National Education Assessment System

NATIONAL ASSESSMENT REPORT
2005

[This title will be replaced]

In collaboration with:

Provincial Education Assessment Centres (PEACEs)
Area Education Assessment Centres (AEACs)
Assessment Training Centres (ATCs)

Government of Pakistan
Ministry of Education, Islamabad

Acknowledgment

Table of Contents

Annexes

List of Tables

List of Figures

List of Abbreviation

AIOU	Allama Iqbal Open University (Islamabad)
ACER	Australian Council for Educational Research
AEAC	Area Education Assessment Centre
ATC	Assessment Training Centre
AJK	Azad Jammu and Kashmir
B.Ed.	Bachelor of Education
CT	Certificate of Teaching
DfID	Department for International Development
DoE	Department of Education
ETS	Educational Testing Service
FCE	Federal College of Education (Islamabad)
FANA	Federally Administrated Northern Areas
FATA	Federally Administrated Tribal Areas
GCET	Government College for Elementary Teachers
IAEA	International Association for the Evaluation of Educational Achievement
ICT	Islamabad Capital Territory
IER	Institute of Education and Research (University of the Punjab)
IPI	International Partnering Institution
IRT	Item Response Theory
M.Ed.	Master of Education
MoE	Ministry of Education
MCQ	Multiple Choice Question
NAT	National Achievement Test
NEAS	National Education Assessment System
NIP	National Institute of Psychology (Quaid-e-Azam University)
NPCC	National Planning and Coordination Committee
NWFP	North West Frontier Province
PITE	Provincial Institute of Teacher Education
PTA	Parent Teacher Association
PTC	Primary Teaching Certificate
PPS	Probability Proportional to Size
PEACE	Provincial Education Assessment Centres
SAT	Scholastic Ability Test
SMC	School Management Committee
SD	Standard Deviation
TA	Technical Assistance
TOEFL	Test of English as Foreign Language
TIMSS	Trends in Mathematics and Science Study

Executive Summary

The Government of Pakistan is committed to improving the quality of education along with its efforts to increase enrolment and access. This commitment is reflected in the government's policy documents (National Education Policy 1998-2010, Education Sector Reforms 2001-05) and in its reaffirmation at international forums to include quality outcomes such as student achievement scores as performance monitoring indicators in the education sector.

Most of the assessment activities in the past have been project driven in Pakistan. There is little institutional base either at the Federal or Provincial level to sustain assessment activities beyond the project timelines. Neither the traditional school examinations nor the external public examinations administered for certification purposes, nor the selection tests for admission to institutions of higher education, sufficiently fulfill the criteria of standardization and comparability required for a monitoring indicator.

National Education Assessment System (NEAS) and its associated centres in the Provinces (Provincial Education Assessment Centres PEACEs) and Areas (Area Education Assessment Centres – AEACs), have been established as a priority programme under the Ministry of Education's (MoE) Education Sector Reform Action Plan. A specific priority within the overall plan is to:

"build assessment capacity at the school, provincial and federal levels to better measure learning outcomes and improve the quality and effectiveness of programme interventions. "

Central to the project is the development of institutional capacity of MoE at the federal level and in the Departments of Education at the provincial and area levels to monitor standards of education nationally.

Within Pakistan, NEAS has established a partnership with the Institute of Education and Research (IER), University of Punjab, Lahore and with the Federal College of Education (FCE) Islamabad as its Assessment Training Centres (ATCs). The International Partnering Institution (IPI) of NEAS is the Australian Council for Educational Research (ACER), the University of Melbourne, Australia and the Educational Testing Service (ETS) in the United States.

NEAS is planned as a sample-based national assessment, conducted at Grade 4 and at Grade 8, in four subjects:
- Language
- Mathematics
- Science
- Social Studies

The objectives of NEAS are:
- **Informing Policy:** the extent to which geography and gender are linked to inequality in student performance.
- **Monitoring Standards:** how well the curricula are translated into knowledge and skills;
- **Identifying correlates of achievement**: the principle determinants of student performance and how resource allocation might be re-directed
- **Directing Teachers' Efforts and Raising Students' Achievements:** Assisting teachers to use data to improve student performance.

This report includes the details of curriculum based test development at NEAS and its associated centres, the basic issues and procedures for selection of national sample, particularly problems of low enrolment in some rural areas and application of sampling weights to rectify the imbalances of the population due to low enrolment in rural areas. Results of the first round of national assessment 2005, in terms of scaled achievement scores and the impact of background and context variables on students' achievement are also presented.

The NEAS national achievement scores are reported on a scale of 1-1000, with a mean set at 500 and a standard deviation of 100. This scale is the same as in TIMSS, SAT, TOEFL etc. The scaled mean scores obtained in 2005 can be used as baseline for reporting trends over time as well as for comparing the average and relative performance of different groups of students. The 2005 baseline scaled mean scores can also be used by Provinces/Areas/ICT to set targets for improvement of learning achievement to be reflected in future rounds of national assessment. On this scale, the Urdu scaled mean score was 369 and the Mathematics mean score was 421, both below the set mean of 500.

The 2005 results show that NEAS achievement test scores are also well below the international average Mathematics score of 495 in TIMSS 2003 but there are four countries (Iran, Philippines, Morocco and Tunisia) whose Mathematics scores are lower than the NEAS Mathematics score. This comparison is only indicative and not entirely valid because of the differences in the NEAS and TIMSS curriculum framework, sampling outcome, test administration conditions and students' learning environment.

The analysis of background and context variables in relation to achievement scores revealed that several variables related to student home background, teaching-learning processes and teaching practices were important in determining students achievement positively or negatively.

Students who reported they were allowed to ask questions in class scored significantly higher than those who were not allowed to ask questions. Similarly, Urdu and Mathematics mean scores of students who reported getting homework were significantly higher than means scores of those students who did not get any homework.

Supporting inputs from family were also associated with higher mean scores. Students who were taught by their father or brother at home scored significantly higher in Urdu and Mathematics compared to students who reported that their mother or sister taught them at home. In both the subjects, students who did not get help from any body scored the lowest.

The educational and occupational profile of fathers of students across urban and rural areas was similar, though the percentage of negative circumstances (fathers' unemployment, fathers not alive, and illiterate fathers) were slightly higher in rural areas. There was a significant percentage of illiterate mothers in both urban (49%) and rural (64%) areas. The majority of mothers in rural as well as in urban areas were reported to be housewives.

Students who remained absent from school due to illness performed significantly better than those who were absent due to babysitting or household and harvesting responsibilities.

Students of female teachers scored significantly higher in Urdu but not in Mathematics. Teachers academic qualifications ranged from secondary to masters level. However, teachers' qualification did not affect students' performance in either subject.

Reported availability of teaching resources (library, teaching kit, textbooks, teaching guides, curriculum document) did not make any significant difference to the mean scores of students in either Urdu or Mathematics.

There were no significant differences in mean scores of students whose teachers reported teaching two or more classes together in the same period compared to those whose teachers taught a single class. The highest percentage of multi-grade teaching was reported by Urdu and Mathematics teachers from FATA (67%) and FANA (53%). The lowest percentage of multi-grade teaching was reported by Urdu and Mathematics teachers from ICT (12%) and Punjab 23%).

Supporting inputs from community (PTA/SMC, supervisory visits) were not significant in determining the students' achievement.

Finally, it would be pertinent to note that the development of a national assessment system is a complex and challenging task. The preparation and implementation of the 2005 assessment was carried out under tight deadlines. The grade 4 assessment instruments were piloted in April 2004 before the TA input started. These were marked and coded by the NEAS team; data entry was outsourced and item analysis was carried out when the TA input started in February 2005.

These constraints were confounded by insufficient staff in the NEAS, PEACEs and AEACs. To ensure that the assessment was conducted in an efficient and timely manner many of the staff were assigned multiple tasks. NEAS and its associated centres should, therefore, try to fill staff vacancies on a permanent basis. The capacity building that

takes place through hands-on work and training with support from Technical Assistance (TA) is dissipated when there is rapid turnover of staff. This situation works against sustainability and institutionalization of the NEAS network. Furthermore, NEAS sample design in 2005 has posed some logistics problems and for 2006 NEAS has employed a modified design which will reduce some of the problems without compromising the representative or random features of the sample.

NEAS has generated a large amount of data which can be used for research to identify specific variables that are associated with high student achievement. It is anticipated that when staff from NEAS and its associated centres proceed for higher studies they will be able to use this year's and future NEAS databases for carrying out such studies. Teacher training institutions and University departments of Education, Psychology and other Social Sciences in Pakistan can also use this database for research.

1

NATIONAL ASSESSMENT

Learning assessment is increasingly being used around the world to identify strengths and weaknesses of the education system. In a national assessment, measures of achievement in key curriculum areas are administered to students of various selected grade levels. Performance of students in selected sub domains can point to strengths and weaknesses in students' learning achievement within the curriculum areas and can show how the intended curricula are implemented in schools. Data on important student, teacher and school background factors affecting the learning outcomes are collected along with student achievement data to help identify areas needing improvement. This information could then help policy makers identify factors over which they can exert some control.

Like many developing countries, Pakistan is also faced with a problem of expanding enrollment while at the same time improving the quality of education remains a challenge. Little evidence however, is available in Pakistan on the assessment of quality of student learning. To fill-in this gap, a well planned and properly executed national assessment is needed. Considering this need, the Government of Pakistan is committed to improving the quality of education along with its effort to increase enrolment and access. This commitment is reflected in the government's policy documents (National Education Policy 1998-2010, Education Sector Reforms 2001-05) and in its reaffirmation at international forums (Jomtien Declaration 1990 and EFA Assessment 2000, Dakar) to include quality outcomes such as student achievement scores as performance monitoring indicators in the education sector.

As a part of Government of Pakistan's commitment, the Federal and Provincial Governments have also taken some assessment initiatives under development projects in the past. Most of these assessment activities have been project driven and there is little institutional base either at the Federal or Provincial level to sustain assessment activities beyond the project timelines. Consequently, there is also little evidence of the impact of project investments in training and human resource development in the field of student assessment. The initiatives taken so far are on a small scale and lack standardization and

comparability. Neither the traditional school examination nor the external public examinations administered for certification purposes, nor the selection tests for admission to institutions of higher education, sufficiently fulfill the criteria of standardization and comparability required for a monitoring indicator.

In the backdrop of the outcomes of these initiatives, National Education Assessment System (NEAS) has been developed as a priority programme under the Ministry of Education's (MoE) Education Sector Reform Action Plan[1] A specific priority within the overall plan is to:

"build assessment capacity at the school, provincial and federal levels to better measure learning outcomes and improve the quality and effectiveness of programme interventions."

Central to the project is the development of institutional capacity of MoE and the Departments of Education (DoE) at the federal, provincial and area levels to monitor standards of education nationally.

The project supports the establishment of a national and cross-provincial baseline of student achievement as measured against the national curriculum. It is anticipated that the national assessment will make available credible data to provide feedback to the education system to facilitate resource allocation, policy formulation, the improvement of the curriculum, textbook and materials development, and teaching, supervisory and management practices, all of which will contribute to the enhancement of the quality of education.

NEAS is planned as a sample-based national assessment, conducted at Grade 4 and at Grade 8, in four subjects:
  i. Language
  ii. Mathematics
  iii. Science
  iv. Social Studies

The project provides the basis for designing and pilot testing the administrative process, the measurement instruments and the information dissemination strategy associated with the assessment of learning outcomes. Objectives of the NEAS are:
  ▪ **Informing Policy:** the extent to which geography and gender are linked to inequality in student performance.
  ▪ **Monitoring Standards:** how well the curricula are translated into knowledge and skills;
  ▪ **Identifying correlates of achievement:** the principle determinants of student performance and how resource allocation might be re-directed
  ▪ **Directing Teachers' Efforts and Raising Students' Achievements:** Assisting teachers to use data to improve student performance.

---

[1] *Education Sector Reform Action Plan (2001-2005) Ministry of Education, Government of Pakistan*

2

NATIONAL ASSESSMENT INSTRUMENTS

To measure students' learning achievement and to obtain information on factors that have been found to affect the quality of student learning, two different kinds of instruments were used for the assessment of grade 4 students in 2005. These were:

- Achievement tests in Language and Mathematics[2], and,
- Student, Teacher and Head teacher Background Questionnaires[3]

Achievement tests were developed on the basis of the National Curriculum, 2002. The following processes were adopted to develop the achievement tests:

- Competencies[4] based on the first three levels of the cognitive domain of Bloom's Taxonomy (Knowledge, Understanding and Application) were developed for four subjects, keeping in view the content areas and objectives of the national curriculum;
- A table of specifications was designed for the tests which included content, learning outcomes, number of items and type of item;
- Multiple Choice and Constructed Response, Completion and Matching items were developed[5]. However it was decided to focus the tests on multiple choice and constructed response items to provide better standardization of test setting and marking;

Separate background questionnaires were developed for Head Teachers, Teachers and Students. These questionnaires contained questions to identify the association of various personal, home, school, teaching and community variables with student achievement. For example, the basic dimensions of this dataset included parental education and occupation; supporting inputs from home and community; students attitude towards school and teachers; the teachers' qualification and teaching practices, and multi-grade teaching etc.

---

[2] Instrument development Annex 2
[3] Instrument development Annex 2
[4] Test Competencies Annex 3
[5] Examples of test items are found in Annex 3

Both the mathematics and language achievement tests and the background questionnaires were piloted in 2004 on a sample of 4593 students[6]. To facilitate systematic conversion of test responses from booklets to computer, these tests were first marked, coded and scored manually followed by data entry in the computer. Classical item analyses were conducted (using ITEMAN software) to select items for the first round of national assessment in 2005. Items were selected on the basis of item statistics (difficulty, discrimination indices) and professional evaluation of item content by subject specialists. Some new items were also developed to improve the item pool and formats. These assessment instruments were administered in May 2005 to a national sample of 11977 students with the objective of establishing a baseline of achievement for grade 4 students in mathematics and language[7].

---

[6] A report on the pilot testing 2004 is available from NEAS
[7] Test Administration Annex 4

3

NATIONAL ASSESSMENT SAMPLE[8]

In order to draw an adequately representative sample of the nation, the list of Government schools in the National Education Management Information System (NEMIS) database was used as the sampling frame. A stratified, random sampling procedure was used with three agreed explicit strata Province/Area, location (rural/urban) and gender (girls/boys) required in the project documents. Data were also analysed by level of school (Mosque, Primary, Middle, Secondary and Higher Secondary) as an implicit stratum to explore if the achievement of grade 4 students was associated with the type of school they attended.

The sample design covered the entire nation (Figure 1). Testing took place in 117 out of 126 districts and federal regions, in itself a major step forward from previous sampling designs, which usually began by selecting a few districts, often for reasons of administrative convenience and financial constraints. The rate of participation, however, was lower than expected. Instead of the desired total of 16,000, the actual sample was 11,977, or 74.86% of the target. Some provinces and areas had relatively high participation rates, others were much less well represented. Also to obtain an estimate of the actual population and to rectify the imbalances in population due to low enrolment in rural areas, particularly in Balochistan sampling weights were applied.

Schools were selected in fixed proportions from the defined groups Province/Area, location (rural/urban) and gender (girls/boys).

---

[8] Item and Background Data Analyses Annex 7

Table 1. National Sample for Grade-4 National Assessment 2005

Province	Number of Schools	Number of students
Balochistan	90	1305
NWFP	99	1630
Punjab	237	4061
Sindh	152	2128
AJK	57	745
FANA	59	719
FATA	60	847
ICT	30	542
Grand Total	784	11977
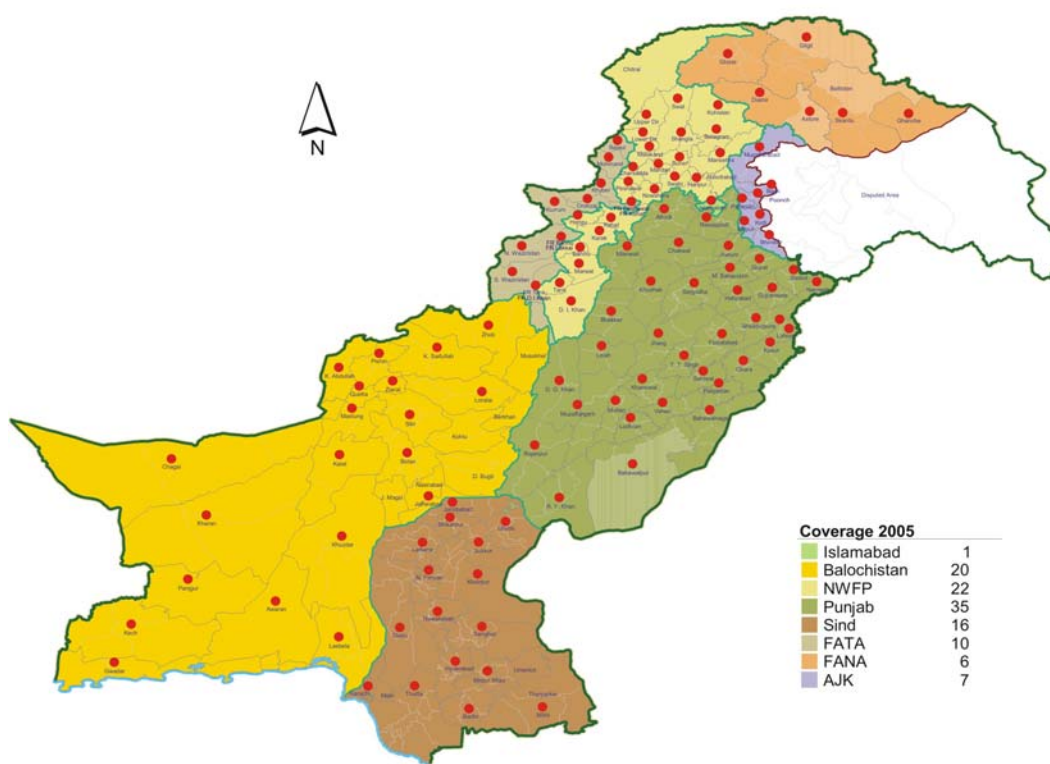
Figure 1. Map of Pakistan Showing the Coverage of the 2005 Sample

Table 2. School Levels in the Sample

School Level	% of NEAS sample	% NEMIS
Higher Secondary	1.1	0.7
Secondary High	17.0	11.3
Middle	17.1	12.9
Primary	63.3	73.7
Mosque	1.6	1.2

4

UNDERSTANDING NEAS RESULTS

Students' achievement is assessed in terms of their scores on the mathematics and language tests. These scores provide a measure of how well the students answer items based on the National Curriculum. Applying IRT (one parameter RASCH model) calibrations, scores on each test were rescaled. The scales were centered on item difficulty in order to have an indicator of the adequacy of average performance and an indication of how well the test takers performed in relation to the desired level of achievement. Scaled scores could also be used to prevent item selection having an undue influence on the scores of students The scaled score used in the NEAS is the same as that used in many international assessments, for example, TIMSS, SAT, and TOEFL. The range of scaled scores is between 0 – 1000. The achievement scales are constructed so that a student achieving 50% of the marks in a test receives a scaled score of 500. The standard deviation, that is, the extent to which the score can vary from 500, is set at 100. This scale will be used to report scores each time the national assessment is conducted and will be used to report trends over time as well as the average and relative performance of different groups of students. The 2005 baseline scaled mean scores can also be used by Provinces/Areas/ICT to set targets for improvement of learning achievement to be reflected in future rounds of national assessment.

The public, policy makers, private sector and schools have an interest in international comparative assessments, using standardized procedures. Trends in Mathematics and Science Studies (TIMSS) are conducted periodically in over 60 countries under the auspices of the International Association for the Evaluation of Educational Achievement (IAEA). It uses national Science and Mathematics curricula for the assessment of learning achievement, and surveys of teachers and students to explore the context in which learning takes place. NEAS is moving towards establishing a credible system of national assessment which can become a part of international assessments in the not too distant future.

Based on IRT analysis, Item Distribution Maps presented in Figures 2 and 3 provide interpretative information about a scaled score in terms of the skills and knowledge students with certain scores (ability) are likely to have. The item names (u1, u2, m1, m2

etc.) are shown located at their calibrations along with the person distribution. An "M" marker represents the location of the mean measure. "S" markers are placed one sample standard deviation away from the mean. "T" markers are placed two sample standard deviations away. Items placed along the scale in the item map demonstrate how skills correspond to levels of performance. For example, Figure 2 shows that in Urdu more students in the lower ability range (-1 to -3 on the person distribution side of the dotted line) found answering questions on a given text easier than writing a story themselves. Similarly in Mathematics (Figure 3) students in the lower ability range found recognition and simple computation items easier than problem solving and fraction items.

Scales are created for each subject independently so even though both language and mathematics have the same numerical range (0 – 1000) average scores should not be compared across subjects.

Significance Levels

To check whether differences in reported scores could have occurred by chance alone, significance tests are reported. A probability of $p < 0.05$ means that the difference could occur by chance alone in only 5 out of 100 cases (usually accepted as a significant level of difference in the Social Sciences). Where p is < 0.01, the difference could occur by chance alone in only 1 out of 100 cases (significant difference); whereas if p is < 0.000 it indicates that the probability of occurrence by chance is less than zero in a 1000 cases (highly significant difference).

Analysis of NEAS Results

The NEAS uses widely accepted statistical procedures in analyzing the data. After completion of the data entry process, data is converted into SPSS file format. SPSS is used for basic descriptive/summary statistics and correlational analysis but because of the limited features of SPSS, the following soft wares are used as the major data analysis tools:

- SYSTAT (Version 9 SPSS Inc.)
- SPSS (Version 12.0(SPSS Inc.)
- ITEMAN (for classical Item analysis) by Assessment Systems (MicroCAT)
- WINSTEPS (introduced in 2005 for IRT)
- AM (for application of tests of significance)

Provincial and Area data generated from the NEAS 2005 assessments have been passed on to all Provincial and Area assessment centres. They can use this data to carry out analyses of their own in more detail for achievement and background variables which are of specific interest to them. This report focuses on the national level analyses with some Provincial and Area analyses which are of general interest.

5

NATIONAL RESULTS

This report presents the National and Provincial/Area results of the NEAS assessment of language and mathematics achievement of a representative sample of grade 4 students in government and community schools.

Grade 4 students achieved a national scaled mean score of 369 in Urdu language, 368 in Sindhi and a national scaled mean score of 421 in mathematics on a national scaled mean set at 500.

Table 3. Scaled Mean Score for Language and Maths

Subject	Scale Mean Score
Urdu	369
Sindhi	368
Maths	421

Throughout the country, all NEAS assessments are administered in Urdu, except for Sindh, where students who study in Sindhi medium schools are administered subject tests in Sindhi and the Sindhi language test instead of Urdu.

National Language Results

Grade 4 students achieved a national scaled mean score of 369 in the Urdu language test. Students' achievement in reading depended upon the cognitive demand of the task. Reading comprehension questions were found to be least difficult by students. Apparently, it seems to be an indication of the emphasis on reading from the textbook as the most common task students are asked to perform during lessons.

On the NEAS reading test lower ability students managed to answer correctly items requiring information or knowledge of facts from a given piece of text. However, students found language questions at the understanding level of Bloom's taxonomy

most difficult to answer correctly. This may be seen as a reflection of the emphasis on rote learning in most government schools.

In writing, the tasks students found most difficult were different kinds of writing (writing stories, describing their own experiences) handwriting and items requiring knowledge about language (grammar and use of idioms, antonyms). Specifically, 17% of students were able to write a simple story while only 9 – 11% could successfully complete the task of writing something for a simple practical purpose.

IRT analysis of each test item along with the competency it tests, is available in NEAS and will form a part of the more technical report under preparation. The item map below summarises students' performance in Urdu by difficulty level and ability.
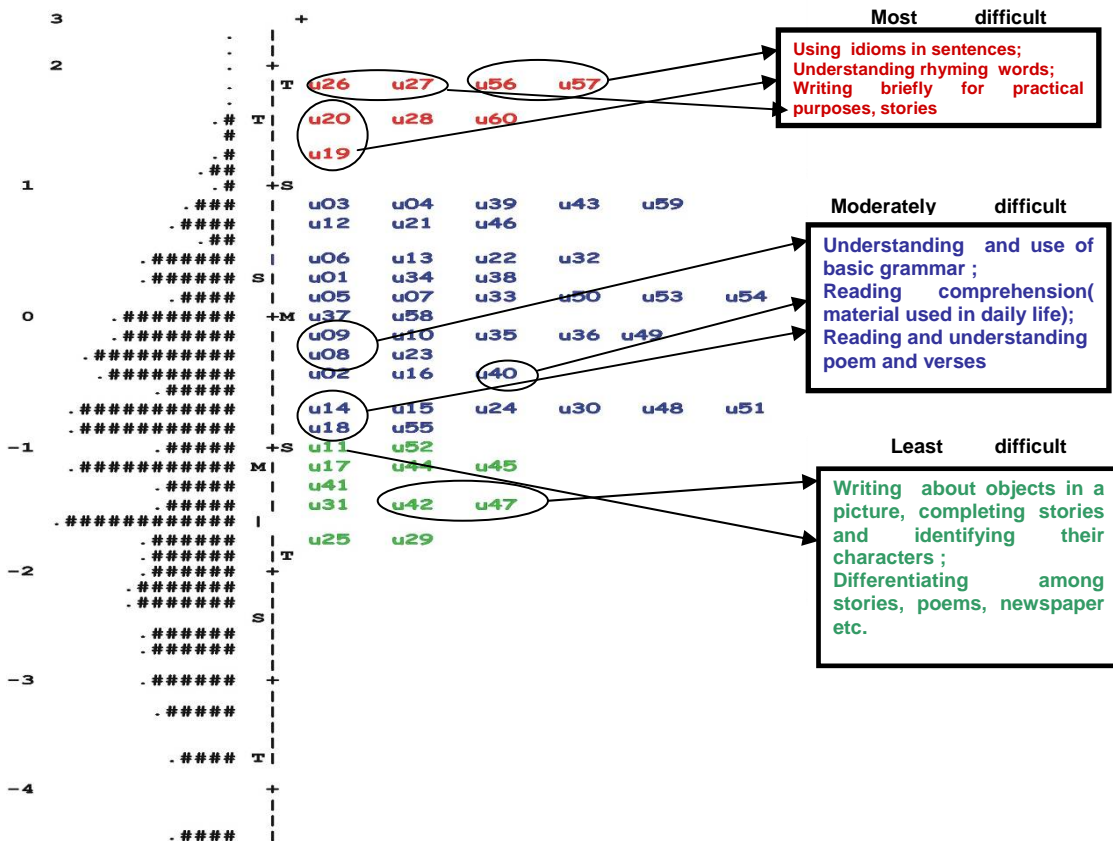
Figure 2. Item Map - Urdu

National Mathematics Results

Grade 4 students achieved a national scaled mean score of 421 in the mathematics test. The students performed well in knowledge based items such as simple computation, questions about place value and estimation of length, capacity and weight. At the understanding level student achieved average results on items of computation, estimation of length and weight while students performed least well in the application

tasks of fractions, odd and even numbers, problem solving, measurement of a line segment, information handling and line graphs.

The item map below summaries students' performance in Mathematic by difficulty level and ability.
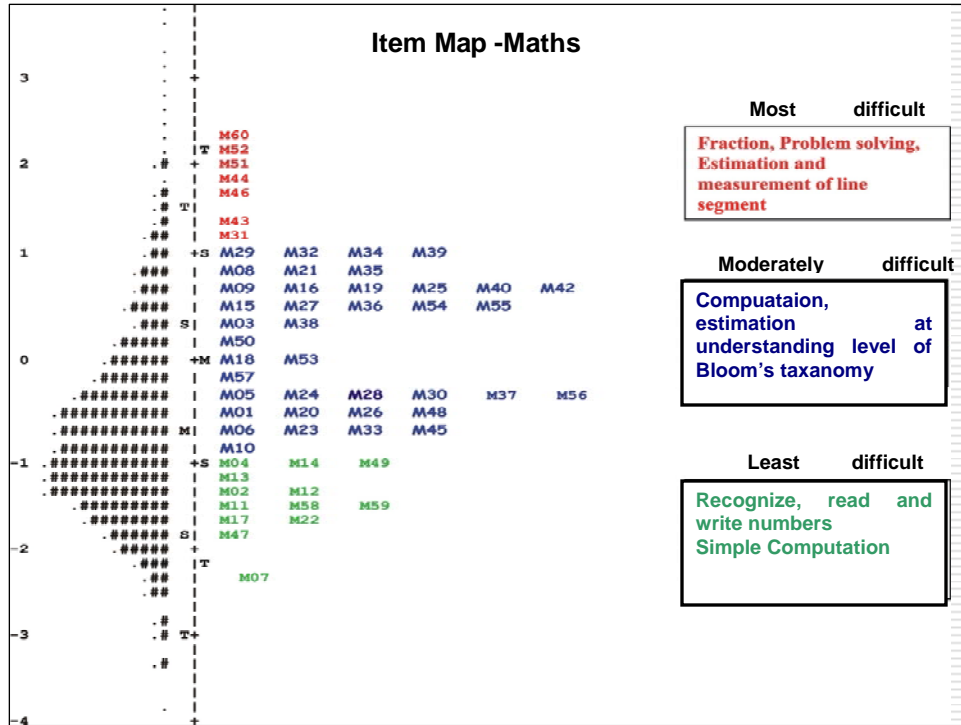
Figure 3. Item Map - Maths

Student Achievement by Location
There was no significant difference between the score of students in urban and rural areas in mathematics. Students in urban areas scored significantly higher than those in rural areas in the Urdu test.

Table 4. Student Achievement by Location

	Rural	Urban	Probability
Urdu	360	399	0.000
Maths	418	430	n.s.

Student Achievement by Gender
There was no significant difference between the scaled mean score of girls and boys in mathematics. However, there was a significant difference between boys and girls' scores in Urdu. Girls achieved a higher scaled mean score than boys.

Table 5. Student Achievement by Gender

	Girls	Boys	Probability
Urdu	405	350	0.000
Maths	424	420	n.s.

Mathematics Results for the Provinces/Areas

Students from ICT scored higher in mathematics as compared to the rest of the nation. Students from Sindh and AJK had significantly lower scores in mathematics than the other Provinces and Areas.

Table 6. Mathematics Results for the Provinces/Areas

Province / Area	Provincial/Area Mean	Rest of Nation	Probability
Balochistan	443	420	n.s.
NWFP	415	426	n.s.
Punjab	430	412	n.s.
Sindh	402	422	0.03
AJK	395	422	0.01
FANA	418	417	n.s.
FATA	451	420	n.s.
ICT	439	433	0.05

Urdu Results for the Provinces/Areas

ICT and Balochistan Urdu scores were significantly higher as compared to the rest of the nation. NWFP scaled scores were significantly lower as compared to the rest of the nation. No significant differences were found in the performance of students from Sindh, AJK, FANA and FATA.

Subsequent to the 2005 assessments, NEAS has been informed that Pushto is used as the medium of instruction in about one third of the schools in NWFP. Perhaps the lower scores of NWFP students in Urdu can be attributed to this situation. A detailed analysis of the extent to which NWFP Urdu scores have been affected by the medium of instruction will be included in the technical report. In the forthcoming round of 2006 assessments, students studying in the Pushto medium will be assessed in Pushto.

Table 7. Urdu Results for the Provinces/Areas

Province / Area	Provincial/Area Mean	Rest of Nation	Probability
Balochistan	403	368	0.000
NWFP	327	378	0.01
Punjab	382	356	0.01
Sindh	367	369	n.s.
AJK	359	369	n.s.
FANA	375	369	n.s.
FATA	345	369	n.s.
ICT	433	368	0.000

Achievement of Sindhi and Urdu Medium Schools in Sindh

There was no significant difference between the achievement of students in Sindhi medium and Urdu Medium schools in Sindh in either Language (Urdu/Sindhi) or in Mathematics.

Table 8. Achievement of Sindhi and Urdu Medium Schools in Sindh

Medium	Language Mean	Probability	Maths Mean	Probability
Sindhi	368	n. s	406	n. s
Urdu	362	n. s	388	n. s

School Level and Mathematics Achievement

The lowest mean score in mathematics is found for students in higher secondary schools. The other school level results do not indicate any significant differences in performance. It should be noted that higher secondary schools formed only 1.1% of the total NEAS sample.

Table 9. School Level and Mathematics Achievement

School Level	% by Level	Mean by Level	Mean: Rest of the nation	Probability
Higher Secondary	1.1	368	422	0.05
Secondary/High	17	421	421	n.s.
Middle	17.1	412	423	n.s.
Primary	63.3	425	412	n.s.
Mosque	1.6	413	421	n.s.

School Level and Urdu Achievement

There was no significant difference in the performance of students belonging to different school levels on the Urdu test.

Table 10. School Level and Urdu Achievement

School Level	Mean by Level	Mean: Rest of the nation	Probability
Higher Secondary	354	369	n.s.
Secondary/High	381	368	n.s.
Middle	368	369	n.s.
Primary	368	371	n.s.
Mosque	362	369	n.s.

6

BACKGROUND AND CONTEXT VARIABLES IN RELATION TO STUDENTS' ACHIEVEMENT SCORES

Findings of previous research in the field of achievement testing and educational research indicate that students' achievement is determined by:

- Student characteristics such as the home background and attitudes towards education;
- The teaching and learning process and teaching practices such as teacher attitude, order and discipline, the variety of teaching strategies, the assignment of homework and providing feedback to students on their work;
- School conditions and climate such as effective leadership, the general facilities of the school, the organized curriculum, flexibility and autonomy;
- Supporting inputs such as parent and community support and effective support from the education system.

Besides testing the students in mathematics and Urdu throughout Pakistan, background data were collected from the students to investigate the impact of background on achievement of students in different subjects. The details of the achievement of students in the subjects of Urdu and mathematics with respect to different aspects of their background are discussed in the following sections.

Student Age and Achievement

The sample included 10.3% 9 year old students and 5% were 8 years old while 7% of the students were aged 14 years and older. Students aged 9 (the desired age level for grade 4) got the highest mean scaled scores on both the tests. Student aged less than 9 years scored the lowest on both the tests. Similarly, students aged 14 and above (6.4%) got low mean scores on both the tests. This finding may have implications for teachers training to develop teaching skills which supports student learning in a wide age range in the same class.

Table 11. Student Age and Achievement

Age	Rural	Urban	Total Sample	Urdu Mean Score	Maths Mean score
Less than 9 years	7.4	7.1	5.1	343.31	386.44
9 years	13.5	11.1	9.4	404.21	436.95
10 years	27.5	25.4	23.2	376.25	427.03
11 years	19.6	17.9	17.4	369.75	422.47
12 years	19.9	22.5	20.8	360.94	422.01
13 years	6.9	9.5	8.9	378.34	423.99
14 years and older	5.2	6.6	6.4	358.42	417.86

Achievement in Relation to Home Language

There was no significant difference in achievement due to different home languages except for Pushto and Seraiki. Pushto speaking students scored significantly lower on both the tests as compared to non Pushto speaking students while Seraiki speaking students performed significantly better on the Urdu test as compared to non-Seraiki speaking students.

Table 12. Achievement in Relation to Home Language

Language	Urdu Mean Score	Maths Mean Score
Urdu (15.4%)		
Yes	376.43	411.08
No	369.84 (p=45, n.s.)	423.93 (p<.065, n.s.)
Pushto (17.38%)		
Yes	336.11	393.25
No	378.33 (p<003)	429.03 (p<000)
Sindhi (12.31%)		
Yes	370.99	405.94
No	370.05 (p<.964, n.s.)	424.15 (p<195, n.s.)
Punjabi (35.54%)		
Yes	380.39	411.08
No	364.91 (p<.14, n.s.)	425.94 (p< .52, n.s.)
Balochi (10.09%)		
Yes	372.43	436.07
No	370.05 (p<.95, n.s.)	420.14 (p< .37 n.s.)
Seraiki (3.26%)		
Yes	441.59	420.47
No	368.43 (p<.000)	461.88 (p< .09 n.s.)

Mathematics Teaching at Home

Mathematics scores appear to be affected by family support. Students whose fathers helped them in Mathematics scored higher than those who did not get any support in studies from fathers. Similarly students getting help from brothers scored higher as compared to students getting help from mothers.

Help from a tutor is also associated with better performance on Mathematics. Students getting help from tutors performed better than students getting help from sisters.

Table 13. Mathematics Teaching at Home

Teaching at Home	Mean Maths Score
Nobody	393.47
Father	446.05
Mother	408.52
Brother	423.69
Sister	421.76
Tutor	429.94
Other	440.71

Nobody vs each other, $p<0.000$; Mother n.s., brother$<.002$, sister$<.006$, tutor<0.05, Father vs mother, $p<0.023$, brother<0.010, sister-n.s

Urdu Teaching at Home

Urdu language achievement is also affected by family support. For language achievement however, sisters support seems to be more effective compared to support from brothers. Fathers also seem to play an important role in improving the learning of Urdu language as compared to mothers.

Students getting help from fathers scored better than students getting support from mothers.

Table 14. Urdu Teaching at Home

Teaching at Home	Mean Urdu Score
Nobody	314.54
Father	397.96
Mother	371.32
Brother	371.12
Sister	399.50
Tutor	388.93
Other	394.46

Note: Nobody vs. each other option: $p<0.000$
Father vs. Mother, $p<0.020$, brother<0.004, sister/tutor n.s.

It will be seen in the section on mothers' education that about half the mothers of urban students and two-thirds mothers of rural students are reported to be illiterate, therefore it is not surprising that they cannot help their children much with their studies.

It should also be noted that students' scores are higher if *someone* helps them with studies at home. Student who have reported getting *no help* with studies at home have lowest scores in Urdu and Mathematics.

Student Absenteeism

Students who were absent from school due to illness performed significantly better than those who were absent due to babysitting or household and harvesting responsibilities. Perhaps the household responsibilities kept students away from school more frequently than occasional illness.

Table 15. Student Absenteeism

Reasons for Absence	Urdu Mean Score	Maths Mean Score
Illness		
Yes (84%)	384	431
No (16%)	323 (p<0.000)	398 (p<0.000)
Baby Sitting		
Yes (15%)	328	393
No (85%)	377 (p<0.000)	429 (p<0.00)
Harvesting		
Yes (20%)	333	403
No (80%)	378 (p<0.04)	427 (p<0.001)

Father's Education in Rural and Urban Areas

Thirty-three percent of fathers in rural areas are illiterate while 24.5% of fathers in rural areas were reported to be educated up to matriculation level. In urban areas only 24.5% of fathers are illiterate while 27.9% of fathers were reported to be educated to matriculation level.

Table 16. Father's Education in Rural and Urban Areas

Education Level	Rural	Urban
Illiterate	32.5%	24.5%
Primary	29.1%	26.6%
Matriculate	24.5%	27.9%
Intermediate	6.4%	8.8%
Bachelor	3.5%	6.8%
Master	2.6%	5.1%
Other	0.5%	0.3%

The educational and occupational profile of fathers of students across urban and rural areas was similar, though the percent of negative circumstances (fathers' unemployment , fathers not alive, and illiterate fathers) were slightly higher in rural areas.

Father's Occupation in Rural and Urban Areas

Table 17. Father's Occupation in Rural and Urban Areas

Occupation	Rural	Urban
Labourer	22%	19%
Government Job	16%	20%
Private Job	19%	18%
Unemployed	9%	9%
Other	30%	30%
Deceased (not alive)	3%	2%

Only lower socio-economic status fathers' appear to send their children to government schools in urban areas as well as in rural areas. Similarity between father's education and occupation in rural and urban areas should be noted.

Mother's Education in Rural and Urban Areas
There was a significant percentage of illiterate mothers in both urban (49%) and rural (64%) areas.

Table 18. Mother's Education in Rural and Urban Areas

Education	Rural	Urban
No Response	< 1%	< 1%
Illiterate	63.7%	48.8%
Primary	24%	26.6%
Matriculation	8.7%	16.1%
Intermediate	1.4%	4.5%
Bachelor's	< 1%	< 1%
Masters	< 1%	< 1%
Other	< 1%	< 1%

Mother's Occupation in Rural and Urban Areas
The vast majority of mothers take care of household duties whether they live in rural or urban areas.

Teachers, Teaching Practices and Students' Achievement

Teacher Gender and Student Achievement
There was a significant difference between teacher gender and student achievement in Urdu tests. Students of female teachers had higher scores in Urdu. However, the majority of girls (80%) and boys (90%) were being taught by teachers of their own gender. More rigorous statistical analysis is needed to determine if the better performance of girls in Urdu is a reflection of the generally reported research finding that girls are better than boys in language skills or the higher scores in this case reflect better teaching by female teachers who may also have better language skills.

Table 19. Teacher Gender and Student Achievement

Gender	Urdu Mean	Probability	Maths Mean	Probability
Male	351.42	0.000	421.90	n. s
Female	401.36		414.61	

No significant difference was observed between teacher gender and students' achievement in mathematics. The generally reported finding of better mathematics achievement of boys in some studies is not supported by these NEAS results.

Teachers' Academic Qualification and Student Achievement
While none of the differences were statistically significant it should be noted that the majority of mathematics teachers were matriculates or graduates. Also 15% of mathematics teachers and 13% of Urdu teachers were master's degree holders. Students of matriculate teachers performed equally well in comparison to students of teachers with master's degrees.

Table 20. Teachers' Academic Qualification and Student Achievement

Teachers Academic Qualification	% Urdu	Urdu Mean	% Maths	Maths Mean
Secondary	36	368.56	35	430.02
Intermediate	19	387.27	20	427.87
Bachelors	30	355.67	30	402.75
Masters	13	369.80	15	420.70

Teachers' Highest Level of Professional Training and Student Achievement
Training received by teachers did not significantly affect students' performance in any of the subjects. None of the mean score differences were significant at 0.05 level.

Table 21. Teachers' Highest Level of Professional Training and Student Achievement

Teachers' Training	% Urdu	Urdu Mean	%Maths	Maths Mean
PTC	66	367	64	416
CT	16	377	18	446
B.Ed.	15	358	16	386
M.Ed.	02	402	02	449

Achievement Scores in Relation to Teaching Experience

Students whose teachers had 16 – 20 years' teaching experience achieved the highest scores in both Urdu and Mathematics while Urdu scores of students were the lowest when taught by teachers whose teaching experience was 5 years or less. However, none of these other differences were statistically significant.

Table 22. Achievement Scores in Relation to Teaching Experience

Teaching Experience (in years)	% Urdu	Urdu Mean	% Maths	Maths Mean
0-5	15	357.08	15	411.82
6-10	14	354.54	16	414.54
11-15	27	374.36	26	421.92
16-20	24	390.07	24	433.95
21-25	11	338.41	9	428.35
26 and above	8	372.91	8	427.73

Multi-Grade Teaching and Achievement

Students whose teachers took more than one class at the same time scored lower than those whose teachers taught single classes. These differences however, were not significant statistically for either the mathematics or Urdu tests.

Table 23. Multi-Grade Teaching and Achievement

Teachers who teach more than one class together	Urdu Mean	Probability	Maths Mean	Probability
No	371.19	n.s.	420.41	n.s.
Yes	360.50		408.48	

Multi-Grade Teaching and Urdu Teachers

Thirty percent of all the Urdu teachers surveyed reported that they had to teach more than one class at the same time. 66.7% of Urdu teachers from FATA and 55.1% Urdu teachers from FANA reported that they had to teach multiple classes in one period. In ICT and Punjab only 12.2% and 22.6% of teachers taught multiple classes.

Table 24. Multi-Grade Teaching and Urdu Teachers

Province/Area	No	Yes
Balochistan	74.3%	25.7%
NWFP	69.7%	30.3%
Punjab	77.4%	22.6%
Sindh	61.8%	38.2%
AJK	60.0%	40.0%
FANA	44.9%	55.1%
FATA	33.3%	66.7%
ICT	87.8%	12.2%
Total	69.6%	30.4%

Multi-Grade Teaching and Mathematics Teachers
Sixty-five percent of the teachers surveyed from FATA and 52.5% of the teachers from FANA reported that they had to teach multi-grade classes in one period. In ICT and Punjab only 14.2% and 23% of teachers taught multiple classes in the same period.

From the information obtained by the questionnaires it can be seen that FATA and FANA have more multi-grade teaching in both subjects taking place than any of the other Provinces and Areas.

Table 25. Multi-Grade Teaching and Mathematics Teachers

Province/Area	No	Yes
Balochistan	74.1%	25.9%
NWFP	66.3%	33.7%
Punjab	77.0%	23.0%
Sindh	66.7%	33.3%
AJK	55.8%	44.2%
FANA	47.5%	52.5%
FATA	35.4%	64.6%
ICT	85.8%	14.2%
Total	69.9%	30.1%

Students' Assessment by Teachers
The survey of the sample schools used in the 2005 National Assessment found that classroom observation techniques to assess students' performance were used by only 40% of the teachers. The most common form of assessment was written tests - 81% of teachers in the sample preferred to use this method. Monthly tests were arranged by 87% of teachers while 71% of the teachers prepared regular progress reports for their students. 89% of the teachers stated that they informed parents of their children's achievement through the progress reports. The question arises as to why student scores in the national assessment are below average in Urdu and Mathematics in spite of this fairly good assessment and reporting practice mentioned by teachers. A more detailed study of how these assessments are carried out and why they do not help in improving student performance is needed.

Teachers' evaluation of students by homework did not affect their performance. There was no significant difference in achievement between students who were evaluated by homework and those who were not evaluated by homework.

Table 26. Teacher Assessment of Students through Homework

	Urdu Mean	Probability	Maths Mean	Probability
No	366.53	n.s.	422.75	n.s.
Yes	375.99		423.50	

Homework and Achievement

Students who reported getting homework scored significantly higher on both Urdu and Mathematics tests.

Table 27. Students Report Getting Homework

	% Urdu	Urdu Mean	Probability	% Math	Maths Mean	Probability
No	4.7	297.94	0.000	4.7	386.93	0.003
Yes	84	376.52		84.1	424.95	

Patterns of Reward and Punishment in Schools

Students who reported getting prizes and incentives scored significantly higher in Mathematics than those who did not get incentives. Similarly, students who reported not getting punishment scored significantly higher in Mathematics than those who got punished. However, neither prizes nor punishment affected performance in Urdu.

Table 28. Prizes, Incentives and Mathematics Achievement

	Mean	Probability
No (33%)	401.004	0.01
Yes (67%)	433.883	

Table 29. Punishment at School and Mathematics Achievement

	Mean	Probability
No (60%)	432.07	0.05
Yes (40%)	405.58	

Participation in Co-Curricular Activities

Students participating in co-curricular activities performed higher on both tests as compared to those who did not participate in these activities. The raw frequencies on the co-curricular activities reveal that majority of students participate in the recitation of the Holy Quran/Na'at and games. Fewer students participated in debates and drama.

Table 30. Participation in Co-Curricular Activities

	% Urdu	Urdu Mean	Probability	% Maths	Maths Mean	Probability
No	22.2	346.81	0.05	20.2	402.45	0.04
Yes	68.3	379.17		68.3	428.46	

Questions Allowed in Class and Achievement

The scores of the students who were not allowed to ask questions scored significantly lower than those who were allowed as questions.

Table 31. Questions Allowed in Class and Achievement

Classroom participation	Urdu Mean	Probability	Maths Mean	Probability
No	287.80	0.03	382.73	0.02
Yes	379.65		435.59	

School Conditions and Climate

Availability and condition of blackboards in schools

From the table below it can be seen that while the majority of the assessment sample schools had blackboards, 55% of the urban sample schools and 52% of the rural sample schools stated that their blackboards needed repair.

Table 32. Availability and condition of blackboards in schools

Availability of Blackboard	Urdu Mean	Probability	Maths Mean	Probability
No	373.49	n.s	398.17	0.03*
Yes	368.85		424.06	
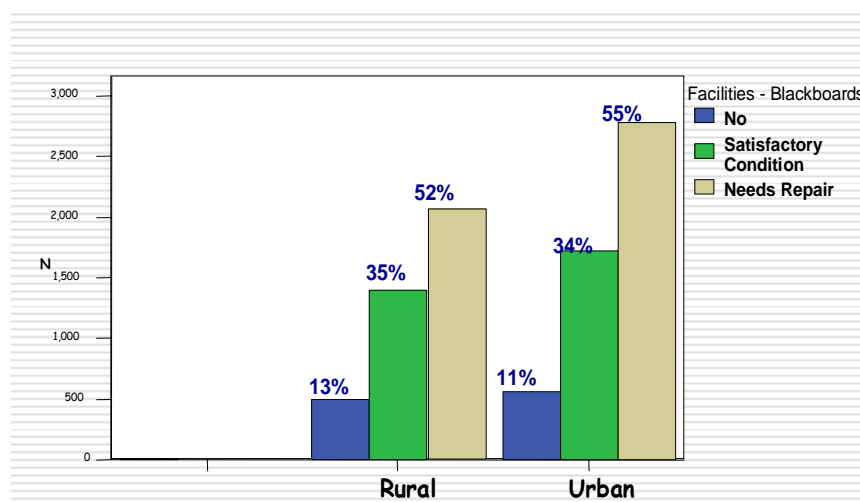
Figure 4. Availability and Condition of Blackboards in Schools

The availability of the blackboard did not have a significant effect on students' Urdu achievement but it had a significant effect on students' mathematics achievement.

Availability of School Library

The presence of a library did not affect student achievement. There was no significant difference in the performance of students on either the Urdu or Mathematics tests in schools which had a library compared to the achievement of students where schools did not have a library.

Table 33. Availability of School Library

Library	Urdu Mean	Probability	Maths Mean	Probability
No	367.45		418.65	
Yes	380.63	n. s	427.98	n. s

More probing questions need to be asked about the use of blackboards and libraries in terms of frequency of use and purpose for which they are used, to determine why such a useful learning resource as a library does not have an impact on student achievement.
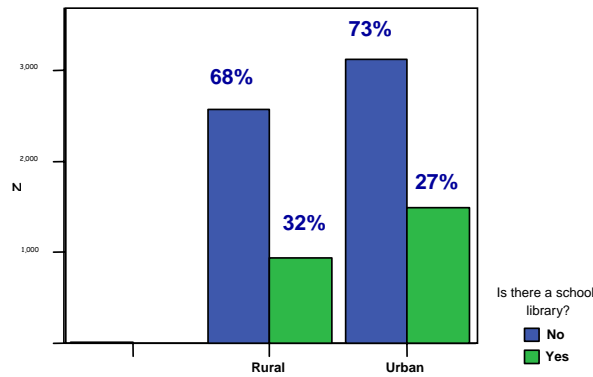
Figure 5. Availability of School Library

Availability of the Ministry Of Education Curriculum (2002)
The reported availability of the 'curriculum' in schools did not contribute toward students' achievement scores in either Urdu or mathematics. However, it is important to note that the 'curriculum' (nisaab) was probably interpreted by most of the schools as the textbooks, because MoE curriculum documents are not distributed to schools and textbooks are.

Table 34. Availability of the Ministry Of Education Curriculum (2002)

Availability of Curriculum		Urdu Mean	Probability	Maths Mean	Probability
No	14%	362.46		411.52	
Yes	86%	371.48	n.s	423.13	n.s
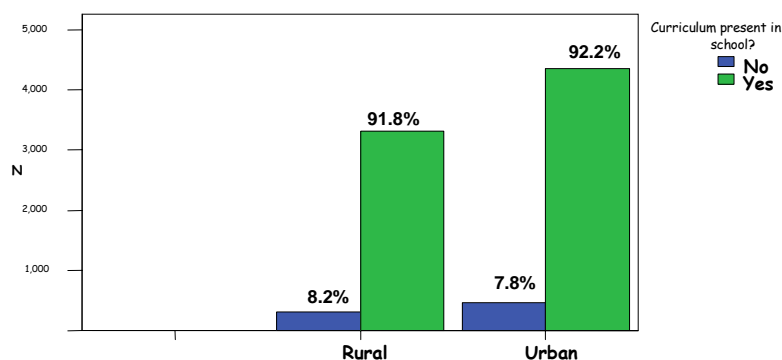
Figure 6. Availability of the Ministry Of Education Curriculum (2002)

As noted in the case of fathers' education and occupation the similarity between facilities or lack of them in rural *and urban* government schools is evident from the above tables.

Supporting Inputs for Schools

Presence of PTA/SMC in the Sample Schools

It can be seen that the majority of the sample schools had PTA/SMC and in the majority of cases the funding (over 90%) came from the Government.

Table 35. Presence of PTA/SMC in the Sample Schools

Location	No %	Yes %	Funds from Government %	Donations %
Rural	10	90	94	6
Urban	8	92	92	8
Total	7	93	93	7

PTA/SMC and Achievement

The presence of the PTA/SMC in a school did not significantly affect the performance of students in either Urdu or Mathematics.

Table 36. PTA/SMC and Achievement

PTA/SMC	Urdu Mean Score	Probability	Maths Mean Score	Probability
Yes	369	n.s.	418	n.s.
No	404		436	

Academic Qualification of Head teachers

Twenty-eight percent of Head teachers in the sample schools in rural areas reported having a Bachelors' Degree as compared to 33% Head teacher in urban schools. Thirty-eight and 31% rural and urban head teachers had postgraduate degrees, respectively.

15% of head teachers in the sample schools in rural areas and 18% of the head teachers in urban sample schools have only a secondary qualification.
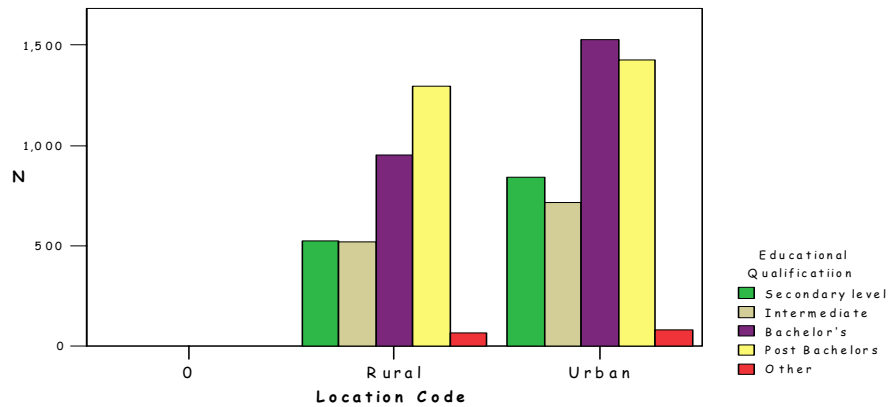
Figure 7. Academic Qualification of Head Teachers

Professional Qualification of the Head Teacher
There was no significant difference in the professional qualifications of Headteachers in the urban and rural sample schools.

Only 43% of the rural and 41% of the urban Head teachers had qualifications of B.Ed. and M.Ed. 57% of head teachers in the rural school sample and 57% of head teachers in the urban school sample had PTC or CT Qualifications.

Again, the educational and professional qualifications of the rural and urban Head teacher are not very different in the government schools sampled here
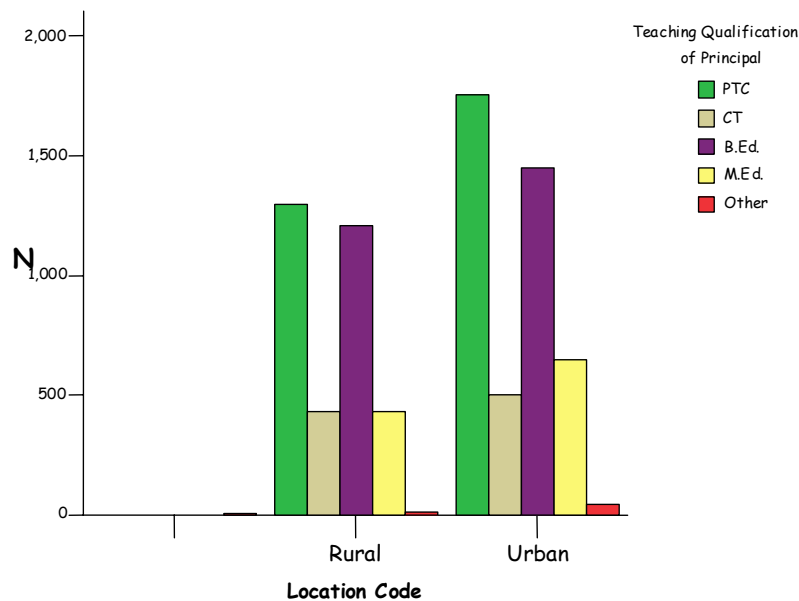
Figure 8. Professional Qualification of the Head Teacher

Supervisor Visits

Fewer visits of supervisors (less than three-monthly) are linked with high mean scores on the mathematics achievement test. There is no significant difference in the Urdu scores associated with the frequency of supervisor's visits to schools. These reports need to be corroborated from other sources and follow up questions need to be asked about what do supervisors do when they visit schools. Frequency of Supervisor visits is mostly similar for rural and urban schools except for more weekly visits in urban areas.

Table 37. Supervisor Visits

Loca-tion	Weekly Visits %	Maths Mean Score	Twice a month Visits %	Maths Mean Score	Monthly %	Maths Mean Score	Three Monthly %	Maths Mean score	Other (Less often) %	Maths Mean Score
Rural	12		13		46		19		9	
Urban	20	410.793	12	420.618	40	414.709	18	415.385	10	451.897
Total	17		13		43		18		9	

7

CONSTRAINTS AND LESSONS LEARNED

Constraints

The development of a national assessment system is a complex and challenging task. It involves the training of staff, the development of test frameworks and items, the organization of the assessment instrument booklets as well as the printing, collating, distribution and collection of the assessment instruments throughout Pakistan under tight deadlines. The grade 4 assessment instruments were piloted in 12 districts of country in April 2004 before the TA input started. These were marked and coded by the NEAS team; data entry was outsourced and item analysis was carried out when the TA input started in February 2005.

Between February 2005 and the 4th of May 2005 the NEAS team with TA assistance reviewed and developed, collated and distributed the assessment instruments and undertook the training of the test administrators and monitors. A cycle of activities that takes at least two years in well establishes assessment centres around the world was completed by NEAS and its associated centres in less than a year from instrument development to the administration of large scale assessments.

The above constraints were confounded by insufficient staff in the NEAS, PEACEs and AEACs to ensure that the assessment was conducted in an efficient and timely manner. Many of the staff were assigned multiple tasks due to this constraint;

Problems of low student enrolment were found especially in rural/hilly areas of Pakistan specifically in girls schools. Many schools had less than 5 students at grade 4 level. In some areas, changes were made in the sample due to inaccessibility and security reasons. This made the representation of the sample somewhat questionable.

The transportation of the assessment materials to arrive in the provinces and areas to enable training and testing to take place was difficult. The return of the assessment instruments after testing had taken place was also difficult particularly from remote areas. The approved flat and uniform rates of transportation allowance did not take into account the availability or non-availability of public transport facilities.

Lessons Learned

NEAS and its associated centres need to fill staff vacancies on a permanent basis. The staff working in the assessment centres should be retained till the life of the project. The capacity building that takes place through hands on work and training with support from Technical Assistance (TA) is dissipated when there is rapid turnover of staff. This situation works against sustainability and institutionalization of the NEAS network..

Rapid staff turnover also mitigates against the timely implementation of national assessment activities. More grades and subjects are scheduled to be added to the NEAS implementation plan. Without stability in the technical and office support staff, delivering yearly national assessments at the required level of quality assurance will become increasingly difficult for NEAS and its associated centres. In addition, there is an ambitious training plan for staff in the NEAS network. If some staff proceed on training there is a vital need for back up staff to be available to continue the implementation of the planned assessment cycles.

NEAS sample design in 2005 has posed some logistics problems and for 2006 NEAS has employed a modified design which will reduce some of the problems without compromising the representative or random features of the sample. Even so, low enrolments in rural and girls' schools are so evident in the NEMIS database that they cannot be excluded for reasons of logistic convenience, if defensible statements are to be made about the achievement of students.

Conclusion

The findings presented in this report are an attempt by NEAS to establish a baseline of student achievement at grade 4 level in Urdu and Mathematics as a starting point for monitoring National and Provincial/Area level trends in the learning achievement of students. The background variables found to be positively or negatively associated with students' Urdu and Mathematics scores are not necessarily causative and need to be considered with caution.

This first round of assessments has generated a pool of data that can be analysed in greater detail and supplemented with rigorous quantitative and qualitative studies to identify more specifically the correlates of achievement in the students' learning environment which can be supported through teacher training and learning materials development. Currently NEAS, PEACEs and AEACs do not have the human and financial resources to undertake such studies. NEAS is working under a time bound implementation plan to deliver several rounds of national assessments at grade 4 and grade 8 levels by June 2008. It is anticipated that when staff from NEAS and its associated centre proceed for higher studies they will be able to use this year's and future NEAS databases for carrying out such studies. Teacher training institutions and University departments of education/psychology can also use this database for research studies.

ANNEXES

NATIONAL TEST SAMPLE

The Proposed Sample Design

Three sample design options were presented to the National Planning and Coordination Committee, and after extended debate, the Committee agreed on the following basic parameters:

a) **32 explicit strata:** Punjab with largest number; other Provinces/Areas approximately pro rata.

b) **Stratum size; Punjab:** 1537 per stratum (total 6148); Sindh 784 per stratum (total 3136); other provinces have 267 per stratum (1068 per province); areas have a stratum size of 97 (388 per area).

c) **Total sample size:** 12,972.

This decision was taken as the basis for the final sample design. However, when the full sampling frame became available, the facts it revealed forced the NEAS team to adapt the basic design in several ways.

Table 38 below outlines the numbers of schools in the various provinces/areas, by "gender" and "rural-urban". It should be noted that there are in fact mixed as well as single-sex schools in Sindh and AJK, as well as ICT. It should also be noted that the "urban" sub-sector is unusually small in AJK and FANA.

Table 38. Numbers of Schools (from NEMIS)

Prov./Area	Rural			Urban			Total
	Female	Male	Mixed	Female	Male	Mixed	
Balochistan	2609	7431	0	400	977	0	11417
NWFP	7802	15960	0	703	1115	0	25580
Punjab	25058	30878	0	3535	3849	0	63320
Sindh	5633	17342	15176	1909	2317	1772	44149
Total Provs.	**41102**	**71611**	**15176**	**6547**	**8258**	**1772**	**144466**
AJK	2597	3134	0	115	102	0	5948
FANA	253	584	622	35	42	54	1590
FATA	1859	3023	0	0	0	0	4882
ICT	116	119	20	45	34	63	397
Total Areas	**4825**	**6860**	**642**	**195**	**178**	**117**	**12817**
Grand Total	45927	78471	15818	6742	8436	1889	157283

Table 39 below outlines the numbers of students in the various provinces/areas. Comparing these two shows that schools in some areas, notably Balochistan, are very small.

Table 39. Class 4 Pupils by Province/Area, Gender and R/U (from NEMIS)

Prov./Area	Rural		Urban		Total
	Boys	**Girls**	**Boys**	**Girls**	
Balochistan	24941	15281	14625	11241	66088
NWFP	175479	89132	24403	18016	307030
Punjab	441103	281374	93380	100797	916654
Sindh	153477	60767	71143	64255	349642
Total Provs.	**795000**	**446554**	**203551**	**194309**	**1639414**
AJK	28020	23772	1798	2077	55667
FANA	7429	4781	925	1183	14318
FATA	28522	11490	0	0	40012
ICT	3466	2876	3941	4283	14566
Total Areas	**67437**	**42919**	**6664**	**7543**	**124563**
Grand Total	862437	489473	210215	201852	1763977

On this basis, the following principles were defined for sample selection:

a) The number of explicit strata would be 38 rather than 32, to allow for the "mixed" schools in Sindh, AJK and ICT.
b) Every explicit stratum was to be represented in the sample.
c) The smallest number to be sampled from any explicit stratum was 100.
d) Where possible, reporting would be based on a minimum of 900 cases in each explicit stratum – which would give a precision of 16% of SD at 95% confidence.
e) Because of the large sample size it was necessary to collapse strata – for example, to report for Rural vs Urban and Boys vs Girls in a Province, rather than Rural Boys vs Urban Boys or Rural Girls.
f) Where strata are collapsed in this way, the target was to be 900 cases in each stratum.
g) Schools were to be selected with Probability Proportional to Size (PPS).
h) The maximum sample size should not exceed 15,000, if at all possible.

On this basis, a revised sample design was agreed. This is given in Table 40 below:

Table 40. Proposed Sample: by Province/Area, Gender and Rural/Urban

Prov./Area	Rural			Urban			Total
	Female	**Male**	**Mixed**	**Female**	**Male**	**Mixed**	
Balochistan	500	500	0	500	500	0	2000
NWFP	500	500	0	500	500	0	2000
Punjab	1200	1200	0	1200	1200	0	4800
Sindh	500	500	500	500	500	500	3000
AJK	500	500	0	100	100	0	1200
FANA	300	300	300	100	100	100	1200
FATA	600	600	0	0	0	0	1200
ICT	100	100	100	100	100	100	600
Total	**4200**	**4200**	**900**	**3000**	**3000**	**700**	**16000**

Even this large design inevitably involved some compromises. For example, no attempt was made to include enough cases in ICT to report separately for explicit strata; no attempt was made to report for urban strata in AJK and FATA.

The Achieved Sample

The total NEAS sample in 2005 was 11,977 grade 4 students. Table 41 provides information regarding the actual numbers of schools and students involved in the 2005 grade 4 national assessment.

Table 41. Achieved Sample by Province/Area

Province/Area	Number of Schools	Number of Students
Balochistan	90	1305
NWFP	99	1630
Punjab	237	4061
Sindh	152	2128
AJK	57	745
FANA	59	719
FATA	60	847
ICT	30	542
Total	784	11977

The sample design achieved covered the entire nation well. Testing took place in 117 out of 126 districts and federal regions. The rate of participation was lower than expected (74.86% of the planned sample).

This in itself is not a major problem; 12,000 is still a good-sized sample; but while some provinces and areas had relatively high participation rates, others were much less well represented. Those Provinces/Areas with above-average participation rates are shown in Table 42 below:

Table 42. Provinces/Areas with above-Average Participation of Students

	Planned	Achieved (Number)	Achieved %
ICT	600	542	90.33
Punjab	4800	4061	84.60
NWFP	2000	1630	81.50
Grand Total	7400	6233	84.23

These are the relatively densely-populated provinces and areas. Provinces and areas with below-average participation rates, as shown in Table 43 below, tended to be those with scattered populations, at least in some areas. It is therefore not at all surprising to discover that the main reason for the shortfall was lower participation in rural areas.

Table 43. Provinces/Areas with below-average Participation of Students

	Planned	Achieved (Number)	Achieved %
FANA	1200	719	59.92
AJK	1200	745	62.08
Balochistan	2000	1305	65.25
FATA	1200	847	70.58
Sindh	3000	2128	70.93
Grand Total	8600	5744	65.79

Participation in rural areas in general was much lower than in urban areas, as summarized in Table 44.below.

Table 44. Sampling Outcome: Rural vs Urban

All Pakistan	Rural	Urban
Intended	9300	6700
Achieved	5684	6293
Achd. As % of Intdd.	61.12	93.93

The impact of population density can be seen in chart 1 below. Overall, almost 20% of class 4 students in Balochistan as a whole are in very small schools (five class 4 pupils or fewer), and a total of about 40% in schools with ten or fewer pupils in class 4. By comparison, Punjab has less than 5% in schools with five or fewer class 4 pupils, and less than 15% in schools with ten or fewer pupils in class 4. Sampling, therefore, is inevitably a far greater challenge in Balochistan than Punjab, because the proportion of small schools is so much higher.

Similarly, the problem is not equally great in rural and urban areas. As chart 2 below shows, the number of class 4 students in Balochistan who are in the very smallest classes (5 or less) is around 20% for rural boys, and 30% for rural girls. By contrast, very few urban boys or urban girls are in such small schools. Therefore, gathering a sample in rural areas poses a far greater challenge than in urban areas, at least in Balochistan. The same is not true in the same way for Punjab, as chart 3 shows. Around 5% or less of Punjab students are in very small schools, even in rural areas.
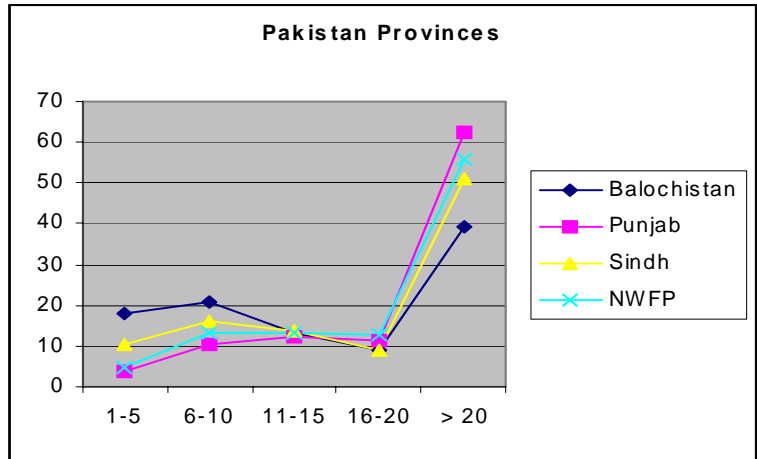
Figure 9. Percentages of Class 4 Pupils in Small Schools, by Province
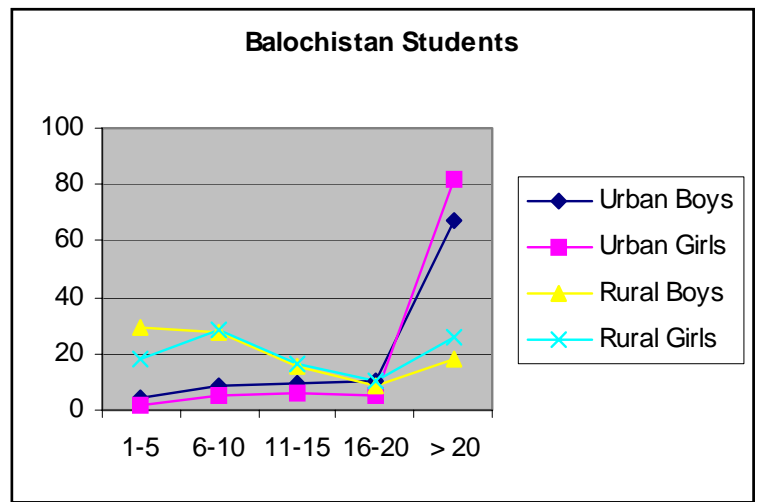
Figure 10. Percentage of Students by Class Size: Balochistan
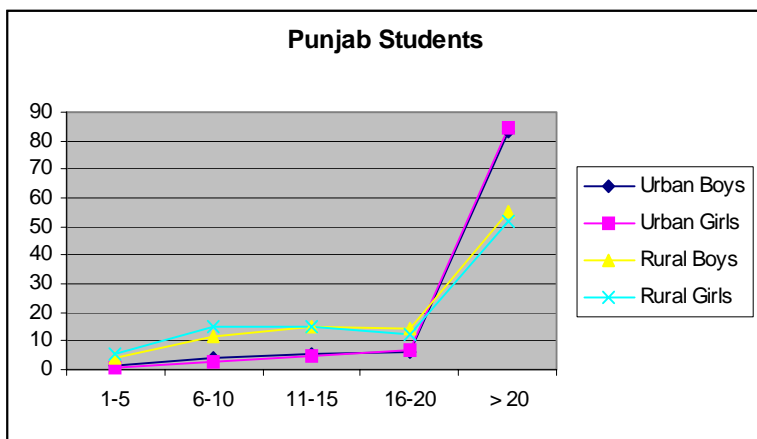
Figure 11. Percentage of Students by Class Size: Punjab

Options for Improving Coverage

Data-gathering problems are essentially rural, and are concentrated in the particularly sparsely-populated provinces and areas. It is however critical that students in sparsely-populated rural areas should be represented in the NEAS sample, because in some provinces and areas they are such a large proportion of the population that the results would be significantly distorted if they are not included. Whatever measures are taken to refine the NEAS sample design, it should respond to this reality. Two specific measures are:

    i.   eliminate the very smallest schools from the frame; and

    ii. increase the number of schools where the average size is small.

INSTRUMENT DEVELOPMENT

Item writing for the 2005 National Assessment was originally conducted in 2002 when a Competency Development and Item Writing Workshop took place.

The workshop objectives were as follows:
- To develop technical capacity and sustainability for test development that did not exist in 2002
- To develop grade 4 Competencies identifying an hierarchy of abilities;
- To develop Tests Specifications according to the weightage identified in the curriculum documents;
- To develop test items for four subjects, Urdu, Mathematics, Science and Social Studies and Sindhi language and Sindhi translation*.

The participants in the Workshop were members of the Curriculum Wing and Subject Specialists from the Provincial and Area Bureaus of Curriculum/Assessment Centres.

The workshop was activity based with all participants being involved according to their subject specialism. The outcome of the workshop was the development of **competencies, test specifications and items** (100 for each subject). These competencies and test items were used for the pilot testing of mathematics and Urdu in 2004 and for large scale testing of mathematics and Urdu in 2005.

The assessment framework developed was the foundation for the national 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 all four subjects (Urdu, mathematics, Science, Social Studies) 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 2004. The same tests were later reviewed and finalised in national level workshops.

---

* Statement about Sindhi language and Sindhi medium tests.

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 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, competency and test specification development **test items** were written. The test items consisted mainly of multiple choice items with a few constructed response items. The main reason for this was that multiple choice items were easily marked and coded while constructed response items required a large amount of training and exemplification of the acceptable answer for a national standard to be achieved.

A workshop was held in February 2003 to develop the **pilot test** in language and mathematics from the pool of items. Two parallel tests were designed. Items which tested the key competencies 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 1995 National Curriculum.

The tests were further reviewed and printed for pilot testing in 2004. Pilot testing was required for the NEAS to ensure that the demands of the tests were appropriate and also to identify items which were reliable, valid and discriminated appropriately.

A sample of 4953 students from across the whole of Pakistan took part in the pilot testing in 2004 in mathematics and Urdu language.

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 May 2005 to a national sample, to establish a baseline for grade 4 languages and mathematics achievement.

The items were then organized into three test booklets. Three test booklets were developed so that students would not be able to copy from each other. The items were ordered in a spiral form.

The **Background Questionnaires** for Head Teachers, teachers and students were also developed. These questionnaires looked at 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.

The **distribution** of assessment materials to the Provincial Education Assessment Centres (PEACEs) and the Area Education Assessment Centres (AEACs), which were managing the further distribution of materials and training of Test Administrators under tight deadlines, was a challenging task. All the materials arrived for the national assessment but there were some delays as the deadlines for delivery were very narrow.

Some of the **difficulties** experienced in instrument development are listed below:
- Lack of continuity of trained personnel;
- Lack of TA support after 2002;
- Delay in analyzing the 2004 pilot test statistics;
- Ensuring that the items address the 2002 National Curriculum. The original items had been based on the 1995 national Curriculum;
- 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;
- The need to review the appropriateness of some of the questions in the light of the 2005 assessment.

COMPETENCIES AND SAMPLE QUESTIONS

Mathematics Class 4 Competencies and Hierarchy

No.	Curriculum Content	Competency The student will be able to:
1	**Number and Algebra**	
1.1	Numbers	Recognise, read and write numbers up to one crore
1.2	Urdu numerals	Read and write Urdu numerals
1.3	Lac, million, crore	Understand the conversion of lac, million and crore
1.4	Odd and even numbers	Identify even and odd numbers
1.5	Addition and subtraction	Add and subtract numbers – mentally up to 50 add two 3 digit and 4 digit numbers Subtract any 4 digit number from a 5 digit number
1.6	Multiplication	Multiply and divide 3 and 4 digit numbers by two digit number
1.7	Distributive property	Understand and apply the distributive property of multiplication over addition and subtraction
1.8	Pakistani currency	Use Pakistani currency notes in simple calculations and problems
1.9	Fractions	Understand different types of fraction (equivalent, proper, improper, mixed)
1.9	Smaller and greater fractions	Identify smaller and greater fractions with the same denominator
1.10	Addition, subtraction and multiplication of fractions	Add, subtract and multiply fractions with the same denominator and with different denominators not greater than 10
1.11	Addition and subtraction of decimal fractions	Add, subtract decimal fractions up to 3 decimal places
2	**Measurement and Geometry**	
2.1	Estimation	Show an understanding by estimating length, capacity and weight
2.2	Computation and problem solving	Compute and solve problems involving similar and compound units of measurement
2.3	Time	Demonstrate knowledge of time using clocks, resources such as timetables, calendars
2.4	Time Problems	Compute and solve problems involving time (simple and compound units) in daily life
2.5	Geometrical shapes	Know and differentiate between square, circle, rectangle
2.6	Line measurements	Estimate, measure a line segment
2.7	Boundaries, interior and exterior regions of common plane figures	Show an understanding of boundaries, interior and exterior regions of common plane figures

No.	Curriculum Content	Competency The student will be able to:
2.8	Perimeter of rectangle and square	Compute the perimeter of rectangle and square
3	Information Handling	
3.1	Line graphs	Identify and interpret line graphs and to show an understanding of uses in daily life

No.	Competency	Knowledge	Understanding	Application
1	**Number and Algebra**			
1.1	Recognise, read and write numbers up to 1 crore	Recognise the numbers	Find the place value	Use real life examples e.g., differentiate between the population of different cities, to demonstrate knowledge and understanding
1.2	Read and write Urdu numerals	Recognise the Urdu numerals	Read and write Urdu numerals	
1.3	Understand the conversion of lac, million, crore	Know that; 1 lac = 1 million 10 million = 1 crore	Convert crore to lac to millions and vice versa	
1.4	Identify even and odd numbers	Define even and odd numbers	Categorise even and odd numbers	Demonstrate that things/objects can be paired and odd things/objects cannot
1.5	Add and subtract numbers - mentally up to 50 add two 3 digit and 4 digit numbers - subtract any 4 digit number from a 5 digit number	Add and subtract mentally up to 50 Add two 3 digit and 4 digit numbers both horizontally and vertically Subtract any 4 digit number from a 5 digit number both horizontally and vertically	Recognise and use appropriate mathematical applications for simple calculations	Use appropriate mathematical applications to solve simple problems in number, money, measurement
1.6	Multiply and divide 3 and 4 digit numbers by a two digit number	Multiply 3 and 4 digit numbers by two digit numbers both vertically and horizontally Divide 3 and 4 digit numbers by two digit numbers both vertically and horizontally	Recognise and use appropriate mathematical applications for simple calculations	Use appropriate mathematical applications to solve simple problems in number, money, measurement
1.7	Understand and apply the distributive property of multiplication and subtraction	Demonstrate knowledge and understanding of distributive property of 'x' over '+' and '-' through completing simple calculations	Solve problems related to daily life	

No.	Competency	Knowledge	Understanding	Application
1.8	Use Pakistani currency notes in simple calculations and problems	Add, subtract using Pakistani currency notes in daily life	Add, subtract, multiply and divide using Pakistani currency notes in daily life	Solve problems related to daily life
1.9	Understand different types of fractions (equivalent, proper, improper, mixed)	Demonstrate knowledge and define common fractions, kinds and decimal fractions	Distinguish between different fractions	Write and convert compound fractions into improper and vice versa
1.10	Identify smaller and greater fraction with the same denominator	Recognise the terms smaller and greater	Understand the difference between greater and smaller fractions	Compute daily life problems relating to fractions
1.11	Add, subtract and multiply fractions with the same denominator and with different denominators not greater than 10	Recognise the different applications in adding, subtraction and multiplying fractions with the same denominator	Use appropriate applications in simple calculations	Use appropriate applications to compute daily life problems relating to fractions
1.12	Add, subtract decimal fractions up to 3 decimal places.	Recognise and define decimal fractions	Convert decimal to common fractions and vice versa	Use appropriate applications to compute daily life problems relating to decimal fractions
2	**Measurement and Geometry**			
2.1	Show an understanding and estimating length, capacity and weight	Recognise the need for formal units of measurement Know that: 1km = 1000m 1m = 100 cm 1kg = 1000gm 100kg = 1 quintal 1 litre = 1000 ml	Demonstrate that objects are of different lengths, capacities and weight Convert and interchange units	Measure the length, capacities and weights Use formal units of measurement in daily life
2.2	Compute and solve problems involving similar and compound units of measurement		Compute simple examples involving similar and compound units of measurement	Solve problems involving formal units of measurement using examples from daily life.
2.3/2.4	Demonstrate knowledge of time using clocks, resources such as timetables, calendars Compute and solve problems involving time (simple and compound units) in daily life	Read the time from clocks Know the number of days in a week and in a month (solar and lunar calendar)	Compute simple examples.	Place events in time sequences Work with timetables, calendars to solve simple problems
2.5	Know and differentiate between square, circle,	Identify and name shapes	Identify the different properties of shapes (shapes which roll,	Classify shapes by simple properties (shapes which roll,

No.	Competency	Knowledge	Understanding	Application
	rectangle		stack, slide, have straight/curved sides)	stack, slide, have straight/curved sides)
2.6	Measure and estimate line segments	Demonstrate the difference between a line and line segment	Measure the given line segment by using a ruler	Solve problems regarding the use of line segments using examples from daily life
2.7	Show an understanding of boundaries, interior and exterior regions of common plane figures.	Show an understanding of the meaning of the terms boundaries, exterior, interior regions of common plane figures Identify geometrical figures	Demonstrate an understanding by identifying examples found in daily life.	Demonstrate knowledge of boundaries exterior and interior of common plane figures in daily life
2.8	Compute the perimeter of rectangle and square	Demonstrate knowledge of the meaning of the term perimeter	Calculate the perimeter using appropriate formulae	Calculate the perimeter of everyday objects e.g. book, table, school, classroom
3	**Information Handling**			
3.1	Identify and interpret line graphs and show an understanding of uses in daily life	Demonstrate knowledge of the difference between pictorial, bar and line graphs	Explain the horizontal and vertical scales Understand the number line	Interpret the information given in a simple line graph

Urdu Language Class 4 Competencies and Hierarchy Levels

No.	Curriculum Content Competency: The student will be able to:
1	**Reading**
1.1	Read silently, simple writing; understand and explain meaning in own words
1.2	Read simple texts aloud
1.3	Read material used in daily life e.g., children's books, handwritten material, newspaper, advertisements and receipts
1.4	Read poems/verses from poems for enjoyment. Memorise favourite verses and show understanding of the meaning
1.5	Differentiate between the different types of writing (story, poem, newspaper, letter)
2	**Writing and Handwriting**
2.1	Copy simple sentences and texts
2.2	Take correct dictation
2.3	Write about things/objects in pictures/incomplete story/picture and identify their important characteristics
2.4	Write, in own words, a read story
2.5	Write briefly for a simple practical purpose, e.g., simple letter, incident, story etc.
2.6	Write briefly about personal experience e.g. diary
3	**Knowledge of Language and Vocabulary**
3.1	Understand and use basic grammar (noun, verb, adjective, subject, object, pronoun, plural)
3.2	Understand and use antonyms, rhyming words, words beginning with the same letter
3.3	Arrange according to alphabetical order
3.4	Understand and use idioms appropriately in sentences

No.	Competency	Level 1	Level 2	Level 3
1	**Reading**			
1.1	Read silently, simple writing; understand and explain meaning in own words.	Read simple sentences silently.	Understand meaning of simple sentences through, e.g., rearranging jumbled sentences; identifying important points in a text; suggesting a title.	Follow instructions given in a text
1.2	Read simple texts aloud	Read simple sentences aloud using correct pronunciation and intonation.	Use punctuation to demonstrate an understanding of meaning	Read with fluency to convey meaning of the text. Able to answer simple questions on the read text.
1.3	Read material used in daily life e.g., children's books, handwritten material, newspaper, advertisements and receipts.	Read simple material silently, with enjoyment.	Understand meaning of simple material through, e.g., rearranging jumbled sentences, identifying important points in a text. .	Follow instructions given in the material
1.4	Read poems/verses from poems for enjoyment. Memorise favourite verses and show understanding of the meaning	Read simple poems/verses from poems aloud and silently, with enjoyment.	Memorise favourite verses	Understand the meaning of a simple poem
1.5	Differentiate between the different types of writing (story, poem, newspaper, letter)	Know the difference between prose and poetry	Differentiate between the different types of writing (story, poem, newspaper, letter)	Understand implications of, e.g., the environment, over population, and make predictions

No.	Competency
3	Knowledge of Language and Vocabulary
3.1	Understand and use basic grammar (noun, verb, adjective, subject, object, pronoun, plural)
3.2	Understand and use antonyms, rhyming words, words beginning with the same letter
3.3	Arrange according to alphabetical order
3.4	Understand and use idioms appropriately in sentences

No.	Competency	Level 1	Level 2	Level 3
2	**Writing**			
2.1	Copy simple sentences and texts.	Copy simple sentences and texts correctly for the most part	Form letters and space words legibly for the most part	Copy simple sentences and texts correctly in good handwriting
2.2	Take correct dictation	Write one sentence correctly	Write at least three sentences correctly with correct punctuation and spelling	Write at least five sentences correctly with correct punctuation and spelling
2.3	Write about things/objects in pictures/ incomplete story/picture, and identify their important characteristics	With the aid of printed words and phrases will write a sentences about things/objects in pictures	With the aid of printed words and phrases will write at least three sentences about things/objects in pictures, identifying the important characteristics.	Write at least five sentences regarding things/objects in pictures. Write an ending to a story. Write a title or heading for a picture story
2.4	Write in own words, a read story.	Write one sentence, using own words, about a read story.	Write, at least three sentences, using own words, about a read story.	Write at least five sentences, in own words, about a read story.
2.5	Write briefly for a simple practical purpose, e.g., simple letter, incident, receipt, invitation etc.	Complete, e.g., a letter, by filling in the gaps.	With the aid of printed words and phrases will write a few sentences.	Write several sentences, in own words
2.6	Write briefly about personal experience, e.g. diary	Write at least one sentence, using own words, about a personal experience, e.g., Eid, visiting a friend. In writing a diary is able to write date etc. correctly	Write, at least three sentences, using own words, about a personal experience, e.g., Eid, visiting a friend. In writing a diary is able to write date etc. correctly	Write at least five sentences, in own words, about a personal experience, e.g., Eid, visiting a friend. In writing a diary is able to write date etc. correctly

No.	Competency
3	Knowledge of Language and Grammar
3.1	Understand and use basic grammar (noun, verb, adjective, subject, object, pronoun, plural)
3.2	Understand and use antonyms, rhyming words, words beginning with the same letter
3.3	Arrange according to alphabetical order
3.4	Understand and use idioms appropriately in sentences

Mathematics Grade 4 Test Examples

Example 1:

Competency 1.4   Number and Algebra
                    Odd and even Numbers (knowledge)

سوال: دیئے گئے اعداد میں طاق عدد ہے۔

20 ا

23 ب

22 ج

24 د

Example 2:

Competency 1.1   Number and Algebra
                    Recognise, read and write numbers up to one crore
                    - Place value (understanding)

سوال: ایک اکائی+3 دہائیاں+7 سینکڑے برابر ہوتے ہیں۔

731 ا

371 ب

713 ج

317 د

Example 3:

Competency 2.2   Measurement and Geometry
                    Compute and solve problems involving similar and compounds of
                    measurement (understanding)

سوال: ایک پنسل کی قیمت 3 روپے ہے۔ 12 پنسلوں کی قیمت کیا ہوگی؟

i   ایک پنسل کی قیمت _____ روپے۔

ii   12 پنسلوں کی قیمت _____ روپے۔

Urdu Grade 4 Test Examples

<u>Example 1:</u>

Competency 3.1   Knowledge about language
Understand and use basic grammar
(Noun, Verb, Adjective, Subject, Object, Pronoun, Plural)

سوال: شاہد اسکول گیا۔

اس جملے میں "شاہد" قواعد کے لحاظ سے کیا ہے؟

ا	فعل
ب	فاعل
ج	صفت
د	مفعول

<u>Example 2:</u>

Competency 3.1   Knowledge about language
Understand and use basic grammar
(Noun, Verb, Adjective, Subject, Object, Pronoun, Plural)

سوال: کونسا جملہ درست ہے؟

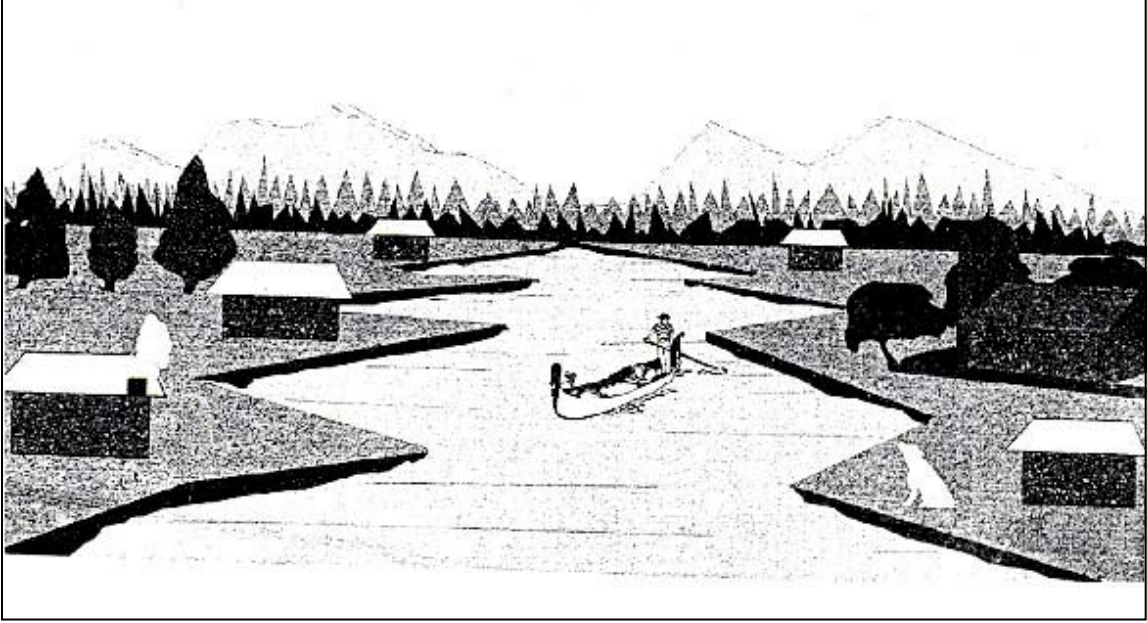|	
ا	اسلم نے بازار سے سبزی خریدی
ب	اسلم کو بازار نے سبزی خریدی
ج	اسلم سے بازار نے سبزی خریدی
د	اسلم نے بازار پر سبزی خریدی

Example 3:

Competency 2.3   Write about a picture and identify the important characteristics (level 1)

نیچے دی گئی تصویر کو غور سے دیکھئے اور دیئے گئے سوالات کے جوابات دیجئے۔

سوال: اس تصویر میں کس جگہ کا منظر دکھایا گیا ہے؟

جواب:

سوال: تصویر میں کتنے گھر نظر آرہے ہیں؟

جواب:

Example 4:

Competency   1.3   Read and understand material used in daily life (level 2)

نیچے دیئے گئے الفاظ سے بامعنی جملہ بنائیے۔

سوال: پودوں ہے پانی مالی کو روزانہ دیتا۔

جواب: _____

سوال: نے اللہ اسلم کیا ادا کا شکر

جواب: _____

51

Example 5:
Competency   1.4   Read poems/verses and understand their meaning (level 3)

<div dir="rtl">

درج ذیل نظم پڑھ کر نیچے دے گئے سوالات کے جوابات دیجئے۔

ایک دو تین چار

بس سے یوں نہ لٹکو یار

پانچ چھ سات

چلنا دائیں ہاتھ

آٹھ نو دس

ڈرئیور چلائے بس

سوال: نظم میں سے دو ہم قافیہ الفاظ لکھیے۔

جواب: _____

سوال: ہمیں سڑک کے کس طرف چلنا چاہیے۔

جواب: _____

</div>

TEST ADMINISTRATION

The tests were administered in May 2005, by teachers in the Provinces and Areas according to the instructions in the Test Administration Guideline Booklet[9] developed. The teachers were trained by the Master Trainers from the PEACEs and AEACs, FCE, IER, AIOU, NIP who had received training in NEAS. The Lead Trainers as well as the NEAS staff monitored the test administration.

For uniformity of test administration, NEAS trained Master Trainers at Federal level. The Master Trainers then trained the Test Administrators (Teachers, Subject Specialists, GCET and PITE faculty) who were to administer the assessment instruments in the Provinces/Areas. The master trainers trained by NEAS were as follows:

Table 45. Number of Master Trainers Trained by NEAS

PEACEs/AEACs	Number of Master Trainers
Balochistan	5
NWFP	4
Punjab	10
Sindh	7
AJK	2
FANA	7
FATA	2
ICT	8
FCE	1
IER	1
AIOU	1
NIP	1
Grand Total	**49**

---

[9] The Test Administrator's Guidelines are available from NEAS

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

PEACEs/AEACs	No of Master Trainers
Balochistan	200
NWFP	200
Punjab	480
Sindh	300
AJK	120
FANA	120
FATA	120
ICT	60
Grand Total	**1600**

Some of the **difficulties** identified in the test administration were as follows:
- There was a delay in the delivery of the Test Administration Manual to the Provinces/Areas, resulting in a delay in the training/non-training of the test administrators;
- 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.

MONITORING NATIONAL ASSESSMENT

Monitoring was undertaken in two areas:
- Monitoring the Administration of National Assessment, and,
- Monitoring the marking and coding of the assessment instruments

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

The monitors consisted of the Master trainers from NEAS, PEACEs and AEACs, FCE, IER, AIOU and NIP. The monitors monitored the assessment activity and reported back to the NEAS on how the test administrators followed the guidelines given during the test administration training.

Monitoring the Marking and Coding of the Assessment Instruments
Marking and coding of all the scripts in the 2005 National Assessment was undertaken in NEAS. Instructions regarding the marking and coding were given to the personnel involved.

The responsibility of the monitors was to:
- Ensure that the marking and coding was conducted in an efficient and fair manner;
- Provide support to those markers and coders who were having some difficulty;
- Identify marking and coding discrepancies and correct them where possible;
- Take a 10% sample of the marking and coding sheets to identify the validity of the marking and coding;
- Ensure that the correct code was being used on the coding sheets;
- Ensure that no marker malpractice is taking place such as the deliberate alteration of a mark to inflate or deflate a student's original mark
- Ensure that the assessment instruments were complete and returned appropriately when the marking and coding was completed.

Some of the difficulties identified in these activities are found below:
- There was some duplication of assessment instrument monitors in the schools – a need for improved communication between the monitors;
- Test Administration Monitors need to be given specific instructions as to how they should behave/react if they observe that the assessment instruments are not being administered appropriately;

- There was a need for the training of the markers and coders 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 marking 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; identical mistakes and peculiarities in the scripts of students (for open ended questions) sitting close to each other; changes in patterns of response to questions, for example, well-written answers with few errors alongside hurriedly written answers with many grammatical and syntactical errors.

MARKING AND CODING ASSESSMENT INSTRUMENTS

For the 2004 pilot, marking and coding methodologies were developed by each of the subject specialists including rubrics for the constructed response items (mostly for language). For the constructed response items (mainly in language) exemplars for the correct answers were provided and agreement was reached as to what was acceptable for a mark to be allocated.

Coding sheets for each subject were further developed in 2005 using 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 not answered a code of 8 was given and where a student has not yet reached the question a code of 9 was given.

Students from the Federal College, Islamabad and teachers from the Federal Directorate of Education, Islamabad were hired for the manual marking and coding. This involved marking and coding of approximately 45,000 assessment items. 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. Due to the large volume of work to be completed in a limited time, delays and problems in marking and coding and later in data entry were encountered.

The staffs of the NEAS, PEACEs, AEACs and ATCs were involved in the monitoring of this activity, checking 10% of the data. Checking the data involved taking every sheet and checking two out of the 20 students on each sheet. This was too onerous a task so it was not possible to check every single sheet. 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 rigors 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.

It was noted that when some of the NEAS staff re-checked the data:
- There were mistakes in the manual data entry. The codes were not accurately entered;
- As only two students out of 20 were checked on each sheet, 18 students' data entry was not checked. Also where there were, for example, only 18 students on a sheet or 50 students on a sheet, only two students were checked. This does not give an accurate representation of a 10% sample.
- There were also some mistakes in the data entry of the scores but these were negligible.

- When the SPSS data was checked there were also some additional mistakes made. Data cleaning in itself also resulted in mistakes. Even some missing figures were found to have been entered.

Marking and coding test booklets is an onerous task. The difficulties that have arisen from the methodology used for marking and coding should be reviewed. More provincial/area participation should be encouraged as recommended in the World Bank Aide Memoir, July 2005 and 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. As well as this the monitoring of the scoring and coding should be much more rigorous and manageable with 10 -20% of the sheets being monitored rather than 10% of 20 students. More monitors should be appointed so that the ratio of monitors to coders and scorers is 1:5.

ITEM AND BACKGROUND DATA ANALYSES

Data analyses at NEAS takes place in three Phases:
- Phase-I consists of Microsoft Excel for sample selection;
- Phase II includes item and reliability analysis of the pilot administration for selection of test items; while,
- Phase-III consists of achievement and background data analysis to determine the correlates and determinants of achievement among the students population.

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

Phase I: Microsoft Excel
This is 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.

Phase II: Item Analysis
Item analysis is carried out using the following programs
a) **ITEMAN** is used for conducting "classical item analysis" for the purpose of item selection. This program provides information on two parameters; item difficulty and item discrimination that could be used in item selection along with information on performance of distractors for each item.
b) **SPSS** is used mainly for data input and some basic descriptive and graphical analysis to supplement ITEMAN outputs.

Phase III: Achievement and Background Data Analysis
c) **WINSTEPS** is based on Rach IRT analysis. For the NAT 2005 WINSTEPS was used for one parameter (item difficulty) IRT analysis as a basis for scaling. IRT parameters are sample independent and provide the basic statistical information about each individual item.
d) For two and three parameter IRT analysis, either **Parscale** or **Bilog-MG** is used. These programs have not been introduced at NEAS yet but they could be used for linking the NEAS results to international assessments such as TIMSS.
e) **SYSTAT** is used to check the consistency of data, editing of data and for some basic descriptive analysis. Because of its more flexible options (e.g., bootstrap) and better all round performance it is sometimes preferred over SPSS.
f) **AM** software by American Research Institute is used for a limited range of analysis at NEAS. Particularly AM is used for application of tests of significance, and computation of raw frequencies. The program has a provision for automatic weighting and estimation of standard errors using Jackknife.