Student Learning Assessment Series

Reform Support Unit

Education & Literacy Department

Government of Sindh

Sindh Students Assessment Results Report, 2009

Mathematics Grade 4

Provincial Education Assessment Centre (PEACE)

Bureau of Curriculum and Extension Wing, Jamshoro, Sindh

May 2010

Acknowledgement

The Sindh Students Grade 4 Mathematics Assessment was designed, developed, administered, analyzed and the report written by the following:

- Diagnostic Assessment tools were developed and designed by PEACE Subject Specialists, Jamshoro in collaboration with members of the Bureau of Curriculum and Extension Wing, PITE, mathematics teachers and GECE teachers.
- Sampling was conducted by the PEACE staff. The sample was verified by Education District Officers
- Tests and Background Questionnaires were administered by test administrators, school teachers (teachers of GECEs and PITE), trained by Master Trainers and Master Trainers trained by PEACE staff
- Focal persons in the districts managed the distribution of the assessment instruments and the collection of the completed test booklets
- Data entry and cleaning was undertaken by PEACE
- Data analysis was conducted by PEACE
- Report Writing was undertaken by PEACE, PITE, BoC and RSU
- DPM PEACE, RSU
- PEACE Monitoring Officer, RSU

The PEACE team consists of:

Mr Daud Memon	PEACE Coordinator
Mr Aftab Ali	Deputy Director, BoC/Research Officer, PEACE
Mr Tanweer Khan	Science Subject Specialist
Ms Tasheeen Ansari	Language Subject Specialist
Mr Khalis Mehmood Memon	Social Studies Subject Specialist
Ns Zahida Banguish	Urdu Subject Specialist
Ms Attia Bhutto	Mathematics Subject Specialist
Mr Zahid Abro	Computer Operator

PEACE Activities were conducted under the auspices of the Reform Support Unit, Education and Literacy Department, Government of Sindh.

All activities were supported by EU, SER-TA.

Table of Contents

List of Tables

Acronyms

BoC	Bureau of Curriculum
BQ	Background Questionnaire
EU	European Union
GEC	Government Elementary College
GIS	Geographic Information System
ITEMAN	Item Analysis Program
MCQ	Multiple Choice Question
MOS	Measure of Size
NEAS	National Education Assessment System
PEACE	Provincial Education Assessment Centre
PITE	Provincial Institute for Teacher Education
PPS	Probability Proportional to Size
PSU	Primary Sampling Unit
QA	Quality Assurance
RSU	Reform Support Unit
SAS	Statistical Analysis System
SEMIS	Sindh Education Information System
SERP	Sindh Education Reform Programme
TA	Test Administrator

Executive Summary

There is general agreement of the need for more consistent attention to be made to improvements in learning quality and to the measurement of learning outcomes. The Sindh Education Reform Programme (SERP) aims to do this.

To support improvements in learning quality a mathematics study has been undertaken to provide education decision makers with systematic information about the status of students' learning and the extent to which they attain pre-defined standards and competencies as identified in the 2006 National Curriculum. It enables Sindh province to identify its needs for focused interventions for the improvement of mathematics teaching and students' learning and their learning environment.

The main objectives of this study are summarized below:

- To assess what students in Grade 4 know and can do in mathematics
- To provide information on performance in areas identified by the National Education Assessment System as weak in relation to the competencies defined in the 2006 Mathematics National Curriculum
- To obtain information regarding students' attitudes to mathematics and mathematics teaching
- To obtain information from Head Teachers and Teachers regarding their attitudes, school policies and mathematics teaching practices

Students at Grade 4 mathematics (number, fractions, measurement and geometry) obtained a mean percentage score of 44.7%. From this study it can be seen that much work is required to improve the teaching and learning of mathematics in the subject areas tested. The results of these tests are found in detail in Section 7.2.

The results of the tests have implications for Sindh Province regarding the need to:

- Identify requirements and strategies and plan for improvements in student learning;
- Interpret the National Curriculum according to the needs of Sindh province;
- Make assessment **for, as** and **of** students' learning central to the development of improved teaching and learning methodologies;
- Develop supplementary materials to support student difficulties and teachers' teaching;
- Improve textbook development in line with the 2006 National Curriculum standards and competencies;
- Improve Teacher Training and Teacher education Development;

- Improve the roles of management in the districts to mentor and advise teachers in a supportive manner.

It is hoped that these implications will be further discussed and integrated into the existing SERP programme.

A full Technical Report is available from the RSU and PEACE.

1. Introduction

The improvement of the quality of education requires a multi directional approach in order to improve the effectiveness of the learning process by improving teaching techniques, supplying better learning aids, motivating students to attend school etc. Student assessments which are credible and objective play a pivotal role in this process by providing critical feedback on what and how well students are learning[1].

PEACE, Sindh was initially developed as a unit to support the development of national assessments conducted by National Education Assessment System (NEAS). Between 2004 and 2008, four national assessment surveys were conducted in languages, mathematics, science and social studies. PEACE personnel obtained capacity building in item writing and test development, test administration and training of test administrators, verifying the NEAS sample, identifying policy issues for background questionnaire development, marking and scoring and statistical analysis and report writing and the dissemination of results through this development.

Further training and support has been provided to PEACE under the SERP by EU technical assistance. PEACE has now established its own role through the development of provincial assessments aimed at identifying in each district the strengths and weaknesses of the achievements of students in relation to the National Curriculum, and correspondingly the strengths and weaknesses of the teaching process and textbooks used in the classroom.

The differences between PEACE Provincial Assessments and the NEAS assessments are as follows:

- PEACE used a much larger sample than NEAS to obtain representation of every district in Sindh;
- PEACE assessments were based on four areas of the National Curriculum (number, fractions, measurement and geometry) while the national assessment was based on the whole of the Grade 4 National Curriculum;
- PEACE assessments covered the National Curriculum from Grade 1 – 4 in each of the areas tested to test the progression achievement of students while the national assessment was based on the Grade 4 curriculum only.

Mathematics, languages, science and social studies have been identified as subjects to be tested by PEACE. Mathematics was chosen to be tested first as mathematics provided concrete concepts and was easier to test than language. Through this assessment **of** mathematics learning (using assessment to make sound judgments about

[1] The World Bank Report, December 2008

learning and school effectiveness) sound judgments can be made about learners' achievements. This assessment also supports assessment **for** learning (using assessment to support classroom teaching) and assessment **as** learning (using assessment to promote autonomy in learning) by ensuring that standards are understood and shared by teachers, the learners themselves, their parents and other adults who are either directly engaged in or are supporting the learning process.

The process of the preparation of the report of the 2009 Grade 4 Mathematics Provincial Assessments has involved all key stakeholders. An Assessment Working Group was constituted and assigned responsibility for reporting in detail on one or more sections which would ultimately form the report. These contributions were subsequently compiled into a first draft at the RSU. A list of the members of this group is found in Annex 1. The contributions were subsequently compiled into a first draft at the RSU.

The draft was further refined and additional material provided by the PEACE has been incorporated into the 2009 Grade 4 Mathematics Assessment Report.

Further information regarding any aspect of the provincial tests can be obtained from PEACE, BoC.

2. Sindh Province Testing Model

The testing model used for the Sindh Provincial Assessment is based on the need to identify student achievement. This is different from the limiting psychometric model which emphasizes ranking and statistically derived distributions since it involves a shift away from a norm referenced approach towards one where what students can and cannot do is stated. This requires descriptions of performance as found in the Mathematics 2006 National Curriculum for Pakistan. The model also looks at the different levels of achievement of students according to the requirements of the Grades 1 – 4, National Curriculum.

The use of performance descriptions has implications for reporting the results. While the use of a single overall figure gives us some notion of student achievement there is also a need to provide qualitative descriptors what students can do according to the areas tested and by denoting the levels attained within the subject areas assessed.
This model was therefore developed through analyzing and mapping the Grade 4 Mathematics, 2006 National Curriculum and identifying the competencies to be tested. From this two domains were identified for testing, namely, context and cognitive domains and test specifications were written to develop test items to assess these specific domains according to the weightage indicated in the National Curriculum.

Test items were written and classical item analysis (ITEMAN) was used to identify item difficulty and the ability of each item to discriminate between students of different abilities. Following this test items were selected for use in the large-scale testing. A two parameter model was used in the final data analysis to identify item difficulty and the ability of each item to discriminate between students of different abilities. This provided results according to, for example, location, gender, districts. A regression analysis (linear/logistic) was used to identify such aspects as the impact of teachers on learning.

The sampling model used was a two parameter model based on district (23 districts) by location. This resulted in 46 strata (e.g. rural, urban) for the province.

3. Mathematics Assessment 2009

The Provincial Education Assessment Centre (PEACE) was established in 2002 to provide objective information on student performance available to policy makers, the Bureau of Curriculum, teacher training institutions and officers at provincial, district and taluka, Union Council and school levels and to provide stakeholders with an evaluation of the condition and progress of education. This information is only related to academic achievement and the effect of different background variables and attitudes on student

achievement. The privacy and therefore the identity of individual students and families as well as the identities of the participating schools are not released.

Sindh Province Grade 4 Mathematics Assessment, 2009 is a provincially representative and continuing assessment of what Sindh Province students know and can do in various aspects of Mathematics. In 2009 the first Mathematics Provincial Assessment was conducted in all the districts of Sindh in 4004 schools (Primary Sample Units) and with 28,866 students who wrote 106,716 tests.

PEACE along with working school teachers, BoC, PITE and GECE staff developed a Mathematics Assessment Framework in 2008. This framework identified the two dimensions of mathematics in the National Curriculum to be tested, namely subject content domain (defines the subject matter covered by the assessment) and the cognitive domain (knowing facts and procedures, using concepts, solving problems, reasoning). This was the foundation of the provincial assessment and the basis for item development. The content domain areas to be tested were f identified in the Assessment Framework as number, measurement, fractions and geometry. Each content domain has several topic areas (e.g., number is further categorized by whole numbers, integers, and ratio, proportion, and percent etc.).

A test specification was developed to describe the number of questions and the weightage of each area to be tested. The areas to be tested were identified by the National Education Assessment System, (NEAS) in 2005, 2006, 2007 and 2008. These were areas where students had below average achievement. It ensured that the curriculum content of these areas was more likely to be assessed in a balanced way.

The test specifications on which the Grade 4 mathematics assessments were based are found in Annex 2.

4. Assessment Instruments

To measure the students' learning achievements as well as getting factors which affect the quality of education in Sindh Province two different assessment instruments were developed to test Grade 4 students in 2009. The two types of instrument developed were:

- Mathematics Diagnostic tests in number, fractions, measurement and geometry;
- Background Questionnaires for head teachers, teachers and students.

Mathematics Diagnostic tests were developed on the basis of the National Curriculum, 2006. The following processes were adopted to develop the achievement tests:

- Competencies based on the National Curriculum were identified for the grade 4 mathematics subject areas to be tested;
- A table of specifications was designed for the mathematics tests which included content, learning outcomes and the number of items;
- Multiple Choice questions were developed to provide objective results and easier standardization of test setting and marking;

Some examples of the Mathematics diagnostic test items are found in Annex 4.

Background Questionnaires were developed for Head Teachers, Teachers and Students. These questionnaires contained questions to identify the effects of, for example, teacher qualifications on student results; the effect of giving homework on student achievement; the effect of living in a rural area as compared with an urban area, and, student attitudes regarding mathematics and mathematics teaching.

Both the mathematics diagnostic tests and the background questionnaires were piloted in 2008 on a sample of 1150 students. These tests were then marked, coded and entered into a computer base. From the results of the statistical analysis of the pilot items (using ITEMAN software) items were selected and additional items developed where required. Formats for large scale testing were developed and administered in April 2009 to a provincial sample of 28,866 Grade 4 students to establish a baseline of achievement and to identify specific problem areas in learning for Grade 4 students in mathematics.

Additional information on the development of the assessment instruments is found in Annex 3.

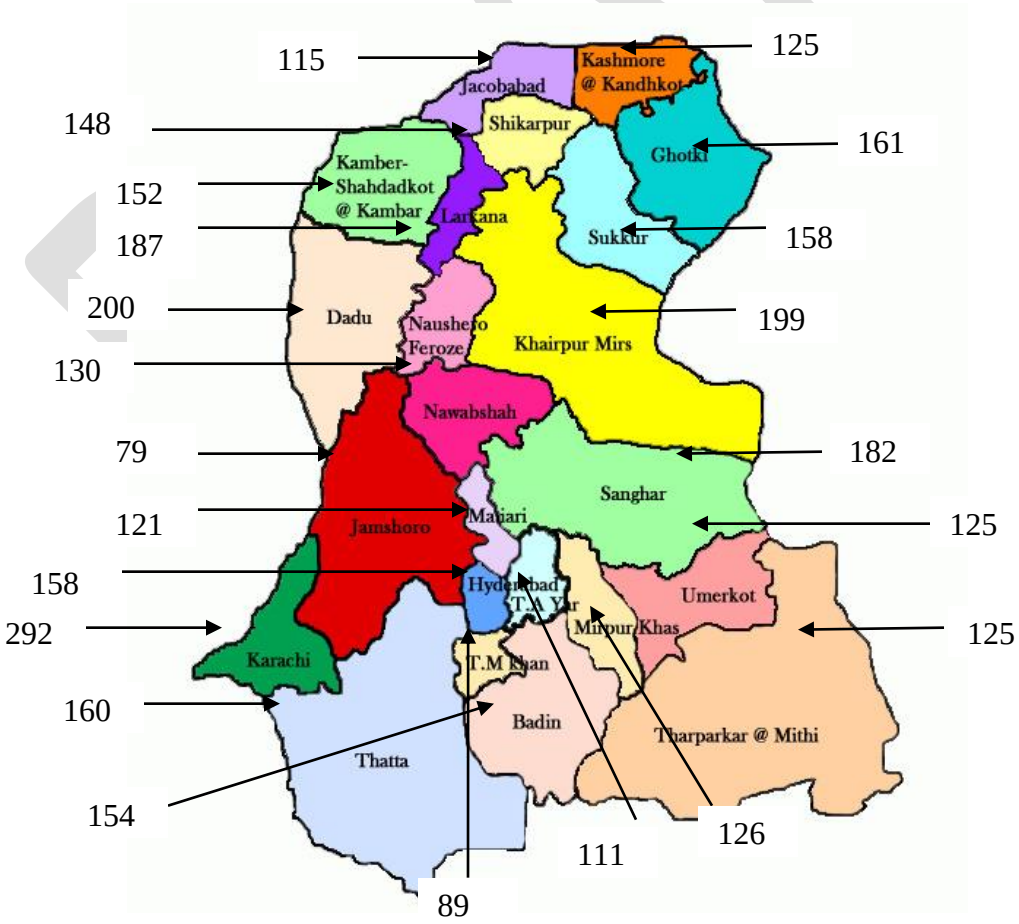
5. Assessment Sample

It was proposed that in 2008 a representative sample (15% of the Grade 4 population) of approximately 40,000 Grade four students (4004 PSUs) should participate in the assessment. Schools were selected, using Proportional Probability Sampling Techniques (PPS), in fixed proportions from the defined groups (districts; rural/urban; male/female). The coverage of the provincial sample for Grade 4 Mathematics is found below. The actual sample used was less than the originally defined sample (28,866 students and 4004 PSUs). This was as a result of verification of the status of the schools by District Officials. The final sample used for analysis was 3, 746 schools. This was a result of the data cleaning exercise conducted by PEACE.

The results obtained from this assessment provide Sindh Province with a benchmark for all future results to be judged. The sample was selected according to district; rural/urban; male/female categories covering the whole province. More detail regarding the development of the sample is found in Annex 4.

A map showing the district wise coverage of the Grade 4 school sample is found below.

District wise coverage of the Grade 4 School Sample



More detailed information is found in Annex 5.

6. Test Administration, Monitoring and Coding, Data Entry and Cleaning.

6.1 Test Administration

The tests were administered in April, 2009 by teachers in the districts according to the instructions in the Test Administration Guideline Booklet developed. The teachers were trained by Master Trainers who had previously received training by PEACE. The PEACE staff monitored the test administration.

Ninety-six master trainers from all districts were trained by PEACE, at two centres, Hyderabad and Sukkur.

These master trainers trained test administrators throughout the Province. The number of test administrators trained was as follows:

Districts	No of Test Administrators Trained
Badin	198
Dadu	217
Hyderabad	175
Thatta	183
Mirpurkhas	143
Tharparkar	166
Sanghar	226
Karachi	326
Jacobabad	150
Larkana	219
Shikarpur	158
Khairpur	231
N Feroze	207
Nawabshah	170
Sukkur	181
Ghotki	217
Umerkot	133
Jamshoro	123
Matiari	148
T-Allahyar	123
TM-Khan	103
Kashmore	161
Kambar	172
Total	**4130**

Some of the **difficulties** identified in the test administration were as follows:

- Some of the test administrators did not always appreciate the need for the assessment to be conducted in a rigorous manner;
- Test administrators did not always use the examples in the test booklets to familiarize the students with the test methodology;
- Test administrators did not always follow the guidance given in the guidance booklet.
    - ✓ Test administrators found the use of the random number table (used to identify 10 students in a class which had more than 10 students) as well as the skip interval difficult to understand and practice
    - ✓ There was also a lack of understanding of the methodology for entering the correct information for "split" schools
- The school enrollment was not recorded during the test administration for some of the sampled schools

6.2 Test Monitoring

Monitoring was undertaken in two areas:

- Monitoring the PEACE Assessment Administration, and,
- Monitoring the marking and coding of the assessment instruments

Monitoring the Provincial Assessment Administration

The main objective of monitoring the provincial assessment was to ensure the validity of the Provincial Assessment data. It is important that all aspects of the provincial Assessment is standardized, including the administration of the assessment instruments.

The monitors consisted of 23 district focal persons who ensured the distribution and collection of the test instruments. PEACE, BoC and PITE were involved in the overall monitoring. The monitors monitored the assessment activity and reported back to PEACE on how the test administrators followed the guidelines given during the test administration training.

Monitoring the Marking and Coding of the Assessment Instruments

Coding and data entry of all the scripts in the 2009 Provincial Assessment was undertaken in eight centres of Sindh Province namely, Hyderabad, Mirpurkhas, Dadu, Larkana, Sukkur, Karachi, Khairpur, Shikarpur. Instructions regarding the coding and data entry were given to the personnel involved.

The responsibility of the monitors was to:

- Ensure that the coding and data entry was conducted in an efficient and fair and transparent manner;
- Provide support to those coders and data entry personnel who were having some difficulty;
- Identify coding and data input discrepancies and correct them where possible;

- Take a 10% sample of the coding sheets to identify the reliability of the coding and data entry;
- Ensure that the correct code was being used on the coding sheets;
- Ensure that the assessment instruments were complete and returned to PEACE when the coding and data entry input was completed.

Some of the difficulties identified in these activities are found below:

- In one monitored school evidence was found of some duplication of monitors– a need for improved communication between the monitors;
- Test Administration Monitors need to be issued with specific instructions as to how they should behave/react if they observe that the assessment instruments have not being administered in an appropriate manner;
- There was a need for the training of the coders and data entry operators to be more rigorous;
- There was a need for more extensive monitoring of both the assessment booklets and coding sheets – 10% of the number of students on a coding sheets was not sufficient to ensure that the coding was being conducted efficiently and fairly;
- Little information was available of any evidence of irregularities in test administration such as identical but unlikely patterns of response in multiple choice answers in more than one booklet.

6.3 Coding and Data Entry

For the 2009 mathematics test marking and coding methodologies were developed by the subject specialists on paper sheets and then transferred into the Excel program. Each possible answer was given a specific code.

Manual test marking and coding was conducted by elementary college and general school teachers and private school teachers at eight centres in the province

Checking the data was an onerous a task so it was not possible to check every single sheet. The data of two students out of 10 students on a scoring sheet was checked by pairs of the elementary college, general school teachers and private school teachers and PEACE specialists super checked one out of 10 students on each scoring sheet in Hyderabad and the focal persons super checked two out of 10 students on each scoring sheet in the other centres. Where discrepancies were identified these were rectified where possible or the data was discarded due to its lack of reliability.

The cleaned data was then entered into the SAS program ready for data cleaning. More detail on the coding, data entry and cleaning can be found in Annex 6.

To ensure quality assurance, there is a need to centralize the marking and coding at one place, such as Hyderabad for monitoring to be done by the persons involved in the analysis workshop.

7. Item and Background Data Analyses

The analysis of test results is complex if the results are to be valid and reliable.

Data analyses at PEACE involved five phases:

- Phase-I consisted of using Microsoft Excel for sample selection;
- Phase II included Classical item and reliability analysis of the pilot test items for the selection of test items for the large-scale testing;
- Phase III consisted of the cleaning of the large-scale testing scores and background data information using SAS
- Phase IV consisted of IRT analysis using Bilog-MG.
- Phase V consisted of the use the WesVar program to analyse the data.

The following is a description of softwares used for each phase of analysis.

Phase I: Microsoft Excel

This was used at the stage of sample selection and is quite helpful in basic arrangements of sampling frames, calculations of sampling intervals, generating random number seeds and sorted lists of sampled schools. However as it involved mainly manual input it did not provide the necessary checks and balances required for such large scale testing. In future sample selection SAS will be used.

Phase II: ITEMAN

Item analysis of the pilot test items was carried out using **ITEMAN.** This is “Classical item analysis” for the purpose of item selection. This program provided information on two parameters; item difficulty and item discrimination and was used in item selection along with information on the performance of distractors for each item.

Phase III: SAS Program

SAS program was used to clean all the data from the large-scale testing and Background Questionnaires. The SAS program provides a complete, comprehensive set of tools that meets the data analysis needs required for provincial assessment.

Phase IV: Bilog-MG3 Program

Bilog-MG3 was used to provide Item Response Theory (IRT) information. It models the response of each student of a given ability to each item in the test. Bilog-MG3 converts students’ raw scores on a test or versions of a test to a common scale that allows for a numerical comparison between students. The PEACE testing programs used multiple versions of a test, and the scale is used to control slight variations from one version of a test to the next. Scaled scores are particularly useful for comparing test scores over time, such as measuring semester-to-semester and year-to-year growth of individual students or groups of students in a content area.

Phase V: WesVar Data Analysis

WesVar was used for the analysis of the survey data and for regression analysis.

8. The Results

8.1 Understanding the Assessment Results

PEACE uses widely accepted statistical standards in analyzing the data. More detailed statistical information can be found in the comprehensive technical report produced by the Government of Sindh.

Results are presented as mean percentage scores. The achievement of students is discussed in Section 7.2 under the following headings:

- Overall student achievement
- Student Achievement by Location
- Student Achievement by Gender
- District results
- Differences between student achievement in each district and the rest of the districts
- Student overall achievement according to Content Domain
- Student overall achievement according to Content Domain and Location
- Student overall achievement according to Content Domain and Gender
- Student achievement According to Content Domain and District
- Student overall achievement according to Cognitive Domain
- Student overall achievement according to school size and Location
- Performance of students for each mathematics area tested according to Context and Non-context items
- Difference in Achievement According to Test item Type and Location
- Achievement of Students by Gender and Location
- Background Questionnaire Findings

To make the scores meaningful and to establish a relationship between student achievement and through the various variables, classical analysis and significance levels are explained below.

Classical Analysis

Classical analysis was used to identify “good” test items from the pilot testing. The items were Classical as “good” if they demonstrated good reliability and validity and if they were able to discriminate between students of different abilities. Classical analysis was also conducted to enable stakeholders to relate more easily to the actual achievement scores of the students as related to the competencies tested. This Classical analysis is presented in percentage form for the four mathematical areas tested.

Significance Levels

To check whether differences in reported scores could have occurred by chance alone, significance tests are reported. A probability where is < 0.05 means that the difference

could occur by chance alone in only 5 out of 100 students; where it is < 0.01 the difference could occur by chance alone in only 1 out of 100 students (significant difference) and where it is 0.00 there is a highly significant difference.

8.2 Findings

Two sets of findings were obtained from the assessment namely:

- Assessment Test findings
- Background Questionnaire findings.

8.2.1 Assessment Test Results

Overall Student Achievement

This report presents the Sindh Provincial results of the assessment in mathematics (number, measurement, fractions, and geometry). On the 2009 Grade 4 mathematics tests students achieved a provincial percentage mean score of 44.7% in mathematics. This was below the mean score of 50% and is similar result to that of the national results produced by NEAS in 2008 where Sindh Province Grade 4 students (on items based on the whole of the National Curriculum) achieved a mean score of 33.6%

Table 1: Provincial Mean Mathematics Score

MEAN SCORE
44.7%

The students performed best on number and measurement test items obtaining the highest scores on procedural knowledge test items (57.7%), followed by conceptual understanding items (52.36%) and being weakest in their achievement on problem solving items (43.8%).

The items in the tests tested students' achievement on their mathematics content knowledge and mathematical skills (cognitive domain). From the analysis of the test data, students found all **fraction** test items used in the tests moderately difficult or difficult. None of the fraction test items, according to student achievement, could be classified as easy. This was reflected in the student overall mean score for fractions of 38.5%. In Geometry the majority of the test items were classified as moderately difficult or difficult. Only 14 test items out of 100 could be classified as easy according to student achievement. In Number, only four test items out of 100 could be classified as easy according to student achievement and in Measurement 10 items were classified as easy according to student achievement.

The results showed that, in general,

- students in rural areas achieved higher scores in mathematics than students in urban areas
- boys achieved higher scores than girls
- boys in rural areas achieving higher scores than boys in urban areas
- girls in rural areas achieving higher scores than girls in rural areas.
- girls in rural areas also achieved higher scores than boys in urban areas

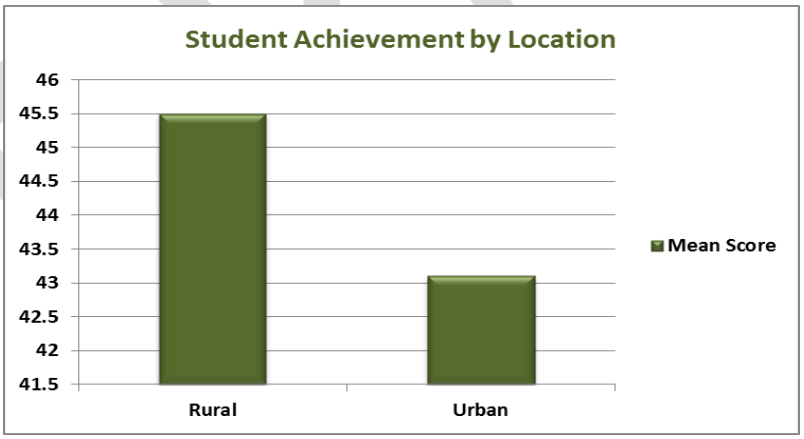
The students of the districts of Kashmore, T M Khan, Matiari, Ghotkhi, N Feroze, Jacobabad, Sanghar, Tharparkar, Mirpurkhas,and Dadu performed best in the province.

Student Achievement by Location

It was noted that students in rural areas performed better in the mathematics areas tested compared with students in urban areas. The difference between rural and urban student performance is highly significant.

Table 2: Student Achievement by Location

SUBJECT	RURAL	URBAN	PROBABILITY	Sig / N-Sig
Mathematics	45.5	43.1	0.00	Highly Significant

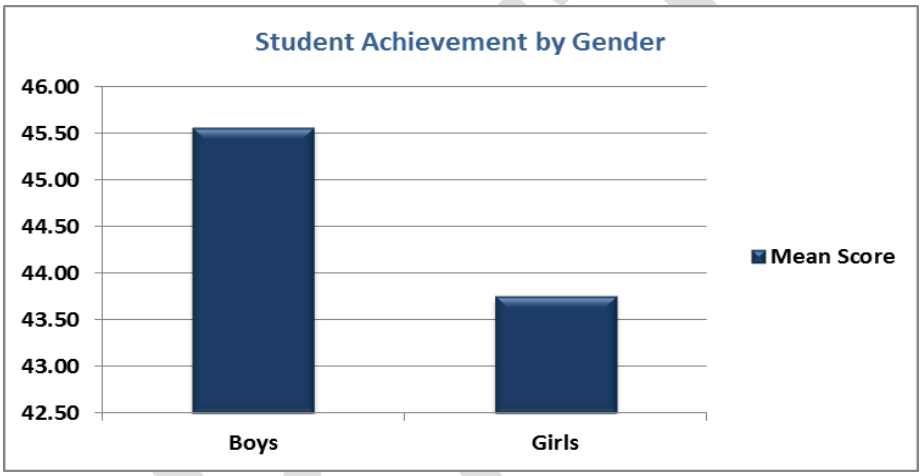


Student Achievement by Gender

Overall, boys performed better than girls in mathematics in Sindh Province. The difference between boys and girls' performance was highly significant.

Table 3: Student Achievement by Gender

SUBJECT	BOYS	GIRLS	PROBABILITY	Sig / N-Sig
Maths	45.6	43.7	0.00	Highly Significant

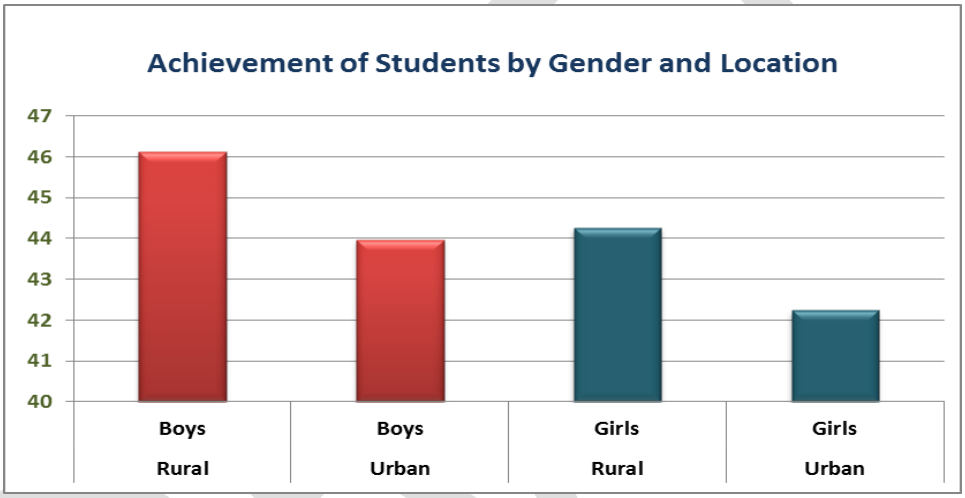


Student Achievement by Location and Gender

Both rural boys and girls performed better than urban boys and girls.

Table 4: Achievement of Students by Location and Gender

Location	Gender	Percentage Score	Probability	Sig./ N-Sig
Rural	Boys	46.12	0.00	Highly Significant
Urban	Boys	43.96		
Rural	Girls	44.26	0.01	Significant
Urban	Girls	42.24		



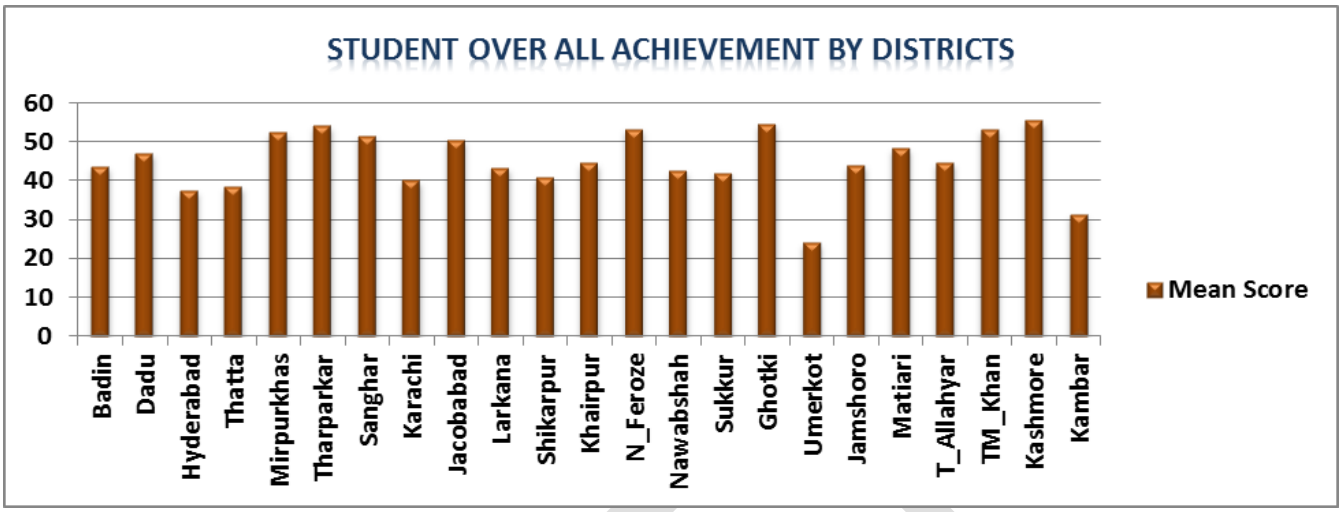
Mathematics Results for the Districts

Students from 10 districts achieved significant and significantly higher mean scores in mathematics as compared with the rest of the province. Students from 12 districts had significantly lower scores in mathematics than the rest of the province.

Table 5: Mathematics Results for the Districts

District ID	District Name	% Mean Score
1	Badin	43.57
2	Dadu	47.00
19	Ghotki	54.62
3	Hyderabad	37.50
12	Jacobabad	50.48
22	Jamshoro	44.09
27	Kambar	31.34
8	Karachi	40.00
26	Kashmore	55.39
15	Khairpur	44.47
13	Larkana	43.39
23	Matiari	48.31
5	Mirpurkhas	52.40
16	N_Feroze	53.32
17	Nawabshah[2]	42.51
7	Sanghar	51.40
14	Shikarpur	40.77
18	Sukkur	41.96
24	Tando_Allahyar	44.69
6	Tharparkar	54.00
4	Thatta	38.51
25	TM_Khan	53.20
20	Umerkot	24.07
Sindh Province Mean Score		**44.75**

[2] At the time of testing it was known as Nawabshah; it is now known as Shaheed Benazirabad

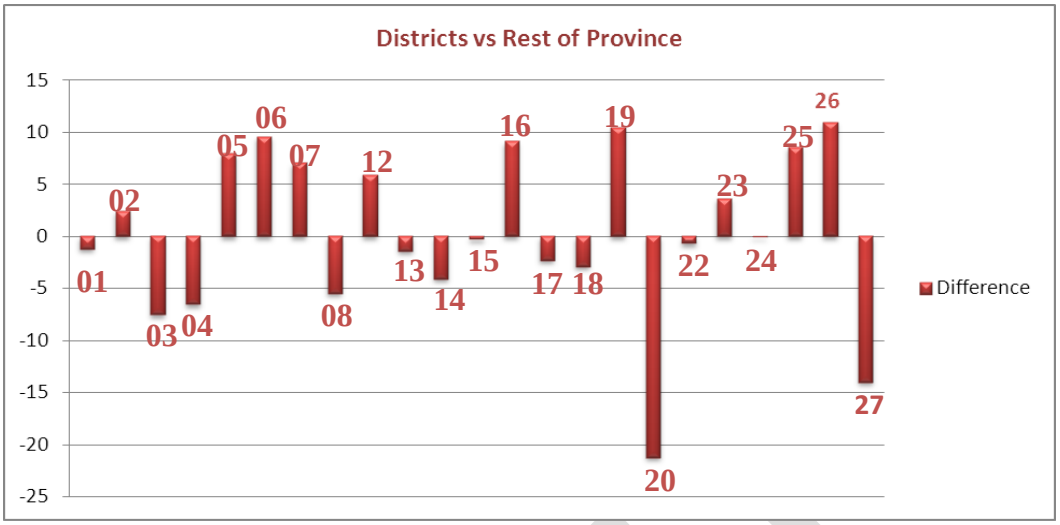

STUDENT OVER ALL ACHIEVEMENT BY DISTRICTS
60
50
40
30
20
10
0
Badin
Dadu
Hyderabad
Thatta
Mirpurkhas
Tharparkar
Sanghar
Karachi
Jacobabad
Larkana
Shikarpur
Khairpur
N_Feroze
Nawabshah
Sukkur
Ghotki
Umerkot
Jamshoro
Matiari
T_Allahyar
TM_Khan
Kashmore
Kambar
Mean Score

Mathematics Results: Difference between Each District and Rest of the Province

From this it can be seen that the students representing the districts of Kashmore, T M Khan, Matiari, Ghotkhi, N Feroze, Jacobabad, Sanghar, Tharparkar, Mirpurkhas,and Dadu performed best in the province. Students in the districts of Umerkot and Kambar showed the least achievement as compared with the rest of the province.

Table 6: Mathematics Results: Difference between Each District and Rest of Province

District Name	Student Score	Rest of Province	Difference	Probability	Sig/ N-Sig
Badin	21.78	22.4	-0.62	0.24	Not Significant
Dadu	23.52	22.3	1.23	0.06	Significant
Ghotki	27.31	22.07	5.24	0.00	Highly Significant
Hyderabad	18.75	22.52	-3.77	0.00	Highly Significant
Jacobabad	25.24	22.29	2.95	0.00	Highly Significant
Jamshoro	22.04	22.38	-0.34	0.77	Not Significant
Kambar	15.67	22.68	-7.01	0.00	Highly Significant
Karachi	20	22.75	-2.75	0.00	Highly Significant
Kashmore	27.7	22.22	5.47	0.00	Highly Significant
Khairpur	22.23	22.38	-0.15	0.75	Not Significant
Larkana	21.69	22.41	-0.72	0.13	Not Significant
Matiari	24.15	22.33	1.82	0.00	Highly Significant
Mirpurkhas	26.2	22.24	3.96	0.00	Highly Significant
N_Feroze	26.66	22.1	4.56	0.00	Highly Significant
Nawabshah	21.25	22.42	-1.17	0.18	Not Significant
Sanghar	25.7	22.18	3.52	0.00	Highly Significant
Shikarpur	20.38	22.44	-2.06	0.01	Significant
Sukkur	20.98	22.43	-1.45	0.04	Significant
T_Allahyar	22.35	22.37	-0.03	0.96	Not Significant
Tharparkar	27.01	22.21	4.79	0.00	Highly Significant
Thatta	19.25	22.51	-3.26	0.00	Highly Significant
TM_Khan	26.6	22.31	4.28	0.00	Highly Significant
Umerkot	12.04	22.65	-10.61	0.00	Highly Significant



LEGEND

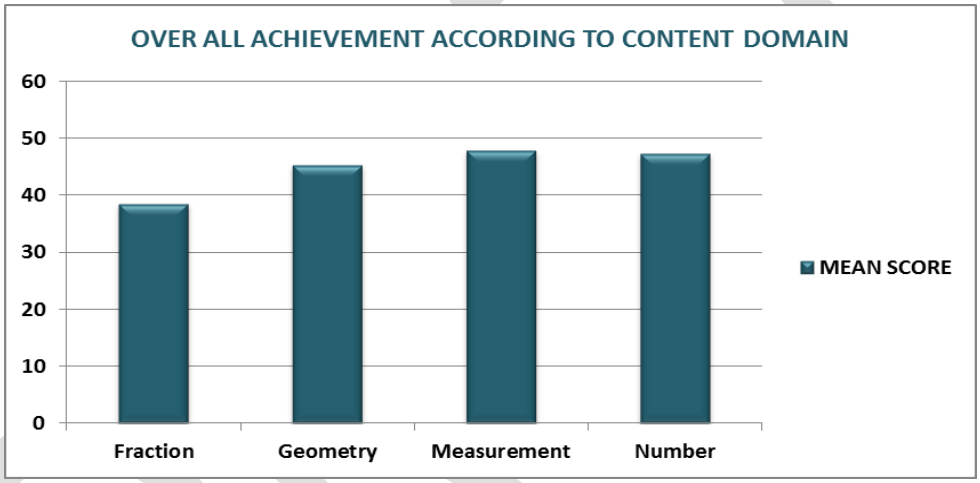
1= Badin	2= Dadu	3= Hyderabad	4= Thatta	5= Mirpurkhas
6= Tharparkar	7= Sanghar	8= Karachi	12= Jacobabad	13= Larkana
14= Shikarpur	15= Khairpur	16= N_Feroze	17= Nawabshah	18= Sukkur
19= Ghotki	20= Umerkot	22= Jamshoro	23= Matiari	24= T-Allahyar
25= TM-Khan	26= Kashmore	27=Kambar		

Student Overall Achievement according to Content Domain

Students performed best in measurement and number followed by the areas of geometry and fractions

Table 7: Student Overall Achievement According to Content Domain

CONTENT DOMAIN	% MEAN SCORE
Fraction	38.5
Geometry	45.34
Measurement	47.9
Number	47.3

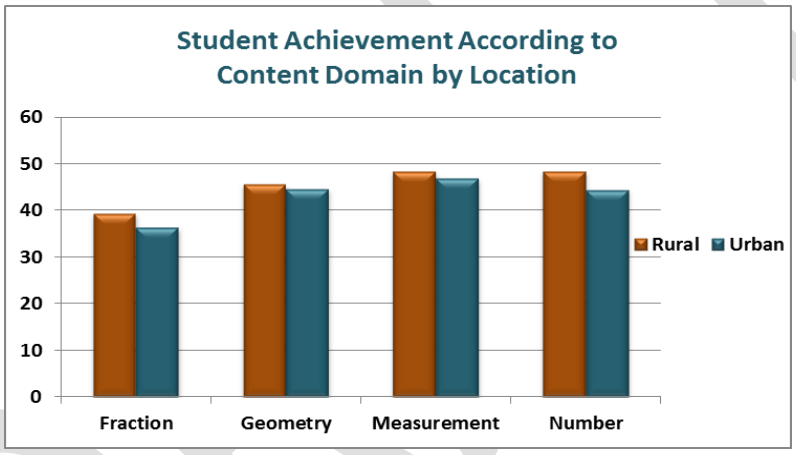


Student Overall Achievement According to Content Domain and Location

Overall and in the different aspects of mathematics students performed better in rural areas than urban areas.

Table 8: Student Overall Achievement According to Content Domain and Gender

CONTENT DOMAIN	LOCATION	
	Rural %	Urban %
Number	48.44	44.48
Fraction	39.4	36.46
Measurement	48.38	46.86
Geometry	45.64	44.7

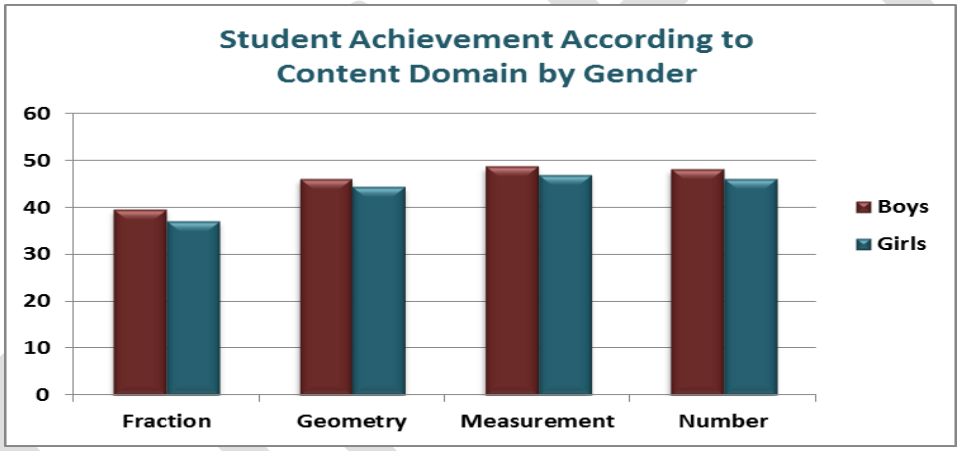


Student Overall Achievement According to Content Domain and Gender

Overall boys performed better than girls in mathematics in all aspects of the areas tested.

Table 9: Student Overall Achievement According to Content Domain and Gender

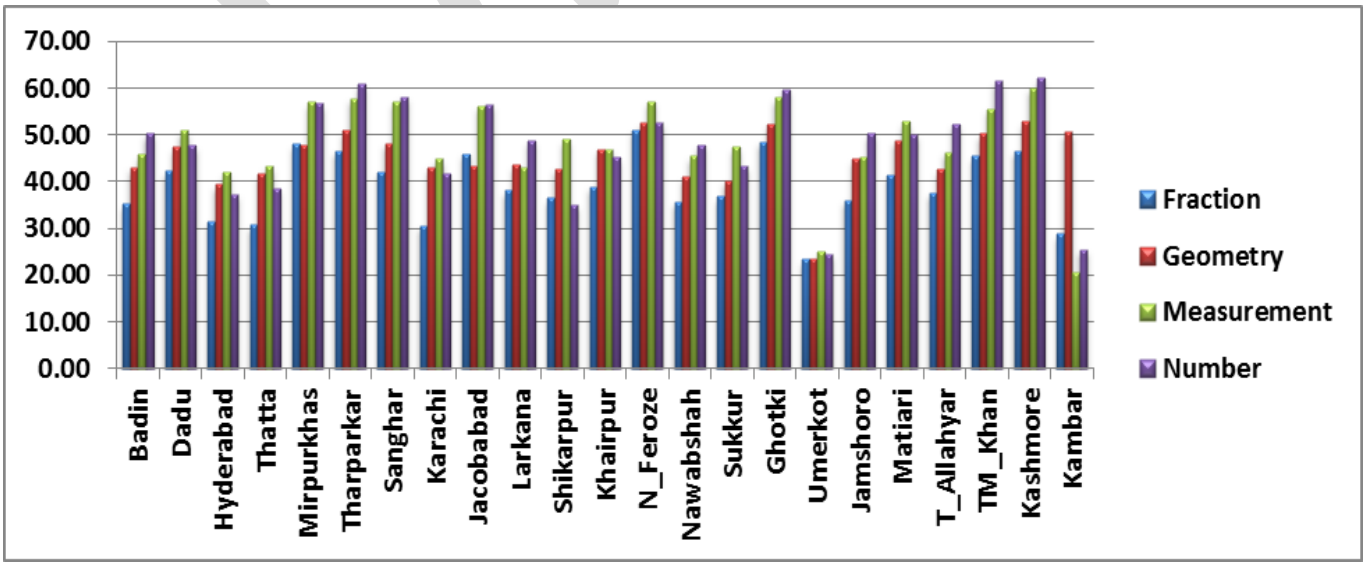
CONTENT DOMAIN	GENDER	
	Boys %	Girls %
Number	48.04	45.96
Fraction	39.5	37
Measurement	48.66	46.78
Geometry	46.04	44.28



Achievement by District According to Content Domain

Table 10: Achievement By District According to Content Domain

District Name	Fraction	Geometry	Measurement	Number	Mathematics
Badin	35.20	43.02	45.72	50.34	43.56
Dadu	42.24	47.32	50.88	47.74	47.04
Ghotki	48.50	52.36	57.92	59.68	54.62
Hyderabad	31.54	39.34	41.88	37.24	37.50
Jacobabad	45.92	43.18	56.24	56.56	50.48
Jamshoro	35.96	44.84	45.20	50.32	44.08
Kambar	28.86	50.60	20.50	25.38	31.34
Karachi	30.54	43.08	44.80	41.60	40.00
Kashmore	46.64	52.84	59.96	62.12	55.40
Khairpur	38.94	46.86	46.92	45.16	44.46
Larkana	38.16	43.50	43.14	48.76	43.38
Matiari	41.50	48.88	52.76	50.08	48.30
Mirpurkhas	48.10	47.78	57.00	56.68	52.40
N_Feroze	51.06	52.50	57.08	52.64	53.32
Nawabshah	35.62	40.96	45.52	47.92	42.50
Sanghar	42.18	48.14	57.16	58.14	51.40
Shikarpur	36.42	42.68	48.92	35.04	40.76
Sukkur	36.86	40.16	47.44	43.34	41.96
Tando_Allahyar	37.56	42.60	46.24	52.40	44.70
Tharparkar	46.64	50.88	57.76	60.78	54.02
Thatta	30.72	41.64	43.16	38.50	38.50
TM_Khan	45.44	50.28	55.58	61.44	53.20
Umerkot	23.34	23.50	24.96	24.48	24.08
Sindh	38.50	45.34	47.90	47.22	44.74

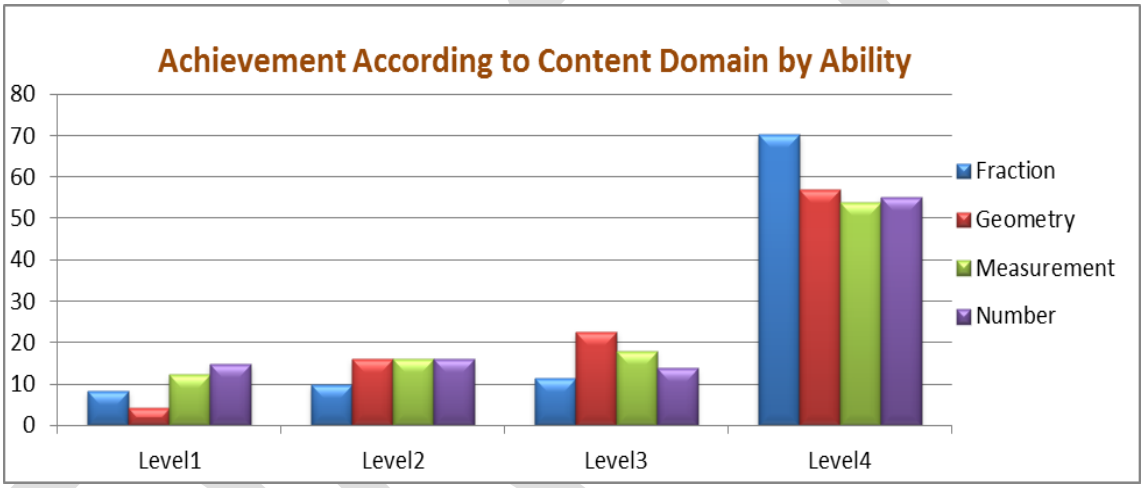


Achievement According to Content Domain and Ability

It can be seen from the table and graph that the majority of the students are achieving at Level 4 (having a score between 0% and 49%); at Level 3 students are achieving a score of between 50% and 64; at Level 2 they are achieving 65% to 78% and at Level 1 they are achieving scores of 80% and over.

Table 11: Achievement According to Content Domain and Ability

Content Domain	Level 1	Level 2	Level 3	Level 4
Fraction	8.35	10	11.36	70.29
Geometry	4.28	16.14	22.67	56.91
Measurement	12.2	16.17	17.92	53.7
Number	14.89	16.18	13.77	55.16

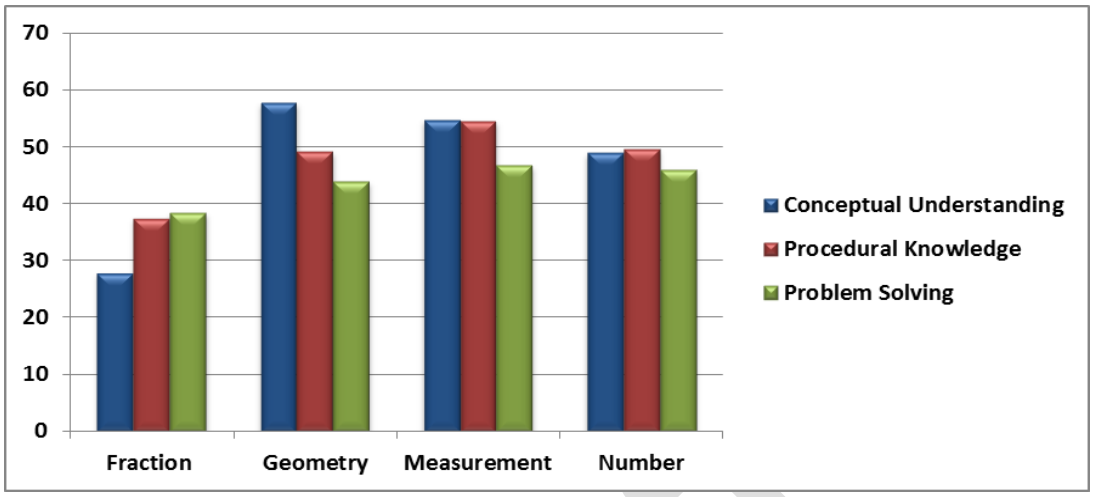


Student Overall Achievement According to Cognitive Domain

Overall students performed better on procedural knowledge test items, followed by conceptual understanding. Students' achievement was weakest in problem solving.

Table 12: Achievement According to Cognitive Domain

COGNITIVE DOMAIN	MATHEMATICS AREAS TESTED			
	Fractions	Geometry	Measurement	Number
Conceptual Understanding	27.85	57.79	54.83	48.98
Procedural Knowledge	37.49	49.19	54.55	49.59
Problem Solving	38.42	43.93	46.74	46.09

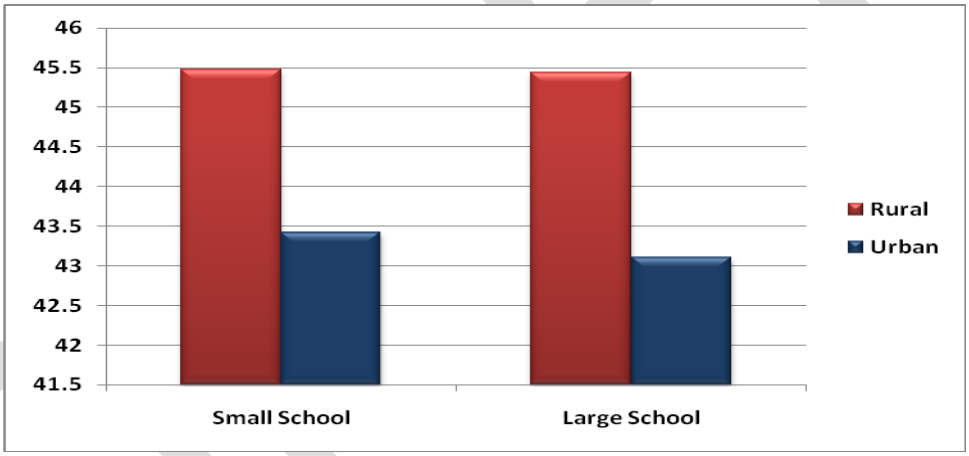

70
60
50
40
30
20
10
0
Fraction
Geometry
Measurement
Number
Conceptual Understanding
Procedural Knowledge
Problem Solving

Student Overall Achievement According to School Size and Location

Grade size whether in rural or urban areas did not appear to have much effect on student achievement.

Table 13: Achievement According to School Size and Location

Location	Small School	Large School	Difference	Probability	Sig/ N-Sig
Rural	45.48	45.44	0.04	0.96	Not significant
Urban	43.42	43.1	0.32	0.804	Not significant
Sindh	45.3	44.46	0.84	0.23	Not significant

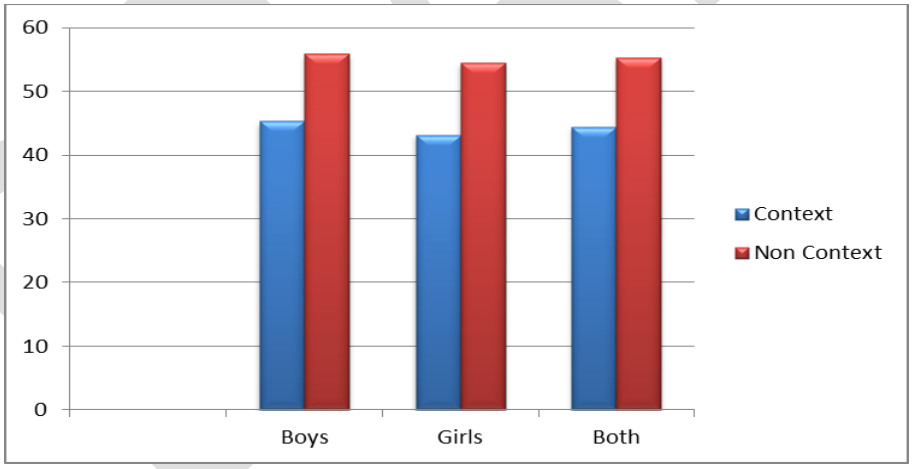


Achievement According to Test Item Type – Context or Non Context and Gender

Test items which were written in context were found to be the most difficult for the students, both boys and girls.

Table 14: Achievement According to Test Item Type- Context or Non-Context and Gender

Gender	Context	Non Context	Probability	Sig/ N-Sig
Boys	45.29	55.9	0.00	Highly Significant
Girls	43.18	54.51	0.00	Highly Significant
Both	44.45	55.35	0.00	Highly Significant

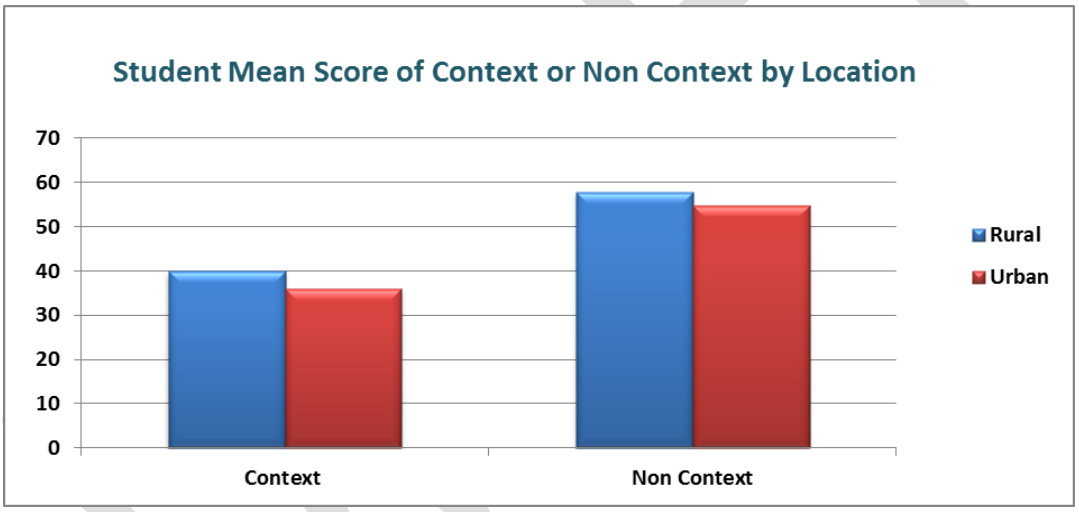


Achievement According to Test Item Type – Context or Non Context and Location

The difference between students' performance according to test item type was found to be highly significant with rural areas having a positive impact on student achievement of test items

Table 14: Achievement According to Test Item Type- Context or Non-Context and Location

Location	Context	Non Context	Difference	Probability	Sig/ N-Sig
Rural	39.96	57.88	-17.91	0.00	Highly Significant
Urban	35.99	54.86	-18.87	0.00	Highly Significant



Achievement of Test Item Competencies – Number

From the table and graph below it can be seen that students had greatest success in:

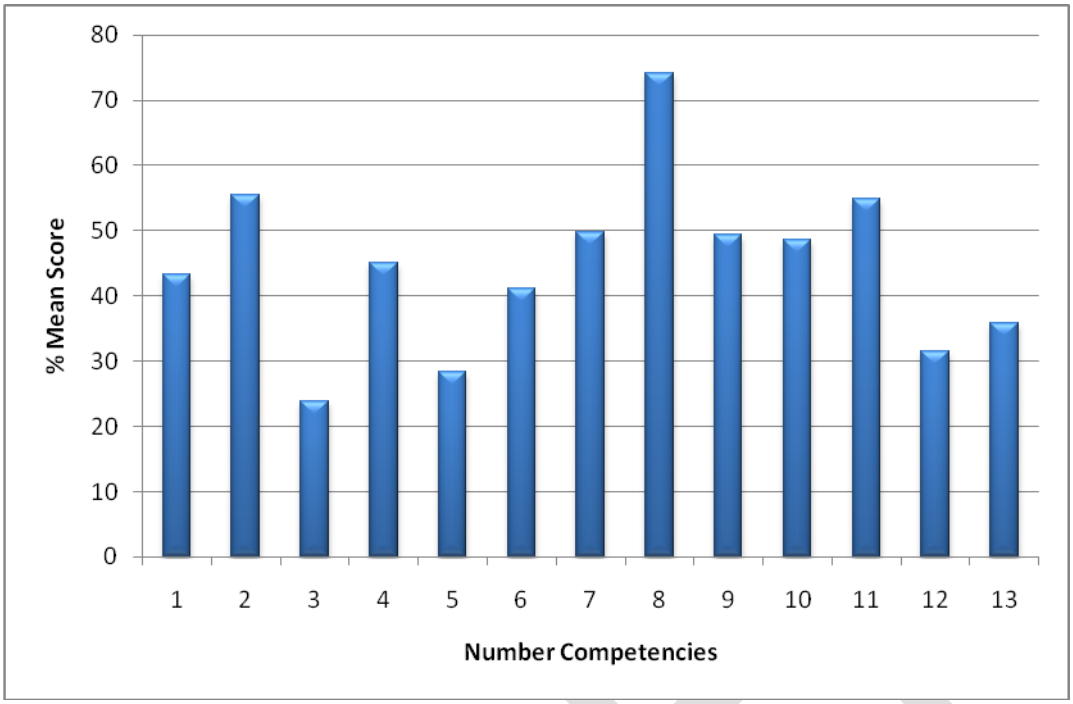
- adding together two numbers of three, four or five digits
- demonstrating an understanding of prime, even and odd numbers
- dividing two numbers of three, four or five digits

Students were least able to:

- demonstrate an understanding of the concept of a number line and how it is constructed
- sequence numbers from smaller to larger and larger to smaller
- estimate products and quotients
- solve simple word problems involving addition, subtraction, multiplication and division

Table 16: Performance of Students According to Number Competencies

	Number Competencies	% Mean Score
1	Student is able to read natural and Roman numbers and convert natural to Roman numbers and vice versa	43.23
2	Student is able to demonstrate an understanding of prime, even and odd numbers	55.55
3	Student is able to demonstrate an understanding of the concept of a number line and how it is constructed	23.86
4.	Student is able to identify smallest and largest numbers	45.09
5	Student is able to sequence numbers from smaller to larger and larger to smaller	28.30
6	Student is able to identify the place value of numbers within a figure (units, tens, hundreds, thousands; tenths, hundredths)	41.16
7	Student is able to translate written numbers into digital forms and digital numbers into written form	49.82
8	Students is able to add together two numbers of three, four or five digits	74.07
9	Students is able to subtract two numbers of three, four or five digits	49.44
10	Students is able to multiply one, two, three digit number by a single digit number	48.68
11	Students is able to divide two numbers of three, four or five digits	59.97
12	The student will be able to estimate products and quotients	31.45
13	The student will be able to solve simple word problems involving addition, subtraction, multiplication and division	35.74


80
70
60
50
40
30
20
10
0
% Mean Score
1
2
3
4
5
6
7
8
9
10
11
12
13
Number Competencies

Achievement of Test Item Competencies – Fractions

From the table and graph below it can be seen that students had greatest success in:

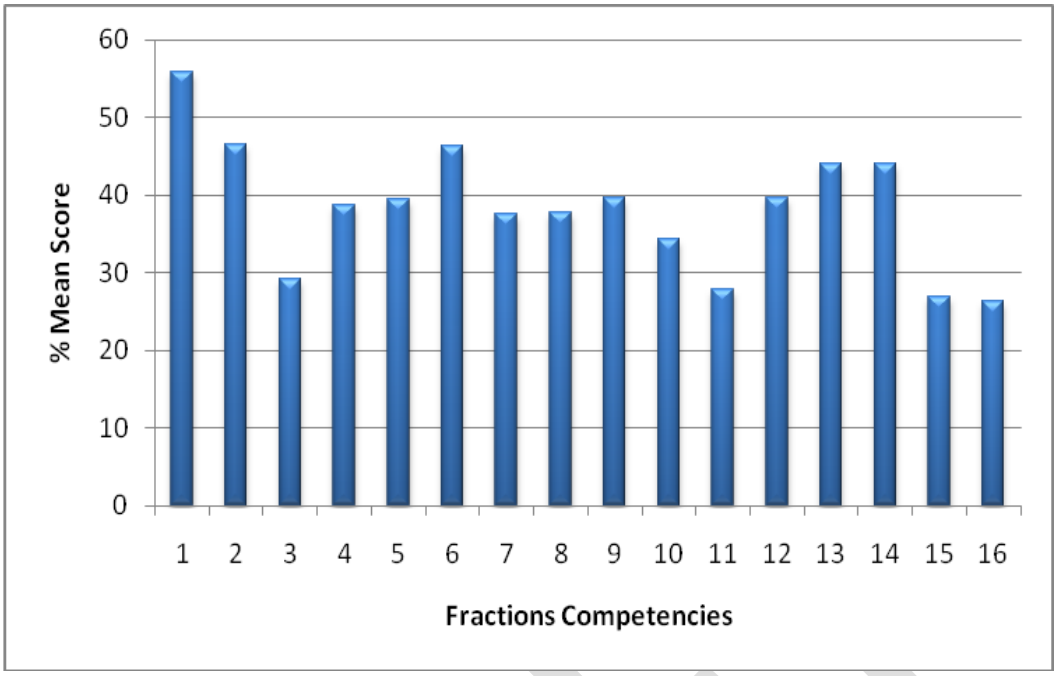
- identifying a fraction of a diagram that was shaded
- ordering fractions from smallest to largest
- simplifying fractions

Students were least able to:

- order fractions from largest to smallest
- divide numbers both fractions of a whole number, mixed fractions and decimal fractions
- solve simple word problems involving fractions of whole numbers
- solve simple word problems involving some combination of addition, subtraction, multiplication (fraction of a whole number) of fractions

Table 17: Performance of Students According to Fractions Competencies

	Fractions Competencies	% Mean Score
1	Student is able to identify a fraction of a diagram that is shaded	55.76
2	Student is able to order fractions from smallest to largest	46.44
3	Student is able to order fractions from largest to smallest	29.10
4	Student is able to identify equivalent fractions	38.69
5	Student is able to identify proper improper and, mixed fractions and decimal fractions	39.41
6	Student is able to simplify fractions	46.30
7	Student is able to add fractions together	37.60
8	Student is able to subtract simple and mixed fractions	37.74
9	Student is able to find a fraction of a whole number	39.63
10	Student is able to multiply numbers both fractions of a whole number and mixed fractions	34.40
11	Student is able to divide numbers, fractions of a whole number and mixed fractions	27.83
12	Student is able to convert fractions to decimal fractions	39.57
13	Student is able to add decimal fractions	43.95
14	Student is able to divide decimal fractions by whole numbers	43.95
15	Student is able to solve simple word problems involving fractions of whole numbers	26.88
16	Student is able to solve simple word problems involving some combination of addition, subtraction, multiplication (fraction of a whole number) of fractions	26.34


60
50
40
30
20
10
0
% Mean Score
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Fractions Competencies

Achievement of Test Item Competencies – Measurement

From the table and graph below it can be seen that students had greatest success in:

- adding units of money
- converting notes to coins and coins to notes
- adding simple weights
- identifying a.m. and p.m. in analogue time

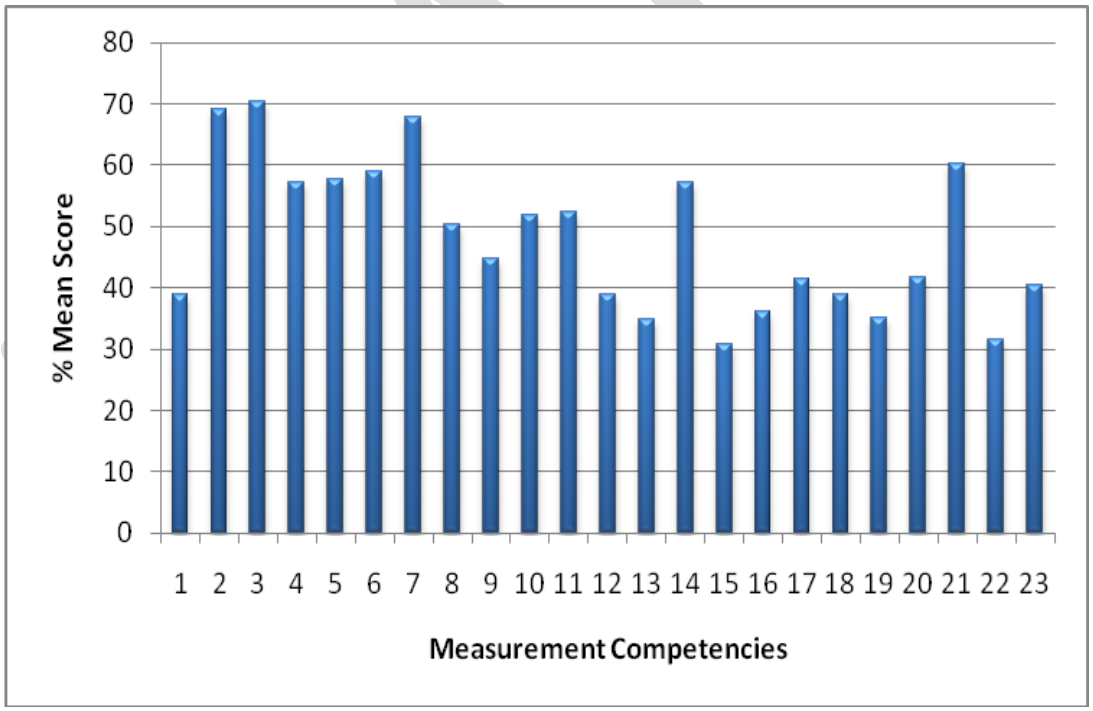
Students were least able to:

- identify the relationship between different forms of measurement
- choose correctly an appropriate unit for measurement of length
- convert millilitres to litres and vice versa

Table18: Performance of Students According to Measurement Competencies

	Measurement Competencies	% Mean Score
1	Student is able to identify different forms of currency	39.01
2	Student is able to convert notes to coins and coins to notes	69.20
3	Student is able to add units of money	70.32
4	Student is able to subtract units of money	57.16
5	Student is able to identify time according to analogue presentation	57.75
6.	Student is able to convert minutes to hours and hours to minutes; days to hours and hours to minutes	58.91
7	Student is able to identify a.m. and p.m. in analogue time	67.73
8	Student is able to the meaning of fast and slow in relation to time	50.20
9	Student is able to translate 12 hour time to a 24 hour clock	44.83
10	Student is able to identify the number of days in a week, months in a year and the number of days in each month	51.86
11	Student is able to calculate the differences between times in simple problems	52.46
12	Student is able to estimate the length and height of an object	39.00
13	Student is able to choose correctly an appropriate unit for measurement of length	34.74
14	Student is able to conduct simple multiplications of	57.27

	measures	
15	Student is able to identify the relationship between different forms of measurement	30.82
16	Student is able to convert lengths from millimetres to centimetres and from cm. to mm.	36.00
17	Student is able to add different lengths of the same measure	41.40
18	Student is able to subtract different lengths of the same measure	38.84
19	Student is able to convert millilitres to litres and vice versa	35.00
20	Student is able to add and subtract litres and millilitres	41.57
21	Student is able to add simple weights together	60.33
22	Student is able to convert km to grams and km to quantiles and choose appropriate procedure	31.45
23	Student is able to solve simple word problems involving money, time, length, weight, liquids	40.40



Achievement of Test Item Competencies – Geometry

From the table and graph below it can be seen that students had greatest success in:

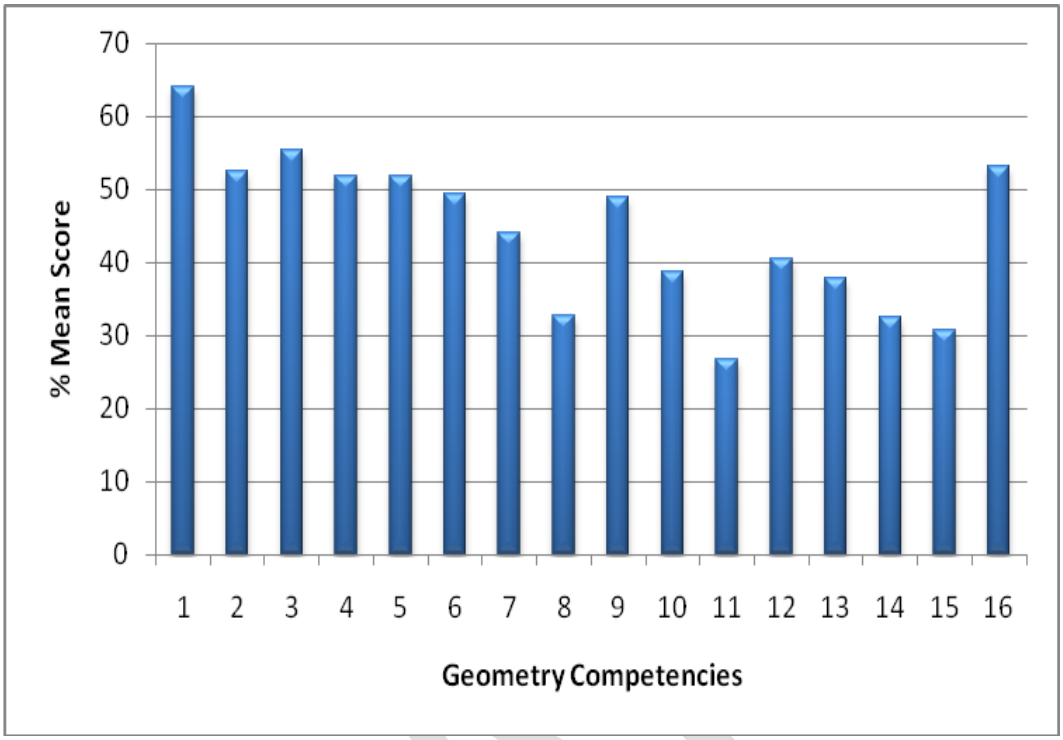
- identifying common shapes
- identify flat and curved faces on solid figures
- understanding the use of geometric instruments
- identifying similar triangles

Students were least able to:

- understanding the steps in drawing a square and rectangle
- measuring the radius and diameter of a circle
- estimating the size of angles

Table 19: Performance of Students According to Geometry Competencies

	Geometry Competencies	% Mean Score
1	Student is able to identify common shapes	64.12
2	Student is able to identify a right angle	52.51
3	Student is able to identify flat and curved faces on solid figures	55.36
4	Student is able to identify similar triangles	51.79
5	Student is able to identify different types of triangles	51.77
6	Student is able to identify types of angles - right angles, acute and obtuse angles	49.36
7	Student is able to compare angle size	44.02
8	Student is able to estimate the size of angles	32.61
9	Student is able to identify characteristics of a line, ray, line segment, vertical and horizontal lines	48.88
10	Student is able to demonstrate an understanding of perimeters and their measurement	38.72
11	Student is able to identify geometric shapes	26.77
12	Student is able to measure angles	40.53
13	Student is able to identify the angle properties of triangles	37.90
14	Student is able to measure the radius and diameter of a circle	32.53
15	Student is able to demonstrate an understanding of the steps in drawing a square and rectangle	30.60
16	Student is able to demonstrate an understanding of the use of geometric instruments	53.25


70
60
50
40
30
20
10
0
% Mean Score
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Geometry Competencies

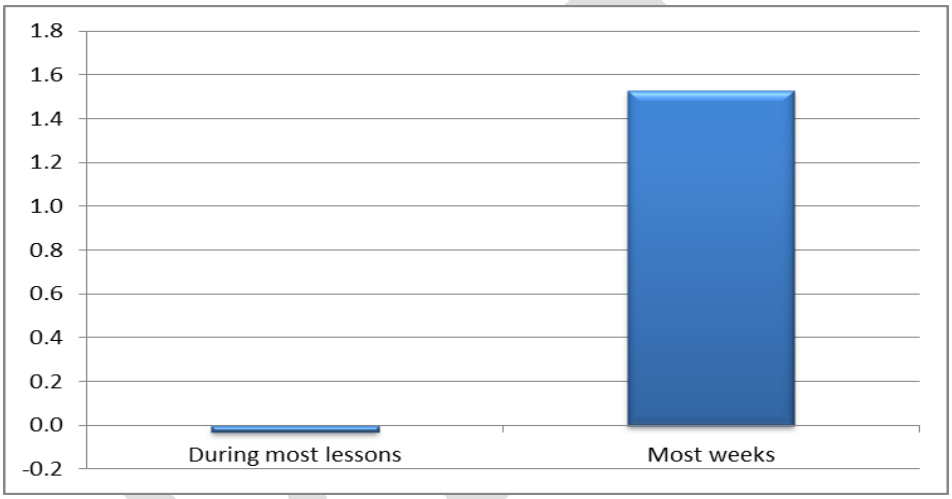
6.2.2 Background Questionnaire Findings

Through data cleaning much of the information found in the Background Questionnaires was discarded. However the following information was found to be highly significant.

The amount of time spent by students in small groups being taught by the teacher

Table 20: Amount of Time Spent Being Taught in Groups

STUDENT RESPONSE	% MEAN SCORE
During most lessons	0.0
Most weeks	1.53



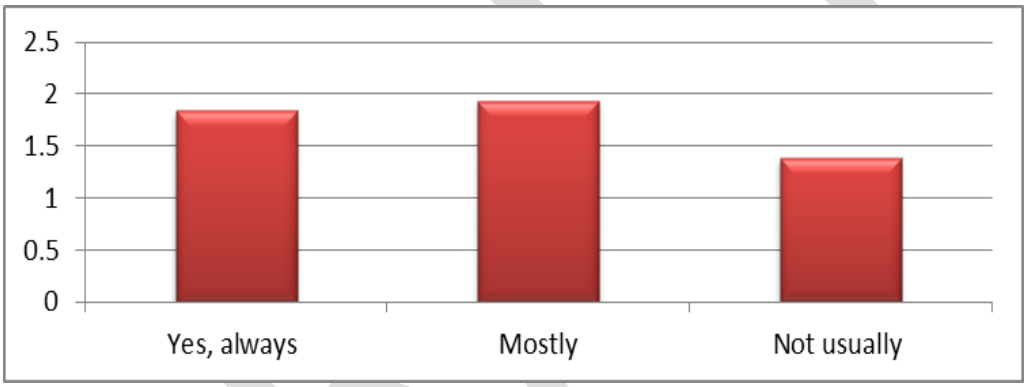
From this it can be observed that being taught in groups always had a negative effect on learning achievement while being taught in groups on a weekly basis was more likely to result in higher achievement.

Students asked to explain their answers during mathematics lessons

It can be observed that students asked to explain their answers achieved better scores in the tests than students who were not questioned regarding their answers.

Table 21: Students Being Asked to Explain their Answers

STUDENT RESPONSE	MEAN SCORE
Yes, always asked about the process to get the answer	1.84
Mostly asked about the process to get the answer	1.94
Not usually asked about the process to get the answer	1.39



Other Background Information

Information was obtained regarding the following:

- Time spent with parents or adults reading and talking about books
- Time spent with parents playing sport or games or keeping fit
- Mathematics lessons where mathematics is done on short tests
- Mathematics lessons where regular homework was provided
- Information regarding the family's perception of the importance of mathematics
- Time spent working on their own
- Using real-life examples to work with

However none of this information proved to be significant for student achievement.

9. Constraints

The development in implementing and analyzing a provincial assessment is a complex and challenging task. It involves the training of staff in curriculum analysis and mapping, the development of test frameworks and items, and the organization of the assessment instrument booklets as well as the printing, collating, distribution and collection of the assessment instruments after testing under tight deadlines. The Grade 4 assessment instruments were piloted in November 2008. These were marked and coded by the PEACE team. Data entry was done by the Bureau of Curriculum and item analysis was carried out by the PEACE subject specialists. Between November 2008 and the April 2009 the PEACE team in Jamshoro, with the help of EU technical assistance, reviewed and developed, collated and distributed the assessment instruments and undertook the training of the test administrators and monitors. After the large scale testing in April 2009, the assessment instruments were returned to the district offices where marking and coding took place and the data was sent to PEACE. Data cleaning and statistical analysis was undertaken with the help of EU technical assistance between April 2009 and March 2010.

The **main constraints** identified are listed below:

The Sample

- Difficulties of using the Grade 3 enrolment from the 2007/08 Census of Schools as the measure of size (MOS) for the PPS sampling of schools. The Grade 3 enrolment was used as measure of size because the current Grade 4 students were in Grade 3 at the time of the 2007/08 Census of Schools. However, this did not provide accurate information. For example it resulted in some schools which had been identified in 2007/08 Census having 0 enrollment and also schools which were identified as closed, not being included in the school sample despite having an appropriate student population.
- Due to time constraints PEACE was not able to carry out sufficient checking/monitoring of the sample selection.

The Assessment Instruments

- PEACE did not have the services of a full-time In-Page composer for producing the Urdu tests in camera ready form. An In-Page composer is necessary to develop the test items in Urdu. Much of the composing work was done by the subject specialists and by the provision of short term support of a composer from the Bureau of Curriculum.
- There was a delay in the printing of the materials due to the time constraint in which all the activities had to be completed;

- The Mathematics test had some misprinting due to the time constraints and the lack of monitoring of the instruments resulting in no corrections being made. This had implications for the statistical analysis of the data.

Statistical Analysis

- PEACE did not have the services of a full-time education psychometrician for analyzing the data. This meant that the subject specialists conducted the analysis through on the job training. This meant that it was time consuming and there were some delays in the finalization of the analysis. This was the first time the group had undertaken such an analysis from a complex survey.

Logistics

- Time constraint. It was difficult to ensure that all the tasks were completed on time in a rigorous manner;
- There was a delay in the statistical analysis due to the delay in cleaning the assessment data with the SAS software;
- The transportation of the assessment materials to arrive in the districts for testing to take place was difficult. The districts also experienced difficulty in ensuring that the assessment instruments reached all the sample schools on time for testing to take place. This was especially true in the remote areas of the province; where insufficient materials were received photocopying of the tests was done to overcome the difficulty.
- The collection of the assessment instruments after testing had taken place was also difficult especially in remote areas;
- The test administration training in many centres was hampered due to the delay in the arrival of the training materials;
- There was a lack of appreciation for the need for the assessment materials to be kept in a secure location and for all the assessment instruments to be returned to PEACE;

Test Administration

- There was little monitoring and accountability undertaken by the districts or PEACE regarding the number of booklets that had been provided to each school and district for testing and the number of booklets that had been returned to the districts or PEACE.
- There was a need for written Guidelines to be provided to the focal persons to ensure that they fully understood their role in the test administration

Management

- The monitoring tool was too simple for effective monitoring to take place– it required only a yes or no answer and issues were only rarely identified;
- District focal persons did not always carry out their duties to the full even after briefing by PEACE; guidelines had not been provided for them.

Staffing

- There was not sufficient staff in PEACE to ensure that the assessment was conducted in an efficient and timely manner. Many of the staff members were assigned multiple tasks due to this constraint;

- Some of the Executive District Officers (schools) demonstrated little interest in the ongoing assessment process: they did not respond to correspondence made in connection with the nomination of test administrators and the test administrators were not informed in sufficient time to attend the training;

Equipment

- There was insufficient equipment (computers, photocopiers) available in PEACE for composing, the duplication of materials for training purposes etc.;
- In marking and coding the scripts there was not sufficient space available for storing the assessment materials.

10. Lessons Learnt

- The tightness of the testing programme requires the availability of a full time In-Page and Sindhi composer
- The need for the employment of an independent editor to review the tests before they are printed and also after the first sample printing has been completed in order to ensure that there is no misprinting;
- The employment of back translators to ensure that the Urdu and Sindhi test instruments and Background Questionnaires are testing the same aspects;
- The training and assessment instruments should be distributed to the district offices and training centres in a timely manner to enable the administration of the assessment instruments to take place efficiently;
- The need to emphasise that the assessment instruments should be kept securely and that all the assessment instruments, whether used or unused, should be returned to PEACE.
- The need for monitors to be more able to address day to day problems;

- Quality Assurance procedures are required to be put in place for all testing activities;
- The need to disseminate information and results regarding testing to all the stakeholders to enable them to understand the importance of conducting provincial assessments and to encourage them to assist in the activity.

11. Implications

The results of the tests have implications for the development of a quality education in Sindh province. Strategies and action plans need to be developed to:

- Identify requirements and strategies and plan for improvements in student learning;
- Interpret the National Curriculum according to the needs of the province;
- Make assessment **for, as** and **of** students' learning central to the development of improved teaching and learning methodologies;
- Develop supplementary materials to support student difficulties and teachers' teaching;
- Improve textbook development in line with the 2006 National Curriculum standards and competencies;
- Improve Teacher Training and Teacher education Development;
- Improve the roles of management in the districts to mentor and advise teachers in a supportive manner.

ANNEXURES

Annex 1

List of Provincial Working Group Members

Chairman: Secretary Education

Members:		
	Secretary Education	Chairman
	Director Bureau of Curriculum	Member
	Additional Secretary (A&T) Education	Member
	DG PITE	Member
	Chairman of the Textbook Board	Member
	Deputy Programme Manager, PEACE, RSU	WG Secretariat
	Registrar, University of Karachi	Member
	Coordinator, PEACE	Member
	Deouty Director BoC	Member
	Subject Specialists, PEACE	Member*

EU SERTA Quality Education Adviser (Ms L. Jones)
EU SERTA Statistical Analysis Adviser (G H Choudhry)
WB Education Specialist (A. Norley)
WB Education Specialist (Ms Umbreen Arif)
USAID PRESTEP Education Specialist (Ms Ratan)

*Depending on subject and/or availability

Annex 2

Mathematics Test Specifications

Below are found three test specifications.
Table 1 shows the specifications obtained from the National Curriculum 2006.
Table 2 shows the test items developed for pilot testing.
Table 3 shows the test specifications according to content areas and abilities

Table 1: Table of Specifications According to the National Curriculum 2006

Mathematical Abilities	National Curriculum Content Areas				
		Numbers (%)	Fractions (%)	Measurement (%)	Geometry (%)
	Conceptual Understanding	40	40	41	50
	Procedural Knowledge	45	53	43	45
	Problem Solving	15	7	16	5
	Total	100	100	100	100

Table 2: Number of Items Developed for Pilot Testing

Mathematical Abilities	Curriculum Content Areas				
		Numbers	Fractions	Measurement	Geometry
	Conceptual Understanding	55	59	156	98
	Procedural Knowledge	50	67	122	55
	Problem Solving	7	13	25	10
	Total Number of Items	112	139	303	163

Table 3: Test Specifications According to Content Areas and Abilities for Large-Scale Testing

Mathematical Abilities	Curriculum Content Areas				
		Number (%)	Fractions (%)	Measurement (%)	Geometry (%)
	Conceptual Understanding	36	30	35	47
	Procedural Knowledge	49	50	45	42
	Problem Solving	15	20	20	11
	Total % of Items	100	100	100	100

Annex 3

Test Instrument Development

Item writing for the 2009 Provincial Assessment was conducted in 2008 when Mapping and Item Writing Workshops were held.

The workshop objectives were as follows:

- To develop technical capacity and sustainability for test development in PEACE;
- To map the National Curriculum in the areas of number, measurement, fractions and geometry
- To identify and understand the Grade 4 competencies and the hierarchy of abilities;
- To develop Tests Specifications according to the weightage identified in the curriculum documents;
- To develop test items for Mathematics

The participants in the Workshops were members of the Bureau of Curriculum, PITE, mathematics teachers and PEACE subject specialist.

The workshops were activity based with all participants being involved. The outcome of the workshops was the development of **mathematics maps, test specifications and items** (350 for each aspect of mathematics to be tested). These competencies and test items were used for the pilot testing of mathematics and for large scale testing of mathematics in 2009.

The **assessment framework** developed was the foundation for the provincial assessment and was the basis for all item development. The assessment framework consisted of two organising dimensions – the content dimension and the cognitive dimension. The content domains define the specific subject matter covered by the assessment, and the cognitive domains define the sets of behaviours expected of students as they engage with the subject content (e.g. knowing facts and procedures, using concepts, solving problems, reasoning). Each content domain has several topic areas (e.g., number is further categorized by whole numbers, fractions and decimals, integers, and ratio, proportion, and percent; reading is further categorised by reading for information, reading to find the main idea, identifying genres, vocabulary, punctuation etc.).

On the basis of these domains and the competencies developed, test specifications for mathematics were prepared and test items developed. The specification table provided a guideline to the development of a comprehensive reliable, valid and practical test for the

pilot testing in 2008. The same tests were later reviewed and finalised by a mathematics working group.

The development of test specifications and framework are essential if the testing activity is going to measure the elements for which it is being constructed. A test has to have a clearly stated purpose and should clearly describe the content areas and the Grade level for which it has been developed. Also the length of time required for the test should be determined as this would have a direct effect on the number of items in the test and also the breadth of the curriculum to be tested.

The development of a test specification ensured that the test measured a representative sample of the curriculum content for the areas to be tested and its objectives. It ensured that the curriculum content was more likely to be assessed in a balanced way. Writing a test specification required:

- ✓ A list of all the instructional objectives and hierarchy. The framework is developed on the basis of three levels of achievement – knowledge, comprehension, application according to Bloom's Taxonomy. Knowledge consists of the simple recall of specific information; comprehension enables the student to demonstrate their understanding of the knowledge obtained and application is the student's ability to use their knowledge and understanding in different contexts and situations.
- ✓ The content of the subject areas.
- ✓ The weightage to be given to each of the instructional objectives.

From this framework, curriculum competencies and test specification development, **test items** were written. The test items consisted of multiple choice items. The main reason for this was that multiple choice items are objective test items and are easily marked and scored. Examples of the test items used in the tests are found in Annex 7 along with a statement of the competencies they assess.

Pilot test in mathematics were developed from the pool of items. Two tests, of 50 test items each for each area to be tested were designed. Items which tested the key competencies in the identified mathematical areas and those competencies that were able to be tested in a pencil and paper test were to be included. The weightage given to the specific content areas was according to the weightage given in the 2006 National Curriculum.

The tests were further reviewed and printed for pilot testing in 2008. Pilot testing was required for the PEACE to ensure that the demands of the tests were appropriate and also to identify items which were reliable, valid and discriminated appropriately between the different abilities of students.

For the piloting, 46 schools were selected from 23 districts (2 schools from each district)

A sample of 1150 students took part in the pilot testing in 2008 in mathematics.

From the result of the **statistical analysis** of the pilot items (using ITEMAN software), items were selected and additional items developed to "fill in the gaps". After item selection and writing additional items, formats for large scale testing were developed and administered in April 2009 to a provincial sample, to establish a baseline for Grade 4 mathematics achievement.

The items were then organized into two test booklets. Two test booklets were developed so that students would not be able to copy from each other. The items were ordered in each sub-content area, from easy to difficult.

The **Background Questionnaires** for Head Teachers, teachers and students were also developed. These questionnaires focused on such things as school conditions and climate; teachers and teaching practices; Supporting Inputs for Schools; and, students' home backgrounds. Difficulties were found in constructing some of the questions, as well as in ensuring sufficient coverage of each background and context variable in relation to the length of the questionnaires and the time it would take for the personnel in the sample schools to complete them.

Marking and coding and data entry was completed using manual recording of scores, Excel and SAS software to finalise the data for statistical analysis.

Statistical analysis was completed using the SAS, BILOG_MG and WesVAr software.

Some of the **difficulties** experienced in instrument development are listed below:

- Ensuring that there were sufficient items to cover the major part of the curriculum;
- Difficulties in spiraling items so that students in the different booklets were given similar items at the same stages in the booklets;
- The time taken to answer the background questionnaires;
- Difficulties in the distribution of the booklets, Test Administrator Guides and conducting the TA training

Annex 4

Developing the Provincial Test Sample

A stratified two-stage sample design was used for selecting the sample of students who were enrolled in Grade 4 in the government schools in the province of Sindh during the 2008/2009 academic year. The objective of the study was to conduct analyses simultaneously for the Province, location type (Rural/Urban), boys/girls within the province, and the 23 Districts in the province. Therefore, the stratification was defined by cross-Gradeification of District by location type (Rural/Urban) resulting in 46 design strata. At the first stage, schools (or clusters of schools) with Grade 4 Gradees were selected with Probability Proportional to Size (PPS) systematic sampling, and at the second stage students which were the ultimate sampling units were selected with systematic sampling.
The measure of size (MOS) for the PPS sampling of schools was the Grade 3 enrolment from the 2007/08 Census of Schools. The Grade 3 enrolment was used as measure of size because the current Grade 4 students were in Grade 3 at the time of the 2007/08 Census of Schools. Since the 2007/08 Census data was only one year old, good correlation between the current Grade 4 enrolment and the MOS from the Census was expected.

All government schools in the province of Sindh with Grade 4 Gradees during 2008/09 were part of the target population. Although, the *desired target population* was the population comprising all Grade 4 students in the government schools in the province, it was not cost effective to sample very small schools, in particular in the rural areas where travel costs are very high. Therefore, very small schools although in the target population were consciously excluded. The remaining schools formed the *survey population*. Exclusions were kept to a minimum and used as a means to reduce cost while still selecting as close as possible to a representative sample. International studies have routinely set the upper limit of exclusions at 5.0 percent of the desired target population. Since the analysis was needed by rural/urban schools, the cut-off value used the criteria that the percentage of students excluded from the survey was less than 5 percent both in rural and urban schools. The survey population was therefore all government schools in the province for which Grade 3 enrolment from the 2007/08 Census of schools was greater than or equal to 4 students. The survey population was approximately 447,000 students with almost 70 percent in the rural schools.

As required, the analysis from the 2008/09 survey of Grade 4 students was conducted at the Province, location type (Rural/Urban) within the province, and District levels. The sample of schools was allocated to the 46 design strata defined by cross-Gradeification of the 23 districts and location type (Rural/Urban). In order to conduct analyses for the Province, Rural/Urban type and the 23 Districts the sample size was roughly 4,000 schools (PSUs) with 10 students selected per sampled school resulting in an overall

sample of about 40,000 students. The above sample size was arrived at by using the criteria that there was a need for an effective sample of about 300 students for the smallest district. The effective sample size is defined as the actual number of students selected divided by the design effect[3]. The typical design effect for education studies is around 3 to 4. Therefore, the actual sample for the smallest district would be more than 1,000 students. The sample allocation to the two stages of sampling was determined on the basis of cost and variance consideration. First, the total sample was allocated to the 46 strata defined by cross-Gradeification of district and Rural/Urban type. Then, the sample within each stratum was allocated to the two sampling stages, i.e. number of schools to be selected, and number of students to be selected from each sampled school (or cluster of schools).

The number of students enrolled in Grade 3 obtained from the 2007/08 Census was the basis for allocating the sample across districts and Rural/Urban type. Since the sample had to be allocated simultaneously to the Province, Rural/Urban type within the province, and the 23 Districts in the province, a compromise allocation was used to allocate the total sample. This was aimed at striking a balance between conducting analyses simultaneously for the Province, districts and the rural/urban type.

The optimum number of students to be sampled per school was at most 10 students per school. The number of schools to be sampled from each primary stratum (i.e., Rural/Urban within each district) was computed by dividing the sample size in terms of number of students (obtained by raking procedure) by 10. The district with the smallest sample size was T.A. YAR with sample size of 103 schools. The sample size in terms of number of students was approximately 1,030 students in the smallest district.

The sample of the required number of schools was selected from each stratum with probabilities proportional to size (PPS), using the systematic sampling algorithm described in Hansen, Hurwitz, and Madow[4] (1953). The measure of size (MOS) to be used for sample selection was the number of students in Grade 3 determined from the 2007/08 Census of Schools. The number of students in Grade 3 was used as the MOS because these students would be in Grade 4 at the time of testing and Grade 4 was the target population.

It was important that the schools were sorted by Tehsil and Gender (Boys vs. Girls schools) within strata (the Rural and Urban parts of the Districts), and then by MOS by alternating between “ascending” and “descending” orders from one Gender type to the next. It should be noted that a school with 50 percent or more boys was defined as a Boy’s school, and the one with less than 50 percent boys was defined as Girl’s school.

[3] The design effect is defined as the ratio of the variance under simple random sampling and the variance under the design that was actually implemented when the sample sizes are the same.

[4] Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953), *Sample Survey Methods and Theory,* John Wiley and Sons

As mentioned previously the schools with MOS (i.e. number of students in Grade 3 from the 2007/08 Census of Schools) less than or equal to 3 were not included in the survey population as it would not be very cost effective to sample very small schools. Ideally, the schools with MOS between 4 and 9 should have been collapsed. But, collapsing too many schools would result in operational issues, e.g. transporting students from several small schools to one test centre. As a compromise, the schools with MOS equal to 4, 5 or 6 were collapsed with neighboring schools within the same Union Council before sampling but the schools with MOS equal to 7, 8 or 9 were not collapsed. Thus, our primary sampling unit (PSU) was a cluster of schools instead of an individual school such that the PSU would have a minimum MOS of 7 students. The collapsing of the small schools would have been done using the criteria of minimum distance if Geographic Information System (GIS) was available. In the absence of GIS the small schools were collapsed within the Union Councils. A PSU that was a cluster of schools was treated as if it was a single school.

Conversely, if it happened that a school was so large that the corresponding selection probability became greater than one (selection probability must always be less than 1) it was decided to divide the original large school into a number of pseudo schools by a "conceptual split" where each pseudo school would be considered to be of the same size. The school was still one "physical" school and a 2nd stage sample of 10 students was selected from each sampled pseudo school. A "weight adjustment" was to be applied to account for the "conceptual split" because the original school would now represent two or more pseudo schools. The weight adjustment factor was equal to inverse of the number of pseudo schools the "large" school was split into. If two or more pseudo schools got selected from the same "physical" school, then a separate sample of students was selected from the same "physical" school to represent each sampled pseudo school.

This sampling methodology of "conceptual split" was implemented so that the same survey processing system could be used for all schools including the "large" schools. It should be noted that a "large" school is not only large relative to other schools in the stratum but it also depends on the number of schools to be sampled from the stratum. Therefore, a school with certain MOS could be "large" in one stratum but a school in another stratum with even a greater MOS may not be "large" in that stratum.

The sample was designed using "EXCEL" and the sample of schools was selected using EXCEL as well.
The samples of students were selected by systematic sampling procedure by sorting the list of students in Grade 4 by section and by roll number within a section. Where the Grade list was less than or equal to 10 students, all students were selected. Otherwise, a sample of 10 students was selected with systematic sampling procedure. The systematic sampling procedure was implemented by providing the test administrators with random number tables with 10 sequence numbers out of the list of sequence numbers of the

Grade IV students in the school. The random number tables were generated from 11 up to some maximum number of students (e.g. 400 students) in Grade 4 in the sampled school.

After the data collection and editing phases of the survey, the sampling weights for the data collected from the sampled students were constructed so that the responses could be properly expanded to represent the entire population of Grade 4 students in the government schools in the province of Sindh. The weights were the result of calculations involving several factors, including original selection probabilities, adjustment for non-response including both school non-response and student non-response, post-stratification adjustment based on the Grade 4 population of boys and girls within urban and rural parts in each district obtained from the 2008/09 Census of Schools.

Non-response is always present in any survey operation, even when participation is not voluntary. Thus, weight adjustment was necessary to account for the non-respondent schools and students. The non-response adjustment for the non-respondent schools was applied at the stratum level (calculated as the ratio of the MOS of schools (or clusters of schools) selected from the stratum and the MOS of those that participated in the assessment tests); whereas the non-response adjustment for the non-respondent students was applied at the school level (the weight adjustment was the ratio of number of sampled eligible students and the number that actually completed the assessment tests).

The base weight (or design weight) for each student was equal to the reciprocal of its probability of selection. The conditional selection probability of the student was equal to the number of students sampled divided by the number of students enrolled in Grade 4 in the school. An adjustment was made for the “large” schools that were split into pseudo schools. The adjustment factor to account for the “conceptual” split was equal to the number of pseudo schools that the large school was split into.

The post-stratification adjustment was applied by benchmarking the survey estimates for boys and girls within each primary stratum (District by rural/urban) to the enrollment obtained from the 2008/09 Census of schools.

The final survey weights for the respondent students were obtained as the product of the base weight, the two adjustment-factors for non-response (i.e. adjustment factor for non-respondent schools and adjustment factor for non-respondent students), and the post-stratification adjustment.

All survey estimates were obtained as domain estimates. The estimation domain was either a geographic domain (e.g., a district) or a characteristic domain (e.g., boys/girls). The estimation domain could also be the intersection of two or more geographic and/or

characteristics domains, e.g. all boys in a particular district who achieved more than 80 percent scores. An indicator variable was used so that all estimates were expressed as "province" level estimates. The indicator variable automatically excluded those students that were not part of the estimation domain. The indicator variable technique ensures the proper estimation of variance.

Because the estimates were based on sample data, they differ from figures that would have been obtained from complete enumeration of the population of students using the same instrument. Results were subject to both sampling and non-sampling errors. Non-sampling errors included biases from inaccurate reporting, processing, and measurement, as well as errors from non-response and incomplete reporting. The non-sampling errors occurred at various phases of the survey process. However, to the extent possible, each error was minimized through the procedures used for data collection, editing, quality control, and non-response adjustment. The variances of the survey estimates were used to measure the sampling errors.

Quality assurance procedures were recommended and training was provided but these quality procedures were never implemented at all phases of the survey process. The main steps recommended for implementation for the QA procedure were: concept of a batch; verification of a sample of units from the batch; criteria to accept or reject the batch based on the observed error rate. Finally, the in-coming and out-going error rates were to be computed from the verification of the QA sample.

Some of the sampling issues identified were:

- The sample was selected using EXCEL. This did not automatically identify discrepancies such as, duplications of SEMIS Codes, replication of schools etc. During the identification of the sample no difficulties were identified. Few checks were conducted by the PEACE staff due to the lack of time available as the testing programme was required to be conducted within strict times with on-the-job training. Also PEACE did not have sufficient capacity or the available budget to enable rigorous monitoring and quality control.

- Problems with the sample were identified through information provided by the Test Administrators and through using SAS for analysis.

- This lack of implementation of quality procedures resulted in various types of errors being introduced:
  - ✓ Some sample schools which showed 0 enrollment were found to have sufficient number of students for testing
  - ✓ Some sample schools which showed high enrollment were found to have 0 enrollment

- ✓ Some schools which were required to be split because of large Grade enrolment, were not identified
- ✓ Approximately three schools which had been identified as requiring to be collapsed were found to be too large after collapsing and were then split (this should not have been done)
- ✓ Some school not in the identified sample were tested and the data received by PEACE
- ✓ Some schools which were in the sample did not send their data to PEACE
- ✓ Some schools which had been identified in 2007/08 Census having 0 enrollment and also schools which were identified as closed, were not part of the school sample despite having an appropriate student population

All of these errors resulted in the database being reduced.

The following recommendations are made for improvement:

- SAS software should be used for selecting the sample of schools as SAS code can be easily developed to identify and flag any discrepancies in the data.
- The technique developed during the sampling workshop to identify enrollment discrepancies, where the ratio of the student enrolment at the time of survey and the MOS is lower than 0.5 or higher than 1.5 should be used for further follow-up must be implemented. Where the ratio is found to be <0.5 or > 1.5 additional field checks will need to be made to ensure the reliability of the enrolment of the sampled schools.
- If the measure of size, which is the basis of the sample design, was larger than some threshold value, say 20, and it has changed greatly at the time of test administration it should be investigated. For example, if the MOS was 20 and it changes to 50, or if the MOS was 100 and changes to 20, these are serious discrepancies and should be investigated.

Annex 5

The Grade 4 Mathematics Analysed Provincial Sample

The following table identifies the proposed number of schools to be sampled, the allocation from sample design, the sample schools selected, and the number of PSUs in the analysis. From this table it can be seen that the total number in the proposed sample is 4004 schools while the actual number of schools used in the statistical analysis after data cleaning was 3476, a reduction of 528 schools.

District_ID	District_Name	Location	Allocation from Sample Design	Sample Selected	PSUs in the Analysis	Diff	% Diff
1	Badin	Rural	124	120	110	14	11.3
		Urban	54	54	44	10	18.5
2	Dadu	Rural	144	144	136	8	5.6
		Urban	68	68	64	4	5.9
3	Hyderabad	Rural	30	28	26	4	13.3
		Urban	142	146	132	10	7.0
4	Thatta	Rural	144	144	127	17	11.8
		Urban	43	43	33	10	23.3
5	Mirpurkhas	Rural	92	92	81	11	12.0
		Urban	68	58	45	23	33.8
6	Tharparkar	Rural	135	133	90	45	33.3
		Urban	23	23	22	1	4.3
7	Sanghar	Rural	117	117	103	14	12.0
		Urban	84	85	79	5	6.0
8	Karachi	Rural	18	4	3	15	83.3
		Urban	305	318	289	16	5.2
12	Jacobabad	Rural	83	83	66	17	20.5
		Urban	58	56	49	9	15.5
13	Larkana	Rural	85	85	74	11	12.9
		Urban	117	117	113	4	3.4
14	Shikarpur	Rural	96	96	87	9	9.4
		Urban	62	62	61	1	1.6
15	Khairpur	Rural	174	172	153	21	12.1
		Urban	51	51	46	5	9.8
16	N_Feroze	Rural	153	153	141	12	7.8
		Urban	53	53	51	2	3.8
17	Nawabshah	Rural	110	109	87	23	20.9
		Urban	67	63	43	24	35.8

18	Sukkur	Rural	75	71	61	14	18.7
		Urban	93	99	97	-4	-4.3
19	Ghotki	Rural	170	170	129	41	24.1
		Urban	37	37	32	5	13.5
20	Umerkot	Rural	120	120	106	14	11.7
		Urban	20	19	19	1	5.0
22	Jamshoro	Rural	64	64	48	16	25.0
		Urban	56	56	31	25	44.6
23	Matiari	Rural	98	98	89	9	9.2
		Urban	34	33	32	2	5.9
24	Tando_Allahyar	Rural	79	79	75	4	5.1
		Urban	36	38	36	0	0.0
25	TM_Khan	Rural	71	71	62	9	12.7
		Urban	32	32	27	5	15.6
26	Kashmore	Rural	102	102	90	12	11.8
		Urban	41	41	35	6	14.6
27	Kambar	Rural	111	97	95	16	14.4
		Urban	65	65	57	8	12.3
Total			**4004**	**3969**	**3476**	**528**	**13.2**

Annex 6

Coding, Data Entry and Cleaning

For the 2009 mathematics test marking and coding methodologies were developed by the subject specialists on paper sheets and then transferred into the Excel program. Each possible answer was given a specific code. The markers did not mark questions right or wrong**.** If the first possible answer was chosen a code of 1 was given; for answer 2 a code of 2 was given; for answer 3 a code of 3 was given for answer 4 a code of 4 was given. Where a student had been given a misprinted test booklet or where the possible printed answers were not clear and the student has not answered the question, a code of 5 was allocated. Where a student marked two or more of the possible answers, a code of 7 was given; a code of 6 was given if the student was given a booklet of the incorrect language medium (Sindhi or Urdu). Where a student had not answered a code of 8 was given and where a student has not yet reached the question a code of 9 was given. This is in line with procedures developed in the National Education Assessment System.

Manual test marking and coding was conducted by elementary college, general school teachers and private school teachers at eight centres in the province. This involved marking and coding of approximately 800 assessment items for each student as well as 11 items in the Head Teacher's Background Questionnaire, 54 items in the teacher questionnaire and 50 items in the students' questionnaire. They were instructed on how to enter the data on the coding sheets before the start of the marking and coding process. They were paid for the completion of each booklet. The marking and coding was conducted in a timely manner.

Checking the data was an onerous a task so it was not possible to check every single sheet The data of two students out of 10 students on a scoring sheet was checked by pairs of the elementary college, general school teachers and private school teachers and PEACE specialists super checked one out of 10 students on each scoring sheet in Hyderabad and the focal persons super checked two out of 10 students on each scoring sheet in the other centres. Where mistakes were found the students/teachers employed were asked to recheck their sheets and correct the mistakes. There appeared to be a lack of understanding of the need for rigour in this work and it appears that the majority of the scorers and coders were mainly interested in the quantity of booklets they could complete rather than in doing the task well.

Twenty persons were involved in entering the data in Excel, 10 from the Bureau of Curriculum and 10 privately hired. They were instructed to enter data exactly as it was found on the scoring sheet. The data entered was checked in pairs by the data entry operators and super checking was conducted by three PEACE specialists and the

remaining inconsistencies in the manual checking were identified and rectified. The Excel data was then converted to SAS software and SAS files were created.

The Excel data was placed in four separate spread sheets for each of the 23 districts, namely, Booklet A rural, Booklet A urban; Booklet B rural, Booklet B urban. Recoding of the variable name and the removal of variables identified as not being useful for analysis purposes such as student starting and finishing times, enrollment, booklet serial number, book version etc. and the addition of a column for split schools, was undertaken. After converting these files into SAS files, fractional parts as found in the Excel files were re-coded from A, B, …F to 1, 2, …6. Again many errors were found. Finally all SAS data files were combined to make a whole province file for each of the four areas of mathematics tested and for each version (A and B). This resulted in eight data files. These C and D files were then compared to the A and B files of the sample frame files.

After this the SAS files were ready for the statistical analysis to be conducted.
At different stages of the process errors, inconsistencies and duplications were found.

The following mistakes were identified in the **manual entry** of information on to the score sheets:

- SEMIS Codes sometimes had digits missing or digits transposed or digits duplicated and split schools were often not entered accurately
- Coding errors were found regarding location (rural/urban)
- A few districts' names were not correctly entered
- Checking the gender code revealed some discrepancies
- Some schools' results were not found

Few mistakes were made in the **data entry** of the scores.

SAS software further identified the following issues:

- Duplication problems regarding the SEMIS Code and student roll numbers
- Split schools, fractional part was found to be incorrect and sometimes schools selected from the split school was without the fractional part and on occasion fractional parts were different from that in the sample
- The gender coding in two districts was found to be incorrect
- Discrepancies were found in the number of tests actually completed by students; these discrepancies ranged from one test to seven tests
- Schools did not match the sample

Marking and coding test booklets is an onerous task. Greater training needs to take place to ensure the validity of the information provided for analysis. The difficulties that have arisen from the methodology used will hopefully not occur in future assessments. SAS is

now used and this program flags up any discrepancies immediately after data entry. The training of the markers and coders should be more thorough – besides explaining the methodology, trial runs of entering the data should take place and where the scorers and coders have difficulty their participation should be discontinued.

Background Questionnaires information was also inputted. The data was entered and cleaned on Excel and then converted to SAS. Some variables were found to be inconsistent and these were deleted. After converting to SAS a file was created for the whole province and data quality checked.

It was found that only nine questions out of the 54 questions in the student questionnaire were in usable form for analysis. The head teacher and teacher background questionnaires did not provide sufficient quality data for analysis and therefore were not used.

Some of the difficulties identified in the questionnaires were as follows:

- Some of the questions were translated incorrectly and provided little information
- There was a lack of specificity in some of the questions
- Many of the background questionnaires had non-response

For improvements in the response to the questions in the background questionnaires it is necessary for:

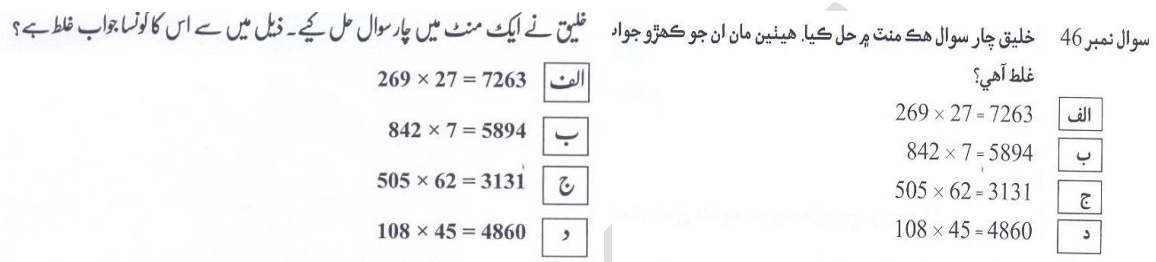
- Greater training to be imparted to Test Administrators
- More time to be provided for the completion of the questionnaires
- Review of the questionnaires to ensure improvements in the specificity of the questions
- Back translations to be conducted to ensure that the same questions are asked in the Sindhi and Urdu questionnaires

Annex 7

Examples of Test Items

Question 1

Competency Assessed: Students will be able to solve simple number problems


سوال نمبر 46 خلیق چار سوال هک منٹ ۾ حل کیا. هيئين مان ان جو کھڑو جواب غلط آهي؟

الف 269 × 27 = 7263

ب 842 × 7 = 5894

ج 505 × 62 = 3131

د 108 × 45 = 4860

خلیق نے ایک منٹ میں چار سوال حل کیے۔ ذیل میں سے اس کا کونسا جواب غلط ہے؟

الف 269 × 27 = 7263

ب 842 × 7 = 5894

ج 505 × 62 = 3131

د 108 × 45 = 4860

Question 2

Competency Assessed: Student is able to subtract two numbers of three, four or five digits


سوال نمبر 10 7000 - 1067 = ...

الف 6099

ب 5933

ج 5007

د 5003

7000 - 1067 = ...

الف 6099

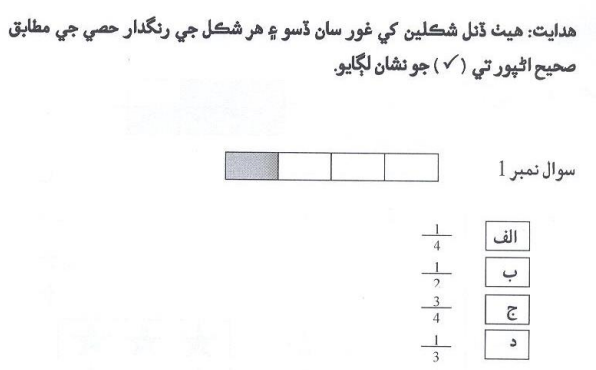
ب 5933

ج 5007

د 5003

Question 3

Competency Assessed: Student is able to identify a fraction of a diagram that is shaded


هدایت: هيٺ ڏنل شڪلين کي غور سان ڏسو ۽ هر شڪل جي رنگدار حصي جي مطابق صحيح اٿپور تي (✓) جو نشان لڳايو.

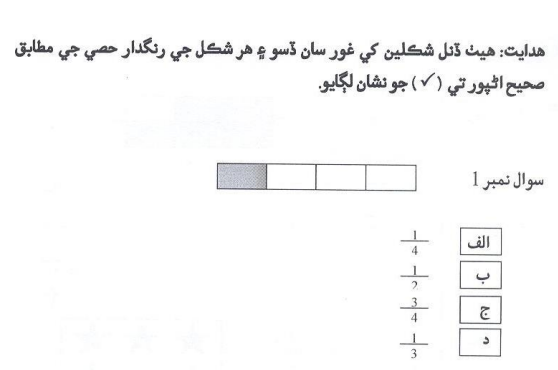
سوال نمبر 1

الف $\frac{1}{4}$

ب $\frac{1}{2}$

ج $\frac{3}{4}$

د $\frac{1}{3}$


هدایت: هيٺ ڏنل شڪلين کي غور سان ڏسو ۽ هر شڪل جي رنگدار حصي جي مطابق صحيح اٿپور تي (✓) جو نشان لڳايو.

سوال نمبر 1

الف $\frac{1}{4}$

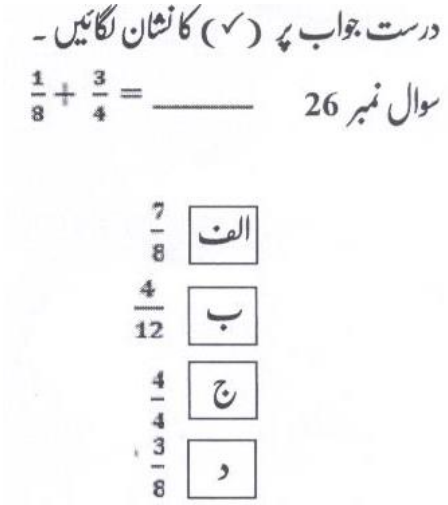
ب $\frac{1}{2}$

ج $\frac{3}{4}$

د $\frac{1}{3}$

Question 4

Competency Assessed: Student is able to add fractions together


درست جواب پر (✓) کا نشان لگائیں۔

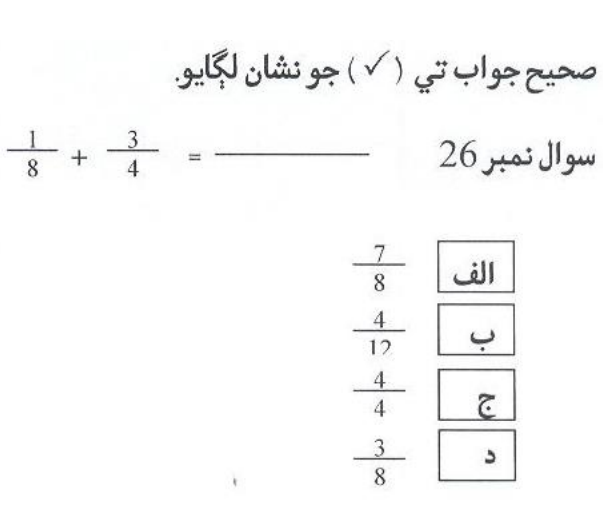
سوال نمبر 26 $\frac{1}{8} + \frac{3}{4} =$ _____

الف $\frac{7}{8}$

ب $\frac{4}{12}$

ج $\frac{4}{4}$

د $\frac{3}{8}$


صحیح جواب تي (✓) جو نشان لڳايو.

سوال نمبر 26 $\frac{1}{8} + \frac{3}{4} =$ _____

الف $\frac{7}{8}$

ب $\frac{4}{12}$

ج $\frac{4}{4}$

د $\frac{3}{8}$

Question 5

Competency Assessed: Student is able to tell the time according to analogue presentation

سامنے دی گئی گھڑی 15 منٹ پیچھے ہے۔ تو صحیح وقت بتائیں۔

الف 8:50

ب 8:35

ج 8:20

د 8:05

سوال نمبر 15 سامهون ڏنل گھڙيال 15 منٽ پويان آهي تہ صحيح وقت ٻڌايو.

الف 8:50

ب 8:35

ج 8:20

د 8:05

Question 6

Competency Assessed: Student is able to identify the relationship between different forms of measurement

سوال نمبر 38 کراچيءَ کان اسلام آباد جي مفاصلي ماپڻ لاءِ مناسب ترين ايڪو ڪهڙو آهي؟

الف ميٽر

ب ڪلوميٽر

ج ملي ميٽر

د سينٽي ميٹر

سوال نمبر 38 کراچی سے اسلام آباد کا فاصلہ ناپنے کے لئے موزوں ترین اکائی کونسی ہے؟

الف ملی میٹر

ب سینٹی میٹر

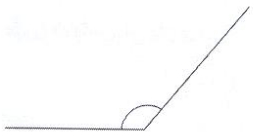
ج میٹر

د کلومیٹر

Question 7

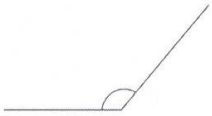
Competency Assessed: Student is able to estimate the size of angles

سوال نمبر32 ڏنل ڪنڊ جي لاءِ هيٺين مان ڪهڙي ماپ صحيح ٿيندي؟



الف	70°
ب	80°
ج	90°
د	110°

دیئے ہوئے زاویے کے لئے ذیل میں سے کونسی پیمائش درست ہوگی؟

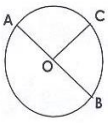


الف	70^0
ب	80^0
ج	90^0
د	110^0

Question 8

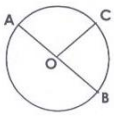
Competency Assessed: Student is able to measure the radius and diameter of a circle

سوال نمبر49 جيڪڏهن $m\overline{AB} = 16cm$ آهي تہ $\overline{OC}$ جي ڊيگهہ ڪيتري ٿيندي؟



الف	5 cm
ب	8 cm
ج	10 cm
د	16 cm

اگر $m\overline{AB}=16$ cm تو $\overline{OC}$ کی لمبائی کتنی ہے؟



الف	5 cm
ب	8 cm
ج	10 cm
د	16 cm

Glossary of Terminology

Assessment	Assessment is the process of documenting, usually in measurable terms, knowledge, skills, attitudes and beliefs of students
Assessment Instruments	These consist of a test framework and specifications, test booklets, rubrics, background questionnaires, guides for test administration, coding, data input and monitoring.
Background Questionnaires	Provide a context for reporting student performance. Student questionnaires collect information on students' demographic characteristics, classroom experience and educational support; teacher questionnaires gather data on teacher training and instructional practices; head teacher questionnaires gather information on school policies and characteristics
Bilog	It models the response of each student of a given ability to each item in the test. Bilog-MG3 converts students' raw scores on a test or versions of a test to a common scale that allows for a numerical comparison between students.
Coding	Each possible answer is given a specific code. For example, if the first possible answer was chosen a code of 1 was given; for answer 2 a code of 2 was given; for answer 3 a code of 3 was given for answer 4 a code of 4 was given.
Collapsed School	A school with an MOS of 4, 5 or 6 is collapsed with a neighboring school to form PSUs before sampling takes place
Cognitive Domain	Cognitive domain is knowledge or *mind* based. It has three practical instructional levels including knowing facts and procedures, using concepts, solving problems and reasoning
Content Domain	Consists of number, measurement, fractions and geometry
Data Cleaning	The detection and correction or removal of corrupt or inaccurate data from a data set by identifying incomplete, incorrect, inaccurate, irrelevant etc data and replacing, modifying or deleting the dirty data to make a data set that is consistent with other data sets in the task or system.
Data Entry	Data entry is the procedure of transcribing the test data from hard copy to a computer using a specific computer program
Diagnostic	Identifying specific areas of difficulty in students' conceptual knowledge and understanding and their ability to solve problems.
Item	A single question which is written for a specific purpose to measure a specific objective
Multiple Choice Questions (MCQ)	Items which provide a phrase or a stem, and four possible answers from which students select the one correct answer.
National Assessment	Large-scale, sample surveys which assess the performance of the education system and not individual students.
National Curriculum	The curriculum to be followed by all the schools in Pakistan
Not significant	Identifies whether differences in reported scores could have occurred by chance alone, significance tests are reported. A probability where $p < 0.65$ means that the difference could

	occur by chance alone in 65 out of 100 students. This means that the results are not significant. They cannot provide reliable information for conclusions to be drawn.
Pilot	Where test items and tests are given to a small sample of representative students to see their effectiveness as assessment tools.
Probability	This is the measure of certainty
Provincial Assessment	These are large-scale, sample surveys which assess the performance of students according to specific content and cognitive domains across a range of levels
Primary Sampling Unit (PSU)	A sampling unit created by collapsing small schools and splitting large schools into two or more pseudo schools
p-value	P value is a probability, with a value ranging from zero to one. If, for example, the p value is 0.03, it means that the difference could occur by chance in only 3 out of 100 students.
Questions	Usually a group of test items written to test the same objective
Random Number Table	A table listing random numbers generated by computer software according to specifications provided by PEACE. If the number of children in Grade 4 is 10 or less then all children in Grade 4 are tested; where the number in Grade 4 is found to be 11- 29 the random table is used to select the 10 students to be tested; where the number of students is greater than 30, the skip interval techniques is used to select the 10 students.
Sample	A representative group of students in Sindh Province; representative of each district, rural/urban area, gender (boys/girls)
Significant	Identifies whether differences in reported scores could have occurred by chance alone, significance tests are reported. A probability where $p < 0.05$ means that the difference could occur by chance alone in only 5 out of 100 students; where $p < 0.01$ the difference could occur by chance alone in only 1 out of 100 students (significant difference); where p is 0.000 there is a highly significant difference.
Skip Interval	The ratio of the number of students enrolled in Grade IV divided by the number of students to be sampled from the school, rounded top an integer.
Split Schools	Dividing a large school into a number of pseudo schools by a “conceptual split”
Survey	Same as provincial assessment
Test Framework	Provides the concept behind the testing and details of what will be tested and how it will be tested
Test Specifications	Provides specific information regarding the content and cognitive domains to be tested, item wise
WesVar	A data analysis tool designed for the analysis of a complex survey using replication and regression analysis