is statistically, highly significant ( $p < .001$ ) on the General Competency (total test score) and three sub-competencies (Subtraction, Fractions and Decimals); increase on two other competencies (Number and, Geometry and Measurement) is also statistically significant ( $p < .05$ ). On three sub-competencies (Addition, Multiplication and Division), however, the increase did not reach the level of statistical significance of ( $p \leq .05$ ).

On the whole, the percent correct gains vary from 8.5%, the lowest, on Addition, through 31.52%, the highest on Subtraction.

#### II.1.2.1.2.3 Comparative Performance of UNRWA G4 Students in 1998 and 2004 on the Math Test 1 (Understanding and Application of Procedures)

The results of independent unequal sample t-test analysis for the UNRWA fourth grade students' performance in 1998 and 2004 on the general competency of Understanding and Application of Procedures, and its eight sub-competencies are presented in Table (9).

**Table (9)**  
**Comparative Performance of UNRWA 4<sup>th</sup> Graders on Math Test 1 (Understanding and Application of Procedures) in 1998 and 2004**

Competency	Year	N	Mean	Std. Deviation	STd. Error Mean	Mean Diff.	df	Sig 2-tailed	t
TOTMP	2004	62	36.08	21.21	2.69	-3.48	179	.33	-.98
	1998	119	39.55	23.22	2.13				
NUMBERP	2004	62	26.45	26.06	3.31	-7.33	179	.31	-1.69
	1998	119	33.78	28.49	2.61				
NUMADP	2004	62	73.39	34.73	4.41	1.12	179	.10	.200
	1998	119	72.27	36.07	3.31				
NUMSUBP	2004	62	46.37	36.17	4.59	-.69	179	.84	-.13
	1998	119	47.06	33.69	3.09				
NUMMULP	2004	62	20.56	32.49	4.12	-1.49	179	.90	-.32
	1998	119	22.06	28.78	2.64				
NUMDIVP	2004	62	24.19	38.13	4.84	-3.54	179	.75	-.60
	1998	119	27.73	37.22	3.41				
FRACTIOP	2004	62	49.35	27.51	3.49	-1.12	179	.55	-.23
	1998	119	48.24	32.14	2.95				
GEUETVP	2004	62	39.78	29.47	3.74	-4.47	179	.82	-.94
	1998	119	44.26	30.72	2.82				
CECIMALP	2004	62	24.19	26.40	3.35	-9.25	179	.35	-2.13
	1998	119	33.45	28.42	2.60				

Negative signs on all the differences except on one in the, "Mean Difference" column of Table (9), point towards a declining trend in UNRWA students' achievement in Math. None of the differences is statistically significant; nevertheless, the pattern itself signals alarm.

#### II.1.2.1.2.4 Comparative Performance of Private Schools Grade 4 Students on Math Test 1 General and Sub-competencies in 1998 and 2004

The results of independent unequal sample t-test on Math Test 1 (Understanding and Application of Procedures) and the associated sub-competencies comparing students' performance in 1998 and 2004 are presented in the following Table (10).

**Table (10)**  
**Comparative Performance of Private School 4<sup>th</sup> Graders on Math Test 1**  
**(Understanding and Application of Procedures) in 1998 and 2004**

Competency	Year	N	Mean	Std. Deviation	Std. Error Mean	Mean Diff.	df	t	Sig 2-tailed
TOTMP	2004	45	53.26	18.14	2.70	1.64	187	.47	.64
	1998	144	51.62	20.83	1.74				
NUMBERP	2004	45	44.48	28.65	4.27	-4.56	187	-.88	.36
	1998	144	49.44	30.83	2.57				
NUMADP	2004	45	90.00	25.23	3.76	4.93	187	1.06	.27
	1998	144	85.07	27.78	2.32				
NUMSUBP	2004	45	66.11	25.65	3.82	2.22	187	.47	.64
	1998	144	63.89	28.01	2.33				
NUMMULP	2004	45	36.67	29.00	4.32	3.51	187	.68	.49
	1998	144	33.16	30.84	2.57				
NUMDIVP	2004	45	33.33	36.93	5.50	-1.39	187	-.22	.82
	1998	144	34.72	36.10	3.00				
FRACTIOP	2004	45	33.22	25.52	3.80	4.00	187	.84	.41
	1998	144	62.22	28.79	2.40				
GEUETVP	2004	45	35.56	26.01	3.88	-16.53	187	-3.43	.001
	1998	144	52.08	28.89	2.41				
CECIMALP	2004	45	55.56	26.93	4.01	14.31	187	2.76	.01
	1998	144	41.25	31.40	2.62				

As expected, private schools performance is better than that of other education authorities; their achievement in Maths, nevertheless, has not shown improvement over the last five years (1998-2004). As can be seen from the Mean Difference and 2-tailed Significance columns in Table (10), there is statistically significant ( $p < .01$ ) improvement on only one of the eight subcompetencies (Decimals) which, unfortunately, has been negated by a stronger decline in the achievement of subcompetency (Geometry and Measurement) statistically significant at the level ( $p < .001$ ).

The mean Difference column in Table (10) displays a mixture of plus and minus signs indicating no clear pattern of progress or recession in the Math achievement of Private Schools' 4<sup>th</sup> grade students from 1998 to 2004.

#### II.1.2.1.3 Differential Progress in Urban and Rural Schools in Math Test1 (General Competency in Understanding and Application of Procedures)

In general, it is known that students studying in urban schools perform better than their rural school cohorts on achievement tests. The progress made from 1998 through 2004 in the Math's area of general competency in Understanding and Application of Procedures is the subject of inquiry here. It is anybody's guess which way the winds of improvement blew in Urban and Rural locations. We have presented a summary of the results of t-test between 1998 and 2004 on the general competency of Understanding and Application of Procedures and its subcompetencies, conducted separately for Urban and Rural 4<sup>th</sup> Graders in Table (11).



Glancing over the columns "Mean Difference" under "Urban" and "Rural" in Table 11 clearly shows the disparity in Math progress of G4 students from 1998 through 2004 in Urban and Rural schools in Jordan.

**Table (11)**  
**Differences on Math Test 1 (Understanding and Application of Procedures) Achievement in 1998 and 2004 and Their Statistical Significance in Urban and Rural Schools**

Competency	Year	Urban			Rural		
		Mean	Mean Diff.	Sig 2-tailed	Mean	Mean Diff.	Sig 2-tailed
TOTMP	2004	41.36	2.24	.11	33.63	-2.80	.13
	1998	39.12			36.44		
NUMBERP	2004	33.53	-.95	.62	26.00	-6.18	.01
	1998	34.48			32.18		
NUMADP	2004	74.71	2.44	.31	65.31	-5.08	.13
	1998	72.27			70.40		
NUMSUBP	2004	52.75	6.06	.01	40.31	-2.31	.43
	1998	46.69			42.62		
NUMMULP	2004	25.94	4.96	.01	18.28	-2.1	.92
	1998	20.98			18.49		
NUMDIVP	2004	25.29	2.41	.26	15.63	-3.86	.12
	1998	22.88			19.48		
FRACTIOP	2004	53.06	5.06	.01	44.00	-.96	.73
	1998	48.00			44.96		
GEUETVP	2004	37.96	-5.68	.00	38.54	-4.13	.11
	1998	43.64			42.67		
DEC/MALP	2004	35.84	1.97	.28	29.50	-1.58	.51
	1998	33.87			31.08		

For the Urban schools we note that almost all the differences are positive, i.e., there is a growth trend, however slight, on all the subcompetencies and the general competency with two exceptions out of nine. The two subcompetencies on which there is some decline, though statistically nonsignificant on Number is significant on Geometry and Measurement ( $p < .00$ ), from 1998 to 2004. On a more positive note, the increase in test performance is statistically significant ( $p < .01$ ) on three out of eight subcompetencies, viz., Subtraction, Multiplication and Fractions.

In the field of Geometry and Measurement, however, there is a decline of the magnitude of nearly six percentage points from 1998 to 2004.

In a sharp contrast, in the Rural schools there is consistent pattern of falling scores in all the subcompetencies and, consequently, in the broad competency measured by the total score on all the subcompetencies. The decrease, however, is nowhere statistically significant at ( $\alpha \leq .05$ ) level, except on one subcompetency, Number, where the decline was over 5 percentage points ( $p < .02$ ).

#### II.1.2.1.4 Differential Progress of Male and Female 4<sup>th</sup> Graders on Math Test 1 (Understanding and Application of (Procedures) from 1998 to 2004

Male students doing slightly better than their female cohorts on Math tests has been a common observation of many researchers in this field, throughout the world. Jordanian researchers, on the other hand, have reported research findings that contradict this assertion. A number of research studies conducted over the past decade in Jordan have reported female students outperforming their male cohorts across the board in all subjects including Maths.

The main objective of this report is not to investigate the differential performance of Girls and Boys on the Math achievement tests. The primary purpose of this study is to monitor the progress of 4<sup>th</sup> grade students in Math. In order to eliminate the confounding effects of differential male/female progress we present, in the following paragraphs, the improvement made by male and female students in their achievement in mathematics, separately, first for Males and then for Females.

##### II.1.2.1.4.1 Male 4<sup>th</sup> Graders Performance on Math Test 1 in 1998 and 2004.

The results of the statistical analysis (t-test) conducted on the male Grade 4 students' achievement scores on Math Test 1 in 1998 and then after six years in 2004 are presented in Table (12).

Table (12)  
Male 4<sup>th</sup> Graders' Differential Achievement on Math Test 1 (Understanding and Application of Procedures) in 1998 and 2004

Competency	Year	N	Mean	Std. Deviation	Std. Error Mean	Mean Diffe	df	Sig 2-tailed	t
TCTMP	2004	258	34.91	22.09	1.38	-2.22	1009	.159	-1.41
	1998	753	37.13	21.77	.79				
NUMBERP	2004	258	27.82	27.03	1.68	-4.81	1009	.019	-2.39
	1998	753	32.64	28.96	1.05				
NUMADP	2004	258	67.05	39.78	2.47	-3.86	1009	.157	-1.39
	1998	753	70.92	38.04	1.39				
NUMSUBP	2004	258	44.38	34.20	2.13	.49	1009	.840	.20
	1998	753	43.89	33.31	1.21				
NUMMULP	2004	258	20.64	27.51	1.71	+1.18	1009	.525	+.64
	1998	753	19.46	25.17	.92				
NUMDIVP	2004	258	19.57	31.42	1.96	-.61	1009	.788	-.27
	1998	753	20.19	31.65	1.15				
FRACTIOP	2004	258	43.95	32.17	2.00	-2.82	1009	.205	-1.27
	1998	753	46.77	30.34	1.10				
GEUETVP	2004	258	35.79	28.32	1.76	-6.80	485*	.001	-3.24
	1998	753	42.59	31.16	1.13				
CECIMALP	2004	258	29.53	29.32	1.83	-1.17	1009	.566	-.57
	1998	753	30.70	27.84	1.01				

\* df for unequal variance t-test.

The column "Mean Difference" in Table 12 shows a consistent array of negative signs which indicate a decline in the male 4<sup>th</sup> grade students' achievement in the general competency of Understanding and Application of Procedures over the past five years from 1998 to 2004. The decline though consistent over seven sub-competencies out of eight, is not statistically significant at  $\alpha=.05$  level except in Geometry and Measurement sub-competency in which a decrease of almost 7%-tage points is statistically significant ( $p<.001$ ).

#### II.1.2.1.4.2 Female 4<sup>th</sup> Graders Performance on Math Test 1 in 1998 and 2004

Independent unequal sample t-Test procedure was applied to study the change in female 4<sup>th</sup> grade students' math achievement from 1998 to 2004. The results of this analysis are presented in Table (13).

**Table (13)**  
**Female 4<sup>th</sup> Graders' Differential Achievement on Math Test 1 (Understanding and Application of Procedures) in 1998 and 2004**

Competency	Year	N	Mean	Std. Deviation	STd. Error Mean	Mean Diff	df	Sig 2-tailed	t
TOTMP	2004	248	43.09	21.61	1.37	4.33	965	.01	2.75
	1998	719	38.76	21.32	.80				
NUMBERP	2004	248	34.60	29.92	1.90	.30	965	.89	.14
	1998	719	34.30	29.22	1.09				
NUMADP	2004	248	76.61	36.79	2.34	4.64	965	.10	1.71
	1998	719	71.98	36.71	1.37				
NUMSUBP	2004	248	53.42	33.11	2.10	7.53	965	.002	3.11
	1998	719	45.90	32.84	1.22				
NUMMULP	2004	248	26.51	31.58	2.00	6.21	965	.002	3.07
	1998	719	20.30	25.92	.97				
NUMDIVP	2004	248	25.00	33.67	2.14	2.40	965	.32	1.00
	1998	719	22.00	32.00	1.19				
FRACTIOP	2004	248	56.69	28.91	1.83	10.18	965	.00	4.79
	1998	719	46.51	28.87	1.08				
GEUETVP	2004	248	40.59	31.76	2.01	-3.27	965	.14	-1.47
	1998	719	43.86	29.68	1.11				
CECIMALP	2004	248	38.31	28.55	1.81	3.68	965	.07	1.81
	1998	719	34.63	27.21	1.01				

Female students outperform their male cohorts almost across the board in achievement in school subjects, is a well-recognized fact in Jordan. Glancing over the "Mean Difference" column in Table (13) shows that all the signs are positive except one. This is opposite of what appears in Table (12) for male students' results where all signs are negative except one.

Plus signs in the "Mean Difference" column indicate improvement in test performance. Apart from the total score, on three of the eight sub-

competencies female students' progress is statistically significant ( $p < .002$ ). The areas of significant improvement are: Subtraction, Multiplication and Fractions.

### **II.1.2.2 Math Test2: General Competency in Mathematical Thinking and Communication**

Math Tested 2, designed to measure the General Competency in Mathematical Thinking and Communication consisted of 24 items. The general competency was divided into three subcompetencies: Mathematical Thinking (6 items), Communication and Thinking (14 items), and Mathematical Communication (4 items). The first two subcompetencies viz. Mathematical Thinking and Communication and Thinking were applied to three content areas: Natural Numbers, Fractions, and Measurement and Geometry. The third subcompetency (Mathematical Communication) was applied to only two topics (Natural Numbers, Fractions).

The Math Test 2 was administered to a subsample of 1435 4<sup>th</sup> graders in 1998 and to a subsample of 504 4<sup>th</sup> graders in 2004.

For each student percent correct scores were computed on the total test and on each subtest measuring a subcompetency in a content area.

The primary focus of this report is on studying the changes in primary pupils' achievement in Math and Arabic Language from 1998 to 2004.

#### **II.1.2.2.1 National Competency Levels of Grade 4 Students on Math Test 2 (Mathematical Thinking and Communication) in 1998 and 2004**

In 1998 it was decided to group students into three broad competency level categories each defining a successive mastery level. Three levels of competency were defined as follows:

<u>Level 1:</u>	Indicating a very low degree of competency and No Mastery. Students scoring 30% or less on the competency test are classified in Level 1 category.
<u>Level 2:</u>	Indicating a modest degree of competency or partial Mastery defined by a score ranging from more than 30% and less than 70% on a competency test.
<u>Level 3:</u>	Indicating high degree of competency or Mastery level defined by a score of 70% or above on a competency test.

The following Figure (2) exhibits the different competency levels reached by students in 2004 against those in 1998.

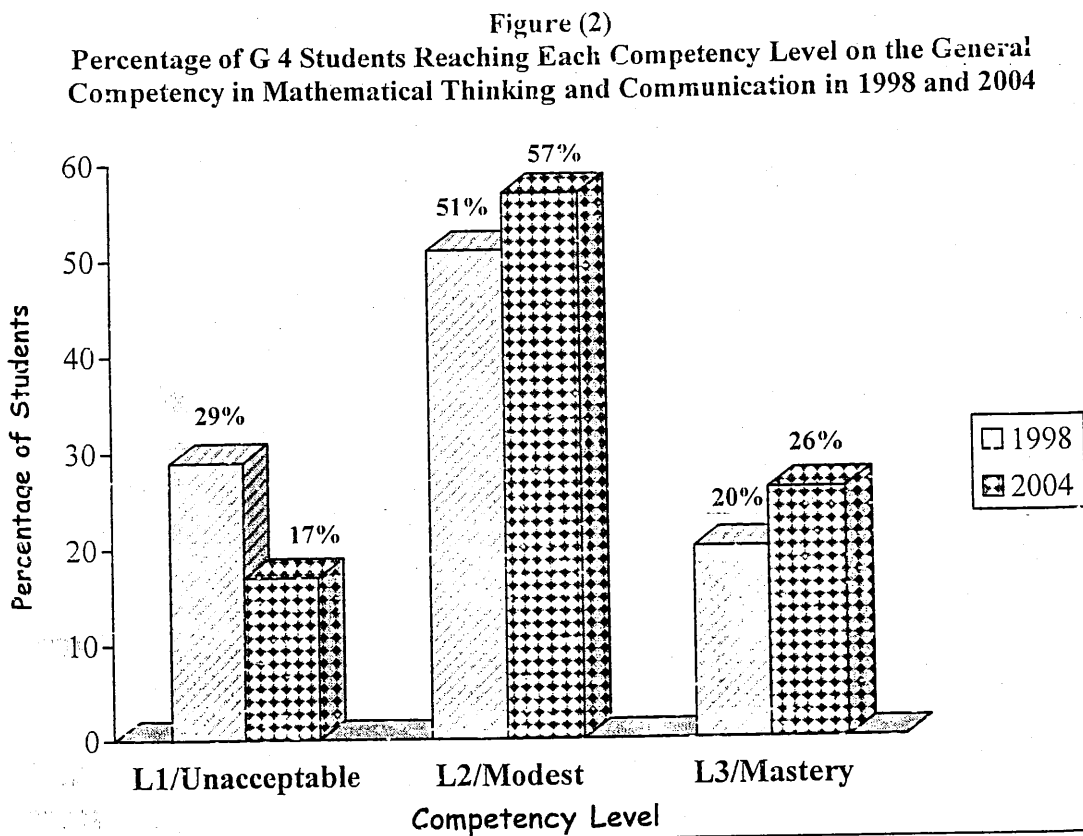


Figure (2) displays comparative percentages of 4<sup>th</sup> grade students reaching successive competency levels in 1998 and 2004.

Evidently, there is improvement in student achievement clearly indicated by 12 percentage point decrease at the lowest level and six percentage point increase at both Level 2 and Level 3.

While seeing some improvement in terms of a small proportion of students moving up the ladder of competency level of the broad competency in Mathematical Thinking and Communication Table 14 presents the progress at the sub-competency and more specific competency levels.

**Table (14)**  
**Mean, Mean Difference, df, t-value and 2-Tailed Significance of the Mean Difference**  
**Between 1998 and 2004 Performance of Grade 4 Students on Math Test 2 in Jordan**

Competency	Year	N	Mean	STd. Error Mean	Mean Diffe	df	t	Sig 2-tailed
THINKP	2004	504	62.62	1.19	.70	1937	.48	.63
	1998	1435	61.92	.75				
COMTHINP	2004	504	38.36	.99	.17	971*	.14	.89
	1998	1435	38.17	.66				
COMP	2004	504	56.67	1.41	1.45	1937	.87	.39
	1998	1435	55.21	.86				
NUMBERP	2004	504	47.78	1.20	1.59	918*	1.12	.26
	1998	1435	46.19	.75				
FRACTIOP	2004	504	59.84	1.34	1.85	1937	1.18	.28
	1998	1435	57.99	.80				
GOEMETRP	2004	504	43.25	.92	.63	1937	.55	.58
	1998	1435	42.62	.60				
NUMTHINP	2004	504	60.42	1.67	2.23	1937	1.13	.26
	1998	1435	58.19	1.02				
NUMCOTHP	2004	504	39.91	1.28	1.29	996*	.85	.39
	1998	1435	38.62	.81				
NUMCOMP	2004	504	54.79	1.74	1.68	1937	.83	.41
	1998	1435	53.10	1.04				
GEOTHINP	2004	504	58.69	1.38	.65	1937	.39	.70
	1998	1435	58.05	.87				
GECOTHP	2004	504	36.54	2.09	.16	1937	.07	.95
	1998	1435	36.38	1.27				
FRACOTHP	2004	504	37.54	2.09	1.16	1937	.47	.64
	1998	1435	36.39	1.27				
FRACOMP	2004	504	58.54	1.70	1.22	1937	.62	.53
	1998	1435	57.32	1.00				
PTOTM2	2004	504	48.71	.96	2.19	941*	1.92	.06
	1998	1435	46.52	.61				

\* df for unequal variance t-test.

The column "Mean Difference" in Table (14) shows the gain-mean scores on each sub-competency and specific competency at the national level. The trend is definitely positive, a healthy sign, the gains, however, are miniscule, not a single one of them reaches anywhere near statistical significance at ( $\alpha \leq .05$ ) level. It is, nevertheless, heartening to note the consistency of increment on every specific subscale of the Math Test2. The mean percent score gain varies from one specific competency to another from .17%, the lowest on Mathematical Communication and Thinking to 2.23%, the highest on Number Thinking (Mathematical Thinking Applied to Natural Numbers); the increase on the total test score (PTOT M2) is 2.19%, the only one that could have reached statistical significance had it been set at ( $\alpha \leq .06$ ) level.

#### **II.1.2.2.2 Fourth Grade Students' Achievement on Math Test2 (General Competency in Mathematical Thinking and Communication) in 1998 and 2004 in Each Education Authority**

In the following sections we discuss the changes in Grade 4 students' achievement on Math Test 2 as a whole and on its sub-scales from 1998 to 2004 within the jurisdiction of each education authority.

#### II.1.2.2.2.1 MOE 4<sup>th</sup> Graders' Achievement on the General Competency in Mathematical Thinking and Communication in 1998 and 2004

The following Table (15) displays the results of statistical analysis comparing the 1998 4<sup>th</sup> grade students' performance on Math Test 2 with that of their 2004 cohorts in the Ministry of Education Schools.

**Table (15)**  
Means, Mean Difference, Degrees of Freedom, t-value and their 2-tailed Significance, on the Math Test 2 Scores of the 4<sup>th</sup> Grade MOE students in 1998 and 2004

Competency	Year	N	Mean	SD	Std. Error Mean	Mean Diff.	df	t	Sig 2-tailed
THINKP	2004	377	61.94	26.91	1.38	.92	1519	.54	.58
	1998	1144	61.02	28.81	.85				
COMTHINP	2004	377	36.95	21.62	1.11	-.71	1519	-.49	.62
	1998	1144	37.66	25.07	.74				
COMP	2004	377	55.77	32.62	1.68	3.00	1519	1.55	.12
	1998	1144	52.77	32.51	.96				
NUMBERP	2004	377	46.28	26.75	1.37	1.53	1519	.92	.35
	1998	1144	44.75	28.02	.82				
FRACTIOP	2004	377	58.96	29.91	1.54	2.54	1519	1.41	.16
	1998	1144	56.42	30.43	.89				
GOEMETRP	2004	377	42.56	20.34	1.04	.31	1519	.23	.30
	1998	1144	42.25	22.97	.67				
NUMTHINP	2004	377	58.45	38.11	1.96	1.33	1519	.58	1.33
	1998	1144	57.12	38.44	1.13				
NUMCOTHP	2004	377	38.40	28.27	1.45	1.03	1519	.58	1.02
	1998	1144	37.37	30.27	.89				
NUMCOMP	2004	377	53.79	38.75	1.99	2.97	1519	1.2	2.96
	1998	1144	50.83	39.20	1.15				
GEOTHINP	2004	377	58.64	30.40	1.56	1.30	1519	.67	1.29
	1998	1144	57.34	33.15	.98				
GECOTHP	2004	377	34.50	46.10	2.37	-1.86	1519	-.65	-1.85
	1998	1144	36.36	48.12	1.42				
FRACOTHP	2004	377	35.50	46.10	2.37	-.86	1519	-.30	-.85
	1998	1144	36.36	48.12	1.42				
FRACOMP	2004	377	57.76	38.77	1.99	3.04	1519	1.33	3.04
	1998	1144	54.72	38.10	1.12				
PTOTM2	2004	377	47.69	21.34	1.09	2.14	693*	1.64	2.13
	1998	1144	45.55	23.26	.68				

\* df for unequal variance t-test.

From the "Mean Difference" column in Table (15) we see that on 11 out of 13 sub-competencies there is slight gain in scores while on two of them there is a little loss. Not a single difference, nevertheless, is statistically significant which is evident from the "Sig". column in Table (15). On the total score also the gain is not statistically significant at the prescribed level of ( $\alpha=.05$ ). On the whole, the trend is in the right direction.

#### II.1.2.2.2 UNRWA Grade 4 Students Progress in Math Test2 (General Competency in Mathematical Thinking and Communication) from 1998 to 2004

United Nations Relief and Works Agency (UNRWA) looks after (among other things) the education of Palestinian refugees in Jordan through its education system. In the past UNRWA schools were considered to be better equipped with educational facilities and qualified teachers. On achievement tests also UNRWA students usually performed slightly better than their MOE cohorts. Here, the objective is to assess the 4<sup>th</sup> grade UNRWA students' improvement in Math achievement on Math Test2 (General Competency in Mathematical Thinking and Communication) from 1998 – 2004.

The results of the independent, unequal sample t-test analyses conducted to study the difference between students' test scores in 1998 and 2004, are presented in Table (16).

Table (16)  
Mean, SD, Mean Difference, df, t-Valu, Std. Err. Of the Difference and 2-Tailed Significance of the Mean Difference for UNRWA 4<sup>th</sup> Graders between 1998 and 2004

Competency	Year	N	Mean	Std. Devi	Std. Error Mean	Mean Diff.	df	t	Sig 2-tailed
THINKP	2004	65	57.64	28.14	3.49	-1.96	181	-.45	.64
	1998	118	59.60	27.51	2.53				
COMTHINP	2004	65	36.97	25.35	3.14	.20	181	.05	.96
	1998	118	36.77	26.03	2.39				
COMP	2004	65	46.61	26.32	3.26	-9.74	60.19*	-2.16	.03
	1998	118	56.35	33.68	3.10				
NUMBERP	2004	65	42.05	26.93	3.34	-2.90	181	-.65	.51
	1998	118	44.96	29.31	2.69				
FRACTIOP	2004	65	54.15	32.00	3.96	-1.35	181	-.027	.78
	1998	118	55.50	31.20	2.87				
GOEMETRP	2004	65	40.85	22.62	2.80	-1.36	181	-.38	.70
	1998	118	42.21	23.09	2.12				
NUMTHINP	2004	65	60.50	37.70	4.67	7.95	181	1.30	.19
	1998	118	52.54	40.21	3.70				
NUMCOTHP	2004	65	35.53	29.22	3.62	-2.51	181	-.53	.59
	1998	118	38.05	30.80	2.83				
NUMCOMP	2004	65	39.92	35.65	4.42	-14.73	181	-2.48	.01
	1998	118	54.66	39.75	3.65				
GEOTHINP	2004	65	52.20	33.34	4.13	-5.98	181	-1.22	.22
	1998	118	58.19	30.87	2.84				
GECOTHP	2004	65	34.76	46.51	5.76	6.80	181	.96	.33
	1998	118	27.96	45.07	4.14				
FRACOTHP	2004	65	35.76	46.51	5.76	7.80	181	1.10	.26
	1998	118	27.96	45.07	4.14				
FRACOMP	2004	65	53.30	37.94	4.70	-4.74	181	-.80	.42
	1998	118	58.05	38.65	3.55				
PTOTM2	2004	65	44.52	22.71	2.81	-.93	181	-.25	.79
	1998	118	45.46	24.20	2.22				

\* df for unequal variance t-test.



Scanning the sign pattern in column "Mean Difference" of Table (16) we find that nine out of 13 differences are negative and four are positive. Two of the nine negatives are statistically significant ( $p \leq .03$  and  $p < .02$ ). The negative difference indicates declining achievement while positive difference indicates improvement. From the information displayed in Table (16) we conclude that UNRWA school students have shown no sign of improvement in the area of general competency in Mathematical Thinking and Communication during the period from 1998 to 2004. There is statistically significant decrease in student achievement on the general competency in Mathematical Communication; the decline is more pronounced in the specific competency in Communication of Numbers.

#### **II.1.2.2.2.3 Comparative Achievement of the MOD 4<sup>th</sup> Graders on the Math Test2 (General Competency in Mathematical Thinking and Communication) in 1998 and 2004**

In the Hashemite Kingdom of Jordan the Ministry of Defense has its own schools which cater for the education of children of the military personnel. They are typically located in the cantonment areas where the families of the military people may reside. They are mainly boarding schools where most students stay in. The students study the same national curricula and go through the same examination procedures taking the same tests. The disciplinary regulations and their character as boarding schools add to them a special feature. On Math Test 1 we saw MOD schools made better progress, relatively, than schools under other education authorities.

The comparative performance of the MOD 4<sup>th</sup> graders on Math Test 2 in 1998 and 2004 is discussed in the following section.

Table (17)  
Mean, SD, Std. Error of the Mean, Mean Difference, df, t-Value and 2-Tailed Sig. of  
the Mean Difference (2004-1998) on Math Test 2 (General Competency in Mathematical  
Thinking and Communication) of the MOD Grade 4 Students

Competency	Year	N	Mean	STd. Devi	STd. Error Mean	Mean Diff.	df	t	Sig 2-tailed
THINKP	2004	12	85.33	14.21	4.10	32.26	43.21*	4.77	.00
	1998	38	53.07	33.07	5.36				
COMTHINP	2004	12	50.02	16.43	4.74	21.57	48	3.21	.00
	1998	38	28.44	21.30	3.45				
COMP	2004	12	77.00	28.20	8.14	27.65	48	3.06	.00
	1998	38	49.34	26.94	4.37				
NUMBERP	2004	12	65.03	15.32	4.42	29.94	31.87*	4.92	.00
	1998	38	35.08	25.77	4.18				
FRACTIOP	2004	12	86.00	26.82	7.74	34.35	48	3.66	.00
	1998	38	51.97	28.70	4.65				
GOEMETRP	2004	12	53.75	13.49	3.89	19.22	29.67*	3.69	.00
	1998	38	34.52	21.31	3.45				
NUMTHINP	2004	12	83.83	24.61	7.10	29.88	34.14*	2.96	.00
	1998	38	53.94	44.08	7.15				
NUMCOTHP	2004	12	54.00	21.74	6.27	28.21	48	3.28	.00
	1998	38	25.78	27.07	4.39				
NUMCOMP	2004	12	73.83	45.01	12.99	34.35	13.96*	2.48	.03
	1998	38	39.47	28.84	4.67				
GEOTHINP	2004	12	81.77	21.71	6.26	35.28	30.50*	4.16	.00
	1998	38	46.49	35.15	5.69				
GECOTHP	2004	12	79.00	45.22	13.05	60.57	48	4.49	.00
	1998	38	18.42	39.28	6.37				
FRACOTHP	2004	12	80.00	45.22	13.05	61.57	48	4.56	.00
	1998	38	18.42	39.28	6.37				
FRACOMP	2004	12	80.16	33.28	9.64	20.95	48	1.76	.08
	1998	38	59.21	36.49	5.92				
PTOTM2	2004	12	64.41	14.3	4.15	26.77	48	4.11	.00

\* df for unequal variance t-test.

The "Mean Difference" column in Table (17) shows relatively outstanding gains of the MOD fourth grade students on each and every subcompetency and specific competency measured under the rubric of General Competency in Mathematical Thinking and Communication. The mean percent correct scores show an increase of percentage points ranging from 19.23%, the lowest on Geometric Reasoning (GEMOETRP) through 61.58%, and outstanding gain the highest, on Mathematical Thinking applied to Fractions (FRACOTHP).

The average percent correct score mean increase represented by the gain on total test score (PTOTM2) is 20.96% percentage points.

Needless to emphasize, all the Mean Differences are highly statistically significant, most of them at ( $p < .001$ ). On one specific competency, Communication of Fractions (FRACOMP) the difference was not statistically significant.

#### II.1.2.2.2.4 Comparative Achievement of the Private Grade 4 Students on the Math Test2 (General Competency in Mathematical Thinking and Communication) between 1998 and 2004

In general private schools used to be better equipped and better staffed than the MOE schools. Moreover, they are located in cities and cater for students coming from relatively higher socio-economic stratum of the society. It is well-recognized that private school students perform better than students studying in schools under the jurisdiction of other education authorities. This, however, does not necessarily imply higher rate of improvement in student test scores over time. The following Table (18) exhibits the 4<sup>th</sup> grade Private School students' achievement gains on Math Test2 (General Competency in Mathematical Thinking and Communication) over the years from 1998 to 2004.

Table (18)  
Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value and 2-Tailed Significance of the Mean Difference (2004-1998) on Math Test2 (General Competency in Mathematical Thinking and Communication) of the Fourth Grade Private School Students

Competency	Year	N	Mean	STd. Devi	STd. Error Mean	Mean Diff.	df	t	Sig 2-tailed
THINKP	2004	50	63.66	22.08	3.12	-5.28	183	-1.42	.56
	1998	135	73.95	22.54	1.94				
COMTHINP	2004	50	47.90	23.19	3.28	1.31	183	.35	.73
	1998	135	46.58	22.14	1.90				
COMP	2004	50	71.50	24.89	3.52	-4.98	183	-1.18	.24
	1998	135	76.48	25.69	2.21				
NUMBERP	2004	50	62.33	25.35	3.58	-.17	183	-.04	.97
	1998	135	62.51	24.89	3.14				
FRACTIOP	2004	50	67.50	25.29	3.57	-7.69	183	-2.03	.04
	1998	135	75.18	21.91	1.88				
GOEMETRP	2004	50	49.06	20.37	2.88	.71	183	.22	.83
	1998	135	48.35	19.80	1.70				
NUMTHINP	2004	50	69.50	33.33	4.71	-3.83	183	-.70	.48
	1998	135	73.33	32.78	2.82				
NUMCOTHP	2004	50	53.60	30.03	4.24	.34	183	.07	.94
	1998	135	53.25	29.08	2.50				
NUMCOMP	2004	50	77.00	32.32	4.57	2.18	183	.39	.70
	1998	135	74.81	34.48	2.96				
GEOTHINP	2004	50	62.00	30.72	4.34	-5.16	183	-1.04	.30
	1998	135	67.15	29.64	2.55				
GECOTHP	2004	50	44.00	49.48	6.99	-4.89	88.71	-.59	.55
	1998	135	48.88	50.17	4.31				
FRACOTHP	2004	50	45.00	49.48	6.99	-3.89	88.71	-.47	.64
	1998	135	48.88	50.17	4.31				
FRACOMP	2004	50	66.00	32.34	4.57	-12.15	183	-2.53	.01
	1998	135	78.14	27.72	2.38				
PTOTM2	2004	50	58.11	19.60	2.77	-.02	183	-.01	.99
	1998	135	58.13	19.20	1.65				

If we scan the last two columns, "Mean Difference" and "Sig 2-tailed" in Table 18 we clearly see that on nine of the thirteen sub-competencies and specific competencies the negative mean difference indicates decrease rather than increase in the student test scores from 1998 to 2004. Moreover, the decline in performance is statistically significant ( $p < .05$ ) in two cases, viz., Fractions (GRACTIOP) and Communication as applied to Fractions (FRACOMP). The negligible increase on the four sub-competencies and specific competencies is virtually nil.

#### **II.1.2.2.3 Urban / Rural School Differences in Gain Scores (2004-1998) on Math Test2 (General Competency in Mathematical Thinking and Communication)**

The phenomenon of urban schools outperforming the rural schools on tests of achievement is well recognized by Jordanian educators. We decided to study the change in Grade 4 students, achievement from 1998 to 2004 in urban schools and in rural schools separately.

##### **II.1.2.2.3.1 Urban Fourth Graders' Performance on Math Test 2 (General Competency in Mathematical Thinking and Communication) in 1998 and 2004**

The following Table (19) presents the results of t-tests using independent, unequal samples of urban school students from 1998 and 2004.

**Table (19)**  
**Mean, SD, Mean Difference, Std. Error of Mean Difference, t-Value, df and Two-tailed Significance of the Difference in Urban 4<sup>th</sup> Grade Students' Gain Scores on Mathematical Thinking and Communication Competency and its Sub-competencies from 1998 to 2004**

Competency	Year	N	Mean	STd. Devi	STd. Error Mean	Mean Diff.	df	t	Sig 2-tailed
THINKP	2004	342	65.30	25.92	1.40	2.08	1152	1.18	.24
	1998	812	63.21	28.13	.98				
COMTHINP	2004	342	39.93	22.17	1.19	.75	716.70*	.50	.61
	1998	812	39.19	27.97	.87				
COMP	2004	342	59.60	29.97	1.62	3.01	712.06*	1.50	.13
	1998	812	56.58	33.53	1.1				
NUMBERP	2004	342	50.72	25.61	1.38	2.53	711.21*	1.48	.14
	1998	812	48.19	28.62	1.00				
FRACTIOP	2004	342	62.13	29.39	1.58	3.51	1152	1.80	.07
	1998	812	58.62	30.58	1.07				
GOEMETRP	2004	342	44.53	20.50	1.10	1.15	703.96*	.84	.40
	1998	812	43.38	22.66	.79				
NUMTHINP	2004	342	63.80	36.40	1.96	3.52	1152	1.45	.15
	1998	812	60.28	38.33	1.34				
NUMCOTHIP	2004	342	42.47	28.25	1.52	2.01	690.42*	1.08	.28
	1998	812	40.46	30.60	1.07				
NUMCOMP	2004	342	58.26	37.73	2.04	2.84	674.88*	1.15	.25
	1998	812	55.41	39.90	1.40				
GECTHINP	2004	342	61.50	30.85	1.66	2.60	1152	1.26	.21
	1998	812	58.90	32.50	1.14				
GECOTHOP	2004	342	39.96	48.06	2.59	3.01	1152	.97	.33
	1998	812	36.94	48.29	1.69				
FRACOTHP	2004	342	40.96	48.06	2.59	4.01	1152	1.29	.20
	1998	812	36.94	48.29	1.69				
FRACOMP	2004	342	60.94	36.97	1.99	3.18	1152	1.31	.19
	1998	812	57.75	38.16	1.33				
PTOTM2	2004	342	50.79	20.79	1.12	3.06	716.99*	2.20	.09
	1998	812	47.72	23.42	.82				

\* df for unequal variance t-test.

To start with, it is clear from the Mean Difference column in Table (19) that there is consistent trend of improvement over all the sub-competencies and the general competency from 1998 to 2004 but gains are miniscule, mostly, statistically nonsignificant with a welcome exception of statistical significance ( $p < .04$ ) and ( $p < .03$ ) on the specific competency of Communication applied to Fractions and the total competency score, respectively.

#### II.1.2.2.3.2 Gain Scores of Rural School 4<sup>th</sup> Graders on Math Test 2 (General Competency in Mathematical Thinking and Communication) from 1998 to 2004

The results of the independent unequal sample size t-tests conducted on rural schools' 4<sup>th</sup> grade students' Math Test2 scores in 1998 and 2004 are reported in Table (20).

Table (20)  
Mean, SD, Mean Difference, Std. Error of Mean Difference, t-Value, df and Two-tailed Significance of the Difference in Rural 4<sup>th</sup> Grade Students' Gain Scores on Mathematical Thinking and Communication Competency and its Sub-competencies from 1998-2004

Competency	Year	N	Mean	Std. Devi	Std. Error Mean	Mean Diff.	df	t	Sig 2-tailed
THINKP	2004	162	56.93	27.54	2.16	-3.28	783	-1.29	.20
	1998	623	60.21	29.05	1.16				
COMTHINP	2004	162	35.01	22.69	1.78	-1.86	270.95*	-.91	.36
	1998	623	36.87	24.92	.99				
COMP	2004	162	50.45	34.43	2.70	-2.95	234.86*	-.99	.32
	1998	623	53.41	31.30	1.25				
NUMBERP	2004	162	41.54	28.82	2.26	-2.02	783	-.82	.41
	1998	623	43.57	27.70	1.10				
FRACTIOP	2004	162	55.00	31.02	2.43	-2.18	783	-.82	.41
	1998	623	57.18	29.83	1.19				
GOEMETRP	2004	162	40.54	20.73	1.62	-1.08	783	-.55	.58
	1998	623	41.62	22.83	.91				
NUMTHINP	2004	162	53.27	39.10	3.07	-2.18	783	-.64	.52
	1998	623	55.45	38.69	1.55				
NUMCOTHP	2004	162	34.49	29.42	2.31	-1.72	783	-.65	.52
	1998	623	36.21	30.27	1.21				
NUMCOMP	2004	162	47.44	40.57	3.18	-2.63	240.47*	-.74	.46
	1998	623	50.08	38.19	1.53				
GEOTHINP	2004	162	52.76	30.21	2.37	-4.16	783	-1.44	.15
	1998	623	56.92	33.30	1.33				
GECOTHP	2004	162	29.30	43.61	3.42	-6.32	271.09*	-1.61	.11
	1998	623	35.63	47.93	1.92				
FRACOTHP	2004	162	30.30	43.61	3.42	-5.32	271.08*	-1.35	.18
	1998	623	35.63	47.93	1.92				
FRACOMP	2004	162	53.46	40.10	3.15	-3.27	783	-.98	.33
	1998	623	56.74	37.43	1.49				
PTOTM2	2004	162	44.32	22.60	1.77	-.62	783	-.31	.76
	1998	623	44.94	22.94	.91				

\* df for unequal variance t-test.

In a sharp contrast with gain score trend in the urban school students' achievement on Math Test2 (General Competency in Mathematical Thinking and Communications), all the differences without exception in the rural 4<sup>th</sup> graders' achievement scores are negative. Differences were computed in this study subtracting 1998 mean scores from 2004 mean scores (20004 mean score-1998 mean score).

This, unfortunately, gives an indication of general decline in rural school 4<sup>th</sup> grade students' achievement in this area of mathematics. The decrease in mean scores, consistent over all the subtests, ranges from -1.08 through -6.33.

#### II.1.2.2.4 Differential Achievement Gains of Male and Female 4<sup>th</sup> Graders

Slightly better performance of female students over their male cohorts on achievement tests in school subjects is known to educational researchers in Jordan.

The subject of comparative achievement gains over the past six years is studied separately for Male and Female students in the following sections.

#### II.1.2.2.4.1 Male 4<sup>th</sup> Graders Progress in Math Test2 Achievement from 1998 to 2004

The following Table (21) presents the results of independent unequal sample t-tests on mean percent correct test scores of students in 1998 and 2004.

**Table (21)**  
Mean, SD, Mean Difference, Degrees of Freedom, t-Value, Std Error of the Difference of Male Students Achievement in Math Test 2 (General Competency in Mathematical Thinking Communication) from 1998 to 2004

Competency	Year	N	Mean	STd. Devi	STd. Error Mean	Mean Diff.	df	t	Sig 2-tailed
THINKP	2004	262	59.38	27.85	1.72	-2.51	993	-1.20	.23
	1998	733	61.89	29.50	1.09				
COMTHINP	2004	262	35.85	22.17	1.36	-1.45	521*	-.87	.38
	1998	733	37.30	25.33	.94				
COMP	2004	262	53.62	32.82	2.03	-.91	993	-.38	.70
	1998	733	54.53	33.61	1.24				
NUMBERP	2004	262	44.45	26.49	1.64	-1.41	993	-.70	.48
	1998	733	45.86	28.39	1.05				
FRACTIOP	2004	262	56.44	29.81	1.84	-.25	993	-.11	.91
	1998	733	56.68	31.22	1.15				
GOEMETRP	2004	262	41.14	21.23	1.31	-.79	993	-.49	.63
	1998	733	41.93	22.95	.84				
NUMTHINP	2004	262	56.41	38.32	2.36	-2.86	993	-1.01	.30
	1998	733	59.27	39.32	1.45				
NUMCOTHP	2004	262	37.16	27.54	1.70	-.44	504*	-.21	.82
	1998	733	37.61	30.41	1.12				
NUMCOMP	2004	262	50.70	39.56	2.44	-2.35	993	-.82	.40
	1998	733	53.06	39.73	1.46				
GEOTHINP	2004	262	55.52	31.62	1.95	-2.31	993	.99	.32
	1998	733	57.84	32.65	1.20				
GECOTHP	2004	262	28.80	43.27	2.67	-6.66	505*	-2.07	.03
	1998	733	35.47	47.87	1.76				
FRACOTHP	2004	262	29.80	43.27	2.67	-5.66	505*	-1.76	.07
	1998	733	35.47	47.87	1.76				
FRACOMP	2004	262	56.53	39.12	2.4	.53	993	.19	.84
	1998	733	56.00	38.65	1.42				
PTOTM2	2004	262	45.93	21.50	1.32	.067	504*	.04	.96
	1998	733	45.86	23.74	.877				

\* df for unequal variance t-test.

In the Mean Difference and Sig (2-tailed) columns of Table (21) we find that almost all the differences are negative and statistically nonsignificant. Only the difference on one specific competency (Application of Thinking and Communication to Geometry and Measurement content) is statistically significant ( $p < .04$ ).

#### II.1.2.2.4.2 Female 4<sup>th</sup> Graders Progress in Math Test2 Achievement from 1998 to 2004

The results of t-test analyses on mean test scores of the 4<sup>th</sup> grade female students in 1998 and 2004 are presented in Table (22).

Table (22)  
Mean, SD, Mean Difference, Degrees of Freedom, t-Value, Std. Error of the Mean Difference and 2-Tailed Significance Comparing Female 4<sup>th</sup> Graders Performance in (General Competency in Mathematical Thinking and Communication) in 1998 and 2004

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	Mean Diff.	df	t	Sig 2-tailed
THINKP	2004	242	66.11	25.02	1.60	4.17	457.38*	2.18	.03
	1998	702	61.94	27.57	1.04				
COMTHINP	2004	242	41.07	22.46	1.44	1.95	454.53*	1.14	.25
	1998	702	39.11	24.58	.92				
COMP	2004	242	59.95	30.23	1.94	4.04	942	1.73	.08
	1998	702	55.91	31.55	1.19				
NUMBERP	2004	242	51.37	27.14	1.74	4.84	942	2.32	.02
	1998	702	46.52	28.24	1.06				
FRACTIOP	2004	242	63.53	29.99	1.92	4.17	942	1.90	.05
	1998	702	59.36	29.17	1.10				
GOEMETRP	2004	242	45.53	19.77	1.27	2.20	472.32*	1.44	.15
	1998	702	43.33	22.52	.84				
NUMTHINP	2004	242	64.75	36.34	2.33	7.70	942	2.76	.00
	1998	702	57.05	37.72	1.42				
NUMCOTHP	2004	242	42.88	29.96	1.92	3.21	942	1.41	.15
	1998	702	39.67	30.62	1.15				
NUMCOMP	2004	242	59.19	37.87	2.43	6.06	942	2.11	.03
	1998	702	53.13	38.75	1.46				
GEOTHINP	2004	242	52.12	29.76	1.91	3.86	461.23*	1.69	.09
	1998	702	58.26	33.07	1.24				
GECOTHP	2004	242	44.90	49.26	3.16	7.58	942	2.09	.03
	1998	702	37.32	48.40	1.82				
FRACOTHP	2004	242	45.90	49.26	3.16	8.58	942	2.36	.01
	1998	702	37.32	48.40	1.82				
FRACOMP	2004	242	60.71	36.98	2.37	2.02	942	.73	.46
	1998	702	58.68	36.94	1.39				
PTOTM2	2004	242	51.72	21.30	1.36	4.52	942	2.71	.07
	1998	702	47.20	22.72	.85				

\* df for unequal variance t-test.

The column, "Mean Difference" in Table (22) shows all the differences invariably positive signifying 4<sup>th</sup> grade female students' increased achievement on Math Test2 (General Competency in Mathematical Thinking and Communication) from 1998 to 2004. On six sub-competencies and the total test score the gains are statistically significant ( $p$  values ranging from .005 through .000). The gain score (Mean Difference ranges from 2.0 the lowest through 8.6, the highest. To say the least, female students' results are somewhat encouraging. They show an unbroken positive trend and statistically significant progress.



### II.1.2.3 MATH TEST 3: PROBLEM SOLVING

This test comprising 20 items covered the General Competency in Math Problem Solving in two content areas, Numbers and, Measurement and Geometry. The problems included one step problems involving one operation and two or more steps problems involving two or more operations.

The following sections of this part of the report will describe the change in the 4<sup>th</sup> grade students' performance on Math Test 3 which measured the general competency in Mathematical Problem Solving, from 1998 to 2004. First we compare the competency levels (classification of students in three levels of competency) reached by students in 2004 vis-à-vis in 1998. Then, we study improvement in student Math achievement (Problem Solving Competency) in each Education Authority followed by the rates of progress in Urban and Rural Schools, and in Male and Female sub-samples of students, from 1998 through 2004.

#### II.1.2.3.1 Comparative Competency Levels of Grade 4 Students on Math Test3 (Problem Solving)

As stated earlier, on the basis of their total test score on the general competency, students were grouped into three successive levels of competency. Students scoring 30% or less were put in Level1; those scoring more than 30% but less than 70% were put in Level2; and those scoring 70% or more were put in Level 3 on the three band competency level scale.

The following Figure (3) exhibits the comparative competency attained by percentages of 4<sup>th</sup> graders on the general competency in Math Problem Solving in 1998 and 2004.

Figure (3)  
Comparative National Competency Levels of Grade 4 Students in Mathematical  
Problem Solving in 1998 and 2004

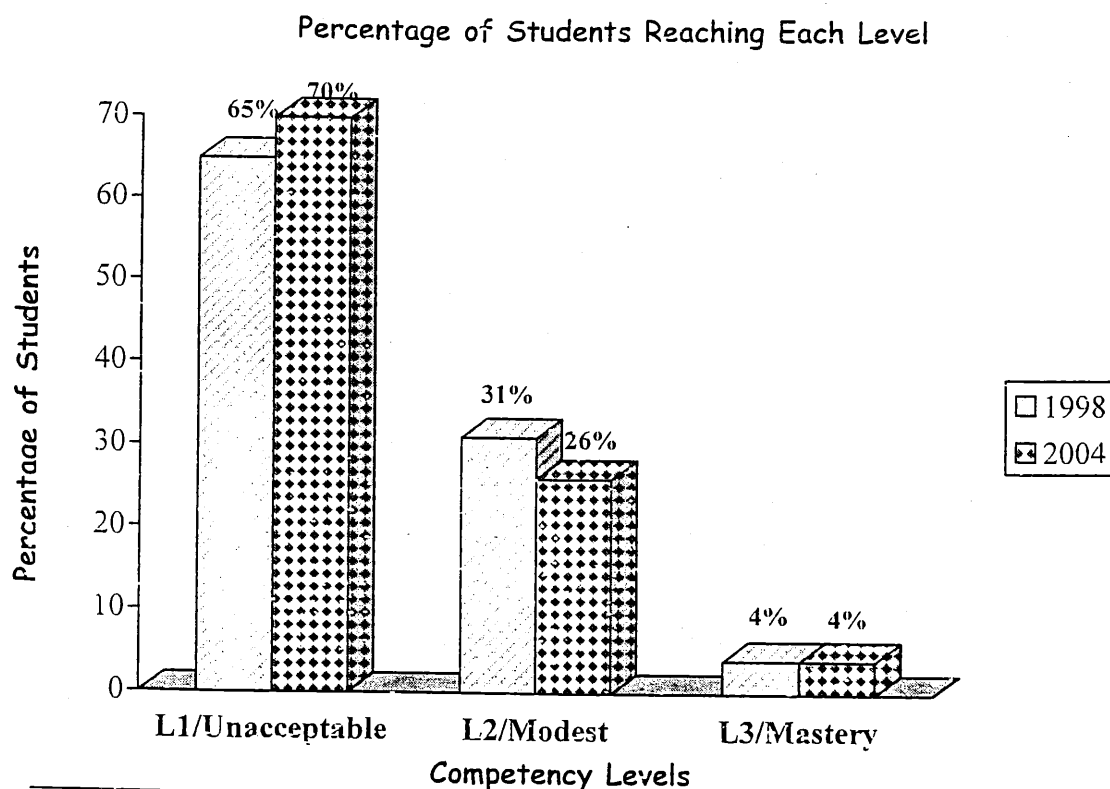


Figure (3) depicting different competency levels certainly paints a picture that does not look rosy in terms of improvement in student performance in the area of Mathematical Problem solving from 1998 to 2004.

At the Level 1, the unacceptable standard of competency, the number of students has increased by 5% from 65% in 1998 to 70% in 2004. In the same vein at the Level 2, a rather modest level of attainment, the number has fallen by 5% from 31% in 1998 to 26%.

On the Level 3, the mastery level of competency, there is no change recorded in percentages of 4<sup>th</sup> graders on the general competency in Math Problem Solving in 1998 and 2004.

Only 4% of the grade 4 students in the whole country reached the Mastery status, Level 3, on the General Competency in Math Problem Solving, in both 1998 and 2004. In all, increase at the lowest level, decrease at the middle level and no change at the top level does not bode well for the progress in Math Problem Solving.

Having described the Competency Level scene, we examine the achievement of 4<sup>th</sup> grade students on each subscale measuring a sub-

competency in problem solving in 1998 as well as in 2004 intending to monitor the change in scores from 1998 to 2004.

#### II.1.2.3.2 National Progress in Math Problem Solving Competency of the 4<sup>th</sup> Grade Students from 1998 to 2004

Percent correct scores of the fourth graders obtained from national samples in 1998 and 2004 were analysed using independent, unequal sample t-test procedures in order to study the change in student performance on the test over the years from 1998 to 2004.

The statistics produced by the analysis are presented in the following Table (23).

Table (23)  
Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. Of the Difference (2004-1998) on Math Test 3  
(General Competency in Math Problem Solving) at the National Level

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
TOTM3P	2004	417	30.08	16.81	.82	.06	770.36*	.95	.06	.97
	1998	1396	30.02	19.25	.52					
NUMBERP	2004	417	35.32	18.37	.90	.50	782.49*	.61	.54	1.07
	1998	1396	34.79	21.38	.57					
GEOMETRP	2004	417	17.84	19.91	.98	-.95	1811	.34	-1.06	1.12
	1998	1396	18.90	20.08	.54					
GEOSTE1P	2004	417	18.10	28.62	1.40	-2.02	746.25*	.04	-3.31	1.64
	1998	1396	21.41	31.67	.85					
GEOSTE2P	2004	417	17.71	20.95	1.03	.06	1811	.95	.07	1.15
	1998	1396	17.64	20.54	.55					
NUMSTE1P	2004	417	48.42	23.47	1.15	1.59	766.36*	.11	2.12	1.35
	1998	1396	46.26	26.73	.2					
NUMSTE2P	2004	417	25.50	19.48	.95	-.60	749.09*	.55	-.67	1.12
	1998	1396	26.18	21.64	.58					
STEP1P	2004	417	40.84	22.26	1.09	.61	762.38*	.54	.78	1.28
	1998	1396	40.05	25.20	.67					
STEP2P	2004	417	21.37	16.31	.80	-.53	743.04*	.59	-.50	.93
	1998	1396	21.86	17.96	.48					

\* df for unequal variance t-test.

In essence, Table (23) indicates no progress in the area of Mathematical Problem Solving at the 4<sup>th</sup> grade level. In the Mean Difference column in Table (23), four differences are positive and four negative, negative

differences are relatively larger than the positive ones, and one of the negative differences is statistically significant ( $p < .05$ ).

### II.1.2.3.3 Differential Change in Performance of Male and Female 4<sup>th</sup> Graders on Math Test 3 (Mathematical Problem Solving) from 1998 to 2004

The following sections present the progress made by sub-samples of male and female G4 students from 1998 to 2004.

#### II.1.2.3.3.1 Male 4<sup>th</sup> Graders Progress in Math Problem Solving Competency from 1998 to 2004

The following Table (24), presents the results of independent, unequal sample t-test on Grade 4 male students' Math Test 3 scores in 1998 and 2004.

Table (24)  
Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) for Male Fourth Graders in 1998 and 2004

Competency	Year	N	Mean	Std. Dev.	Std. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
TOTM3P	2004	217	31.16	16.23	1.10	-.21	456.15*	.83	-.29	1.35
	1998	703	31.45	20.87	.78					
NUMBERP	2004	217	36.09	18.16	1.23	-.08	446.43*	.93	-.12	1.50
	1998	703	36.21	22.87	.86					
GEOMETRP	2004	217	19.66	19.55	1.32	-.41	918	.68	-.68	1.65
	1998	703	20.34	21.77	.82					
GEOSTE1P	2004	217	20.96	29.03	1.97	-1.16	406.32*	.25	-2.72	2.34
	1998	703	23.68	33.29	1.25					
GEOSTE2P	2004	217	19.00	21.11	1.43	.20	918	.84	.34	1.67
	1998	703	18.67	21.69	.81					
NUMSTE1P	2004	217	49.03	23.25	1.57	.96	431.73*	.34	1.82	1.91
	1998	703	47.21	28.35	1.06					
NUMSTE2P	2004	217	26.38	19.45	1.32	-1.00	414.61*	.31	-1.59	1.57
	1998	703	27.96	22.77	.85					
STEP1P	2004	217	42.02	22.01	1.49	.38	434.3*	.70	.69	1.81
	1998	703	41.33	26.99	1.01					
STEP2P	2004	217	22.54	16.01	1.08	-.720	424.43*	.47	-.94	1.31
	1998	703	23.48	19.19	.724					

\* df for unequal variance t-test.

### II.1.2.3.3.2 Female 4<sup>th</sup> Graders' Progress in Math Problem Solving Competency From 1998 to 2004

The test data from the administrations in 1998 and 2004 were analysed using independent, unequal sample t-test procedures to study changes in students' performance on the test over the past five years. The results of the analysis are presented in the following Table (25).

**Table (25)**  
Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) for Female Fourth Graders in 1998 and 2004

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df.	Sig 2-tailed	Mean Diff.	Std. Error Diff.
TOTM3P	2004	200	28.91	17.38	1.22	.24	891	.81	.34	1.39
	1998	693	28.57	17.34	.65					
NUMBERP	2004	200	34.50	18.60	1.31	.74	891	.46	1.16	1.56
	1998	693	33.34	19.65	.74					
GEOMETRP	2004	200	15.87	20.14	1.42	-.99	297.93*	.32	-1.57	1.58
	1998	693	17.44	18.10	.68					
GEOSTE1P	2004	200	15.00	27.90	1.97	-1.81	340.80*	.07	-4.12	2.27
	1998	693	19.11	29.79	1.13					
GEOSTE2P	2004	200	16.31	20.75	1.46	-.19	891	.85	-.30	1.57
	1998	693	16.61	19.27	.73					
NUMSTE1P	2004	200	47.75	23.75	1.67	1.23	891	.22	2.44	1.98
	1998	693	45.31	24.96	.94					
NUMSTE2P	2004	200	24.56	19.50	1.37	.12	891	.90	.19	1.61
	1998	693	24.36	20.28	.77					
STEP1P	2004	200	39.56	22.51	1.59	.43	891	.67	.80	1.85
	1998	693	38.76	23.20	.88					
STEP2P	2004	200	20.10	16.57	1.17	-.10	891	.92	-.13	1.32
	1998	693	20.23	16.47	.62					

\* df for unequal variance t-test.

The "Mean Difference" column in Table (25) shows that five of the eight sub-competencies mean score differences are negative and four are positive. The difference on the total test score is also positive. Also, from the Sig (2-tailed) column in this table it is clear that none of the differences is statistically significance. These results indicate, in plain words, total lack of progress in this area of mathematics during the past five or six years among the 4<sup>th</sup> grade female students, while the male students performance is no better in any sense of the word.

#### II.1.2.3.4 Urban and Rural Schools' 4<sup>th</sup> Graders' Progress in Math Problem Solving Competency from 1998 to 2004

Following the practice of describing the progress of each segment of grade 4 students' population separately, in the following sections of this report we discuss the achievement gains made by Urban School students and Rural School students under separate headings.

##### II.1.2.3.4.1 Urban School 4<sup>th</sup> Graders' Progress in Math Problem Solving Competency From 1998 to 2004

Math Test 3 (Problem Solving in Math) was administered to national samples of Grade 4 students in 1998 and then in 2004. After the 2004 administration we have analysed the scores from the two samples to study the progress from 1998 to 2004, separately for different subpopulations of students. The following Table (26) gives the results of the tests of the Urban School students.

Table (26)  
Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) between 1998 and 2004 Testings of the Urban School Grade 4 Students

Competency	Year	N	Mean	Std. Dev.	Std. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
TOTM3P	2004	292	30.84	16.49	.96	.43	605*	.66	.51	1.19
	1998	794	30.33	19.41	.68					
NUMBERP	2004	292	36.00	17.81	1.04	.65	626*	.52	.84	1.29
	1998	794	35.16	21.68	.76					
GEOMETRP	2004	292	18.80	20.02	1.17	-.18	1084	.85	-.25	1.37
	1998	794	19.05	19.94	.71					
GEOSTE1P	2004	292	19.00	28.53	1.67	-1.63	578*	.11	-3.28	2.02
	1998	794	22.29	32.06	1.13					
GEOSTE2P	2004	292	18.70	21.35	1.24	.88	1084	.38	1.26	1.42
	1998	794	17.44	20.70	.73					
NUMSTE1P	2004	292	49.20	23.55	1.37	1.70	589*	.09	2.85	1.68
	1998	794	46.34	27.00	.95					
NUMSTE2P	2004	292	26.11	18.84	1.10	-.49	596*	.62	-.67	1.35
	1998	794	26.77	21.87	.77					
STEP1P	2004	292	41.65	22.24	1.30	.83	588*	.41	1.32	1.58
	1998	794	40.33	25.43	.91					
STEP2P	2004	292	22.04	16.16	.94	-.13	578*	.89	-.15	1.14
	1998	794	22.19	18.15	.64					

\* df for unequal variance t-test.

Of the eight subtests, measuring specific sub-competencies in Problem Solving, five show positive differences while three show negative differences. Not a single one of them is statistically significant at the significance level ( $p \leq .05$ ). The magnitude of the mean differences is rather miniscule. In practical terms there is no improvement at all in Grade 4 students' achievement in the general competency area of Math Problem Solving during the period from 1998 through 2004.

#### II.1.2.3.4.2 Rural School Fourth Grade Students' Progress in Math Problem Solving Competency from 1998 to 2004

The results of the independent unequal sample t-test analyses testing the difference between Rural Schools' 4<sup>th</sup> grade students' performance on Math Test 3 in 1998 and 2004 are presented in Table (27).

The "Mean Difference" column in Table (27) has all the signs negative, indicating decrease rather than increase in the test scores. There is only one positive sign showing only .43% percentage point gain in the mean score. All the differences are statistically nonsignificant. The pattern of decrease in all the sub scores even if it is not statistically significant should raise an alarm bell.

Table (27)  
Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) between 1998 and 2004 Testings of the Rural School Grade 4 Students

Competency	Year	N	Mean	Std. Devi	Std. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
TOTM3P	2004	125	28.30	17.46	1.56	-.71	725	.47	-1.31	1.84
	1998	602	29.61	19.03	.77					
NUMBERP	2004	125	33.74	19.60	1.73	-.27	725	.79	-.55	2.04
	1998	602	34.29	20.98	.85					
GEOMETRP	2004	125	15.60	19.54	1.74	-1.56	725	.12	-3.10	1.98
	1998	602	18.70	20.28	.82					
GEOSTE1P	2004	125	16.00	28.82	2.57	-1.48	189.10*	.14	-4.26	2.87
	1998	602	20.26	31.13	1.26					
GEOSTE2P	2004	125	15.40	19.89	1.77	-1.26	725	.21	-2.52	1.99
	1998	602	17.91	20.34	.82					
NUMSTE1P	2004	125	46.60	23.28	2.08	.17	725	.86	.43	2.54
	1998	602	46.15	26.39	1.07					
NUMSTE2P	2004	125	24.10	20.89	1.86	-.62	725	.54	-1.29	2.09
	1998	602	25.39	21.32	.86					
STEP1P	2004	125	38.95	22.26	1.99	-.31	725	.76	-.74	2.41
	1998	602	39.69	24.92	1.01					
STEP2P	2004	125	19.80	16.60	1.48	-.95	725	.34	-1.63	1.72
	1998	602	21.43	17.71	.72					

\* df for unequal variance t-test.

### II.1.2.3.5 Grade 4 Students Achievement on Math Test 3 (General Competency in Math Problem Solving) in Each Education Authority

Four education authorities namely, MOE, UNRWA, Private and MOD govern, in a centrally guided education system, the schools in their jurisdictions. In general, while the curricula and examination system are uniform all over the country, schools under different education authorities diverge in respect of the provisions of educational facilities, the quality of management, instructional quality, teacher qualifications and incentives, etc.

In view of this, the following sections examine the change in Math Test3 (Problem Solving) achievement of the 4<sup>th</sup> graders from 1998 to 2004 within each Education Authority.

#### II.1.2.3.5.1 MOE 4<sup>th</sup> Graders' Performance on Math Test3 (Problem Solving) in 1998 and 2004

The results of independent, unequal sample t-tests are presented in the following Table (28).

Table (28)  
Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) Scores of the MOE Grade 4 Students in 1998 and 2004

Competency	Year	N	Mean	Std. Dev.	Std. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
TOTM3P	2004	313	28.57	15.65	.88	-1.26	605.15*	.21	-1.32	1.06
	1998	1113	29.89	19.31	.57					
NUMBERP	2004	313	33.87	17.42	.98	-.54	598.98*	.59	-.63	1.17
	1998	1113	34.51	21.27	.63					
GEOMETRP	2004	313	16.18	18.31	1.03	-2.45	549.93*	.01	-2.95	1.20
	1998	1113	19.13	20.44	.61					
GEOSTE1P	2004	313	15.97	27.15	1.53	-2.60	568.94*	.01	-4.69	1.80
	1998	1113	20.66	31.44	.94					
GEOSTE2P	2004	313	16.29	19.71	1.11	-1.57	528.69*	.12	-2.08	1.28
	1998	1113	18.37	21.04	.63					
NUMSTE1P	2004	313	46.88	22.85	1.29	.78	567.30*	.44	1.17	1.51
	1998	1113	45.70	26.38	.79					
NUMSTE2P	2004	313	24.12	18.94	1.07	-1.58	568.04*	.11	-1.99	1.25
	1998	1113	26.11	21.89	.65					
STEP1P	2004	313	39.15	21.45	1.21	-.20	569.79*	.84	-.29	1.42
	1998	1113	39.44	24.88	.74					
STEP2P	2004	313	19.80	15.40	.87	-2.24	580.75*	.03	-2.30	1.03
	1998	1113	22.10	18.22	.541					

\* df for unequal variance t-test.

Studying the Mean Difference" and "Sig. 2-tailed" columns in Table (28) reveals that: (i) on all the subtests and also on the total test score (TOTM3P)



the differences are negative indicating the decline in stead of increase in the test scores with one exception of One Step Number Problems (NUMSTE1P) where the increase is insubstantial for all practical purposes as well as statistically nonsignificant at the ( $\alpha \leq .05$ ) level; (ii) decrease on two subscales, Geometry and Measurement Problems (GEOMETRP) and Geometry One Step Problems (GEOSTE1P) is statistically significant ( $p < 0.02$ ) and ( $p < 0.01$ ) respectively; (iii) mean differences range from -4.69 through 1.18% (the only positive difference among all the negative ones).

#### II.1.2.3.5.2 UNRWA 4<sup>th</sup> Graders Performance on Math Test3 (Problem Solving) in 1998 and 2004

The results of the statistical analyses (independent, unequal sample t-tests) conducted on the 4<sup>th</sup> grade UNRWA school students' Math Test3 scores obtained in 1998 and 2004 are given in Table (29).

Table (29)  
Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3 (Problem Solving) Scores of the UNRWA Grade 4 Students in 1998 and 2004

Competency	Year	N	Mean	Std. Dev.	Std. Error Mean	t	df	Sig 2-tailed	Mean Diff.
TOTM3P	2004	40	33.93	16.20	2.56	1.99	153	.05	6.76
	1998	115	27.17	19.24	1.79				
NUMBERP	2004	40	41.51	19.06	3.01	2.41	153	.02	9.41
	1998	115	32.11	21.95	2.04				
GEOMETRP	2004	40	16.25	17.99	2.84	1.8	153	.86	.60
	1998	115	15.65	18.30	1.70				
GEOSTE1P	2004	40	21.25	27.47	4.34	-.09	153	.94	-.49
	1998	115	21.73	31.85	2.97				
GEOSTE2P	2004	40	13.75	17.63	2.78	.53	153	.73	1.14
	1998	115	12.60	17.63	1.64				
NUMSTE1P	2004	40	55.62	25.83	4.08	2.59	153	.01	12.73
	1998	115	42.89	27.13	2.53				
NUMSTE2P	2004	40	30.93	16.50	2.60	1.85	153	.07	6.92
	1998	115	24.02	21.46	2.00				
STEP1P	2004	40	47.03	24.26	3.83	2.01	153	.05	9.42
	1998	115	37.60	25.88	2.41				
STEP2P	2004	40	23.95	12.99	2.05	1.73	153	.09	5.19
	1998	115	18.76	17.37	1.62				

Right after examining the state of progress in MOE schools, where almost all the differences were negative showing a declining trend, looking at a positive trend in the "Mean Difference" column in Table (29)

is somewhat refreshing. Regardless of the fact whether the improvement is statistically significant or not and the question of effect size where it is statistically significant, it is heartening to see the general trend going in the right direction in UNRWA schools.

To start with, the "Mean Difference" column in Table (29) shows all the signs plus with just one exception (One Step Geometric Problems (GEGSTE1P)) where the difference is less than half a percentage point. Looking at the increments we find the maximum gain 12.73% points to the sub-competency in Solving One-Step Number Problems (NUMSTE1P) statistically significant ( $p < .010$ ). The gain percentage mean scores range from 0.60% on Geometry and Measurement Problems (GEOMETRP), the lowest through, 12.73% on (NUMSTE1P), the highest.

#### **II.1.2.3.5.3 Private School 4<sup>th</sup> Graders' Performance on Math Test3 (Problem Solving) in 1998 and 2004**

Private schools, in general provide expensive education, relatively. Students generally come from upper socioeconomic stratum of society. Schools are better equipped with educational facilities and instructional materials, and teachers are better qualified and more proficient. On the tests of achievement private school students, on the average, score better than students from other education authorities. The following Table (30) displays the results produced by independent, unequal sample t-tests conducted on the Private schools' Grade 4 students' Math Test 3 scores obtained in 1998 and 2004.

**Table (30)**  
**Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of**  
**the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test3**  
**(Problem Solving) Scores of the Private Grade 4 Students in 1998 and 2004**

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
TOTM3P	2004	56	35.49	21.24	2.83	.17	187	.86	.53	3.11
	1998	133	34.96	18.79	1.62					
NUMBERP	2004	56	38.32	21.19	2.83	-.86	187	.39	-2.94	3.43
	1998	133	41.27	21.71	1.88					
GEOMFTRP	2004	56	28.86	26.20	3.50	2.23	80.34*	.03	8.63	3.87
	1998	133	20.23	18.97	1.64					
GEOSTE1P	2004	56	28.57	35.49	4.74	-.14	187	.89	-.75	5.40
	1998	133	29.32	33.21	2.88					
GEOSTE2P	2004	56	29.01	26.44	3.53	3.42	79.46*	.00	13.32	3.89
	1998	133	15.69	18.82	1.63					
NUMSTE1P	2004	56	51.93	24.72	3.30	-.89	187	.37	-3.89	4.35
	1998	133	55.82	28.33	2.45					
NUMSTE2P	2004	56	28.12	22.39	2.99	-.66	187	.51	-2.23	3.35
	1998	133	30.35	20.48	1.77					
STEP1P	2004	56	46.09	24.38	3.25	-.75	187	.45	-3.11	4.12
	1998	133	49.20	26.47	2.29					
STEP2P	2004	56	27.52	20.83	2.78	1.21	86.15*	.23	3.81	3.14
	1998	133	23.71	16.69	1.44					

\* df for unequal variance t-test.

In the Private Schools, contrary to expectations, progress, shown by the "Mean Difference" column in Table (30) is rather haphazard. On five of the eight subscales including Number Problems, One Step Problems in Geometry, One Step Problems in the Numbers, Two Step Problems in Numbers and One Step Problems in all areas, the performance shows a declining trend; none of the negative differences is statistically significant, though. On the positive side on three of the eight subtests Private school students show increased performance. On two of them (Two Step Geometry Problems, and Geometry and Measurement Problems) the gains are statistically significant ( $p < .12$ ) and ( $p < .000$ ) respectively. On the total test score (TOTM3P), there is small, statistically nonsignificant, increase. On the two step problems, in general, there is 4%-tage point increase which did not reach statistical significance.

#### II.1.2.3.5.4 MOD Schools' 4<sup>th</sup> Graders' Performance on Math Test 3 (Problem Solving) in 1998 and 2004

MOD Schools, governed by the Ministry of Defence, cater for the education of the children of military personnel. Typically, they are boarding schools where students live in boarding houses. MOD schools also differ, perhaps, in the provision of instructional facilities and a more disciplined environment.

The results of independent, unequal sample t-tests conducted on the MOD grade four students' scores on Math Test3 in 1998 and 2004 are presented in the following Table (31).

**Table (31)**  
**Mean, SD, Std. Err. of the Mean, Mean Difference, df, t-Value, 2-Tailed Significance of**  
**the Mean Difference and Std. Err. of the Difference (2004-1998) on Math Test 3**  
**(Problem Solving) Scores of the MOD Grade 4 Students in 1998 and 2004**

Competency	Year	N	Mean	Std. Dev	Std. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
TOTM3P	2004	8	32.18	20.89	7.38	1.17	41	.25	7.54	6.42
	1998	35	24.64	15.30	2.58					
NUMBERP	2004	8	40.17	23.74	8.39	1.77	41	.085	12.32	6.97
	1998	35	27.85	16.31	2.75					
GEOMETRP	2004	8	13.54	16.62	5.87	-.52	41	.60	-3.60	6.88
	1998	35	17.14	17.72	2.99					
GEOSTE1P	2004	8	12.50	23.14	8.18	-.16	41	.87	-1.78	10.88
	1998	35	14.28	28.62	4.83					
GEOSTE2P	2004	8	14.06	18.22	6.44	-.66	41	.51	-4.51	6.83
	1998	35	18.57	17.24	2.91					
NUMSTE1P	2004	8	47.91	21.24	7.51	1.04	41	.30	9.11	8.73
	1998	35	38.80	22.50	3.80					
NUMSTE2P	2004	8	34.37	26.51	9.37	1.51	8.13*	.17	14.73	9.74
	1998	35	19.64	15.54	2.62					
STEP1P	2004	8	39.06	20.25	7.16	.78	41	.44	6.38	8.16
	1998	35	32.67	20.95	3.54					
STEP2P	2004	8	26.56	19.91	7.04	1.50	41	.14	9.18	6.10
	1998	35	17.38	14.52	2.45					

\* df for unequal variance t-test.

The "Mean Difference" column in Table (31) shows positive change on the Total Test Score (TOTM3P) along with five of the eight sub-competencies including Number Problems (NUMBERP) in general, Number Step1 Problems (NUMSTE1P), Number Step2 Problems (NUMSTE2P), Step1 Problems in general (STEP1P) and Step2 Problems in general (STEP2P). Despite the improvement in student test performance in 2004, the gains are not statistically significant except on 2 Step Number Problems (NUMSTE2P) at ( $p<.05$ ) level of significance.

On three of the eight sub-competencies, namely, Geometrical Problems (GEOMETRP), 1 Step Geometry Problems (GEOSTE1P) and 2 Step Geometry Problems (GEOSTE2P), the performance shows a declining trend, the decrease, however, is not statistically significant on any of the subtests.

---

## **II.2 ARABIC LANGUAGE TEST**

The Arabic language test was designed to assess students' mastery in four basic competencies: Comprehension, Grammar, Writing and Dictionary Utilization. Each of these competencies comprised of one or more sub-competencies.

### **II.2.1 Psychometric Properties**

The Arabic language test consisted of two different forms. Form 1 contained 28 items assessing four competencies (viz. Comprehension, Grammar, Writing and Dictionary Use). The Alpha reliability coefficient for this form was .89 in 1998 and .88 in 2004. The discrimination coefficients ranged from .11 through .59 in 1998 and .06 to .64 in 2004. The item means taken as indicators of difficulty of items varied from .24 through 1.01 with a mean of .61 in 1998 and from .24 to 1.13 with a mean of .63 in 2004.

The second form consisting of 24 items assessed "Grammar" and "Writing" competencies. The Alpha reliability coefficient for this form was .92 in both 1998 and 2004. The discrimination coefficients ranged from .35 through .66 in 1998 and .35 to .73 in 2004. Item means taken as a measure of item difficulty ranged from .20 through 1.31 with a mean value of .67 in 1998 and from .25 to 1.37 with mean value of .70 in 2004.

### **II.2.2 Arabic Language Competency Levels of Grade 4 Students in 1998 and 2004**

Four general competencies: Comprehension, Grammar, Writing and Use of Dictionary were measured by the two forms of the Arabic Test in both 1998 and 2004. Three levels of competency were set such that students obtaining 30% or less will be placed on Level 1 which indicates a very low or unacceptable level of competency; those obtaining scores more than 30% but less than 70% will be placed on Level 2 which indicates a modest level of competency; and those who obtain 70% or higher will be placed at Level 3 indicating the mastery level.

The following sections of this report discuss the progress made by 4<sup>th</sup> grade students in Jordanian schools on the four general competencies in Arabic Language from 1998 to 2004.

### II.2.2.1 Arabic Language Comprehension Competency Levels of 4<sup>th</sup> Grade Students in 1998 and 2004

The percentage of students reaching each level of competency in 1998 and 2004 are shown in the following Figure (4).

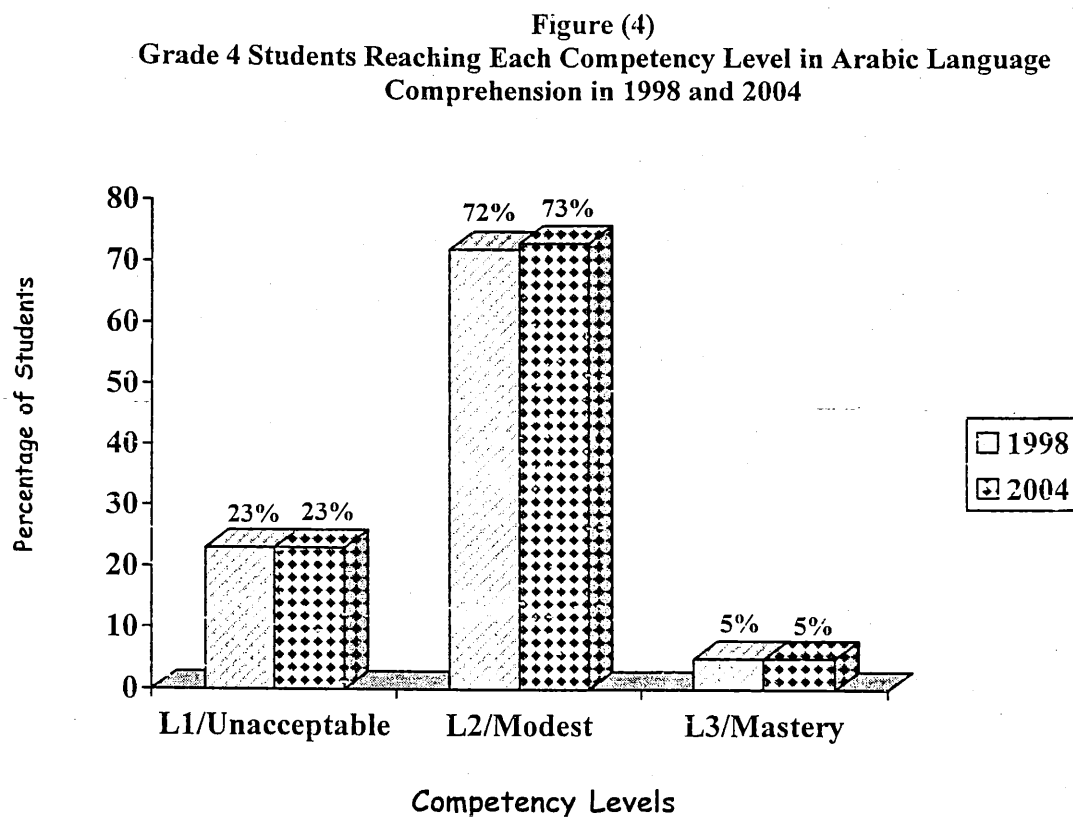
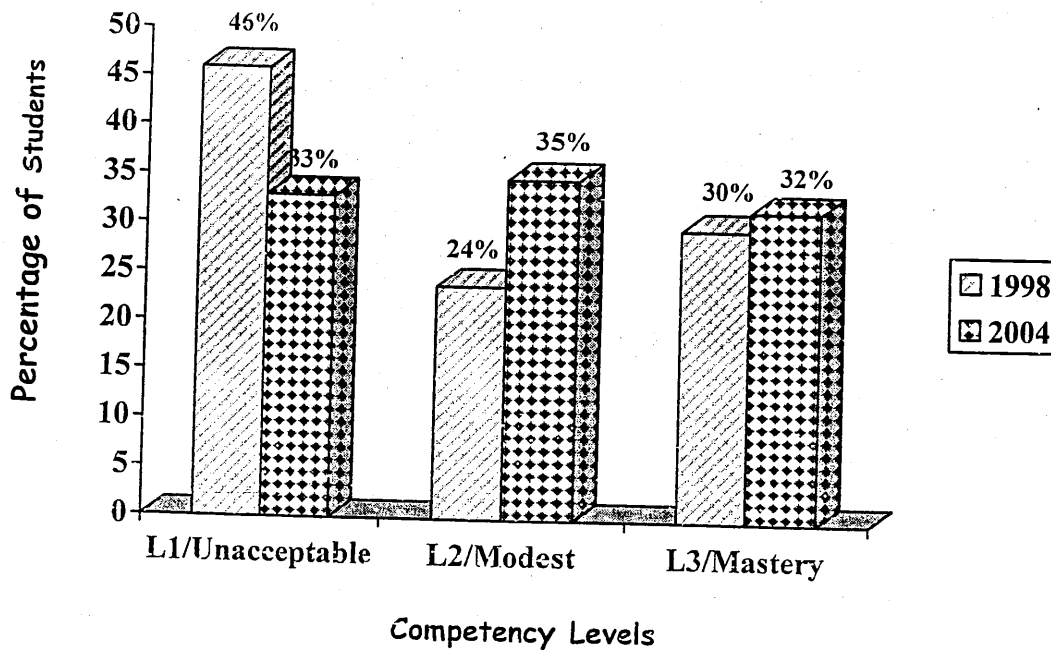


Figure (4) clearly demonstrates absolute lack of progress of the grade 4 students on their competency in Arabic Language Comprehension from 1998 to 2004, if we measure progress by increased number of students in successively higher levels of competency on the scale.

### II.2.2.2 Arabic Writing Competency Levels of Grade 4 Students in 1998 and 2004

Figure (5) displays the percentage of students attaining each level of competency in 1998 as well as in 2004.

**Figure (5)**  
**Grade 4 Students Reaching Each Competency Level in**  
**Arabic Writing in 1998 and 2004**

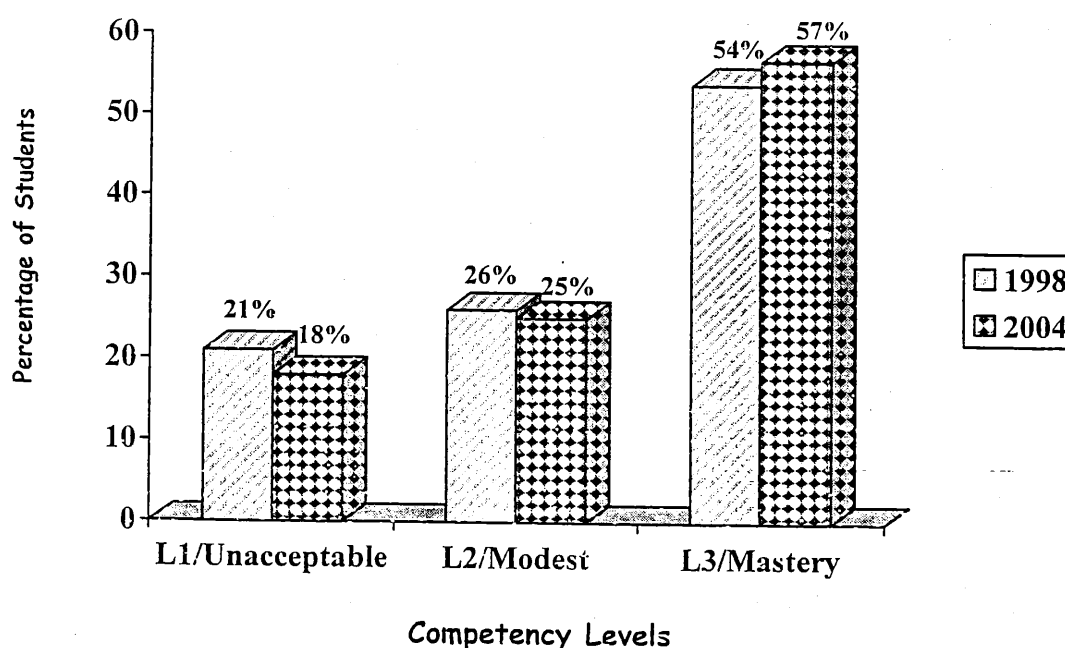


From Figure (5) we can see, there is some improvement in 4<sup>th</sup> grade students writing skills. The percentage of unacceptable level students has decreased by 13 percentage points from 46% in 1998 to 33% in 2004, while the percentage in the Level 2 (Modest) category has increased by 11 percentage points from 24% in 1998 to 35% in 2004, at the same time, the percentage in the mastery category (L3) has also shown a 2% increase from 30% in 1998 to 32% in 2004.

#### **II.2.2.3 Arabic Grammar Competency Levels of Grade 4 Students in 1998 and 2004**

The percentages of students placed at each level of Competency in Arabic Grammar in 1998 and 2004 are exhibited in Figure (6).

**Figure (6)**  
**Percentage of 4<sup>th</sup> Grade Students at Each Competency Level of Arabic Grammar in 1998 and 2004**



In the 4<sup>th</sup> grade students' competency in Arabic Grammar there is little improvement; 3% increase in the mastery level (L3) band and 3% decrease in the (L1), the unacceptable level.

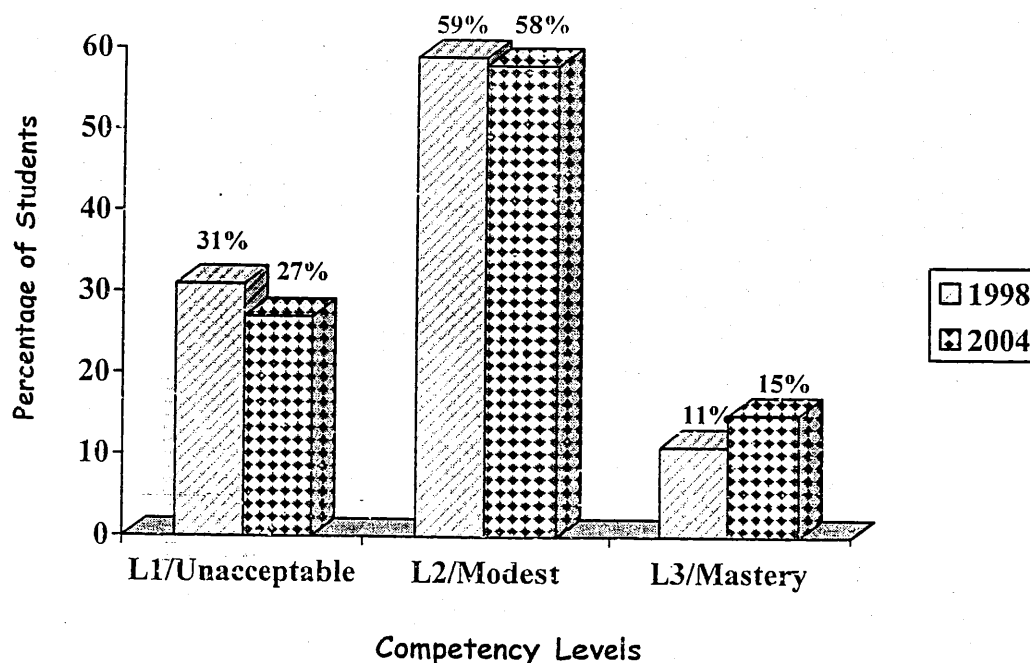
#### **II.2.2.4 Grade 4 Students' Competency Levels on the Use of Dictionary in 1998 and 2004**

Percentages of 4<sup>th</sup> grade students reaching each competency level in the use of Arabic Dictionary in 1998 and 2004 are presented in Figure (7).

On the competency in using dictionary there is minimal improvement indicated by 4% increase from 11% in 1998 to 15% in 2004 at the Mastery Level (L3) accompanied by 4% decrease at (L1), unacceptable and 1% decrease at the Modest Competency level (L2).



**Figure (7)**  
**Percentage of 4<sup>th</sup> Grade Students at Each Level of Competency in the Use of Arabic Dictionary in 1998 and 2004**



### II.2.3 Grade 4 Students' Arabic Language Competencies in 1998 and 2004

The Arabic test measured four sub-competencies of the general competency in Arabic Language. The four sub-competencies included Grammar, Comprehension, Writing and Using Dictionary whereas the general linguistic competency was assessed by aggregating the four subtest scores. The primary objective of this report is to track progress in the primary education from 1998 through 2004, particularly the achievement of 4<sup>th</sup> graders in Math and Arabic subjects. In order to compare students' performance on Arabic Language Competency test in 1998 and 2004, independent, unequal sample t-tests were conducted. The results are reported in Table (32).

Table (32)  
Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference of the 4<sup>th</sup> Grade Students' Performance on the Arabic Test in 1998 and 2004

Competency	Year	N	Mean	STd. Devi	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
GRAMMAR	2004	1348	65.07	30.09	.81	3.21	2445*	.00	3.10	.96
	1998	3806	61.97	31.23	.50					
WRITEP	2004	1348	45.54	35.19	.95	2.34	2431*	.01	2.64	1.12
	1998	3806	42.90	36.31	.58					
COMPP	2004	685	36.00	22.29	.85	.25	2610*	.80	.26	1.02
	1998	1927	35.75	23.11	.52					
DICTRYP	2004	685	37.59	43.75	1.67	3.94	1128*	.00	7.52	1.91
	1998	1927	30.07	40.56	.92					
TOTSCRIP	2004	1348	47.96	24.94	.67	1.83	5152	.07	1.48	.81
	1998	3806	46.48	25.73	.41					

\* df for unequal variance t-test.

The positive mean differences on all the four subtests and on the whole test, though relatively small, augur a rising trend in student achievement in Arabic Language. At least, on three of the four subscales the improvement is statistically significant ( $p \leq .02$ ) in case of Writing, ( $p \leq .01$ ) in case of Grammar, and ( $p < .000$ ) for Dictionary Use.

#### II.2.4 4<sup>th</sup> Graders' Progress in Arabic Language Within Each Education Authority from 1998 to 2004

Schools in Jordanian education system operate under the jurisdiction of four different Education Authorities with diverse resources and student populations. In the following sections we present comparative achievement of Grade 4 students on Arabic subtests in 1998 and 2004 separately for each Education Authority.

##### II.2.4.1 MOE 4<sup>th</sup> Graders' Progress in Arabic Language from 1998 to 2004

The following Table (33) presents the results of independent, unequal samples t-test analyses on fourth grade MOE school students' Arabic test scores in 1998 and 2004.

**Table (33)**  
**Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between the 1998 and 2004 Arabic Test Scores of the MOE 4<sup>th</sup> Grade Students**

Arabic	Year	N	Mean	STd. Dev	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
Grammar	2004	1014	63.47	30.84	.96	3.33	1793*	.00	3.76	1.12
	1998	3018	59.70	31.88	.58					
WRITEP	2004	1014	43.23	34.74	1.09	2.76	1788*	.00	3.51	1.27
	1998	3018	39.72	35.80	.65					
COMPP	2004	516	35.57	22.53	.99	1.17	2042*	.24	1.36	1.15
	1998	1528	34.21	22.82	.58					
DICTRYP	2004	516	37.40	44.35	1.95	3.69	820*	.00	8.14	2.20
	1998	1528	29.25	40.27	1.03					
TOTSCRIP	2004	1014	46.60	25.22	.79	2.45	4030	.01	2.27	.92
	1998	3018	44.33	25.68	.46					

\* df for unequal variance t-test.

The "Mean Difference" column in Table (33) shows that all the mean differences between 2004 and 1998 taken as (2004-1998) are positive signifying a trend toward improvement in students' performance on Arabic Language achievement tests. Moreover, on three of the four sub-competencies the improvement is highly, statistically, significant: ( $p<.001$ ) for Grammar, ( $p<.01$ ) for Writing, and ( $p<.000$ ) for Dictionary Use. The gain on the whole test score is also statistically significant at ( $p\leq.01$ ).

#### II.2.4.2 UNRWA Schools' 4<sup>th</sup> Graders' Progress in Arabic Language from 1998 to 2004

For the purpose of comparing students' Arabic test scores in 2004 to those in 1998; independent, unequal sample t-tests were carried out. The results of the t-test analyses on UNRWA samples are presented in Table (34).

**Table (34)**  
**Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between the 1998 and 2004 Arabic Language Test Scores of the UNRWA School 4<sup>th</sup> Grade Students**

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
GRAMMAR	2004	170	65.53	28.78	2.20	-.60	474	.55	-1.64	2.75
	1998	306	67.18	28.67	1.63					
WRITEP	2004	170	48.63	36.50	2.80	.93	474	.35	3.21	3.43
	1998	306	45.42	35.52	2.03					
COMPP	2004	87	33.63	20.69	2.21	-.67	239*	.50	-1.89	2.83
	1998	154	35.52	21.36	1.72					
DICTRYP	2004	87	36.78	42.08	4.51	1.34	239*	.18	7.56	5.66
	1998	154	29.22	42.28	3.40					
TOTSCRIP	2004	170	47.41	24.29	1.86	-.73	474	.47	-1.70	2.33
	1998	306	49.11	24.44	1.39					

\* df for unequal variance t-test.

In the past, UNRWA school students used to do slightly better than their MOE cohorts on tests of achievement in school subjects. This time the situation seems to be reversing. Comparing "Mean Difference" columns in Tables (33) (MOE) and (34) (UNRWA) we notice, where MOE students registered gains on all the four subtests and the total test score, the UNRWA students' mean difference scores are negative on three out of five counts, signaling a declining trend. In general, the size of mean differences is very small and none of them reached anywhere near statistical significance.

#### **II.2.4.3 Private School 4<sup>th</sup> Graders' Progress in Arabic Language From 1998 to 2004**

Bye and large, Private school students have always performed better on tests of achievement in school subjects than their cohorts from other education authorities in Jordan.

The objective of this exercise is not so much to compare the achievement of students among different education authorities as to monitor the progress in learning achievement made by the grade 4 students from 1998 to 2004 in each education authority. It is quite likely that regardless of the achievement levels of students the degree of improvement may vary among different education authorities.

For the purpose of studying changes in students test scores from 1998 to 2004, independent, unequal sample t-tests were performed on Private schools grade 4 students' Arabic test scores obtained on the same tests in 1998 and 2004. The results of the statistical analyses are displayed in Table (35).

**Table (35)**  
**Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the Private School Grade 4 Students**

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
GRAMMAR	2004	129	75.20	24.59	2.16	-.86	191*	.39	-2.06	2.41
	1998	380	77.26	20.50	1.05					
WRITEP	2004	129	59.32	33.21	2.92	-3.49	507	.00	-10.97	3.14
	1998	380	70.29	30.02	1.54					
COMPP	2004	64	40.97	22.41	2.80	-3.43	255*	.00	-10.49	3.06
	1998	193	51.46	20.78	1.49					
DICTRYP	2004	64	37.50	42.72	5.34	-.56	255*	.57	-3.43	6.07
	1998	193	40.93	41.87	3.01					
TOTSCR	2004	129	58.41	22.11	1.94	-2.81	507	.00	-5.72	2.04
	1998	380	64.14	19.23	.98					

\* df for unequal variance t-test.

Contrary to expectations, the "Mean Difference" column in Table (35) displays a pattern of decline in stead of increase in students' Arabic test scores. Invariably, on all the four subtests and the whole test mean scores have decreased in degrees varying from 2.1% points, the lowest on Grammar, through 10.5% points, the highest on Comprehension. The decrease is statistically significant on Writing, Comprehension and the Whole Test ( $p<.001$ ), ( $p<.001$ ) and ( $p<.005$ ) respectively.

#### II.2.4.4 MOD School Grade 4 Students' Progress in Arabic Language From 1998 to 2004

The results of independent, unequal sample t-tests carried out on MOD students' Arabic test scores gathered in 1998 and 2004 are presented in Table (36).

**Table (36)**  
**Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std.**  
**Err. of the Mean Difference Between the 1998 and 2004 MOD Schools Grade 4 Student**  
**Scores on the Arabic Language Test**

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
GRAMMAR	2004	35	71.75	24.87	4.20	2.85	81*	.00	15.40	5.41
	1998	102	56.35	34.36	3.40					
WRITEP	2004	35	46.50	37.64	6.36	2.72	48*	.00	18.93	6.96
	1998	102	27.57	28.43	2.81					
COMPP	2004	18	42.38	20.33	4.79	3.41	68*	.00	18.95	5.55
	1998	52	23.43	20.29	2.81					
DICTRYP	2004	18	47.22	40.11	9.45	3.38	68*	.00	30.88	9.14
	1998	52	16.34	30.87	4.28					
TOTSCRIP	2004	35	51.60	20.77	3.51	3.23	135	.00	15.10	4.67
	1998	102	36.50	24.77	2.45					

\* df for unequal variance t-test.

In terms of the magnitude of mean gain scores, MOD students have recorded the largest positive differences consistently on all the tests among all the education authorities.

On all the four subscales as well as on the whole test the mean differences are invariably positive, ranging from 15.1%, the smallest, on the Total Test Score through 30.9%, the largest on the Use of Dictionary. What is more, on all the five measures the progress is statistically significant ( $p < .01$ ).

## **II.2.5 URBAN and RURAL Area Schools' Grade 4 students' Progress on Arabic Language Competencies from 1998 to 2004**

Due to a variety of influencing factors students' achievement in Urban and Rural schools often differs. In the following we offer separate analyses for Rural and Urban school grade 4 students' performance on the Arabic Language Test in 1998 and 2004.

### **II.2.5.1 Urban School's Grade 4 Students' Progress in Arabic Language from 1998 to 2004**

Independent, unequal sample t-tests were conducted on the 4<sup>th</sup> grade students Arabic Language Test scores collected in 1998 and 2004, with intent to study change in students' test performance during this period of time.

The results produced by the t-test analyses are presented in the following Table (37).

**Table (37)**  
**Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the Grade 4 Students in Urban Schools**

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
GRAMMAR	2004	916	67.63	28.83	.95	2.64	1806*	.00	3.05	1.15
	1998	2159	64.59	30.31	.65					
WRITEP	2004	916	48.71	34.98	1.15	.43	1800*	.668	.60	1.40
	1998	2159	48.11	36.66	.78					
COMPP	2004	464	38.23	22.12	1.02	.23	1556*	.82	.28	1.26
	1998	1094	37.94	23.00	.69					
DICTRYP	2004	464	40.19	44.03	2.04	3.04	1556*	.00	7.15	2.35
	1998	1094	33.04	41.67	1.25					
TOTSCRIP	2004	916	50.83	24.48	.80	1.52	1807*	.13	1.49	.98
	1998	2159	49.34	25.76	.55					

\* df for unequal variance t-test.

The "Mean Difference" column of Table (37) shows all round improvement in varying degrees across all the five measures of competencies in the Arabic Language which include: Grammar, Comprehension, Writing, and Use of Dictionary and the aggregated score on all these four taken as a measure of overall competency in Arabic Language.

The gain score represented by the Mean Difference ranges from .29%, the lowest, on comprehension through 7.20%, the highest, on the Use of Dictionary. The gain is statistically significant, however, only on two of the four subtests, Grammar ( $p < .01$ ) and Use of Dictionary ( $p < .002$ ).

#### II.2.5.2 Rural Schools' Grade 4 Students' Progress in Arabic Language From 1998 to 2004

Rural School students' test scores were treated the same way as the Urban School students' test scores.

The results of the independent, unequal sample t-tests are exhibited in Table (38).

**Table (38)**  
**Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the 4<sup>th</sup> Grade Rural School Students**

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
GRAMMAR	2004	432	59.63	31.94	1.53	.63	2077	.53	1.09	1.73
	1998	1647	58.53	32.08	.79					
WRITEP	2004	432	38.80	34.72	1.67	1.46	2077	.14	2.73	1.87
	1998	1647	36.07	34.69	.85					
COMPP	2004	221	31.3	21.98	1.47	-.89	1052*	.37	-1.53	1.72
	1998	833	32.87	22.94	.79					
DICTRYP	2004	221	32.12	42.74	2.87	1.99	1052*	.05	5.96	3.00
	1998	833	26.17	38.75	1.34					
TOTSCRIP	2004	432	41.88	24.84	1.19	-.62	2077	.53	-.85	1.36
	1998	1647	42.73	25.21	.62					

\* df for unequal variance t-test.

In the "Mean Difference" column of Table (38) we see, three of the five mean differences are positive signifying improvement and two are negative signaling depletion. The differences, positive or negative, are all too small to be statistically significant. Looking at the results on the whole, one can only conclude that the results show no progress in Rural schools' 4<sup>th</sup> grade students' performance on achievement tests of Arabic Language from 1998 till 2004.

## II.2.6 Male/Female 4<sup>th</sup> Graders Progress in Arabic Language From 1998 to 2004

The phenomenon of female students outperforming their male cohorts on tests of achievement in school subjects is well-recognized by an increasing body of educational research in Jordan and female students get higher marks than their male counterparts on language tests nearly all over the world.

The following pages describe the progress made by male 4<sup>th</sup> graders and then by their female cohorts on competencies in Arabic Language since 1998 and 2004.

### II.2.6.1 Male 4<sup>th</sup> Graders' Progress in Arabic Language from 1998 to 2004

With the purpose to study change in students' performance on Arabic Language competencies' test and subtests from 1998 to 2004, the test



scores from the two testings were statistically analyzed employing independent, unequal sample t-test procedures.

The results provided by these analyses are summarized in Table (39).

**Table (39)**  
Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the 4<sup>th</sup> Grade Male Students

Competency	Year	N	Mean	STd. Dev.	STd. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
GRAMMAR	2004	714	60.97	32.04	1.19	12.3	2658	.02	3.30	1.43
	1998	1946	57.67	32.91	.74					
WRITEP	2004	714	39.08	34.51	1.29	2.20	2658	.03	3.34	1.52
	1998	1946	35.74	34.75	.78					
COMPP	2004	366	32.76	22.53	1.17	1.17	1346*	.24	1.63	1.39
	1998	982	31.13	22.77	.72					
DICTRYP	2004	366	33.19	42.17	2.20	3.99	584*	.00	9.98	2.50
	1998	982	23.21	36.82	1.17					
TOTSCR	2004	714	43.12	24.76	.92	1.19	2658	.23	1.33	1.11
	1998	1946	41.79	25.76	.58					

\* df for unequal variance t-test.

All the five mean differences included in the "Mean Difference" column of Table (39) are positive indicating a betterment trend in male 4<sup>th</sup> graders' performance on the Arabic Language achievement test. In addition, the improvement on two out of four subscales is statistically significant: ( $p < .03$ ) on Writing competency and ( $p < .000$ ) on the competency in the Use of Dictionary. The mean gain score on the other two sub-competencies and the total test score did not attain statistical significance at the preset ( $\alpha \leq 0.05$ ) level. The increase in the Dictionary Competency mean, nearly 10% points, is substantial and significant.

#### II. 2.6.2 Female 4<sup>th</sup> Graders' Progress in Arabic Language from 1998 to 2004

Female grade 4<sup>th</sup> students' Arabic Language Test scores taken in 1998 and then again 2004 were subjected to the same statistical analyses as those of their male cohorts.

A summary of the results produced by independent, unequal sample size t-test is presented in Table (40).

**Table (40)**  
**Mean, SD, Std. Err. of the Mean, t-Value, df, 2-Tailed Sig, Mean Difference, and Std. Err. of the Mean Difference Between 1998 and 2004 Arabic Language Test Scores of the 4<sup>th</sup> Grade Female Students**

Competency	Year	N	Mean	Std. Dev.	Std. Error Mean	t	df	Sig 2-tailed	Mean Diff.	Std. Error Diff.
GRAMMAR	2004	634	69.68	27.02	1.07	2.47	2492	.01	3.22	1.30
	1998	1860	66.46	28.71	.66					
WRITEP	2004	634	52.81	34.56	1.37	1.50	1146*	.13	2.41	1.61
	1998	1860	50.40	36.40	.84					
COMPP	2004	319	39.73	21.46	1.20	-.57	1262*	.57	-.82	1.44
	1998	945	40.54	22.48	.73					
DICTRYP	2004	319	42.63	45.04	2.52	1.89	526*	.06	5.44	2.88
	1998	945	37.19	42.99	1.39					
TOTSCRIP	2004	634	53.42	24.01	.95	1.80	2492	.07	2.03	1.13
	1998	1860	51.39	24.77	.57					

\* df for unequal variance t-test.

Comparing the "Mean" columns of the Tables (39) (males), and (40) (females) we see that they uphold the well-established empirical fact that female students outperform the male students on Language achievement tests. On the other hand, when we compare the "Mean Difference" columns of the two tables, we evidence a different phenomenon.

Female 4<sup>th</sup> graders have four positive and one negative difference. Of the four positives only one, Dictionary Use, is statistically significant ( $p \leq .011$ ), the rest are diminutive in size and nonsignificant, statistically.

On the whole, regarding female 4<sup>th</sup> graders' improvement in Arabic Language test performance, as evidenced by the test results, it seems rather stagnant. Their male cohorts, in comparison, have registered somewhat better progress.

---

### **PART.III. GENERAL CONCLUSION**

In general, on most tests and subtests in Math and Arabic Language, we found no clear-cut, indications of definite improvement in 4<sup>th</sup> grade students' test scores from 1998 to 2004. There are, all the same, encouraging trends of progress marred only by a few instances in specific subpopulations in some areas of achievement where there are signs of declining trend. On the whole, however, such decline is not statistically significant and therefore, statistically speaking we can only conclude that there is no significant change in those cases.

There are differences in progress in achievement among various subpopulations on different tests and subtests. The causes of such differences and of the lack of improvement, in general, need to be investigated systematically on their own right but such an investigation falls outside the scope of this report.

Progress at a glance in the three Math Tests and the Arabic Language Test is presented, respectively in Exhibits 1, 2, 3 and 4.

---

## PART.IV RECOMMENDATIONS

The causes of lack of progress in Math and Arabic Language at the Primary level, are diverse and manifold. Even the results could have been marred by lack of rigor in the study itself. Since there is generally a tradeoff between cost and quality of research, it is paramount that validity of results and the conclusions drawn from them should not be compromised. Based on the cumulative experience of the four studies the following suggestions are offered for serious consideration by the concerned authorities.

6. First and foremost, the authorities should have a firm conviction about the real importance and need for national studies like Monitoring and Assessment of Learning Achievement, and the objectives of such studies should be crystallized and prioritized in order to guide the design and allocation of adequate resources.
7. The choice should be between conducting a rigorous study and relinquishing it altogether because conducting a sloppy study is a sheer waste of resources.
8. If learning achievement is monitored to study the educational realities on the ground and to measure student achievement for guiding improvements where they are needed, then the authorities should openly accept the realities revealed by the valid findings of the study. Only the sincere realization and acceptance of the shortfalls can lead to earnest effort for improvement.
9. The causes of low performance and lack of progress of different segments of primary cycle populations on different tests and subtests should be properly investigated and pin pointed.
10. Having identified the causes of low performance, appropriate measures and practicable actions should be proposed to improve the situation.

# EXHIBIT – 1 : MATH1

National Progress on Math Test1 (General Competency in Understanding and Application of Procedures) in Different Sub-populations of 4<sup>th</sup> Grade Students from 1998 to 2004

## Sub-populations

Competency	Whole Sample	MOE	UNRWA	Private	MOD	Urban	Rural	Male	Female
TOT	+	+	-	+	** +	+	-	-	** +
NUMBER	-	-	-	-	* +	-	** -	* -	+
ADDITION	+	-	+	+	+	+	-	-	+
SUBTRACTION	* +	* +	-	+	** +	** +	-	+	** +
MULTIPLICATION	** +	* +	-	+	+	** +	-	+	** +
DIVISION	+	+	-	-	+	+	-	-	+
FRACTIONS	* +		-	+	** +	** +	-	-	*** +
GEOMETRY + MEAS.	** -	** -	-	** -	* +	** -	-	** -	-
DECIMALS	+	+	-	* +	** +	+	-	-	+

## EXHIBIT – 2 : MATH2

**National Progress on Math Test2 (General Competency in Mathematical Thinking and Communication) in Different Subpopulations of G4 Students from 1998 to 2004**

### Sub-populations

Competency	Whole Sample	MOE	UNRWA	Private	MOD	Urban	Rural	Male	Female
Thinking	+	+	-	-	** +	+	-	-	* +
Thinking + Communication	+	-	+	+	** +	+	-	-	+
Communication	+	+	* -	-	** +	+	-	-	+
Number	+	+	-	-	** +	+	-	-	* +
Fractions	+	+	-	*	** +	+	-	-	* +
Geometry + Measurement	+	+	-	-	** +	+	-	-	+
Number Thinking	+	+	+	-	** +	+	-	-	** +
Number Thinking Communication	+	+	-	+	** +	+	-	-	+
Number Communication	+	+	** -	+	** +	+	-	-	* +
Geometric Thinking	+	+	-	-	** +	+	-	-	+
Geometric Communication	+	-	+	-	** +	+	-	*	* +
Fractions Thinking	+	-	+	-	** +	+	-	-	** +
Fractions Communication	+	+	-	** -	** +	+	-	+	+
Total Test	+	+	-	-	** +	* +	-	+	+

### EXHIBIT – 3 : MATH3

National Progress on Math Test3 (General Competency in Math Problem Solving) in Different Sub-populations of G4 Students from 1998 to 2004

#### Sub-populations

Competency	Whole Sample	MOE	UNRWA	Private	MOD	Urban	Rural	Male	Female
Whole Test	+	-	*	+	+	+	-	-	+
Number Problems	+	-	*	+	+	+	-	-	+
Geometry Problems	-	**	+	+	-	-	-	-	-
Geometry 1 Step	*	**	-	-	-	-	-	-	-
Geometry 2 Step	+	-	+	**	-	+	-	+	-
Number 1 Step	+	+	**	+	+	+	+	+	+
Number 2 Step	-	-	+	-	+	-	-	-	+
Step 1 General	+	-	*	+	+	+	-	+	+
Step 2 General	-	*	+	+	+	-	-	-	-

## EXHIBIT – 4 : ARABIC LANGUAGE

National Progress on Arabic Language General Competencies In Grammar,  
Writing, Comprehension and Dictionary Use in Different Sub-Populations of G4  
Students from 1998 to 2004

### Sub-populations

Competency	Whole Sample	MOE	UNRWA	Private	MOD	Urban	Rural	Male	Female
GRAMMAR	** +	** +	-	-	** +	** +	+	* +	+
WRITING	** +	** +	+	** -	** +	+	+	* +	+
COMPREHENSION	+	+	-	** -	** +	+	-	+	-
DICTIONARY USE	** +	** +	+	-	** +	** +	* +	** +	+
WHOLE TEST	+	** +	-	** -	** +	+	-	+	+