

The Hashemite Kingdom of Jordan



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

**Monitoring Learning
Achievement and Instructional
Quality in the Primary
Grades of the Basic Cycle
In Jordan: Phase III**

A Preliminary Report

By
Dr. Kapur Ahlawat Dr. Tayseer Al-Nhar

Publication Series (69)

1999

Table of Contents

Table of contents	A
List of Figures	C
List of Tables	D
EXECUTIVE SUMMARY	i
▪ Context	i
▪ Objectives	i
▪ Study Design	ii
▪ Instruments	ii
▪ Data	ii
rustles	iii
▪ I Math test	iii
▪ II Competencies in Social Studies	vi
▪ III Arabic Language Test	vii
ontext	1
he Objectives and Outcomes of Phases I and II	2
▪ More specific objectives included	3
▪ The indirect objectives included	4
bjective of the Third Phase of the Study	5
udy Design	6
▪ Population and Sample	6
▪ Instruments	8
▪ Lay Out of Achievement Tests	9
- Structure of the Math Test	10
- Structure of the Social Studies Test	12
- Structure of the Arabic Language Test	13
ata Collection	15
oding, Scoring, and Data Entry	15
ata Analysis	16
ESULTS	18
. Math Tests	18
▪ Psychometric Properties of the Math Tests	18
▪ Students' Performance on Math Tests	19
- Education Authority Comparisons	21
- Rural Schools Vs Urban Schools	24
- Gender Comparisons on Math Test1	26
▪ Math Test2: General Competency in Mathematical Thinking and Communication	27

- Comparative performance Among Education Authorities	29
- Urban/Rural Comparisons	33
- Gender Differences on Mathematical Thinking and Communication Competencies	34
- Math Test3: Problem Solving Competency	36
- Education Authority comparisons on Math Test3: Problem Solving	38
- Urban/Rural Schools Comparisons	41
- Gender Differences on Math Problem Solving	42
II. Competencies in Social Studies	44
▪ Psychometric Properties of the Social Studies Tests	44
▪ National Performance on Social Studies Test	45
- Education Authority Differences in Social Studies Achievement of Grad4 Students	49
- Urban/Rural Differences on Competencies in Social Studies	52
- Gender Differences on Social Studies Competencies	53
III. Arabic Language Test	55
▪ Psychometric Properties	55
▪ Reading Comprehension Competency	55
▪ Grammar competency	56
▪ Writing Competency	56
▪ Dictionary Utilization Competency	57
- Gender Comparisons on Arabic Language Competencies	57
- Comparative Performance of Urban and Rural Schools' Students	58
- Comparisons between Educational Authorities	59

List of Figures

Figure1	National Competency Level in Knowledge and Application of Mathematical procedures	20
Figure2	Comparative Performance of Education Authorities on the General Competency of Understanding and Application of Mathematical procedures	24
Figure3	National Competency Level in the General Competency of Mathematical Thinking and communication	29
Figure4	Comparative performance of Education Authorities in the General Competency of Mathematical Thinking and Communication	32
Figure5	National Competency in Mathematical problem Solving	37
Figure6	Comparative performance of Education Authorities on the General Competency in Problem Solving	40
Figure7	Relative national Performance on the Nine Major Competencies in the Social Studies	48
Figure8	National Overall Competency Level in Social Studies	49
Figure9	Comparative Performance of the Four Education Authorities in Social Studies	51
Figure10	Percentage Frequency Distribution of Percent Correct Score Grammar Competency	56
Figure11	Percentage Frequency Distribution of Percent Correct Score Writing Competency	57
Figure12	Gender Differences on the Arabic Language Competencies	58
Figure13	Urban/Rural Schools Differences on the Arabic Language Competencies	59
Figure14	Education Authority Difference on the Arabic Language Competencies	60

List of Tables

Table1	Sampling Distribution of Schools According to Education Authority, Location and Shift	7
Table2	Sampling Distribution of Students According to Education Authority, School Location and Student Gender	7
Table3	Sampling Distribution of Students by Governorate Student Gender and Schools by Governorate and Location	8
Table4	Table of Specifications for Math Test Form1	11
Table5	Table of Specifications for Math Test Form2: Mathematical Thinking and Communication	11
Table6	Table of Specifications for Math Test Form3: Problem Solving	11
Table7	Table of Specifications for SS Test1 and SS Test2 (Grade4 Tests of Social Studies)	12
Table8	Table of Specifications for the Arabic Language Test (Form1 and Form2)	14
Table9	Competency Levels of Grade4 Students on Math Test1 (Understanding and Application of Procedures)	21
Table10	Comparative Achievement of Different Education Authorities on Math Test1 (Understanding and Application of Procedures)	22
Table11	Comparative Performance of Urban/Rural School Students on Math Test1 (Understanding and Application of Procedures)	25
Table12	Comparative Performance of Male/Female Students on Math Test1 (Understanding and Application of Procedures)	26
Table13	National Level Performance of Grade4 Pupils on Math Test2: Mathematical Thinking and Communication (N=1435)	27
Table14	Comparative Performance of Education Authorities in the General Competency of Mathematical Thinking and Communication (Test2)	30
Table15	Comparative performance of Urban and Rural Schools' Students in the General Competency of Mathematical Thinking and Communication (Math Test2)	33
Table16	Gender Comparisons on Mathematical Thinking and Communication Competencies	35

ble17	National Performance on Problem Solving General competency and Sub-Competencies (N=1396)	36
ble18	Comparative performance of Education Authorities in the General and Specific problem Solving Competencies	39
ble19	Comparison of Urban and Rural School Students' Performance on the General Competency in Problem Solving	41
ble20	Gender Comparisons on Problem Solving Competency and Sub-Competencies	42
ble21	National performance on General Competencies in Social Study	47
ble22	Comparative Performance of Education Authorities on General Competencies in Social Studies	50
ble23	Urban/Rural School Location Differences in Student Performance on Competencies in Social Studies	53
ble24	Gender Differences in Students' performance on Social Studies Competencies	54

EXECUTIVE SUMMARY

Context

Pursuant to the goals of EFA, the Government of Jordan had embarked upon a comprehensive program of educational reform aimed at enhancing the quality, efficiency, effectiveness and relevance of the general education system. The reform program particularly emphasized upon enhancing student achievement levels through promoting critical thinking, problem solving, integration of diverse knowledge and its innovative application to real life situations, and other higher cognitive skills of students.

In the wake of EFA, UNICEF/UNESCO launched a global project for monitoring learning achievement of the goals of EFA. At the initial phase of MLA project only six countries had participated. Jordan was one of them.

The NCHRD in Jordan had charted its own program of assessment of the quality of instruction and monitoring the progress of education on regular basis, which fitted nicely with the objectives of the UNICEF/UNESCO's MLA project.

Since 1992 UNICEF-Amman has cofinanced three phases of the NCHRD's longitudinal program of which the present study is the Third Phase.

Objectives

The major objectives of this study included:

- a. To measure student achievement levels, at the end of 4th grade, in Arabic Language (Literacy), Mathematics (Numeracy), and Social Studies (Life Skills). Achievement was defined in terms mastery levels reached in well-defined general and specific competencies.
- b. To identify pockets of strengths and weaknesses.
- c. To investigate disparities across education authority, student gender and urban-rural location.

- d. To conduct error analyses to identify sources of misconceptions in order to provide for strategies for instructional/ learning improvement.

Study Design

A two-stage stratified random sampling design was used to extract a sample of 216 schools with Grade4 classes. In schools with multiple Grade4 classes one class was randomly selected. The whole sample comprised 5,929 pupils of which: 3,030 were males and 2,899 females; 3,367 studied in urban schools and 2,562 in rural schools.

Instruments

The battery of instruments was comprised of :

- a. Competency based achievement tests of Arabic Language, Mathematics, and Social Studies;
- b. A students questionnaire;
- c. A teacher questionnaire; and
- d. A principal questionnaire.

The Math Test had four forms, three new and one old. Arabic Test had three forms two new and one old; and the Social Studies Test had two new forms because Social Studies was included in this study for the first time.

Data

The data were collected during the last week of May, 1998 by MOE's team of supervisors who had undergone special training in the standardized modalities of the test administration and data collection operation.

The tests were subsequently scored and coded. Data were entered and cleaned and then analysed using appropriate statistical procedures.

RESULTS

I Math Test

General competency of Understanding and Application of Procedures

On this general competency the national average was 37.9 percent correct. 42% of the pupils scored 30% or below and were placed in L1 category defined as unacceptable or nonmastery level. 51% of the sample scored more than 30% correct but less than 70% correct. This was defined as L2 or partial mastery or modest competency level. Only 7% of the pupils scored 70% correct or above and they were placed in L3 or mastery level category.

Education Authority Comparisons

Private school students scored the highest average (51.62% correct), UNRWA students with 39.6% correct average came second, MOE students with 36.31% correct came third while MOD students with approximately 30.92% correct were at the bottom.

Urban-Rural Comparisons

Urban school students' performance was slightly but consistently better than that of their rural school cohorts. The differences, however, were statistically significant on the total test and on only three out of eight sub-competencies. Considering the fact that urban communities enjoy several advantages over rural communities and also that all the private schools are urban the differences between performance of urban and rural school students are rather nominal.

Gender comparisons on the General Competency of Understanding and Application of Mathematical Procedures

Contrary to findings of numerous studies supporting the hypothesis that male students outperform the female students on Math tests, studies in Jordan have reported no significant differences between the achievement of male and female students on Math tests. In this respect the findings of this study are consistent with those of other studies of this type conducted in Jordan during the nineties.

On the total test, as well as, on every one of the eight sub-tests the difference between the male and female students' performance was not statistically significant at $\text{Alpha}=.05$ level. Nevertheless, the female students scored consistently higher than their male cohorts on every subtest, as well as, on the test as a whole.

National Performance on the General Competency in Mathematical Thinking and Communication was 46.3% correct, below the 50% cutscore.

Regarding specific competencies enfolded by the general competency, the national average on Mathematical Thinking was 62% correct; on Communication and Thinking 37% correct; and on Communication in Mathematics it was 55% correct.

According to three mastery levels 24%, nearly a quarter of the pupils, failed to reach the 30% correct limit and fell in the unacceptable/nonmastery category L1. 51% fell in L2, a partial mastery category, whereas, 20% of the students reached or surpassed the 70% correct boundary line for mastery level, L3.

Education Authority Comparisons

Private schools with a rounded mean of 58% correct scored the highest, UNRWA and MOE with their means of 45.5% and 45.3% respectively, came next, while the MOD's mean of 37.6% correct was the lowest.

Urban-Rural Comparisons

On the total test, as well as, on all the 13 subtests, urban school students outperformed their rural school cohort but differences were statistically significant ($\text{Alpha} \leq .05$) only on the total test and six out of thirteen specific competencies.

Gender Comparisons

On the test as a whole, as well as, on all the thirteen specific competencies female students scored consistently but negligibly higher than their male cohort, nevertheless, none of the differences reached statistical significance level of $\text{Alpha} = .05$.

Problem Solving Competency

The national average performance on the general competency in Problem Solving was only 29.6% correct. Nearly two-thirds (64%) of the sample was placed in L1 (noncompetency) category; and 32% of the sampled students were placed in L2 (modest competency or partial mastery) category. Only 4% qualified for the L3 or mastery category.

Education Authority Comparisons on Problem Solving

According to the total test score on problem solving Private Schools' mean percent correct score was 34.4%; MOE's mean was 29.4%; UNRWA's mean was 27% while MOD's mean was 23.6%. This ordering pattern among the four education authorities persisted on all the specific competencies also.

Urban-Rural Comparisons

Perhaps due to the floor effect the difference between the performance of urban and rural schools was nonexistent on the general competency of problem solving as well as on all the specific competencies in this field.

Gender Differences on Problem Solving

Female students outperformed their male cohort on the whole test including every one of the specific competencies except one (One step problems in numbers) on which the difference did not reach statistical significance at $\text{Alpha} = 0.05$ level.

II Competencies in Social Studies

National Performance

The national average on the whole social studies test which entailed nine general Competencies was only 35.5% correct.

Among the nine general competencies the best performance (46.2% correct) was on Competency 7 (Understanding the concepts of social responsibility and good citizenship); while the worst (27.8% correct) was on competency 2 (Understanding the relationship between mankind, environment and resources).

On the mastery scale, only 6% of the students reached the mastery category L3; 50% were placed in partial or modest mastery category L2 whereas, 44% of them remained in the L1 or unacceptable performance category.

Education Authority Comparisons

Private schools obtained a mean of 45.6% correct which was statistically significantly higher than the mean of each of the other three education authorities. UNRWA with a mean of 35.0% and MOE with a mean of 34.6% came bracketed second, and MOD with a mean of 26.8% correct came last.

Urban-Rural Comparisons

On the whole test and on four of the nine general competencies urban schools' students scored higher than their rural schools' cohort; on five general competencies there were no significant differences between the performance of urban and rural schools.

Gender Differences on Social Studies Competencies

Female students outperformed their male cohort on the whole test as well as on eight of the nine general competencies.

III Arabic Language Test

Arabic language test assessed student achievement on four general competencies viz., Reading Comprehension, Grammar, Writing and Using Dictionary.

National Performance Comprehension.

National average on Reading Comprehension was 34.5% correct. Only 5.5% of the sampled students reached mastery level L3; 51% were placed in L2, partial mastery category while, 43.5% remained in the unacceptable performance category L1.

Grammar

The national average on Grammar competency was 61.3% correct. More than half (52.9%) reached the mastery level L3; 26.1% were placed in partial mastery category L2 and 21% remained in the unacceptable performance category L1.

Writing

The national average was 43.6% correct; 34.1% of the students reached the mastery Level L3; 17.9% of the students reached the partial mastery category L2 while 48% of the sample could not cross the unacceptable performance category L1.

Dictionary Utilization

The national average was 39.2% there was only one item. 39.2% of the students passed and 60.8% failed.

Gender Comparisons

The performance of female students was statistically significantly better ($p < .001$) than that of their male cohort on all the four measured general competencies of Arabic Language.

Urban-Rural Comparisons

Urban school students scored higher than their rural school cohort on all the four general competencies of Arabic Language.

Education Authority Comparisons

Private school students' means were the highest on all the four competencies.

On Reading Comprehension and Dictionary Utilization there were no significant differences between MOE and UNRWA but both of them outperformed MOD.

On Grammar there was no significant difference between MOE and MOD while UNRWA scored significantly higher than both of them.

On Writing, Private schools' performance was the best of all. UNRWA did significantly better than both MOE and MOD while MOE did significantly better than MOD.

Context

Never before in human history the role of primary education in national development got such paramount importance as it did during the past two decades.

The worldwide realization of the fact that investment in primary education is positively related to economic growth, social progress and quality of life materialized in the world conference on Education for All (WCEFA) held at Jomtien Thailand in 1990 which mandated the World Declaration on Education for All. In the WCEFA the delegates from all over the world pledged themselves and their Governments to make renewed effort to achieve the moral imperative of Education for All (EFA).

Following the EFA mandate most countries all over the world, especially the developing countries like Jordan, have made relentless effort to provide universal access to primary education for ever increasing numbers of school-age children. To this effect, Jordan right from its independence, has followed a rigorous policy to harness its human resources by means of providing for quality education and training. Because of this policy, education system in Jordan expanded rapidly during 60s through 80s. Enrollment rates in 1995-96 had reached 98% for (6-11)-year-olds and 93% for (12-16)-year-old children; among the highest in the Middle East. The rapid quantitative expansion, however, took its toll in terms of instructional quality.

To address the issues of quality and relevance of education the Government of Jordan in 1988 embarked upon a comprehensive education reform plan aimed to improve the quality of learning and enhance its relevance to country's social and economic needs while maintaining steady growth of the education system to achieve the goal of Education For All. The reform plan emphasized the importance of promoting critical thinking, problem solving, and higher cognitive skills of students so that they would be productive citizens making effective contribution to social and economic development of the country in the coming years of fast emerging global economic order.

The comprehensive education reform revamped every aspect of the education system which included: reconstruction of curricula; redesigning and development of textbooks, teacher guides and other learning/teaching materials; upgrading the qualifications of teachers, principals and supervisors through reforming preservice teacher education programs at the universities and provision of inservice long-term and short-term training programs to effectively use the newly developed curricula and textbooks and to apply modern pedagogy and instructional technology in classrooms. Upgrading school plant and school facilities like library, science labs, AV/TV room, multipurpose workshop; restructuring the examination system; provision of learning/teaching materials to schools; restructuring basic and secondary education system; and establishing learning resources centers in all regions.

In order to institutionalize the reform process and its implementation and monitoring functions NCERD (now NCHRD) was created as an autonomous body in 1990. The NCHRD designed a longitudinal study to monitor the learning achievement of students and assess instructional quality of the basic education.

During the same period, at the international scale, UNESCO and UNICEF following up the WCEFA mandate "to define acceptable levels of learning acquisition for educational programs and improve and apply systems of assessing learning achievement", had launched a joint project, Monitoring Learning Achievement, in September 1992. Jordan was one of the five countries that participated in this project. In this context UNICEF has financially supported the monitoring learning achievement and instructional quality study in Jordan of which this is the report of the Third Phase.

The Objectives and Outcomes of Phases I and II

The broad objectives, of the First and Second Phases of the Jordanian Monitoring Learning Achievement study included establishing baseline bench marks of indicators of learning achievement and of some of the contextual factors to serve as baseline criteria against which educational progress could be compared and reform impact could be measured, and providing formative feedback to guide the implementation of reform projects. Over and above all this, the primary objective of building up national capacity has remained constant throughout the whole project.

More specific objectives included :

- a) Measuring learning achievement of students at the end of the 4th year of basic education cycle in the following key subjects;
 - 1. Arabic Language (Literacy),
 - 2. Mathematics (Numeracy),
 - 3. Science, and
 - 4. Life Skills.
- b) Studying students' family background, parental practices, perceptions, expectations and attitudes related to school and classroom practices and their children's performance
- c) Studying teachers' instructional practices, attitudes, beliefs, perceptions, expectations and background characteristics.
- d) Studying schools' characteristics and principals' administrative practices, beliefs and attitudes.
- e) Benchmarking students' attitudes, beliefs, and perceptions of schools, teachers, and classroom practices.
- f) Estimating distribution parameters of salient school, family principal, teacher and student variables across different regions, student gender, education authorities and rural/urban environments.
- g) Studying relations between various home background, community, school, teacher, classroom, and student related factors and students' achievement.
- h) Identifying strengths and weaknesses of the system, providing formative feedback, deriving policy implications and suggesting plans of action.

The indirect objectives included :

- a) Building up national capacity for monitoring educational progress, and assessment of student achievement and instructional quality.
- b) Institutionalizing the evaluation function by establishing mechanism for monitoring learning achievement at regular intervals.
- c) Establishing channels of communication and promoting dialogue between evaluation research and educational planning to formulate informed policy and effective plans of action.
- d) Establishing Education Management Information System (EMIS) and promoting the utilization of relevant, reliable and timely information to develop useful indicators of access, quality, and efficiency of the education system.

Regarding the delivery of these objectives, NCHRD has published several reports describing students' levels of achievement and the relative impact of school and out-of-school variables on students' achievement. These reports have been circulated to concerned parties and have been discussed systematically in terms of the implications of the findings on the teaching/ learning process in a special seminar sponsored by UNICEF in 1997.

In addition, a special report focusing on monitoring learning achievement of Basic Education for All Goals had been published and widely discussed in UNESCO/UNICEF sponsored regional and international workshops held respectively in Amman and in Paris. The outcomes of the Paris seminar which included the major findings and implications of Jordanian study along with those of other participating countries have been published by UNESCO/UNICEF (Paris).

The objectives of the Second Phase of this study (again financed by UNICEF) were related to assessment of the reform impact. The results of this study also have been published by NCHRD in a series of reports.

Objectives of the Third Phase of the Study

The reform process has been going on for the past 9 years and its full implementation is near completion. A variety of key changes have been introduced in the education system.

It is critical time to monitor the intended effects of these changes. So Phase III of this continual Monitoring Learning Achievement and Assessment of Instructional Quality study is intended to address several current hot issues which have been translated into the following objectives :

- a) To assess learning achievement of Grad4 pupils in Arabic Language, Mathematics, and Social Studies.
- b) To study the trends in the learning achievement levels of pupils at the end of the primary cycle (Grade4) in view of the reform inputs and expected outcomes. A thorough trend analysis would involve examination of both quantitative performance over time and qualitative change in learning competencies.
- c) To assess instructional quality in terms of relevant school related inputs along with cumulative instructional outcomes indicated by levels of desired knowledge, skills and competencies.
- d) To identify areas of relative weaknesses and strengths both in school inputs and learning outcomes.
- e) To investigate if there were pockets of inequality; regional, urban/rural, or gender.
- f) To study relationships between learning outcomes and schooling factors.
- g) To identify students' misconceptions and plausible sources of errors through error analysis.
- h) To provide formative feedback to concerned authorities in the education system.
- i) To continue building up national capacity for monitoring evaluation and educational research.

Study Design

Population and Sample

The population of this study was defined as the population of all students enrolled in 4th grade classes in the mainstream schools of the Jordanian education system, and at the school level the population of all the schools that have a 4th grade class.

A two-stage stratified random sampling procedure was developed to extract the sample of the study. Using Education Authority and Location (Urban/Rural) as stratification variables a random sample of 216 schools was selected at the first stage. Essentially this was the sample which was used in 1993 and 1995 for the monitoring learning achievement study. At the second stage of sampling, one class was randomly selected out of all Grad 4 classes of a school in all sampled schools that contained two or more 4th grade classes. Thus the sample comprised 5,929 male and female students enrolled in 216 Grad 4 classes; one class per school.

Sampling distribution of schools according to Education Authority, school Location, Ownership of school building and Shift was as follows :

In respect of education authority 177 (81.9%) schools belonged to the Ministry of Education (MOE), 6 (2.8%) to Ministry of Defense (MOD), 12 (5.6%) to United Nations Relief and Works Agency (UNRWA) and 21 (9.7%) to Private entities.

Locationwise, 107 (49.5%) schools were situated in urban areas and 109 (50.5%) in rural areas.

In terms of Ownership, 159 (73.6%) schools functioned in buildings owned by the education authority concerned, 43 (19.9%) schools functioned in rented buildings and 14 (6.5%) schools had part owned and part rented buildings.

With regards to Shift, 181 (83.8%) schools operated as single shift schools while 35 (16.2%) operated under double-shift conditions

Table1 presents the three dimensional distribution of schools across Education Authority, School Location and Shift.

Table 1
Sampling Distribution of Schools According to
Education Authority, Location and Shift

Shift	MOE		MOP		UNRWA		Private		Total
	U	R	U	R	U	R	U	R	
1.	95	58	5	1	1	0	20	1	181
2.	5	19	0	0	8	3	0	0	35
Sub-total	100	77	5	1	9	3	20	1	216

Of the 5,929 pupils included in the sample 3,030 were males and 2,899 females; 3,367 came from urban schools and 2,562 from rural schools; while of all the pupils in the sample 4,708 came from MOE schools, 156 from MOD schools, 484 from UNRWA schools and 581 from Private schools. The sampling distribution of pupils by Location by Sex and Governorate is given in Table2

Table 2
Sampling Distribution of Students According to Education
Authority, School Location and Student Gender

Sex	Authority	MOE		MOD		UNRWA		Private		Total
	Location	U	R	U	R	U	R	U	R	
Male		972	1393	40	116	199	37	267	6	3030
Female		1414	929	0	0	174	74	301	74	2899
Total		2386	2322	40	116	373	111	568	80	5929

Sampling distribution of schools and pupils across the 12 Governorates is given in Table 3

Table 3
Sampling Distribution of Students by Governorate
Student Gender and of Schools by Governorate and Location

Governate	Male	Female	Total	Urban	Rural	Total
<i>Amman</i>	826	1139	1965	55	11	66
<i>Madaba</i>	128	142	270	5	6	11
<i>Zarqa</i>	389	456	845	18	8	26
<i>Balqa</i>	135	121	256	4	9	13
<i>Irbid</i>	754	420	1174	11	27	38
<i>Jarash</i>	103	118	221	3	5	8
<i>Ajloun</i>	52	51	103	1	3	4
<i>Mafrq</i>	167	169	336	1	15	16
<i>Tafleh</i>	26	83	109	2	3	5
<i>Karak</i>	221	102	323	3	11	14
<i>Maan</i>	158	98	256	4	8	12
<i>Aqaba</i>	71	0	71	0	3	3
TOTAL	3030	2899	5929	107	109	216

Instruments

Since the study aimed to obtain information on salient contextual factors that influence student learning as well as on the learning achievement itself, the following instruments were constructed, developed and validated.

1. Competency-Based Achievement Tests of :

- a. Arabic Language
- b. Mathematics, and
- c. Social Studies.

2. **Student Questionnaire** was designed to tap information on: students' home background, parental practices, interest, involvement and expectations; students' in-class, out-of-school, and at-home activities and practices; their perceptions of teacher practices, classroom environment and instructional processes; their attitudes, beliefs and feelings.

3. **Teacher Questionnaire** was constructed to obtain information about teacher characteristics, their problems, instructional

practices, interaction with principal and supervisors, efficacy and job satisfaction.

4. **Principal Questionnaire** was designed to get information on principals' characteristics, leadership, administration practices, school-community relationships, problems; school condition, facilities, staff, resources and environment.

Lay Out of Achievement Tests

The following section describes the content and structure of the achievement tests.

In line with the new objectives, curricula and textbooks the achievement tests were redesigned to reflect the changes introduced by the educational reform. The reconstructed educational objectives had incorporated the Education for All (EFA) objectives. Moreover, the modern innovative teaching/learning strategies including the application of appropriate learning technologies and instructional aids were also on the lines of recommendation of the Action Plan of EFA.

In view of the current situation the steering committee for the Monitoring Learning Achievement study decided to reconstruct the achievement tests to measure the achievement of reformulated objectives.

The new strategy entailed several changes :

1. While Arabic and Mathematics subjects were retained, Science was replaced by Social Studies.
2. In order to ensure adequate coverage of skill competencies and subject matter without increasing the length of the tests, more than one independent forms were constructed for each test.
3. Different forms were independent in the sense that each form contained different general and specific competencies in different content areas. The test forms were thus, complementary rather than parallel or equivalent in the sense of classical test theory.
4. Three new forms were developed for Math two for Arabic and two for Social Studies.

5. While new forms were designed to measure the levels of skills and competencies emphasized by the reform program, the old Math and Arabic tests were kept for the purpose of comparing student achievement in 1998 to that in 1995 and 1993.

In this way, Math test had four forms, Arabic test had three forms, and Social Studies test had only two forms because this subject was included in the study for the first time.

The new test forms corresponded to the altered objectives of the study. The primary objective of the 1993 study was to establish baseline bench marks and that of the 1995 study was to evaluate reform impact. The goal of the present study, in addition to measuring learning achievement and tracking educational progress, has focused upon determining levels of general and specific competencies of grade 4 students in prescribed areas of the subject matter.

Besides establishing competency levels and identifying areas of strength and weakness, a more important objective was to identify various kinds of misconceptions, misunderstandings and errors of the students and to link them to their sources so that appropriate teaching/learning strategies could be developed by educators to enhance student achievement levels.

It is commonly recognized that research studies like this one fall short of their objectives if they were unable to provide effective feedback and line of action to the concerned audiences.

The foregoing concerns guided the construction and development of the achievement tests and other instruments.

Structure of the Math Test

Each of the three new forms of the math test was designed to measure a different broad competency. Different general and specific competencies, however, could be applied to the same topic or subject matter. The three broad competencies, each covered by a different test form, are :

Form 1. Understanding and Application of Procedures

Form 2. Mathematical Thinking and Communication

Form 3. Problem Solving

The tables of specifications for the three forms of the Math Test are presented in Tables 4,5, and 6 respectively for forms 1,2, and 3.

Table 4
Table of Specifications for Math Test Form 1

Competency		
Understanding and Application of Procedures		
Content		Items
Natural Numbers	Properties	5
	Addition	2
	Subtraction	4
	Multiplication	4
	Division	2
Fractions		5
Decimals		5
Meas. & Geometry		3
Total		30

Table 5
Table of Specifications for Math Test Form 2: Mathematical Thinking and Communication

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

Table 6
Table of Specifications for Math Test Form 3
: Problem Solving

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

Structure of the Social Studies Test

The national test construction team for Social Studies had translated the Grade4 curricular objectives into a domain of general and specific competencies expected to be acquired by students. Out of the whole domain, nine major competencies were selected for assessment of student achievement in Social Studies. Each of the nine general competencies was composed of several specific competencies.

Test items were constructed to measure each specific competency. Two tests were constructed to measure students' performance on the 9 general competencies in social studies.

The distribution of test items between the two tests was such that items related to some competencies were restricted to only one of the tests while items related to some others were spread over both the tests. The combined table of specifications for both tests is given in Table7.

Table 7
Table of Specifications for SS Test 1 and
SS Test2 (Grade4 Tests of Social Studies)

Serial NO.	Competency	SS Test1 No of Items	SS Test2 No. of Items
1.	Ability to draw geographical maps	8	14
2.	Understanding the relationship between mankind, environment and resources.	9	—
3.	Reading and interpretation of maps	2	—
4.	Understanding systems and institutions of modern state	6	—
5.	Understanding historical events.	2	4
6.	Understanding geographical historical and economic concepts.	7	—
7.	Understanding the concepts of social responsibility and good citizenship	4	3
8.	Understanding political life in Arab and Islamic world	—	5
9.	Understanding pictographs	—	10
	Total	38	36

Structure of the Arabic Language Test

The Arabic Test was designed to measure student achievement levels on 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. Form1 comprised 28 items and Form2 comprised 24 items. The table of specifications for the Arabic Language Test is given in Table8.

Table 8
Table of Specifications for the Arabic Language Test
(Form1 and Form2)

Competency	Sub-competency	Form1	Form2	Number of Items
Reading Comprehension	<ul style="list-style-type: none"> ▪ Give meanings of new words. ▪ Identify main ideas. ▪ Identify implicit ideas. ▪ Justify events. ▪ Give real life situations related to the text. ▪ Determine the expected end. ▪ Ask questions. ▪ Follow events in order. 	Q2: a, b, c Q1: 1, 2 Q1: 3, 4, 5 Q6 Q4: a, b, c Q7 Q5: a, b Q3	(3) (2) (3) (1) (3) (1) (2) (1)	16
		Q10: a, b, c, d Q11: a, b, c _____ _____	Q5: a, b, c, d, e, f Q6: a, b, c, d Q7: a, b, c, d Q9: a, b Q8: a, b, c, d	
Grammar	<ul style="list-style-type: none"> ▪ Use language structures in a proper way. ▪ Apply language patterns. ▪ Fix word endings using grammatical rules. ▪ Relate structure to meaning. 	Q9: a, b, c, d _____ _____ _____	(4) (1) (1) (1)	8
Writing	<ul style="list-style-type: none"> ▪ Use punctuation marks. ▪ Write a story about a picture or a set of pictures. ▪ Write a letter. ▪ Write a composition on a subject using a set of given questions. ▪ Use dialogue in writing a narration. 	Q8	(1)	1
Using Dictionary	Order a set of words according to the first letter.			
Total		28	24	52

Data Collection

Since the study was 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 the school classrooms and supervision and monitoring of the whole data collection operation in the field.

The data were collected from students in the sampled 4th grade classes; from the Arabic, Math and Social Studies teachers of those classes; and from the principals of sampled schools, in the last week of May in 1998.

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 to be 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.

Since erroneous responses to specific test questions lead to thinking processes and response strategies that most likely would have generated particular wrong responses, only experienced professionals were capable of performing the task of error classification and linking errors to their most plausible sources. It was recognized that none else but the teams which constructed the tests would be most competent to perform this specialized task.

Therefore, the test construction teams were recommissioned and charged with the responsibility to : classify responses according to types of errors; clearly define each typological category; relate each

error type to its plausible causal sources; and assign a code to each error type. The purposes of error analysis and the coding process were clarified to the scoring teams and they were properly trained for the tasks at hand.

Then, the teams coded and scored the supply-response items of each form of the achievement tests and classified the responses into distinct categories.

The underlying idea was to code and score the responses in a way that would facilitate conducting error analyses to yield substantive information which, in turn, would provide proper insight into those cognitive processes, problem solving strategies, operational procedures, misconceptions, etc. which produced particular classes of wrong responses; and subsequently suggest appropriate corrective measures and strategies for improvement.

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

Data Analysis

In accord with the 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' (Grad4) general achievement in Arabic, Mathematics and Social Studies?
2. What are the national levels of primary pupils' general and specific competencies in different content and skill areas of Arabic, Mathematics and Social Studies?
3. Are there any Gender, Location and Education Authority differences in levels of various competencies in Arabic, Math and Social Studies?

4. What are the areas of relative strength and weakness in respect of competency levels in the three subjects?
5. What are the major types of errors committed by Grad 4 pupils in each of the three subjects?
6. How does the Arabic and Math tests' performance of G4 pupils in 1998 compare with that in 1995?
7. What are the major sources and operating factors behind different types of errors in Arabic, Math and Science?
8. What actions and measures could be adopted to minimize occurrence of different types of errors and to improve competency levels in each of the three subjects?

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 Education Authorities, Rural/Urban schools and Male/Female students with respect to students' performance on tests and sub-tests measuring general and specific competencies in each subject.

Results of the Math Tests are presented first, then those of the Social Studies followed by the results of the Arabic Test.

I. Math Tests

Psychometric Properties of the Math Tests

The math test had four forms, three new and one old. Since each form measured performance on a different general competency and its constituents (specific competencies) instead of using the terms Form1, Form2, Form3 we have used the terms Test1, Test2, Test3, respectively, in the following description. The old test was used only for studying trends in progress, therefore, its results are not included in this report.

Math Test1 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). Test1, was administered to 1472 students.

The homogeneity or internal consistency reliability index (Coefficient Alpha) of this 30-item test was 0.89 which is considered very good. The item difficulty indices ranged from .08 to .77 with a mean value of .38. The discrimination power of all the items was very good. The Corrected Item Total Correlation Coefficients ranged from .30 to .58.

Math Test2 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.

The internal consistency reliability was quite high ($\text{Alpha} = .87$); difficulty indices ranged from 0.10 to 0.81 with a mean value of 0.48. Discrimination index values ranged from .09 to .65; all of them were above .31 except one which was .09. Positive discrimination index above .30 is considered very good for such tests.

Math Test3 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 or operation while 12 required two or more steps or operations to solve a problem. Alpha coefficient of this 20-item test was 0.93. Item mean score varied from .04 to .77 and discrimination indices ranged from 0.10 to 0.59.

Students' Performance on Math Tests

Test1 : General Competency of Understanding and Application of Procedures

The national average percent correct score on procedures' knowledge and applications is only 37.9 with a minimum score of zero% and maximum of 100% correct. This means that an average 4th 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 you 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.

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

Level 1 (very low competency or no 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

Figure1 shows the percentage of students falling in each category of competency.

Figure1
National Competency Level in Knowledge
And Application of Mathematical Procedures

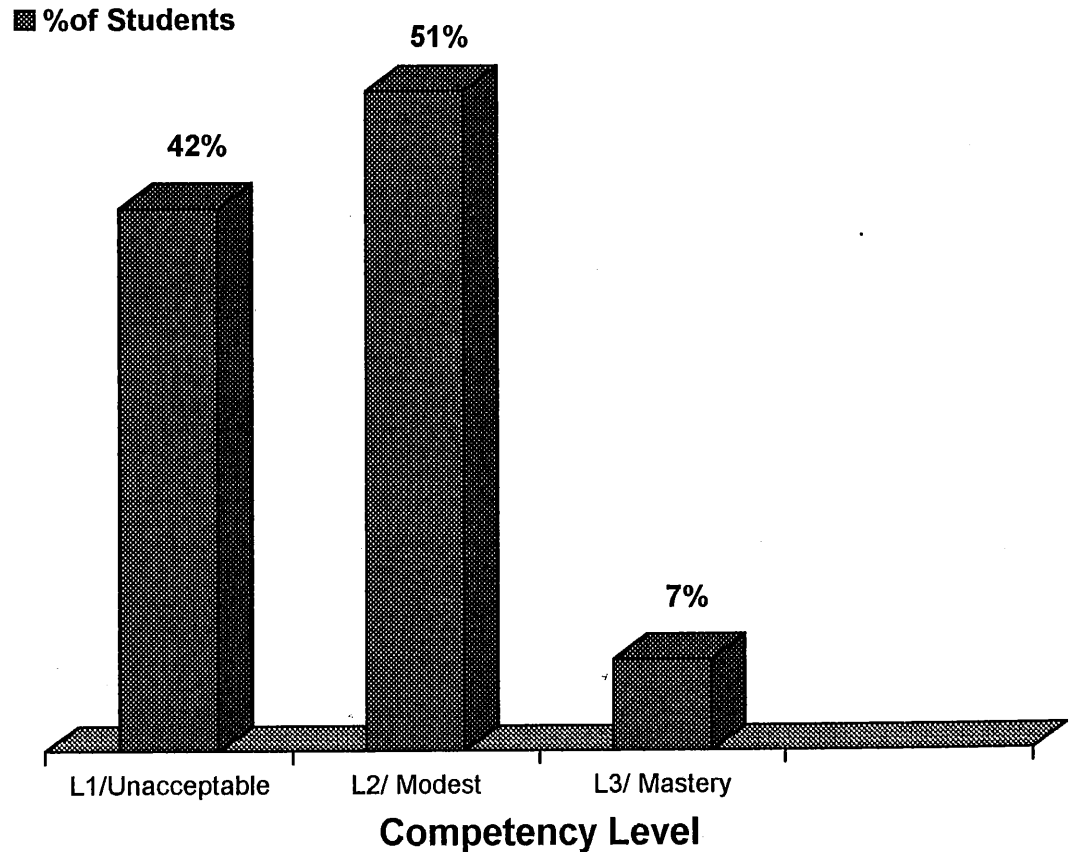


Figure1 doesn't present a pleasant picture. Even when we have set mastery level at 70% only 7% of students have reached or crossed the mastery boundary line leaving the vast majority of 93% below the mastery cut-score of 70% correct. 51% of the Grade4 students have obtained scores, which are more than 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 4th grade students have acquired little competency in the knowledge and application of simple arithmetical procedures.

Table9 presents the national achievement levels on the subcompetencies defined in terms of understanding and application of procedures to specific contents and operations such as properties of Natural Numbers; Addition, Subtraction Multiplication, and Division of Natural Numbers; Fractions; Decimals; and Measurement and Geometry.

The national Mean, Standard Deviation and Standard Error of the Mean and 95% Confidence Interval for the mean of each competency are presented in Table9.

Table 9
Competency Levels of Grade4 Students on
Math Test 1 (Understanding and Application of Procedures)

	N	Mean	SD	Std. Er.	95% CI	
TOTMP	1472	37.93	21.56	.56	36.82	39.03
NUMBERP	1472	33.45	29.20	.76	31.96	34.94
NUMADP	1472	71.43	37.39	.97	69.52	73.35
NUMSUBP	1472	44.87	33.08	.86	43.18	46.56
NUMMULP	1472	19.87	25.53	.67	18.56	21.18
NUMDIVP	1472	21.37	31.83	.83	19.74	22.99
FRACTIOP	1472	46.64	29.62	.77	45.12	48.15
GEOUETVP	1472	12.96	9.13	.24	12.49	13.43
DECIMALP	1472	29.66	25.39	.66	28.37	30.96

Following the same line of interpretation as for the total score (TOTMP) Test1, we see that the national average percent correct score on understanding and application of procedures related to properties of natural numbers is 33.5%.

In the content areas of Addition, Subtraction, Multiplication and Division the mean percent correct scores are 71.4, 44.9, 19.9, and 21.4 respectively.

The mean score on Fractions is 46.6%, on Decimals it is 29.7% and on Geometry and Measurement it is only 13%.

Education Authority Comparisons

Students were sampled from each of the four education authorities: MOE, MOD, UNRWA and Private. It is of interest to know whether, schools under any one authority are more effective than schools under

any other authority, where effectiveness is measured by student achievement level.

Education Authority means on the total Test1 score, and on each subtest score along with their Standard Errors and 95% Confidence Intervals are presented in Table10.

Table 10
Comparative Achievement of Different Education Authorities on
Math Test 1(Understanding and Application of Procedures)

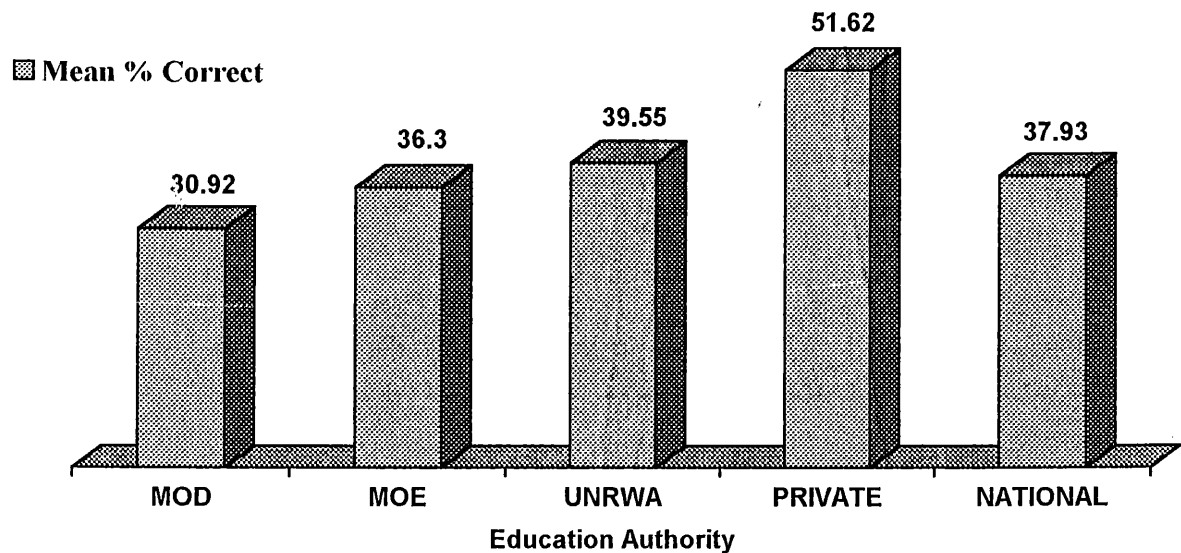
Competency	Authority	N	Mean	SD	Std. Err	95% CI	
TOTMP	MOE	1169	36.31	20.92	.61	35.11	37.51
	MOD	40	30.92	19.18	3.03	24.78	37.05
	UNRWA	119	39.55	23.21	2.13	35.34	43.77
	Private	144	51.62	20.83	1.74	48.19	55.05
NUBERP	MOE	1169	31.75	28.54	.83	30.12	33.39
	MOD	40	24.50	21.48	3.39	17.63	31.36
	UNRWA	119	33.78	28.49	2.61	28.60	38.95
	Private	144	49.44	30.83	2.57	44.36	54.52
NUMADP	MOE	1169	69.71	38.24	1.12	67.52	71.91
	MOD	40	70.00	37.21	5.88	58.10	81.90
	UNRWA	119	72.27	36.07	2.21	65.72	78.81
	Private	144	85.06	27.79	2.22	80.92	89.64
NUMSUBP	MOE	1169	42.54	32.79	.96	40.66	44.42
	MOD	40	38.13	34.89	5.52	26.97	49.28
	UNRWA	119	47.06	33.69	3.09	40.94	53.18
	Private	144	63.89	28.01	2.33	59.27	68.50
NUMMULP	MOE	1169	18.24	24.07	.70	16.86	19.62
	MOD	40	13.13	21.17	3.34	6.35	19.90
	UNRWA	119	22.06	28.78	2.64	16.83	27.28
	Private	144	33.16	30.84	2.57	28.07	38.24
NUMDIVP	MOE	1169	19.25	30.29	.89	17.51	21.00
	MOD	40	16.25	28.62	4.52	7.10	25.40
	UNRWA	119	27.73	37.22	3.41	20.96	34.49
	Private	144	34.72	36.10	3.00	28.76	40.67
FRACTIONP	MOE	1169	44.80	28.96	.85	43.15	46.47
	MOD	40	39.50	27.36	4.33	30.75	48.25
	UNRWA	119	48.23	32.15	2.95	42.40	54.07
	Private	144	92.22	28.79	2.40	57.48	66.96
DECIMALP	MOE	1169	28.45	24.58	.72	27.13	29.95
	MOD	40	23.33	24.11	3.81	15.62	31.04
	UNRWA	119	31.23	25.45	2.34	26.60	35.87
	Private	144	39.24	29.73	2.48	34.33	44.13
GEOMETRP	MOE	1169	12.96	9.12	.26	12.16	13.21
	MOD	40	10.50	9.32	1.50	7.52	13.50
	UNRWA	119	13.28	9.21	.84	11.60	15.05
	Private	144	15.63	8.67	.72	14.20	17.05

It is clear from Table10 that the performance of the private schools' students is the highest of all. The difference between the next highest mean score (that of UNRWA) and private schools' mean score is 12 points; much higher than the differences between second and third, and third and fourth which is 4 points (between UNRWA and MOE) and 5 points (between MOE and MOP), respectively. Another way of comparing the relative performance of each education authority is to compare it with the national performance (38% correct). Private schools' mean score (52%) is 14 points higher than the national norm; whereas, the performance of UNRWA (40%) is only 2 points higher than the national norm. In contract, the performance of MOE and MOD schools is ,respectively, 2 points and 7 points lower than the national norm.

If we rank order the means of the four education authorities on each subtest, we see private schools come consistently first, UNRWA second, MOE third and MOD the last. Figure2 illustrates the comparative means of the four educational authorities along with the national mean.

The mean competency scores of four education authorities along with the national mean are shown in Figure2.

Figure2
Comparative Performance of Education Authorities on
The General Competency of Understanding and
Application of Mathematical Procedures



Rural Schools Vs Urban Schools

It is common knowledge that in Jordan, generally, urban schools have several advantages over rural schools.

Table11 presents the means of rural and urban schools on the total test, as well as, on each of its 8 components. Other statistics like difference between the means (Urban-Rural), the t-value, and 2-tailed Significance of the difference computed for unequal variances are also given in Table11.

Table 11
Comparative Performance of Urban/Rural School Students
on Math Test1 (Understanding and Application of Procedures)

Competency	Location	N	Mean	Std. D.	Std. Err.	Diff. U-R	t-Value	Sig. 2-tailed
TOTMP	Urban	815	39.12	21.93	.77	2.68	2.38	.02
	Rural	657	36.44	21.00	.82			
NUBERP	Urban	815	34.48	29.53	1.03	2.30	1.52	.13
	Rural	657	32.18	28.50	1.11			
NUMADP	Urban	815	72.27	37.40	1.31	1.87	.96	.34
	Rural	657	70.40	37.38	1.46			
NUMSUBP	Urban	815	46.69	32.95	1.15	4.07	2.35	.02
	Rural	657	42.62	33.14	1.29			
NUMMULP	Urban	815	20.98	26.29	.92	2.49	1.88	.06
	Rural	657	18.49	24.51	.96			
NUMDIVP	Urban	815	22.88	32.61	1.14	3.40	2.05	.04
	Rural	657	19.48	30.76	1.20			
FRACTIOP	Urban	815	48.00	30.02	1.05	3.04	1.97	.05
	Rural	657	44.96	29.06	1.13			
DECIMALP	Urban	815	30.84	26.03	.91	2.63	1.99	.05
	Rural	657	28.21	24.52	.96			
GEOMETRP	Urban	815	13.03	9.47	.33	.29	.61	.54
	Rural	657	12.80	8.70	.34			

From the significance column of Table11 we find that on the total test score urban school students' performance is somewhat better than the performance of rural school students. The difference is only 2.7 percentage points which, however, is statistically significant at .05 level ($p < .02$).

According to the 8 subtests, Urban-Rural difference in student achievement is statistically significant ($\alpha < .05$) on three out of the eight. The difference occurs on Number Subtraction, Number Division, and Decimals. Despite the fact that the highest performing private schools and relatively better performing UNRWA schools are situated in urban areas, the difference between urban and rural school students performance on the measured competencies is only marginal.

Gender Comparisons on Math Test1

Plethora of research from all over the world has generally shown that male students outperform their female cohorts on math tests.

On the contrary, almost all the research studies conducted in Jordan over the past 10 years have consistently reported either no significant differences or female students' performance slightly better than the performance of their male cohorts.

The results of the Male-Female comparison on the test of competencies in Understanding and Application of Math Procedures are presented in Table12.

Table 12
Comparative Performance of Male/Female Students
On Math Test1 (Understanding and Application of Procedures)

Competency	Student Sex	N	Mean	Std. D.	Std. Err.	Diff. M-F	t-Value	Sig. 2-tailed
TOTMP	Male	753	37.13	21.77	.79	-1.63	-1.4	.15
	Female	719	38.76	21.31	.80			
NUBERP	Male	753	32.64	28.96	1.06	-1.65	-1.09	.28
	Female	719	34.30	29.22	1.10			
NUMADP	Male	753	70.92	38.04	1.39	-1.06	-.54	.59
	Female	719	71.98	36.71	1.36			
NUMSUBP	Male	753	43.89	33.31	1.21	-2.01	-1.16	.25
	Female	719	45.90	32.84	1.22			
NUMMULP	Male	753	19.46	25.17	.92	-.85	-.64	.52
	Female	719	20.30	25.92	.97			
NUMDIVP	Male	753	20.19	31.65	1.15	-2.41	-1.46	.15
	Female	719	22.60	32.00	1.19			
FRACTIOP	Male	753	46.77	30.34	1.11	.263	.17	.86
	Female	719	46.51	28.87	1.08			
DECIMALP	Male	753	28.46	26.12	.95	-2.46	-1.86	.06
	Female	719	30.92	24.56	.34			
GEOMETRP	Male	753	12.78	9.35	.34	-.38	-.80	.42
	Female	719	13.16	8.90	.33			

As the significance column in Table12 clearly shows neither on the total test nor on any of its eight components the difference was statistically significant at the .05 Alpha level of significance. On the other hand, female students' mean score was consistently higher than male students' mean on the total test, as well as, on all the components except one on which it was exactly the same for both groups.

Math Test2: General Competency in Mathematical Thinking and Communication

Math Test2 was administered to a national subsample of 1435 pupils. It measured 4th grade pupils' general competency in Mathematical Thinking and Communication. The general competency was ramified into three subcompetencies: Mathematical Thinking (6 items), Communication and Thinking (14 items), and Mathematical Communication (4 items). The first two subcompetencies were applied to three content areas viz., Natural Numbers, Fractions, and Measurement and Geometry; but the last (Communication) was applied to only the former two topics viz., Natural Numbers and Fractions. Along the three content areas, 9 items belonged to Natural Numbers, 4 items to Fractions, and 11 items to Measurement and Geometry.

For each student a percent correct score was computed on the total test, as well as, on each subtest representing a subcompetency in each content area.

The mean percent correct scores on all the subtests including the total test, the Standard Deviations, the Standard Errors of the Means, and the 95% Confidence Intervals for means are reported in Table13.

Table 13
National Level Performance of Grade4 Pupils on Math Test 2:
Mathematical Thinking and Communication (N=1435)

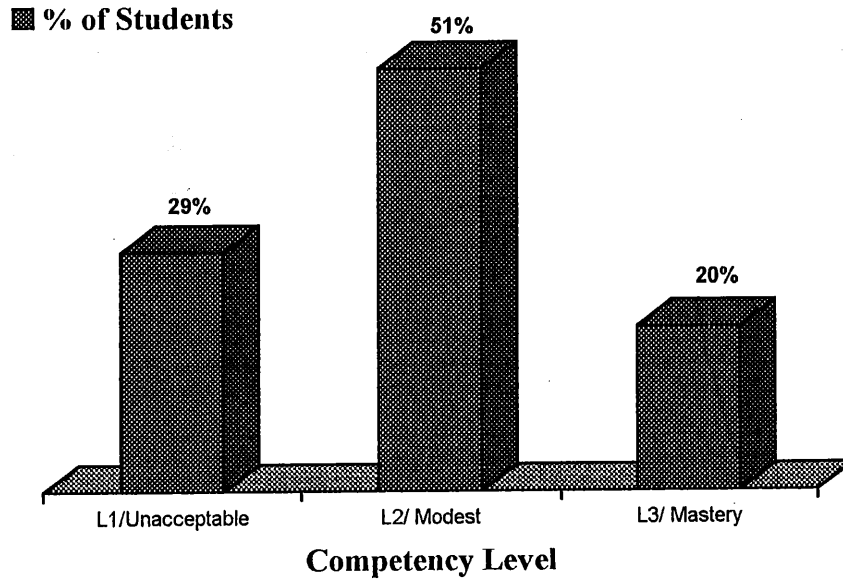
	Mean	Std. Err.	Std. D.	95% CI	
TOTMP	46.34	23.31	.62	45.14	47.55
THINKP	61.92	28.57	.75	60.44	63.40
COMTHINP	37.14	24.14	.64	35.89	38.39
COMP	55.21	32.62	.86	53.52	56.90
NUMPERP	46.02	28.37	.75	44.55	47.49
FRACTIOP	58.00	30.25	.80	56.43	59.56
GEOMETRP	42.38	22.86	.60	41.19	43.56
NUMTHINP	58.18	38.55	1.08	56.19	60.18
NUMCOTHP	38.31	30.65	.81	36.73	39.90
NUMCOMP	53.10	39.25	1.04	51.07	55.13
GEOTHINP	58.05	32.85	.87	56.35	59.75
GECOTHP	36.50	23.67	.62	35.27	37.72
FRACOTHP	80.98	39.26	1.04	78.94	83.02
FRACOMP	57.31	37.84	.10	55.35	59.27

The average percent correct score on the total test shows that the average student could answer only 46.3% test items correctly. With a Standard Deviation of 23 the percent correct score ranged from zero percent to one hundred percent, i.e., there were students who could not get any item right and on the other extreme there were students who got all the 24 items right. The national average competency level of 46% is far from acceptable.

Regarding specific competencies, the average competency level on Mathematical Thinking is 62%; on Communication and Thinking 37%; and on Communication it is 55%.

According to the three levels of competency 20% of the sampled Grade4 students were classified in the L3 (mastery level) category (they scored 70% or above 70% marks), 51% fell in the L2 (partial competency) category, and the remaining 29% of them fell in L1, category, that is, they didn't acquire an acceptable level of competency in the general competency of Mathematical Thinking and Communication. The distribution of students along the three levels of competency is depicted in Figure3.

Figure 3
National Competency Level in the General Competency
of Mathematical Thinking and Communication



Comparative Performance Among Education Authorities

Table14 presents the Means, Standard Deviations, and 95% Confidence Intervals for the Means on the total test score and on subtest scores for each Education Authority..

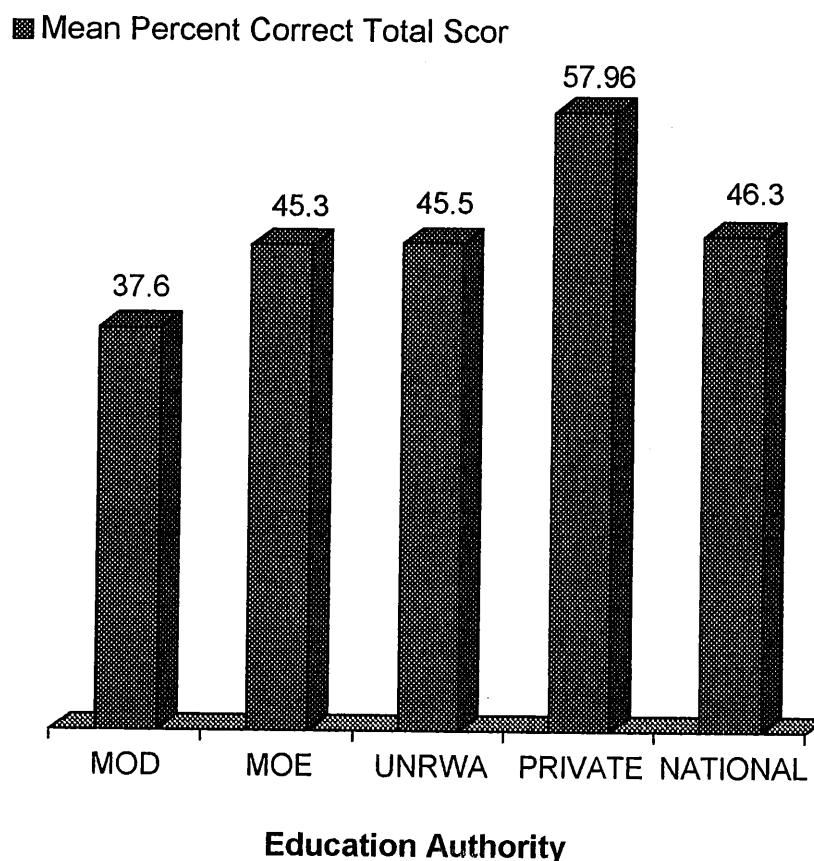
Table 14
Comparative Performance of Education Authorities in
The General Competency of Mathematical Thinking and
Communication (Test 2)

Competency Authority		N	Mean	Std. D.	Std. Err.	95% CI	
TOTMP	MOE	1144	45.35	23.32	.69	44.00	46.70
	MOD	38	37.61	21.12	3.43	30.67	44.55
	UNRWA	18	45.52	24.36	2.24	41.07	49.96
	Private Schools	135	57.96	19.33	1.66	54.67	61.25
THINKP	MOE	1144	61.03	28.81	.85	59.36	62.70
	MOD	38	53.07	33.07	5.37	42.20	63.94
	UNRWA	18	59.60	27.51	2.53	54.59	64.62
	Private Schools	135	73.95	22.55	1.94	70.11	77.79
NUMTHINP	MOE	1144	57.12	38.44	1.14	54.89	59.35
	MOD	38	53.95	44.08	7.15	39.46	68.43
	UNRWA	18	52.54	40.22	3.70	45.21	59.87
	Private Schools	135	73.33	32.78	2.82	67.75	78.91
NUMCOM	MOE	1144	50.83	39.20	1.16	48.56	53.10
	MOD	38	39.47	28.85	4.68	29.99	48.96
	UNRWA	18	54.66	39.76	3.66	47.41	61.91
	Private Schools	135	74.81	34.49	2.97	68.94	80.69
NUMCOTHP	MOE	1144	37.05	30.41	.90	35.28	38.81
	MOD	38	25.79	27.08	4.39	16.89	34.69
	UNRWA	18	37.80	31.03	2.86	32.14	43.45
	Private Schools	135	53.04	29.15	2.51	48.08	58.00
NUMBERP	MOE	1144	44.57	28.08	.83	42.94	46.20
	MOD	38	35.09	25.78	4.18	26.62	43.56
	UNRWA	18	44.82	29.45	2.71	39.45	50.19
	Private Schools	135	62.39	24.96	2.15	58.14	66.63
GEOTHINP	MOE	1144	57.34	33.15	.98	55.42	59.27
	MOD	38	46.49	35.12	5.70	34.95	58.04
	UNRWA	18	58.19	30.87	2.84	52.56	63.82
	Private Schools	135	67.16	29.64	2.55	62.11	72.21
GEOMETRP	MOE	1144	41.96	23.06	.68	40.62	43.30
	MOD	38	34.45	21.67	3.51	27.32	41.57
	UNRWA	18	42.45	23.29	2.14	38.20	46.70
	Private Schools	135	48.08	20.08	1.73	44.66	51.50
GECOTHP	MOE	1144	36.19	23.86	.71	34.80	37.57
	MOD	38	29.93	22.04	3.58	22.69	37.18
	UNRWA	18	36.55	24.69	2.27	32.05	41.05
	Private Schools	135	40.93	21.02	1.81	37.35	44.50
FRACTIOP	MOE	1144	56.42	30.44	.90	54.66	58.19
	MOD	38	51.97	28.70	28.70	42.54	61.41
	UNRWA	18	55.51	31.20	31.20	49.82	61.20
	Private Schools	135	75.19	21.92	21.92	71.45	78.92

Competency Authority		N	Mean	Std. D.	Std. Err.	95% CI	
FRACOTHP	MOE	1144	79.90	40.10	1.19	77.57	82.22
	MOD	38	71.05	45.96	7.46	55.95	86.16
	UNRWA	18	77.97	41.62	3.83	70.37	85.55
	Private Schools	135	95.56	20.68	1.78	92.03	99.07
FRACOMP	MOE	1144	54.72	38.10	1.13	52.51	56.93
	MOD	38	59.21	36.49	5.92	47.22	71.21
	UNRWA	18	58.05	38.66	3.56	51.00	65.10
	Private Schools	135	78.15	27.73	2.39	73.43	82.87
COMTHINP	MOE	1144	36.51	24.16	.71	35.11	37.91
	MOD	38	27.63	21.02	3.41	20.72	34.54
	UNRWA	18	36.38	25.28	2.33	31.77	40.99
	Private Schools	135	45.82	21.67	1.87	42.13	49.51
COMP	MOE	1144	52.78	32.51	.96	50.89	54.66
	MOD	38	49.34	26.94	4.37	40.49	58.20
	UNRWA	18	56.36	33.68	3.10	50.26	62.50
	Private Schools	135	76.48	25.69	2.21	72.11	80.90

The mean competency scores of education authorities along with the national mean on the general competency of Mathematical Thinking and Communication are displayed by Figure4.

Figure 4
Comparative Performance of Education Authorities
in the General Competency of
Mathematical Thinking and Communication



On the general competency (total test score) as well as on each subcompetency private schools have done by far the best; their 58% average on the total test is 13 points higher than both UNRWA's and MOE's averages (45.5% and 45.3% respectively) and 20 points higher than that of MOD schools.

On most competencies also both UNRWA and MOE have performed worse than Private but better than MOD while there seem to be no significant differences between UNRWA and MOE.

Urban/Rural Comparisons

Table 15, presents the Means, Standard Deviations, differences between means, t-values and significance levels for the differences of the groups of urban and rural school students on the total test, as well as, on all the subtests.

Table 15
Comparative Performance of Urban and Rural Schools' Students
in the General Competency of Mathematical Thinking and
Communication (Math Test 2)

		N	Mean	Std.D.	Std. Err.	Diff. (U-R)	t-Value	Sig. 2-tailed
TOTMP	Urban	812	47.54	23.49	.82	2.75	2.22	.027
	Rural	623	44.79	23.01	.92			
THINKP	Urban	812	63.22	28.14	.99	2.10	1.97	.050
	Rural	623	60.22	29.05	1.16			
NUMTHINP	Urban	812	60.28	38.33	1.35	4.83	2.35	.019
	Rural	623	55.46	38.70	1.55			
NUMCOMP	Urban	812	55.42	39.90	1.40	5.34	2.57	.010
	Rural	623	50.08	38.20	1.53			
NUMCOTHP	Urban	812	40.15	30.77	1.10	4.22	2.60	.009
	Rural	623	35.92	30.36	1.22			
NUMBERP	Urban	812	48.02	28.69	1.01	4.61	3.07	.002
	Rural	623	43.41	27.75	1.11			
GEOTHINP	Urban	812	58.91	32.50	1.14	1.98	1.13	.260
	Rural	623	56.93	33.30	1.33			
GEOMETRP	Urban	812	43.11	22.76	.80	1.70	1.40	.163
	Rural	623	41.41	23.00	.92			
GECOTHP	Urban	812	37.19	23.81	.84	1.60	1.27	.204
	Rural	623	35.59	23.48	.94			
FRACTIOP	Urban	812	58.62	30.58	1.07	1.44	.90	.371
	Rural	623	57.18	29.83	1.20			
FRACOTHP	Urban	812	82.10	38.42	1.35	2.40	1.14	.253
	Rural	623	79.61	40.32	1.62			
FRACOMP	Urban	812	57.76	38.16	1.34	1.02	.51	.613
	Rural	623	56.74	37.44	1.50			
COMTHINP	Urban	812	38.23	24.25	.85	2.52	1.96	.050
	Rural	623	35.71	23.93	.96			
COMP	Urban	812	56.59	33.54	1.77	3.18	1.85	.065
	Rural	623	53.41	31.31	1.25			

In all, 14 separate comparisons were made. In each of the 14 cases the observed mean score of the urban school students was lightly better than that of the rural school students. Only seven of them were statistically significant at 5% level of significance or better. The seven variables on which urban school students significantly outperformed their rural school counterparts included: the general competency in Mathematical Thinking and Communication, and six subcompetencies including Mathematical Thinking; Mathematical Thinking applied to Natural Numbers; the content area of Numbers; Thinking and Communication; Communication applied to Numbers; and Communication and Thinking applied to Numbers.

Gender Differences on Mathematical Thinking and Communication Competencies

Table16 presents the mean scores of male and female groups, on the total test scores and on thirteen subtest scores. The Standard Deviations, Standard Errors, differences between means, the t-value and the 2-tailed significance of the difference are also reported in Table16.

A glance over the difference column in Table16 shows that all the 14 differences are negative, that is, all of them are in favor of female students. The differences are so small, however, that none of them is statistically significant at 5% level of significance despite the relatively large sample size (733 males and 702 females). Therefore, in contrast with the findings of gender studies in other countries, we found no gender differences on Mathematical Thinking and Communication in the population of Grade4 students in Jordan.

Table 16
Gender Comparisons on Mathematical Thinking
and Communication Competencies

Competency Student Sex	N	Mean	Std. D.	Std. Err.	Diff. M-F	t-Value	Sig. 2-tailed
TOTMP Male	733	45.73	23.82	.88	-1.24	-1.01	.313
Female	702	40.97	22.77	.85			
THINKP Male	733	61.89	29.50	1.08	-.05	-.033	.973
Female	702	61.94	27.57	1.04			
NUMTHINNP Male	733	59.27	39.32	1.45	-2.22	1.09	.274
Female	702	57.05	37.72	1.42			
NUMCOMP Male	733	53.06	39.73	1.46	-.06	-.031	.975
Female	702	53.13	38.75	1.46			
NUMCOTHP Male	733	37.38	30.50	1.12	-1.9	-1.18	.239
Female	702	39.28	30.79	1.16			
NUMBERP Male	733	45.73	28.43	1.05	-.58	-.39	.699
Female	702	46.31	28.32	1.06			
GEOTHINP Male	733	57.84	32.65	1.20	-.428	-.24	.810
Female	702	58.26	33.07	1.24			
GEOMETRP Male	733	41.75	23.10	.85	-1.26	-1.05	.296
Female	702	43.01	22.60	.85			
GECOTHP Male	733	35.72	23.86	.88	-1.58	-1.26	.207
Female	702	37.30	23.44	.88			
FRACTIOP Male	733	56.68	31.22	1.15	-2.68	-1.68	.093
Female	702	59.36	29.17	1.10			
FRACOTHP Male	733	79.26	40.56	1.49	-3.50	-1.69	.091
Female	702	82.76	37.79	1.42			
FRACOMP Male	733	56.00	38.65	1.42	-2.69	-1.35	.178
Female	702	58.68	36.94	1.39			
COMTHINP Male	733	36.29	24.39	.90	-1.71	-1.35	.178
Female	702	38.01	23.84	.90			
COMP Male	733	54.53	33.60	1.24	-1.38	-.800	.424
Female	702	55.91	31.55	1.19			

Math Test3: Problem Solving Competency

Math Test3 measured the general competency of 4th grade students in solving mathematical problems. Two competency levels were included: level one problems required one step or a single operation to solve the problem; level two problems required two or more steps or operations to solve the problem. Problems were related to two content areas or topics: Numbers and, Measurement and Geometry. The test included 20 questions; 8 one step problems and 12 two or more steps problems. Along the topics, 14 problems belonged to Numbers and 6 to Measurement and Geometry.

Table17 presents the national Means, Standard Deviations, Standard Errors, and 95% Confidence Intervals, of the total percent correct score and of the percent correct subtests' scores of the 1396 Grad4 pupils who took Math Test3.

Table 17
National Performance on Problem Solving General Competency and Sub-Competencies (N=1396)

	Mean	Std. D.	Std. Err.	95% CI	
TOTMP	29.55	20.06	.54	28.50	30.61
STEP1P	39.21	24.74	.66	37.91	40.51
STEP2P	23.11	19.86	.53	22.07	24.16
NUMBERP	34.83	22.66	.61	33.64	36.02
GEOMETRP	17.24	19.46	.52	16.22	18.26
GEOSTE1P	21.42	31.67	.85	19.76	23.08
GEOSTE2P	15.15	19.66	.53	14.12	16.18
NUMSTE1P	45.14	26.11	.70	43.77	46.51
NUMSTE2P	27.10	24.19	.65	25.83	28.37

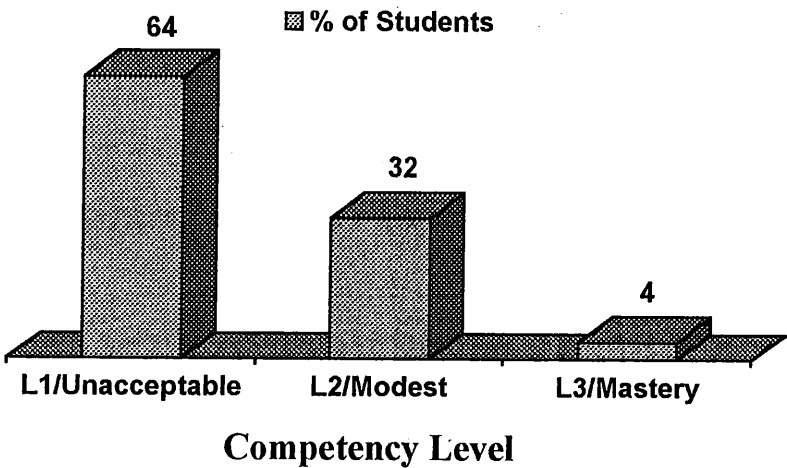
The percent correct score on the total test and on each subtest ranged from zero% to 100%.

The average percent correct score on the total test was 29.6%. This spells dismay. Below thirty percent competency in problem solving presents a gloomy picture of the teaching and learning of mathematics in the primary grades of Jordanian schools.

In terms of the three mastery levels of this general competency we can see from Figure5 that only nearly 4% of all students who took this form have attained the mastery level (L3) by obtaining 70% or higher

marks; 32% have attained the modest level (L2) and nearly two-thirds of them (64%) have failed to reach acceptable level of competency in Problem Solving, especially when developing problem solving skills is one of the most touted objectives of the educational reform.

Figure 5
National Competency in Mathematical Problem Solving



Among subcompetencies the national means on one step problems and two or more steps problems, respectively are 39% correct and 23% correct.

On specific topics the average percent correct score is 35% on Numbers and only 17% on Measurement and Geometry. Across the four content by skill cross sections, performance on One Step Numbers is relatively the best (45%), Two or More Steps Numbers, with the mean of 27%, comes next. National mean on One Step Measurement and Geometry is 21% while on Two or More Steps Measurement and Geometry problems the mean is only 15%. It is disappointing to report that according to students' performance on this test the average Grade4 student in Jordan can solve only 15 problems out of 100 problems from the topic of Measurement and Geometry involving two as more steps solutions.

Education Authority Comparisons on Math Test3: Problem Solving

Table18 presents the Means, Standard Deviations, Standard Errors of the Means, and 95% Confidence Intervals for each Education Authority on the whole test and on each of its components.

Table 18
Comparative Performance of Education Authorities in the
General and Specific Problem Solving Competencies

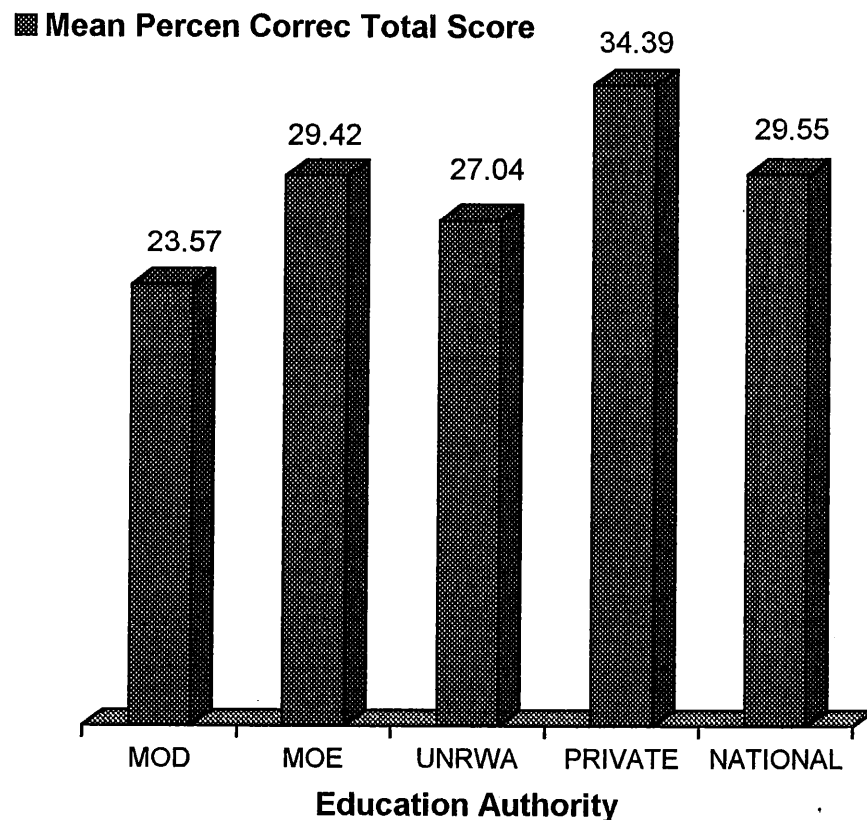
Competence Authority	N	Mean	Std. D.	Std. Err.	95% CI	
TOTMP MOE	1113	29.42	20.06	.60	28.24	30.60
MOD	35	23.57	14.88	2.52	18.46	28.68
UNRWA	115	27.04	20.41	1.90	23.27	30.81
Private	133	34.40	20.14	1.75	30.94	37.85
STEP1P MOE	1113	38.65	24.46	.73	37.21	40.08
MOD	35	32.14	20.63	3.49	25.06	39.23
UNRWA	115	36.52	24.90	2.32	31.92	41.12
Private	133	48.12	26.09	2.26	43.65	52.60
STEP2P MOE	1113	23.27	20.04	.60	22.09	24.45
MOD	35	17.86	14.17	2.39	12.99	22.72
UNRWA	115	20.72	19.85	1.85	17.06	24.39
Private	133	25.25	19.35	1.68	21.93	28.57
NUMBERP MOE	1113	34.56	22.48	.67	33.24	35.88
MOD	35	26.94	16.26	2.75	21.35	32.52
UNRWA	115	32.55	24.03	2.24	28.11	36.99
Private	133	41.14	23.30	2.02	37.14	45.13
GEOMETRP MOE	1113	17.43	19.76	.59	16.27	18.59
MOD	35	15.71	16.64	2.81	10.00	21.42
UNRWA	115	14.20	17.56	1.63	10.96	17.44
Private	133	18.67	19.14	1.66	15.39	21.95
GEOSTE1P MOE	1113	20.66	31.45	.94	18.86	22.51
MOD	35	14.29	28.62	4.83	4.45	24.11
UNRWA	115	21.74	31.85	2.97	15.86	27.62
Private	133	29.32	33.22	2.88	23.66	35.02
GEOSTP2P MOE	1113	15.81	20.07	.60	14.63	16.99
MOD	35	16.43	15.98	2.70	10.94	21.91
UNRWA	115	10.43	16.88	1.57	7.32	13.55
Private	133	13.36	18.86	1.64	10.11	16.58
NUMSTE1P MOE	1113	44.64	25.82	.77	43.12	46.16
MOD	35	39.00	21.98	3.72	30.54	45.64
UNRWA	115	41.45	25.97	2.42	36.65	46.25
Private	133	54.39	27.74	2.41	49.63	59.14
NUMSTE2P MOE	1113	27.00	24.21	.73	25.58	28.42
MOD	35	18.57	16.15	2.73	13.02	24.12
UNRWA	115	25.87	25.74	2.40	21.11	30.62
Private Schools	133	31.20	23.87	2.07	27.10	35.30

Private schools' mean on the whole test (34.4%) is the highest of all. MOE schools with a mean of 29.4% take the next place, UNRWA

with a mean of 27% comes third and MOD schools with the lowest mean score of 23.6% come the last.

Figure6 presents the comparative performance levels of the four education authorities and of the nation as a whole on the general competency of Problem Solving.

Figure 6
Comparative Performance of Education Authorities
on the General Competency in Problem Solving



This rank order pattern persists for all other subtests of the Test3. One, however, cannot help noticing that MOE schools' performance is consistently better than that of UNRWA schools. This is a change from the patterns on Test1 and Test2 wherein UNRWA's means were generally slightly higher than those of the MOE. One cannot draw any conclusions from such data because the differences between the means of MOE and UNRWA were seldom significant, statistically.

Urban/Rural Schools Comparisons

A subsample of 1396 grade4 students took Math Test3; 794 students represented Urban schools and 602 represented Rural schools.

The results of the t-tests conducted on all the variables to compare the mean scores of urban school students and rural school students are presented in Table19.

Table 19
Comparison of Urban and Rural School Students' Performance
on the General Competency in Problem Solving

Competency	Location	N	Mean	Std. D.	Std. Err.	Diff. M-F	t-Value	Sig. 2-tailed
TOTMP	Urban	794	29.82	20.33	.72	.63	.58	.56
	Rural	602	29.19	19.70	.80			
STEP1P	Urban	794	39.47	24.90	.88	.60	.45	.66
	Rural	602	38.87	24.54	1.00			
STEP2P	Urban	794	23.39	20.17	.72	.65	.61	.54
	Rural	602	22.74	19.45	.79			
NUMBERP	Urban	794	35.09	23.05	.82	.61	.50	.62
	Rural	602	34.48	22.17	.90			
GEOMETRP	Urban	794	17.53	19.44	.69	.67	.63	.53
	Rural	602	16.86	19.50	.79			
GEOSTE1P	Urban	794	22.30	32.07	1.14	2.03	1.19	.24
	Rural	602	20.26	31.13	1.27			
GEOSTE2P	Urban	794	15.14	19.92	.71	-0.02	-.012	.99
	Rural	602	15.16	19.34	.79			
NUMSTE1P	Urban	794	45.19	26.34	.93	.12	.086	.93
	Rural	602	45.07	25.82	1.05			
NUMSTE2P	Urban	794	27.52	24.50	.87	.98	.754	.45
	Rural	602	26.54	23.78	.97			

Group Means, Standard Deviations, Standard Errors of the Means, differences between the group means, the t-values, and 2-tailed unequal sample significance levels of the differences are presented in respective columns of table19.

It is clear from the "Difference" and "Significance" columns in Table19 that the differences between the means are totally negligible and not even a single one of them reaches anywhere near the most liberal level of statistical significance.

Gender Differences on Math Problem Solving

Of the 1396 4th grade students who took this test, 703 were males and 693 females. The performance of male and female students was compared by means of independent unequal sample t-tests conducted on the total test score and on each subtest score.

The results of the t-tests including group Means, Standard Deviations, Standard Errors, Differences between the Means, t-values and 2-tailed significances of the differences are given in Table 20.

Table 20
Gender Comparisons on Problem Solving
Competency and Sub-Competencies

Competency Location	N	Mean	Std. D	Std. Err.	Diff. M-F	t-Value	Sig. 2-Tailed
TOTMP Male	703	31.15	21.89	.82	3.22	3.01	.00
TOTMP Female	693	27.93	17.92	.68			
STEP1P Male	703	40.51	26.50	1.00	2.6	1.97	.04
STEP1P Female	693	37.90	22.77	.86			
STEP2P Male	703	24.92	21.58	.81	3.63	3.94	.00
STEP2P Female	693	21.28	17.78	.68			
NUMBERP Male	703	36.47	24.46	.92	3.30	2.73	.00
NUMBERP Female	693	33.17	20.57	.78			
GEOMETRP Male	703	18.75	21.10	.80	3.05	2.94	.00
GEOMETRP Female	693	15.70	17.53	.67			
GEOSTE1P Male	703	23.68	33.29	1.26	4.56	2.70	.00
GEOSTE1P Female	693	19.12	29.79	1.13			
GEOSTE2P Male	703	16.29	20.84	.79	2.29	2.18	.03
GEOSTE2P Female	693	14.00	18.35	.70			
NUMSTE1P Male	703	46.11	27.70	1.04	1.96	1.40	.16
NUMSTE1P Female	693	44.16	24.37	.93			
NUMSTE2P Male	703	29.23	26.01	.98	4.30	3.33	.00
NUMSTE2P Female	693	24.93	21.99	.84			

The "Difference" column in Table20 shows that consistently on each subtest and on the test as a whole male students have scored higher than their female cohorts. The differences, however, though statistically highly significant are not large in magnitude. On the other hand, taking into account the floor effects on the performance of both male and female students and the consistency and statistical significance together these differences take on special importance.

We recall that on Math Test1 (Understanding and Application of Procedures) and Math Test2 (Mathematical Thinking and Communication) there were no significant differences though the trend was in favor of females. When it comes to the general competency of Problem Solving, things have taken a reverse turn; it seems that the Grade4 boys have demonstrated their superiority over girls.

This is an interesting finding which should raise some leading questions.

II. Competencies In Social Studies

In each class about half the students took Test1 and the other half took Test2. Thus, altogether, 2814 students took Test1 and 2786 took Test2.

A competency score was computed for each student on the basis of the percent correct score on the competency items included in the test form he/she took.

Where items related to a general competency were distributed on both tests every student had a percent correct score on the general competency but the competency scores of 2814 students were computed only on those items which were included in Test1 and accordingly, competency scores of 2786 students were based upon those items which were included in Test2.

Psychometric Properties of the Social Studies Tests.

SSTest1 consisted of 38 items, of which 8 items were related to the general competency in Drawing Geographical Maps; 9 item tapped students' Understanding of the Relationship between Mankind, Environment and Resources; 2 items were related to Reading and Interpretation of Maps; 6 items measured Understanding of Systems and Institutions of Modern State; 2 items assessed students' Understanding of Historical Events; 7 items measured Understanding of Geographical, Historical and Economic Concepts; and 4 items were related to students' Understanding of the Concepts of Social Responsibility and Good Citizenship.

The Alpha coefficient (a measure of internal consistency reliability or homogeneity of the test) of this 38-item test was .91; while the difficulty levels as measured by item means varied from 0.02, the lowest, through 0.68, the highest.

SSTest2 consisted of 36 items which tapped five different general competencies: Ability to Draw Geographical Maps (14 items); Understanding Historical Events (4 items); Understanding the Concepts of Social Responsibility and Good Citizenship (4 items); and Understanding Pictographs (10 items).

The Alpha Coefficient of this 36-item test was .91 while the difficulty level as represented by item means varied from 0.08 through 0.68.

In case of the general competencies which had their items confined to only one test form percent correct competency scores of the students who took that particular test were computed on all the items belonging to that competency.

Since each test was administered to half of the students in each class, the mean percent correct scores of some competencies are taken over on a sample of 5600 students, of some on a sample of 2814 students (Test1) and of some on a sample of 2786 students (Test2).

National Performance on Social Studies Test.

Table21 presents the national level performance of Grad4 students on each of the nine general competencies and on the whole test of social studies. For each general competency and for the whole test Table21 gives: the sample size over which the mean and other statistics are based, the percent correct mean score; Standard Error of the Mean; Standard Deviation; Minimum and Maximum scores and 10th, 25th, 50th, 75th and 90th percentiles.

Achievement in the Social Studies was defined in terms of nine general competencies where each competency was measured by a set of items. Thus, the total percent correct score represents overall performance in the subject of Social Studies.

The last Row in Table21 shows that the national Mean Percent Correct Score on the social studies test is only 35.5. This means that an average 4th grade student could answer only 35.5 percent of the items correctly. Another way of looking at the national competency level in this field is that if we define overall competency level on a 100 point scale where each point represents a specific degree on the competency scale then the national competency level is 35.5 degrees or 35.5% level.

The 50th percentile or the Median score of 34.2 shows that about half the population of Grade4 students have competency level below 34.2%. In the same Row of this table if we look in the column of 75th percentile we read the value 52.6. This means 75% of all the 4th grade

students (i.e. three quarters of the students) have scored less than 52.6 out of 100 points.

Now if we divide all the students according to their performance on this test in four quartiles from lowest to highest then the bottom one-fourth of the students have answered less than 17% of the test items correctly; the second quarter from the bottom have obtained scores ranging from 16.7% through 34.2%; the third quarter of the students have scores ranging from 34.2% through 52.6%.

However in the top quarter 15% of the students have obtained from 52.6% through 65.8% correct score. The highest score on the test was 97.2% correct. Only two students in the whole sample have scored above 94.5% correct. The top 10% students in the whole country have obtained scores ranging from 65.8% to 97.2%.

Among the nine general competencies, the best performance average (46.2% correct) is on Competency 7 (Understanding the concepts of social responsibility and good citizenship). The worst performance average (27.8%) is on Competency 2 (Understanding the relationship between mankind, environment and resources).

The means' column in Table21 clearly shows that the national performance on the nine major competencies substantially varies from one competency to another. The worst performance (mean = 28% correct) is on COMPT2 (Understanding the Relationship between Mankind, Environment and Resources) while the highest performance level (mean = 46% correct) is achieved on COMPT7 (Understanding the Concepts of Social Responsibility and Good Citizenship). Figure7 displays the relative levels of national performance on the 9 major competencies in the Social Studies.

Table 21
National Performance on General
Competencies in Social Study

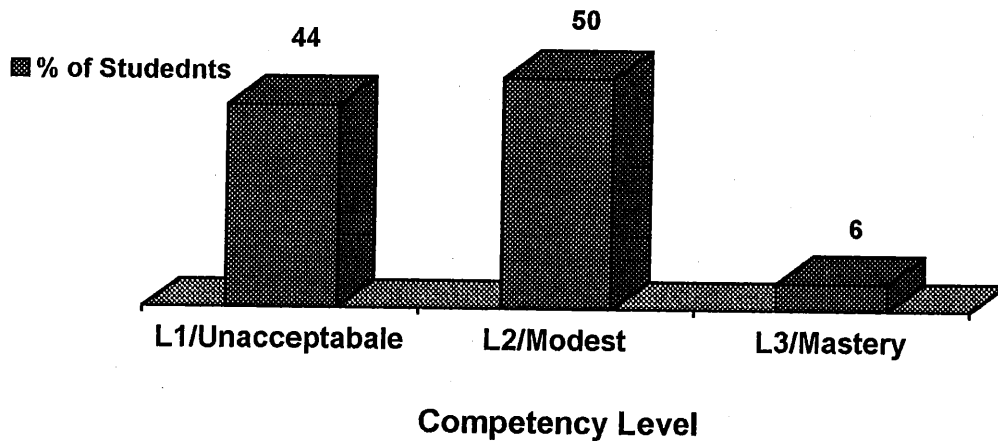
	N	Mean	Std. Div	Std. Err.	Min.	Max.	Percentiles				
							10 th	25 th	50 th	75 th	90 th
COMPT1P	5600	34.23	23.11	.31	.00	100.00	7.14	14.28	35.71	50.00	64.29
COMPT2P	2814	27.83	22.69	.43	.00	100.00	.00	11.11	22.22	44.44	66.67
COMPT3P	2814	29.90	40.84	.77	.00	100.00	.00	.00	.00	50.00	10.00
COMPT4P	2814	30.62	31.69	.60	.00	100.00	.00	.00	16.67	50.00	83.33
COMPT5P	5600	35.97	32.31	.43	.00	100.00	.00	.00	50.00	50.00	75.00
COMPT6P	2814	44.60	29.53	.56	.00	100.00	.00	14.29	42.86	71.43	85.71
COMPT7P	5600	46.23	35.50	.47	.00	100.00	.00	.00	66.67	75.00	100.00
COMPT8P	2786	42.72	35.60	.67	.00	100.00	.00	20.00	40.00	80.00	100.00
COMPT9P	2786	31.22	25.10	.48	.00	100.00	.00	10.00	30.00	50.00	70.00
OTSCR	5600	35.48	21.88	.29	.00	100.00	7.89	16.67	34.21	52.63	65.79

Figure 7
Relative National Performance on the
Nine Major Competencies in the Social Studies

Competency	Average Percent Correct
▪ Concepts of Social Responsibility and Good Utilization	46.2%
▪ Understands Geographical, Historical and Economic Concepts	44.6%
▪ Understand Political Life in Arab and Islamic World	42.7%
▪ Understanding Historical Events	36%
▪ Draw Geographical Maps	34.2%
▪ Interpret pictographs	31.2%
▪ Understand Systems and Institutions of Modern State	30.6%
▪ Reads and Interprets Maps	29.9%
▪ Understands the Relationship between Man Environment and Resources.	27.8%

Now if we define L3, high competency or mastery level, as 70% or above; L2, modest or partial competency level, as greater than 30% and less than 70%, and non-competency as 30% or below; then we can see the disappointing picture. At least 44% of the students are below the acceptable competency level, 50% of the students have attained modest level of competency and only 6% of the Grade4 students in Jordan have attained mastery level of competency in social studies. Figure8, depicts the national competency levels in Social Studies.

Figure8
National Overall Competency Level
In Social Studies



Education Authority Differences in Social Studies Achievement of Grade4 Students

In order to study the differences between the performance of groups of students studying in schools run by different education authorities, we conducted one way ANOVA for each competency score and also for the total test score. Since there are four education authorities pair-wise comparisons were made using Multiple Comparisons test. Herein our main concern is with describing the trend rather than with statistical significance of each difference.

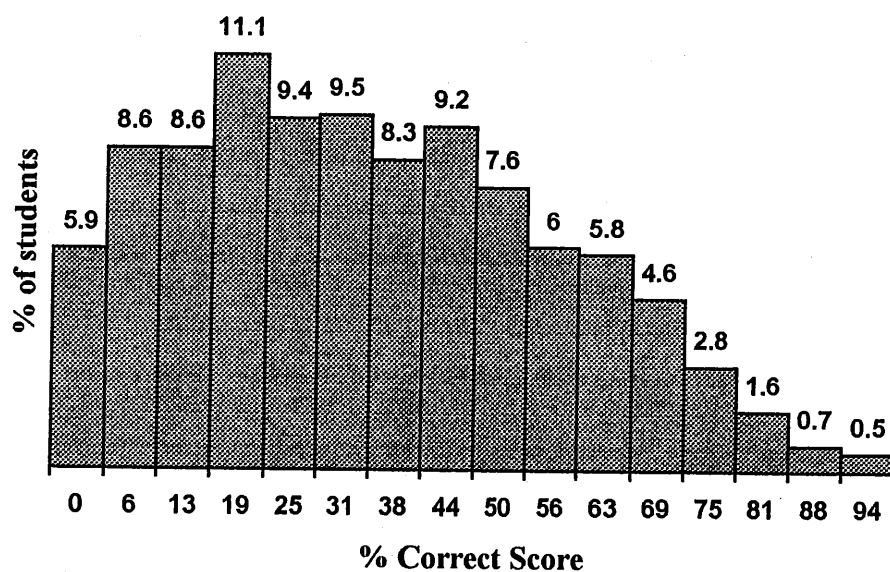
Table22, presents the Means, Standard Deviations, Standard Errors of the Means and 95% Confidence Intervals for the Means of students in each education authority on each competency and on the total test.

Table 22
Comparative Performance of Education Authorities
on General Competencies in Social Studies

Authority		N	Mean	Std. D.	Std. Err.	95% CI	
COMPT1P	MOE	4436	33.54	23.15	.35	32.86	34.23
	MOD	149	28.60	22.43	1.84	24.96	32.23
	UNRWA	462	33.78	23.75	1.11	31.61	35.95
	Private Schools	553	41.67	20.91	.889	39.92	43.42
COMPT2P	MOE	2235	27.01	22.58	.48	26.07	27.95
	MOD	74	16.82	16.68	1.94	12.95	20.68
	UNRWA	232	29.17	22.66	1.49	26.24	32.10
	Private Schools	273	36.43	22.60	1.37	33.73	39.12
COMPT3P	MOE	2235	29.40	40.63	.86	27.71	31.08
	MOD	74	29.73	40.48	4.71	20.35	39.11
	UNRWA	232	25.00	38.08	2.50	20.07	29.93
	Private Schools	273	38.28	43.82	2.65	33.06	43.50
COMPT4P	MOE	2253	30.36	32.00	.68	29.03	31.69
	MOD	74	25.23	31.00	3.60	18.04	32.41
	UNRWA	232	30.68	31.80	2.09	26.56	34.79
	Private Schools	273	34.25	28.91	1.75	30.80	37.69
COMPT5P	MOE	4436	34.88	32.11	.48	33.93	35.82
	MOD	149	24.66	29.49	2.42	19.89	29.44
	UNRWA	462	38.58	32.41	1.51	35.62	41.55
	Private Schools	553	45.61	32.48	1.38	42.90	48.33
COMPT6P	MOE	2253	43.47	29.40	.62	42.25	44.69
	MOD	74	36.87	24.80	2.88	31.13	42.62
	UNRWA	232	44.46	29.11	1.91	40.70	48.22
	Private Schools	273	56.04	29.72	1.80	52.50	59.59
COMPT7P	MOE	4436	45.13	35.79	.54	44.08	46.19
	MOD	149	31.82	33.29	2.28	26.43	37.21
	UNRWA	462	46.77	35.55	1.65	43.52	50.02
	Private Schools	553	58.47	30.49	1.30	55.92	61.02
COMPT8P	MOE	2201	41.48	35.16	.75	40.01	42.95
	MOD	75	34.40	36.55	4.22	25.99	42.81
	UNRWA	230	37.48	34.60	2.28	32.98	41.97
	Private Schools	280	59.00	35.34	2.11	54.84	63.16
COMPT9P	MOE	2201	29.81	24.58	.52	28.79	30.84
	MOD	75	23.20	21.88	2.53	18.18	28.23
	UNRWA	230	30.74	26.01	1.72	27.36	34.11
	Private Schools	280	44.79	25.02	1.50	41.84	47.73
TOTSCR	MOE	4436	34.55	21.85	.33	33.91	35.20
	MOD	149	26.78	19.44	1.59	23.63	29.93
	UNRWA	462	35.01	21.93	1.02	33.01	37.02
	Private Schools	553	45.62	19.62	.83	43.98	47.26

In Table22, in the TOTSCR rectangle we find that Private Schools' mean score on the social studies test (45.6% correct) is the highest of all. UNRWA with a mean of 35.0% correct score comes next, then MOE with a mean of 34.6% follows, and MOD with a mean of 26.8% comes last. Obviously the differences between Private Schools and each of the other three education authorizes are substantial and statistically significant. MOE and UNRWA are neck and neck and there is no significant difference between the two means. On the other hand, MOD's performance is significantly lower than that of MOE and UNRWA. Figure9 shows the differential performance of the four education authorities.

Figure 9
Comparative Performance of the Four Education
Authorities in Social Studies



On six general competencies (viz Comp1, Comp2, Comp5, Comp6, Comp7, and Comp9) out of nine the differences between Education Authorities were of the same type as on the total test score. This common pattern of differences, however was altered regarding performance on Comp3 (Understanding of Maps), Comp4 (Understanding of Systems and Institutions of Modern State), and Comp8 (Ability to Understand Political Life in Arab and Islamic World).

On Comp3 (Understanding of Maps), Private schools' mean was significantly higher than that of the MOE and UNRWA while the differences between MOD and any of the other authorities were not statistically significant at the five percent level of significance. The observed MOD mean was higher than the means of both UNRWA and MOE but lower than that of the Private Schools.

Regarding Comp4 (Understanding of Systems and Institutions of Modern State) the private schools outperformed the schools of the other three education authorities while the differences among them were not statistically significant at Alpha=.05 level.

Regarding Comp8 (Ability to Understand Political Life in Arab and Islamic World) also the private school students outperformed the students from each of the other three education authorities while the differences among them were not statistically significant at the 5% level of significance.

Urban/Rural Differences on Competencies in Social Studies

Comparisons were made between the percent-correct-score means of urban and rural schools' students on each competency score and on the total test score. The results are presented in Table23. On the first five competencies (Compt1 through Compt5) there were no significant differences between the performance of urban and rural schools. On the last four competencies, as well as, on the total test score, however, the differences were statistically significant at the Alpha levels ranging from 0.05 to 0.000. In each case urban school students outperformed their cohort in the rural schools.

Table 23
Urban/Rural School Location Differences in
Student Performance on Competencies in Social Studies

Competency	Location	N	Mean	Std. D.	Std. Err.	Diff. U-R	t-Value	Sig. 2-tailed
COMPT1P	Urban	3177	34.64	22.83	.40	.93	1.50	.13
	Rural	2423	33.71	23.48	.48			
COMPT2P	Urban	1588	27.10	22.33	.56	.15	1.70	.87
	Rural	1226	27.75	23.14	.66			
COMPT3P	Urban	1588	30.23	41.05	1.03	.74	.48	.60
	Rural	1226	29.49	40.57	1.16			
COMPT4P	Urban	1588	30.89	30.89	.77	.59	.48	.63
	Rural	1226	30.30	32.71	.93			
COMPT5P	Urban	3177	36.54	32.03	.57	1.30	.47	.14
	Rural	2423	35.24	32.68	.66			
COMPT6P	Urban	1588	45.57	29.49	.74	2.2	1.49	.05
	Rural	1226	43.45	29.64	.85			
COMPT7P	Urban	3177	48.28	34.90	.61	4.74	1.98	.00
	Rural	2423	43.54	63.12	.73			
COMPT8P	Urban	1589	44.15	36.16	.90	3.33	4.96	.01
	Rural	1197	40.82	34.78	1.01			
COMPT9P	Urban	1589	32.93	24.93	.63	3.99	2.45	.00
	Rural	1197	28.94	25.17	.73			
TOTSCR P	Urban	3177	36.39	21.47	.38	2.10	3.57	.00
	Rural	2423	34.29	22.36	.45			

The competencies on which urban school students did better than rural school students include: Compt6 (Understanding Geographical, Historical and Economic Concepts; Compt7 (Understanding the Concepts of Social Responsibility and Good Citizenship); Compt8 (Ability to Understand Political Life in Arab and Islamic World); Compt9 (Understanding and Interpretation of Pictographs); and the test as a whole.

Gender Differences on Social Studies Competencies

Differences in male and female students' performance on Social Studies Test were investigated using independent unequal samples t-test. Altogether ten t-tests were conducted, one for each of the nine competencies and one for the whole test. The results of the t-tests are presented in Table24.

Table 24
Gender Differences in Students' Performance on
Social Studies Competencies

Competency	Gender	N	Mean	Std. D.	Std. Err.	Diff. U-R	t-Value	Sig. 2-tailed
COMPT1P	Male	2843	33.31	23.85	.45	-1.87	-3.03	.000
	Female	2757	35.18	22.29	.42			
COMPT2P	Male	1434	25.11	22.16	-.59	-5.55	-6.52	.000
	Female	1380	30.66	22.82	.67			
COMPT3P	Male	1434	29.25	40.71	1.08	-1.33	-.86	.38
	Female	1380	30.58	40.97	1.10			
COMPT4P	Male	1434	26.66	31.23	.82	-8.08	-6.81	.000
	Female	1380	34.75	31.66	.85			
COMPT5P	Male	2843	32.92	31.98	.60	-6.20	-7.21	.000
	Female	2757	39.12	32.36	.62			
COMPT6P	Male	1434	42.95	30.20	.80	-3.73	-3.03	.000
	Female	1380	46.31	28.74	.77			
COMPT7P	Male	2843	40.27	35.40	.66	-12.11	-12.95	.000
	Female	2757	52.38	34.57	.66			
COMPT8P	Male	1409	38.20	34.67	.92	-9.15	-6.84	.000
	Female	1377	47.35	35.97	.10			
COMPT9P	Male	1409	29.80	25.66	.68	-2.86	-3.02	.000
	Female	1377	32.67	24.45	.66			
TOTSCR	Male	2843	32.93	22.34	.42	-5.19	-8.94	.000
	Female	2757	38.11	21.08	.40			

From the "Significance" column in Table 24 it is clear that female students scored higher than their male counterparts on the whole test, as well as, on all the competencies except one on which the difference did not meet the criterion of statistical significance ($\alpha = .05$).

All the differences, with the sole exception of Competency 3 (Understanding Maps), were statistically significant at the .003 Alpha level or better.

III. Arabic Language Test

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

Psychometric Properties

The Arabic language test consisted of two different forms. The first form, which was administered to half of the sample, contained 28 items and assessed four competencies (viz. Comprehension, Grammar, Writing and Dictionary Using). The Alpha reliability coefficient for this form was .89 and the discrimination coefficients ranged from .11 through .59. The second form assessed 'Grammar' and 'Writing' competencies and was administered to the other half of the sample. The second form contained 24 items. The Alpha reliability coefficient for this form was .92 and the discrimination coefficients ranged from .35 through .66.

Reading Comprehension Competency

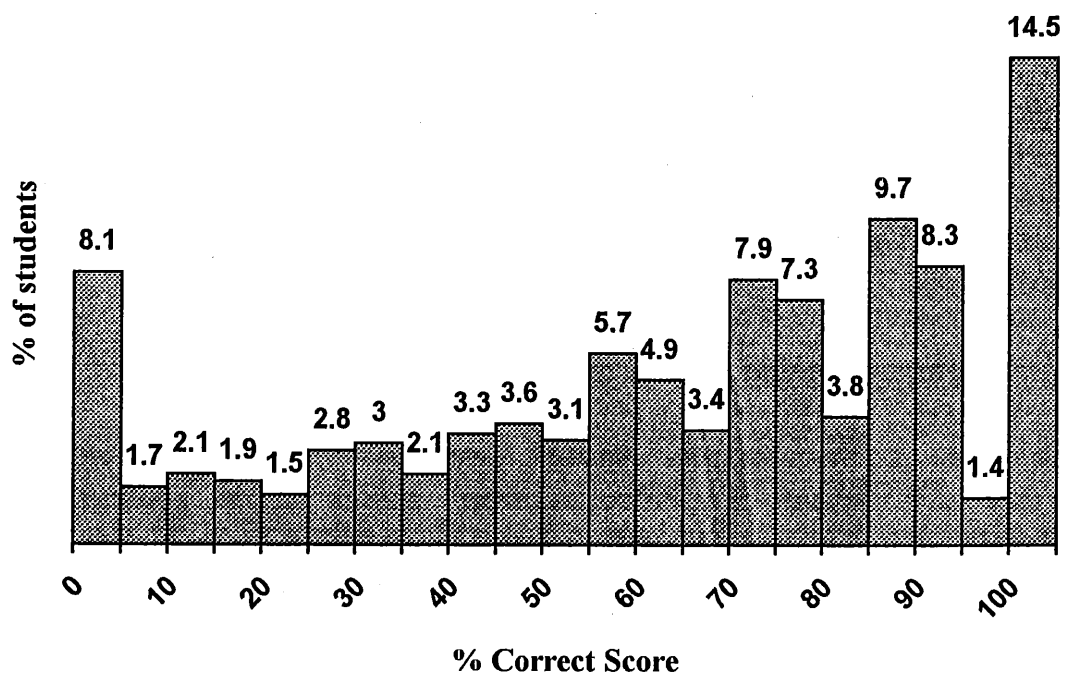
The average percent correct answer was 34.5. The percentage frequency distribution of percent correct scores on the general competency in Reading Comprehension is presented in Figure 9. It is clear from the distribution that around 6% of the sample failed to answer any of the questions correctly. Approximately 30% of students answered more than 50% of the questions correctly. Taking this into consideration one can conclude that only 30% of grade four students pass the cutoff level of 50% on this competency.

When students were categorized according to level of mastery (three levels), only 5.5% of the Grade 4 students reached the mastery level (70% or more of the questions were answered correctly). 51% of students fell in the second or L2 category (more than 30% and less than 70% of the questions were answered correctly). Finally, 43.5% of the students were able to answer only less than 30% of the questions. Taken as a whole, we can conclude that only 6% of the grade four students mastered the comprehension competency, 51% partially mastered this competency and around 44% of them failed to master this competency.

Grammar Competency

The performance of the students on grammar competency test was relatively high where the average percentage of the correct answers was 61.3. Figure10 shows the distribution of percent correct score wherein around two thirds of students passed the cutoff level on this competency (answered more than 50% of the questions correctly). 8% of the sample failed to answer any of the questions correctly.

Figure 10
Percentage Frequency Distribution of Percent Correct Score
Grammar Competency



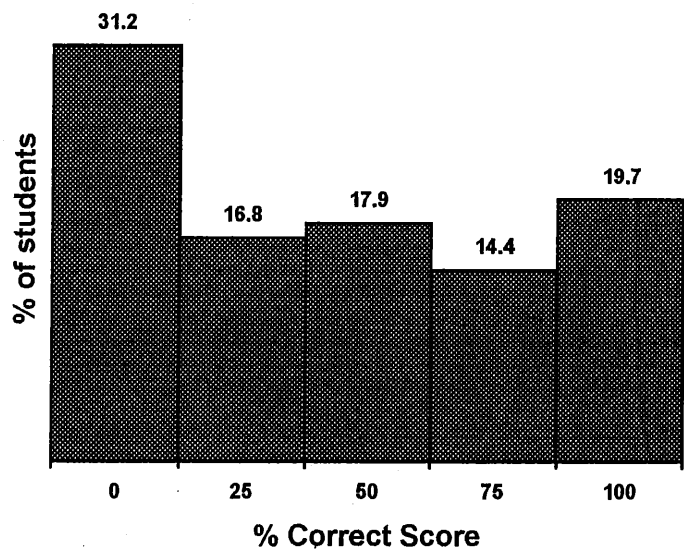
With regard to the three categories of mastery levels, the results showed that more than half of the students (52.9%) mastered Grammar competency and around one quarter (26.1%) partially mastered this competency, while around one fifth (21%) failed to master the competency to an acceptable level.

Writing Competency

The average performance on the writing competency was 43.6%. One can notice from Figure11, which represents the distribution of students' scores (percentage of the correct answers) that less than half of the students passed the cutoff level (50%) and around one third of

the students could not answer any of the 'Writing' competency questions correctly.

Figure 11
Percentage Frequency Distribution
of Percent Correct Score on Writing
Competency



The results showed that around one third of the students (34.1%) reached the mastery level in the 'Writing' competency and around half of the students (48%) did not master this competency wherein the rest of the students (17.9%) were considered to have reached a modest level of mastery in the 'Writing' competency.

Dictionary Utilization Competency

Dictionary utilization competency was measured with one question. The average performance on this question was 39.2%, which means that around 60% of the students could not answer this question correctly.

Gender Comparisons on Arabic Language Competencies

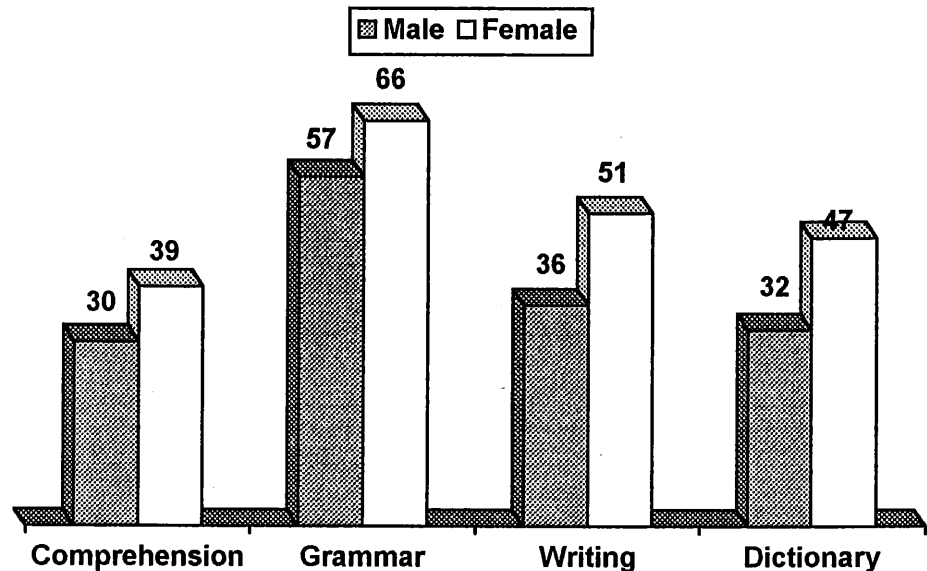
Figure12 shows the performance (mean of the percent correct answers) of the sampled students according to the gender variable on each of the competencies; Comprehension, Grammar, Writing, and Dictionary Utilization. The performance of female students was

statistically significantly ($p < .001$) higher on each of the four competencies than that of the male students.

The order of performance from highest to lowest was the same for both girls and boys. This order was as follows: Grammar, Writing, Dictionary Utilization, and Comprehension.

Figure 12
Gender Differences on the Arabic Language Competencies

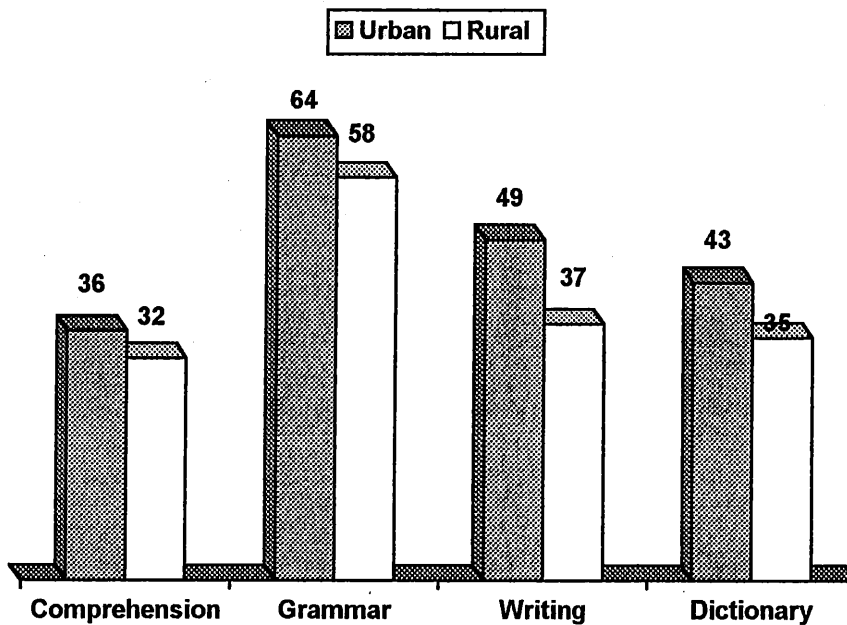
Average % Correct Score



Comparative Performance of Urban and Rural Schools' Students

Figure 13 shows the performance of the sampled students according to be seen that urban school students outperformed rural school students on the four competencies with a difference ranging from 4 to 12 points.

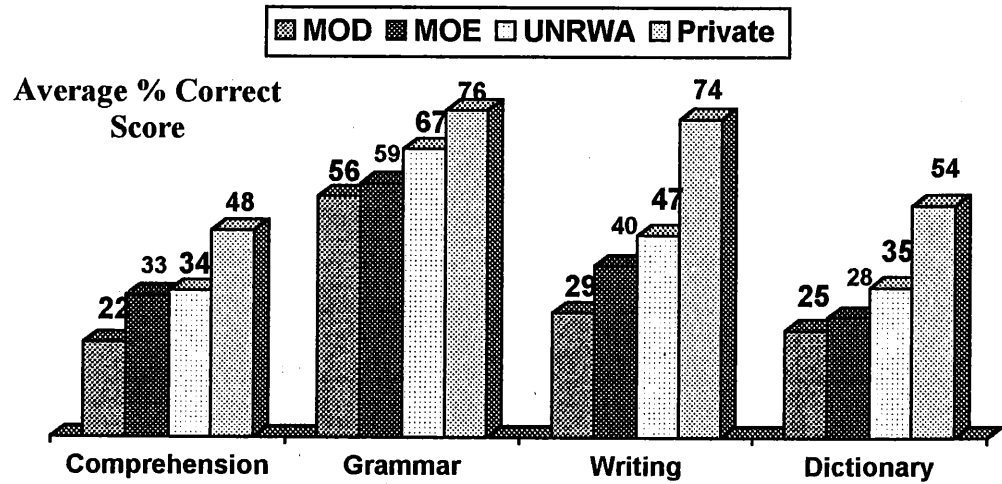
Figure 13
Urban/Rural Schools Differences on the
Arabic Language Competencies



Comparisons between Educational Authorities

Figure14 shows the average percent correct scores on the four competencies in each education authority; Ministry of Education (M.O.E), Ministry of Defense (M.O.D), UNRWA, and Private schools. The graph reveals that in general, on all competencies, the highest performance was of the private schools' students followed by UNRWA students; M.O.E school students came third and the lowest performance was for the M.O.D school students.

Figure 14
Education Authority Difference on the
Arabic Language Competencies



In terms of statistical significance, however, not all the pairwise differences were significant at the .05 Alpha Level.

Private schools' performance was significantly better than each of the other three education authorities on all the four general competencies in Arabic Language.

On two general competencies (Reading Comprehension and Dictionary Utilization) the differences between MOE and UNRWA were not statistically significant at $\text{Alpha} = .05$, but both of them did significantly better than MOD.

On Grammar Competency the difference between MOE and MOD was not statistically significant ($\text{Alpha} = .05$) but UNRWA scored significantly higher than both of them.

On writing Competency all the pairwise differences were statistically significant at the .05 Alpha level; that is, UNRWA outperformed both MOE and MOD, whereas, MOE outperformed MOD.

