Monitoring & Evaluation Division
Directorate of Primary Education
Ministry of Primary & Mass Education

The National Student Assessment 2015

Grades 3 and 5

Government of People's Republic of Bangladesh

Ministry of Primary and Mass Education

Monitoring and Evaluation Division

Directorate of Primary Education

Revised September 2016

CONTENTS

Acknowledgement

List of Acronyms

Acronyms

AIR	American Institutes for Research
ACER	Australian Council for Educational Research
AUEO	Assistant Upazila Education Officer
BRAC	Bangladesh Rural Advancement Committee
BSS	Bangla Scale Score
CELS	Child Education and Literacy Survey
CPE	Compulsory Primary Education
CRQ	Constructed Response Question
DPE	Directorate of Primary Education
ECCD	Early Childhood Care and Development
EFA	Education for All
GOB	Government of Bangladesh
GPS	Government Primary Schools
HSA	High School Attached
IRT	Item Response Theory
KG	Kindergarten
LASI	Learning Assessment in SEQAEP Institutions
MCQ	Multiple Choice Question
MoPME	Ministry of Primary and Mass Education
MSS	Mathematics Scale Score
NAC	National Assessment Cell
NAPE	National Academy for Primary Education
NNPS	Newly Nationalized Primary School
NCTB	National Curriculum and Textbook Board
NSA	National Student Assessment
PECE	Primary Education Completion Examination
PEDP	Primary Education Development Program
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
RNGPS	Registered Non-Government Primary School
ROSC	Reaching Out of School Children
SRQ	Selected Response Question
TIMSS	Trends in International Mathematics and Science Study
URC	Upazila Resource Centre

List of Tables and Figures

EXECUTIVE SUMMARY

The National Student Assessment (NSA) program in the People's Republic of Bangladesh was initiated in 2006 by the Ministry of Primary and Mass Education (MoPME) to assess achievement in primary education. A key purpose of the NSA is to provide accurate and timely data-driven information to support policy and planning, enhance teacher education programs, and improve classroom instruction in order to increase student learning. This report presents the results of the 2015 NSA – the 5th administration of the NSA conducted since 2006 – for Bangla Language and Mathematics in Grades 3 and 5. In addition to assessing student learning outcomes as prescribed by curricula and content standards, the NSA program investigates differences in pupil achievement by key system, school, and student factors. As a monitoring program, the NSA provides an independent and objective source of information for those seeking a clear view on the state of primary education in Bangladesh.

Beyond the results obtained in the NSA 2015, highlights of which are briefly summarized below, 4 major achievements of the current administration of the NSA and the assessment development work that led up to the administration must be emphasized:

1   The NSA has taken important steps to continue to improve quality and meet international standards in sample-based assessment design, development, analysis and reporting;
2   The government institutions involved in the multiple phases of the assessment have made important gains in the levels of technical and management expertise, responsibility, and ownership of many of the key assessment procedures;
3   Greater focus is being placed on the formative, informative, and pedagogical value of the NSA results through reporting procedures and dissemination of results, increasing the relevance and impact of the program;
4   As a consequence of these achievements, more solid foundations have been laid for the consolidation and longer-term sustainability of the NSA towards becoming a full service assessment unit.

Assessment results of a program like the NSA, which attempt to capture learning outcomes of students who represent national and sub-national achievement, must always be understood in the context from which they are derived. In the case of the NSA 2015, it is important to point out the following contextual characteristics which, to a greater or lesser degree, may have had some impact on the results obtained:

* Bangladesh initiated latest curriculum revision efforts in 2011 with a staggered distribution program of curriculum materials that will not be complete until the 2017 school year. As was noted in the NSA 2013, minimal changes or even declines in measured student performance are reflective of the time it takes to have an impact on instruction and see improvements in student learning when curricular and instructional reforms are taking place;
* Some loss of instructional time was experienced during both the 2014 and 2015 school years, mostly due to a variety of national socio-political events in the country;
* Improvements in access to the education system and its resources, typically targeting traditionally marginalized groups of students who represent the lower performance levels in the system, may have had negative effects overall on student performance as measured on the NSA;

- Motivation among schools, teachers, and students to participate on a low-stakes assessment such as the NSA has been noted to be somewhat low, especially in Grade 5 where there is a greater need for more rigorous attention to the high-stakes national primary education completion examinations (PECE);
- There are important differences between what was measured on the NSA 2015 as compared to the PECE, with a greater focus on cognitively more challenging test items on the former. This in part may also have an effect on motivation.

This report highlights NSA 2015 results as well as key performance trends between the 2011, 2013 and 2015 assessments which must be understood in light of the contextual constraints described above. Because the NSA tests across 2011, 2013 and 2015 were equated and placed on a common measurement scale, changes in performance across grade levels can be compared.

Table 1 below summarizes the participation rates for both Mathematics and Bangla Language Grades 3 and 5 among students and schools:

Table 1. Participation Rates of Students and Schools in the NSA 2015

	Students	Schools
Bangla Grade 3	22,889	
Mathematics Grade 3	22,954	1185
Bangla Grade 5	19,388	
Mathematics Grade 5	19,383	

All 3 administrations of the NSA have shown good test reliability (see Table 2 below) with a significant average 5-point increase in reliability shown in the NSA 2015 tests, in other words demonstrating that test items have consistently measured the targeted constructs:

Table 2. Test Reliability Coefficients (Cronbach Alphas) for 2011, 2013, and 2015

	2011	2013	2015
Bangla Grade 3	0.84	0.83	0.91
Bangla Grade 5	0.82	0.81	0.87
Mathematics Grade 3	0.87	0.85	0.89
Mathematics Grade 5	0.88	0.86	0.89

Overall,2015 NSA performance results in terms of scale scores (raw scores that are transformed onto a single scale allowing for comparisons across the different administrations) show negligible to small declines in Bangla Language as compared to 2011 and 2013, and small to moderate declines in Mathematics (see Table 3 below). These differences appeared statistically significant in both grades and both subjects across administrations. However, when differences across administrations are evaluated in terms of effect sizes (using Cohen's *d*) rather than tests for statistical significance which do not take into account the difference in size between groups, Bangla language score differences are evaluated as negligible to small (+0.05 to -0.29, i.e., not particularly significant) while declines in Mathematics are considered as small to moderate (-0.21 to -0.76, i.e., somewhat significant).

These losses across administration years within grade (i.e., horizontal losses) are contrasted with gains from Grade 3 to Grade 5 within administration year (i.e., vertical gains): in Bangla Language ranging from 16 score points in 2011, 11 score points in 2013, to 14 score points in 2015. In Mathematics, we

see a similar pattern: rises of 18 scores points in 2011, 12 in both 2013 and 2015 from Grade 3 to Grade 5. Vertical gains are unsurprising and to be expected but worthwhile mentioning.

Table 3. Test Scale Scores for 2011, 2013, and 2015

	2011	2013	2015
Bangla Grade 3	100.2	104.2	100.8
Bangla Grade 5	116.2	115.2	114.1
Mathematics Grade 3	100.8	103.7	98.4
Mathematics Grade 5	118.6	115.8	110.2

Viewed from the perspective of performance bands – using the bands developed by ACER and described in the 2011 technical report (henceforth referred to as "legacy bands") – results provide a more dramatic picture (see Table 4 below). Bangla Language in Grade 3 remains somewhat stable across administrations, with approximately 68%-75% of students achieving AT or ABOVE grade expectations (i.e., Bands 3-5). In Grade 3 Mathematics, the percentage of students who reach grade level expectations decreases to 41%-57%. Grade 5 results on the "legacy bands" are significantly more dramatic. In Bangla Language Grade 5 only between 25%-29% of students across the 3 NSA test administrations achieved at grade level (i.e., scored at Band 5). In Mathematics Grade 5, the figures range from 11%-32% of students who achieve on grade level.

As we explain in Chapters 2 and 3, we believe that the "legacy bands" do not provide a convincing methodology to accurately capture performance scale achievement, mostly because the same content is used to estimate both Grade 3 and Grade 5 performance. A casual glance at the data in Table 3 indicates that while 2/3 of students in Grade 3 Bangla Language are performing at grade level, and in Mathematics approximately 1/2 perform at grade level, in Grade 5 only 10%-30% of students are performing at grade level in both subject areas. This large difference seems to suggest that the methodology used for determining performance levels is not functioning in a reasonable way.

Table 4. Test "Legacy Band" Percentage Scores for 2011, 2013, and 2015

Administration year…	2011		2013		2015	
Bands…	Below grade (1-2)	On/above grade (3-5)	Below grade (1-2)	On/above grade (3-5)	Below grade (1-2)	On/above grade (3-5)
Bangla Grade 3	32%	68%	25%	75%	35%	65%
Mathematics Grade 3	50%	50%	43%	57%	59%	41%
Bands…	Below grade (1-4)	Grade (5)	Below grade (1-4)	Grade (5)	Below grade (1-4)	Grade (5)
Bangla Grade 5	75%	25%	75%	25%	77%	23%
Mathematics Grade 5	68%	32%	75%	25%	90%	10%

Further details regarding Bangla Language and Mathematics achievement data together with analysis across the 3 NSA test administrations (2011, 2013, and 2015) is provided in Chapters 2 and 3.

Some of the additional noteworthy contextual observations that can help interpretation of the Bangla Language test results discussed in Chapter 2 are summarized here:

- The Bangla Language tests for the NSA 2015 were developed on the basis of a revised test blueprint aligned with the 2011 modified national curriculum. While the development of this curriculum was initiated prior to the 2013 NSA, the pedagogical materials based on the new

- curriculum have only started to influence instruction in the targeted grades in 2015. Distribution of materials for the 2011 curriculum will be completed in 2017;

- The Bangla Language blueprint for both Grades 3 and 5 prioritizes the measurement of communication skills in reading: comprehension of ideas in different types of texts, and the role of grammar and vocabulary in conveying those ideas, as required by the 2011 modified national curriculum;

- There are no significant differences in performance by gender in either grade across all 3 NSA administrations. Performance band scores within each administration vary by 1-3 percentage points between male and female students. The NSA 2015 scores, as with 2011 and 2013, show clear evidence of gender parity– this is not typical of gender-differentiated performance in many other countries in this region and represents an important achievement of the Primary Education Development 3 program (PEDP3);

- In terms of Bangla Language scores by school type on the NSA 2015, Government Primary Schools (GPS) and Newly Nationalized Primary Schools (NNPS), which represent 77% of the total school sample for Grade 3 and Grade 5, are inevitably close to the total mean. KG schools have consistently outperformed all other school types in both grades and in the 2013 and 2015 administrations of the NSA;

- All test items, for both Bangla Language and Mathematics, are designed to measure specific curriculum content at one of 4 cognitive processing levels: Knowledge level (requiring recall of information); Understanding (requiring demonstration of comprehension of concepts); Application level (requiring demonstration of use of knowledge and concepts for the resolution of problems); and Higher order thinking skills (requiring demonstration of an ability to synthesize, analyze and evaluate). In the NSA 2015 Bangla Language Grade 3, the average percentage score on higher order thinking items was 40% while on knowledge items the average score was 70%. In Grade 5, percentage scores rise to 60% on higher order thinking items and close to 80% on understanding items.

- Students in Grade 3 Bangla Language show that they have difficulty responding to open-ended items (items that require a written response) with scores at around 40% in the NSA 2015.By Grade 5, students no longer have this problem with scores on Bangla Language open ended items rising to close to 65%.

Chapter 3 presents the details of results and analysis of student performance in Mathematics. In addition to the information provided above related to test reliability, scale scores in Grades 3 and 5 as well as "legacy band" performance scores in Mathematics, the following key points should be made:

- The Mathematics framework used as the basis for the design of the NSA 2015 tests underwent far fewer changes than those seen in Bangla Language. The framework focuses on the 4 domains of: Number properties and operations; Measurements and units of measurement; Shape and space; and Data. The data domain was only assessed in Grade 5;

- As with Bangla Language, data show gender parity with a 0-2 point difference between genders on Grade 3 and Grade 5 on each of the 2011, 2013, and 2015 iterations of the NSA; this represents a significant achievement for Bangladesh;

- In terms of student performance relative to the different cognitive processing levels assessed, scores on the less complex level (recall of information) were approximately 61% in Grade 3 rising to 68% in Grade 5; at the more complex levels (application and higher order thinking) scores were 50% (Grade 3) and 47% (Grade 5). There was no significant difference by gender.

There clearly is a significant difference between performance on factual recall versus application in both grades;

- In terms of test item type, which is highly correlated with cognitive level since constructed response items are more typically used to measure application and higher order skills, student performance in Mathematics ranged from around 60% on multiple choice items in both grades to between 37%-44% on constructed response items – a significant different of some 20 percentage points, highlighting the need for students to develop greater skills in more complex levels of cognitive processing of information;
- With respect to school type in Mathematics performance we see a similar pattern as with Bangla Language – KG schools (representing about 7.6% of the total sample for both grades in 2015) outperform all other school types in both 2013 and 2015. Schools of the Reaching-Out-of-School Program fell significantly from 2013 to 2015 from high performing to worst performing. Government Primary Schools (GPS), representing about 60% of the sample in 2015, performed around the mean in Grade 3 while outperforming all school types in Grade 5.
- By geographic division, as with Bangla Language Rajshahi scored highest in both 2013 and 2015 in both grades, having been one of the worst performers in 2011. Sylhet consistently performs the worst of all divisions, across all 3 iterations of the NSA. Dhaka made significant improvements in 2015 from 2013, although has shown erratic performance across the 3 NSA assessments since 2011 when it was the top-performing division.

Chapter 4 focuses on contextual factors associated with achievement through survey data derived from all students who participated on the NSA 2015. The following are two highlights that are worthwhile bringing to the fore here, especially given the potential implications that exist for support programs:

- As in 2013, between-school factors continue to account for large differences in student achievement across schools (in excess of 60% in 2013 and close to 60% in 2015). Explanations postulated in 2013 ranged from differences in resources, impediments to implementing improvement programs and increases in enrollment of low performing students into the educational system. With the introduction of the newly modified curriculum beginning to reach schools in 2015 and continuing into 2016 and 2017, it is to be hoped that professional development will accompany resource distribution and begin to have a positive effect on the large disparities in achievement across schools.
- As is to be expected, a statistically significant relationship was found between a mother's educational attainment level and student performance. Mothers who have completed Grade 5 primary education give rise to a 2 point increase in student achievement on the Bangla Language test over those students whose mothers report being illiterate. This difference rises to a 3.7 increase when the mother has completed Grade 8, and to a 5.6 point increase when the mother has completed secondary school. Similar figures, although a little less dramatic, were found for the impact of a father's educational attainment on student performance. This finding would seem to suggest that there is value in focusing on the educational needs of illiterate families.

Chapter 5 looks at data obtained from surveys conducted among teachers and head teachers of sampled schools. The following are brief highlights:

- In terms of demographics, the average age of sampled head teachers was 43, with an average of almost 12 years in service. 30% reported having a post-graduate degree, 36% a first degree, 19% a high school certificate, and 10% a lower secondary certificate or no certificate at all.

77% reported having a degree at some level in education. These data would seem to suggest a mostly well prepared, experienced and reasonably young head teacher profession.

- 3,278 teachers completed the teacher survey, with an average of almost 13 years in service, and average age of 36. 23% had a graduate degree, 33% were graduates of higher education institutions, and 25% had a secondary education certificate while 17% reported having only a lower secondary certificate or no certificate at all. 68% reported having an education degree at some level. 30% of teachers surveyed did not respond to the question on professional education while of those who did 53% had studies in areas other than Bangla Language and Mathematics. 37% of teachers failed to report the subject that they were teaching. The average number of male teachers per school was 3.5 and 4.3 for female teachers. Overall one would conclude that half of the teachers in Grades 3 and 5 (of the 70% of the sample that responded to the question of educational training) are probably not trained in the 2 targeted subject areas of Bangla Language and Mathematics while almost 20% have only a lower secondary certificate or no school certificate at all;
- Class length, in terms of minutes per day, ranged from 50 minutes (36% of total number of teachers reporting) to 35 minutes (30% of the total);
- From the survey data teachers overwhelmingly project a sense of satisfaction with their job with some 91% wanting to remain in the profession or progress to administrator positions. Over 80% believe their work helps to build good citizens, that their working environment is positive, and the profession stable;
- In terms of the relationship between teacher survey data and student performance, students whose teachers had received recent training in Bangla Language instruction scored at statistically significant higher levels than those students of teachers who had not received such training;
- Teacher educational attainment was positively associated with increased student outcomes (using achievement data from Mathematics Grade 5); for example, students of teachers with a university degree scored 4 percentage points higher than students of teachers with no completed secondary school certificate (the reference group), while students of teachers with a completed secondary school certificate scored 3.6 points higher than the reference group.

Chapter 6 provides discussion on the implications of the results of the NSA 2015. It also provides concluding remarks geared toward offering suggestions for improvements for the 2017 NSA to be conducted under post-PEDP3. The recommended improvements to the 2017 NSA are summarized here (see Chapter 6 for a more detailed discussion of these recommendations):

1. Introduce into the NSA 2017 the domain of writing in Bangla Language; because of the complexities associated with such a test, mostly related to scoring, we would recommend conducting an assessment of writing through a sub-sample;
2. Conduct standard-setting using NSA 2017 data – this would enable test results for 2017 to be interpreted in terms of grade-specific performance scales, defined by the content standards measured on the tests. This would provide a useful pedagogical tool for teachers, schools, and districts in their setting of targets, planning of instruction, and monitoring of ongoing achievement of targets. Results from 2015 and earlier can be placed on the performance scale since scores are already scaled from 2011 onwards;
3. Intensify DPE and NCTB capacity-building around key technical areas in order to consolidate and improve skills already acquired;
4. Identify implementation indicators from post-PEDP3 program implementation plans and measure them concurrently with the NSA so that achievement results on the NSA can be correlated with school support initiatives;

5. More preparation and monitoring of the administration of the NSA in the field as well as data cleaning and scoring needs to take place in order to ensure availability of high quality data;
6. Greater focus on formative uses of the results of the NSA should be integrated into the study; teachers, schools and districts can benefit greatly from the availability of data and school reports for instructional planning purposes;
7. Consider the NSA 2017 as a baseline for the implementation of the newly modified national curriculum as well as a baseline for post-PEDP3 programming;
8. Greater promotion of the NSA needs to be conducted so that differences between the PECE and the NSA are understood in the educational profession, what the different goals of both are, and why it is necessary for students and schools to take these tests seriously;
9. Collapse the 2 higher cognitive levels (combining application and higher order thinking) into one single level – distinguishing between these two cognitive levels when designing test items for primary levels, especially at Grade 3, is not easy, although one begins to see greater differentiation in the upper grades.

INTRODUCTION

Over the last decade Bangladesh has made significant progress in increasing access to primary education. Improving educational quality, and not just access, however, has recently become a top strategic priority. Quality education provides students the opportunity to acquire the knowledge and skills needed to participate productively in the growth of their country; improving the educational performance of primary grade students is critical for both economic progress and for the development of active and well-educated citizens.

The National Student Assessment (NSA) program is an initiative of the Ministry of Primary and Mass Education (MoPME) to evaluate achievement in primary education. A key purpose of the NSA is to provide accurate data and information to inform policy planning, enhance teacher education, and improve instruction in order to improve student achievement. The NSA for Grades 3 and 5 was first conducted by the Directorate of Primary Education (DPE) of the MoPME in November of 2006. This was followed by a second administration in November 2008 and two more cycles in 2011 and 2013. A key difference exists between the first two iterations of the NSA (2006 and 2008) and the second two (2011 and 2013); in the latter, a commonly used methodology was employed to link assessments in both years so that valid comparisons across the two administrations by subject and grade could be made. This methodology was not used in the first two years of the NSA and therefore the 2006 and 2008 assessments stand as independent administrations and are not validly comparable between them or with any ensuing assessment. The 2011 NSA served as a baseline for the PEDP3 and was designed with its linking methodology to be able to offer comparative data with 2013, 2015 and beyond in order to monitor student progress over time and relative to PEDP3 indicators.

However, the NSA itself is not designed to capture data relative to any program implementation conducted under the PEDP3 and therefore does not provide sufficient evidence for the effects of the PEDP3 program. It would be good to have data on the degree of exposure of NSA students to any PEDP3 program activities in order to measure its impact.

NSA Objectives and Use of Results

The objective of the NSA program is to provide high quality, reliable assessments from which valid inferences can be drawn about the state of two essential primary school subjects in Bangladesh: Bangla Language and Mathematics in Grades 3 and 5. The results of the NSA provide the government, researchers, educators, and parents with information that, if acted upon in a timely fashion, can lead to improvements in policy making, resource allocation, instruction, and pedagogical program design, just to name a few areas. Foundational questions that can be answered with NSA results include: How well are students learning the various content domains (numbers and operations vs. geometry vs. measurement in Mathematics for example)? Is there evidence of strengths and weaknesses in particular knowledge and skills? How are the various sub-groups performing in the system? What home or school factors are associated with student achievement? What higher cognitive skills are students developing as opposed to merely demonstrating recall or memorization of factual information?

If NSA results are analyzed and reported at the content domain (as well as at finer levels of the domain such as at the strand or student learning outcome levels),and also at the cognitive processing level, the NSA can also provide useful diagnostic and formative information to teachers and school administrators. At the same time, while the NSA 2015 measured national, regional and sub-group

achievement in the specified subjects, it was not designed to report on individual student performance; or to be used for quality evaluation of a particular school or teacher.

What Institutions are Responsible for Developing and Administering the NSA?

The body responsible for managing and conducting the NSA at the Ministry of Primary and Mass Education (MoPME) is the Monitoring and Evaluation Division of the Directorate of Primary Education (DPE). Directly responsible for technical development of the NSA is the National Assessment Cell (NAC) in close collaboration with the National Curriculum and Textbook Board (NCTB), and the National Academy for Primary Education (NAPE).

The content of the tests is determined by specifications provided in assessment frameworks for each subject that describe the specific knowledge and skills to be assessed. The frameworks prescribe curriculum balance and the range and type of test questions that are to be used. They are aligned with the most recent version of the National Curriculum, initiated in 2011. The 2015 NSA design was governed by a recently created document entitled NSA 2015 Assessment Frameworks, created in partnership between the DPE and American Institutes for Research (AIR).

How is Quality of the NSA Ensured?

The design, administration, and analysis of the NSA is led by Bangladeshi assessment experts and content specialists. In the summer and fall of 2015, a rigorous review of NSA curricula expectations, alignment of assessment frameworks with content expectations, and procedures for ensuring assessment quality were all carried out. The focus on the alignment of assessments to the curriculum was particularly important in the NSA 2015 because the national curriculum underwent a major revision beginning in 2011. The effects of this extended revision effort are expected to begin to be seen in 2016 when they will permeate down to the school level. This is described in more detail in Chapter 1 of this report. Particular investment was made to ensure that valid comparative inferences could be made based on NSA results across assessment years. In addition to involving a broad range of experts in the test and item development process, external technical assistance from international specialists from AIR (responsible for the NSA 2015) together with the Australian Council for Educational Research (ACER, who were responsible for the 2011 and 2013 iterations[1]) also assisted with the statistical analysis of pilot and operational test data, scoring and scaling procedures.

How was the 2015 NSA Sample of Students Selected?

The NSA is a learning assessment program that gauges the performance of students in grades 3 and 5 in a nationally representative sample of randomly selected schools. In 2015, samples of 22,954 Grade 3 students and 19,388 Grade 5 students were drawn from a sample of 1185 schools and educational centers to take part in the NSA program. Students from the seven geographic divisions of Bangladesh and seven main types of primary schools from rural and urban regions were chosen to participate in the assessment. As in previous years, in 2015 sampling weights are applied to ensure that any disproportionate representation of student groups did not unduly impact NSA results.

The number of students in each grade sampled from each region was proportional to the total enrolment for that grade in that region. Approximately 50% of the sample is composed of students from the Dhaka and Chittagong Divisions, the most populous divisions in Bangladesh. Figure 1 below presents the breakdown in proportion of students by division for Grade 3.

---

[1] The 2006 and 2008 iterations of the NSA, the first to be conducted, were led by the DPE.

Figure 1. Sample of Participating NSA Students by Division (2015 Grade 3)

2015 Sample by Division (Grade 3)

- Barisal — 6%
- Chittagong — 20%
- Dhaka — 30%
- Khulna — 10%
- Rajshahi — 13%
- Rangpur — 12%
- Sylhet — 9%
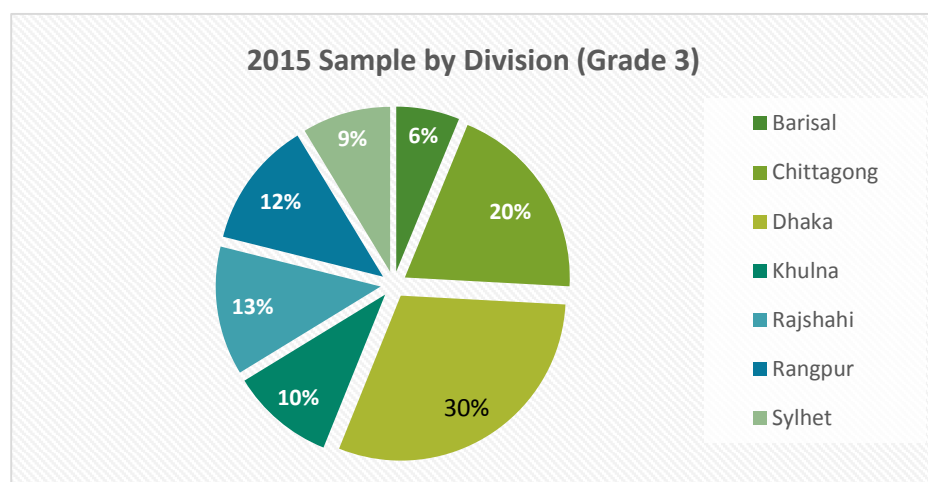
Approximately 77% of all students come from either Government Primary Schools (GPS) or Newly Nationalized Primary Schools (NNPS).[2] Figure 2 below presents the breakdown of the sample by school type for Grade 3.

Figure 2. Sample of Students by School Type (2015 Grade 3)

2015 Sample by School Type

- Government Primary Schools — 59%
- Newly Nationalized Primary Schools — 18%
- Madrasah Schools — 4%
- Kindergarten Schools — 7%
- High School Attached Primary Schools — 3%
- Bang Rural Advancement Committees — 3%
- Reaching Out-of-School-Children — 6%
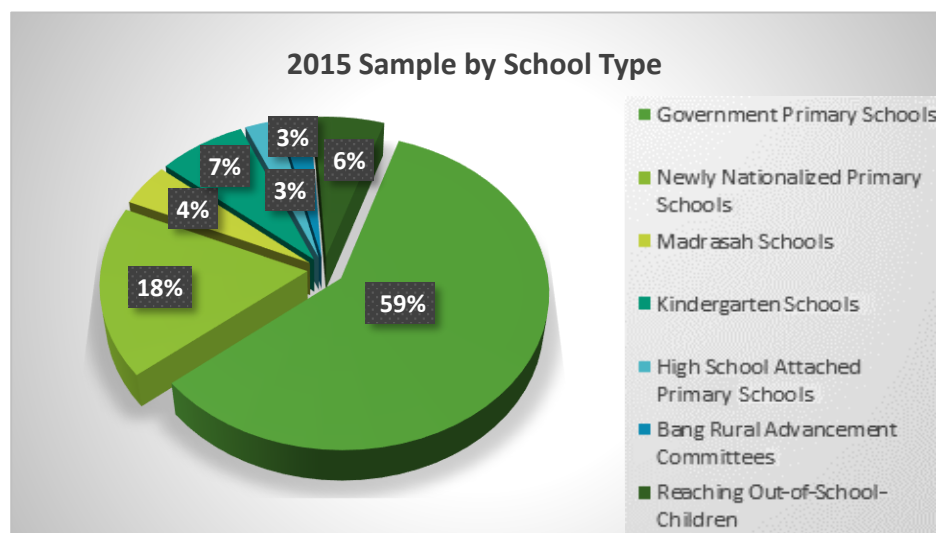
How was the 2015 NSA Administered and Monitored?

The NSA was administered on November 14th, 2015 throughout Bangladesh. The NAC was responsible for all aspects of administration and data collection. As in 2013, students who were selected in the sample at each grade level were expected to sit for both the Bangla Language and Mathematics tests, plus a brief survey to collect background information about their home environment. Test administrators were trained to ensure high levels of consistency of administration across the country.

---

[2] Prior to 2015, Newly Nationalized Primary Schools (NNPS) were referred to as Registered Non-Government Primary Schools (RNGPS).

Teams of quality monitors also visited selected schools during the assessment administration to ensure fair conditions of administration and the highest levels of quality.

How were the NSA Tests Scored?

The majority of the 2015 NSA test items (questions) were selected response questions (SRQ) or items that required the selection of one correct answer from several options (often called 'multiple choice' or MCQ). However, there were also items that required short written responses from students ('constructed response'). The selected response items were scored by computers. To ensure consistency of marking, constructed response questions were marked by teachers specifically trained for the marking task.

How can NSA Results Be Compared from Year to Year?

For the NSA 2015, a methodology based on linking test items was used for vertical equating across grades. NSA tests are equated and placed on a common scale so that the 2015 results can be validly compared with those of 2011 and 2013 as well as with administrations beyond 2015. This enables valid inferences on trends in performance across years and grades.

How was the NSA 2015 Different from Previous NSAs?

While the 2015 NSA is similar in appearance to previous NSA iterations, in 2015 new blueprints for the design of the tests were developed and approved for Mathematics and Bangla Language. Blueprints provide a detailed description of the content and cognitive skills to be measured in a test, and the types of items that can be used to measure knowledge and skills. The standards framework from NSA 2013 was improved for the NSA 2015 in terms of content covered and the articulation of that content, reflecting changes that were prescribed in the new 2011 national curriculum.

Another change this year relates to how test items were piloted. Since inception in 2006, the NSA has piloted test items for operational administrations on an annual basis and separately from the operational administration. For the NSA 2015, an embedded pilot items design was employed, which represents the industry standard in most developed assessment programs. This design assumes that a test is composed of operational items that are used to derive student scores together with a small number of pilot items (4-6) that are embedded in positions among the operational items. This method is more cost efficient and also increases the quality of items by having them piloted tested by students of the targeted grade and at the end of the grade when instruction of the content has been completed.

How were the NSA Results Analyzed and Presented?

This national report presents the results of initial analyses carried out on the NSA 2015 data. Scores and sub-scores are presented by total and sub-score means and by five performance levels or "performance bands" (1, 2, 3, 4, and 5). These performance bands were developed by the Australian Council for Educational Research (ACER) in early iterations of the NSA and a fuller explication of the bands can be found on page 30 (Bangla Language) and page 45 (Mathematics). Performance bands are reported in order to provide a more meaningful interpretation of what students know and can do at each grade level in each subject. NSA scores are further broken down by content domains, cognitive processing levels, and item types. Results are also analyzed by gender, geographic division of residence, and by school type.

Primary NSA Analysis Methods

NSA results for Grades 3 and 5 are presented in this report. General descriptions are provided that enable comparisons across years 2013 and 2015 in terms of proportion of students in the various

performance categories or "band levels," overall mean scores and sub scores, scores by gender, division (region), and school type. Analyses of results by content domain, cognitive processing level, and item type are also presented. Reliability coefficients for Bangla Language and Mathematics in both grades were estimated using Cronbach's alpha, a coefficient of scale reliability, and the Spearman-Brown, split half estimation method.

It should be noted that comparisons across group means are revealing but may be misleading as numerical differences in mean scores in samples are not certain to reflect actual score differences in the total population. Therefore, after mean scores were calculated, any differences in mean scores were tested for statistical significance by conducting independent sample t-tests using SPSS software (significance below 0.05 level). The t-test assumes a null hypothesis of equality of means between the groups under study, such as boys and girls. When comparing means across groups of more than two, it was necessary to employ a univariate analysis of variance model that enabled comparison across three or more groups.

Because tests for statistical significance frequently result in the rejection of the null hypothesis when sample sizes are large, an effect size was also estimated to determine the *practical* significance of the differences in means (Cohen, 1992). Effect size coefficients are expressed in terms of standard deviation units. Effect size values can be interpreted as small (0.2), moderate (0.5 and above), or large (0.8 and above).

In 2015, three surveys were conducted as part of the NSA to collect information about students, teachers, and head teachers, their backgrounds and demographic status. This information was used in various linear regression models to analyze what factors were associated with student achievement.

Structure of the Report

Following the Executive Summary and the Introduction to the report, we begin the body of the report with a focus on highlights in the results of the NSA 2015 (Chapter 1), first of a general nature and then highlights that are specific to each of the two subject areas evaluated, Bangla Language and Mathematics. For those readers who wish to obtain a brief understanding of the NSA 2015 results, without going into the details and more technical explanations, we recommend this first chapter. Chapter 2 focuses on the details of the results on the Bangla Language assessment, first examining the curriculum objectives and expectations measured on the test, followed by the results. Chapter 3 follows the same format for the Mathematics assessment. Chapter 4 looks at those contextual factors that are measured through the student survey and their impact on student achievement, while Chapter 5 examines data derived from the head teacher and teacher surveys correlated with student achievement. Chapter 6 provides a discussion of the implication of the NSA 2015 results for policy makers. For those readers who are looking for a more detailed technical explanation of the assessment development process, data analysis methodology and results, we recommend that they obtain the NSA 2015 Technical Report (see *Bangladesh NSA2015 Draft Technical Report 10 14 16 vs 4* held at the DPE).

CHAPTER 1. HIGHLIGHTS OF RESULTS ON NSA 2015 STUDY

It is increasingly recognized that measuring student learning outcomes can play an important role in monitoring the progress of an educational system. This is seen in greater participation in international testing programs such as the Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS), and the Program for International Student Assessment (PISA), as well as the increase in national sample-based testing programs such as the NSA. Assessment results can reveal the degree of progress made by a specific grade of students, and by targeted sub-groups of those students, compared with the performance of students of the same grade in previous administrations of the same test, psychometrically shown to be of the same content and difficulty level as the current test. Similarly, results may indicate the degree of success of teachers to instruct the assessed content and the degree of success that head teachers and teacher-parent committees have in supporting school progress. Test results, however, must be understood in the broader context from which they are derived. This is no less true of the NSA 2015 in Bangladesh and the entire structure set up to ensure that high quality assessments are developed and administered to deliver test results that are reliable, valid and can be used to improve the educational system at all levels.

Broad-level Achievements of the NSA 2015

Before we examine the broader context for the interpretation of the NSA 2015 results, it is worth highlighting 4 major achievements of the NSA 2015:

- First, and in keeping with goals established in 2006 and observed in successive administrations in 2008, 2011 and 2013, the NSA has made important strides in continuing to improve the quality of its work. Progress includes greater alignment between content standards (which are also more accurately defined) and test items; the inclusion of test items that measure higher order cognitive thinking skills; the adoption of a methodology that allows for pilot-testing of new test items within the operational test forms; greater control over the quality of test item development, among other improvements. In 2015 we can consider that the program is well on the way to being considered on a par with international standards in many of its aspects. These continued improvements have led to the foundations for a quality national assessment system and while there are still important areas to continue to improve upon (for example in the areas of test administration, reporting by reference to a content-defined performance scale, and data analysis using methodologies based on item response theory), the NSA and all of the teams associated with test development and administration have the structures in place to be able to analyze their strengths and weaknesses and plan for continued growth.

- The second achievement of the NSA 2015 relates to the increased capacity of the team members to assume responsibility and acquire technical expertise to carry out the diverse activities that a program like the NSA implies: alignment of test frameworks and blueprints to content standards and curriculum; the design of high quality test items that not only measure lower cognitive thinking skills (for example, memorization and recall) but more importantly the more demanding cognitive skills of analysis, synthesis, and problem solving; the assembly of tests that are balanced for content coverage and difficulty level and permit discrimination of a range of student abilities; the equating of multiple test forms within a test administration as well as across administration years; analysis of test data using methodologies based on

both classical and item response theories; and reporting of test results, disaggregated by targeted sub-domains of the content measured, to diverse stakeholders.

- A third important achievement of the NSA 2015 has to do with the recognition that assessments of this type must underscore the formative, informative, and pedagogical value of assessment results. This is certainly a work in progress and will be significantly richer when the NSA program has conducted a standard-setting process involving the establishment of a performance scale of 3 to 4 levels of achievement, each level defined by the content measured on a test and represented in the curriculum and activity of teachers in the classroom. Such a performance scale will permit the reporting of test results by reference to achievement levels that will provide information about what students can and can't do at each level, disaggregated by sub-domains measured on the test. This becomes a very useful pedagogical tool for teachers, head teachers, school committees, and district supervisors to be able to make data-informed decisions about how to improve learning outcomes and how to set targets for improvement for the following academic school year. For this to happen it is important that the reporting of test results from programs like the NSA is timely and facilitates immediate use by stakeholders at both the central and local levels of the educational system.

- Finally, the NSA 2015 has strengthened the foundations of the government assessment-related institutions, both technically and from a management perspective, toward becoming a full service assessment unit. The goal of this unit must be to ensure its longer term sustainability through continued building of technical and management expertise.

Important Contextual Features Informing the Interpretation of the NSA 2015 Results

Bangladesh initiated new curriculum for grade 1 to 5 in 2011 and developed textbook according to new curriculum in 2012. In January 2013 new textbooks for grade 1 to 5 were distributed throughout the country Teacher guides are expected to be distributed at the beginning of the 2017 academic year. Textbook and teacher guide distribution, together with professional development programs supporting the implementation of a revised curriculum represent impact on instructional activity in the classroom. As can be seen, the process of implementation of a revised curriculum is one that takes place over a number of years and while these curriculum reform efforts in Bangladesh started in 2011 one would not expect to begin to see any measurable impact on teaching and student learning outcomes for a number of years.

The NSA 2013 was developed to measure the prevailing curriculum, that is, the one in place prior to the introduction of curriculum revision efforts which do not begin to have an influence on teacher behavior until the 2014 school year. The NSA 2015, by comparison, was intentionally designed to be aligned with the revised curriculum – it is a curriculum in the process of being implemented. It must be stressed that the implementation of the revised curriculum, by the end of 2015 when the NSA 2015 was administered but no teacher guides were available for grade 3 and 5.

What we see then is a situation in which we have assessment instruments aligned with the new curriculum but classroom practice in many respects continues to be tied to the old curriculum. The consequences are that test scores will inevitably under-represent student ability as defined by the goals of the revised curriculum. We have observed this phenomenon in a number of international sample-based assessment program administrations like the NSA 2015 which occur during curriculum reform movements – initial declines in student scores until implementation is completely consolidated at the classroom level. The same was also noted, but to a lesser degree, in the NSA 2013: " minimal change in student performance over the past 2 years is reflective of the time it takes to see

improvements in student learning as a result of structural curricular and instructional reforms". From this point of view we would prefer to consider the NSA 2015 as a baseline measure of the beginning of the implementation of the revised national curriculum.

In addition to the curriculum issue, there are other important contextual factors at play in the broader Bangladesh educational context that need to be taken into consideration in the interpretation of NSA 2015 test results. These factors are summarized in the following list:

- Some loss of instructional time that affected classrooms during the 2014 and 2015 school years, due to a variety of socio-political events in the country, could have diminished scores on the NSA 2015. Our experience with similar situations in other countries indicated that loss of instructional time adversely affected student scores on external achievement tests.
- As access to the educational system has increased in Bangladesh, the net effect on the student population has been to increase student representation at the lower performance end of the scale, simply because greater access targets students who have traditionally been marginalized from the system. This is likely to have some effect on student performance as measured by the NSA.
- Some constraints were imposed on training activities in preparation for the NSA 2015 which may have had some influence on student performance during testing. It is known that factors such as under-monitored and under-standardized test administration or incomplete training protocols may lead to negative impact on results.
- Anecdotal reports were received of low motivation among school head teachers, teachers and students to participate in low-stakes assessments such as the NSA. This is not unusual in a school year that experiences pressures from more high stakes activities, especially of tests such as the Primary Education Completion Examination (PECE) that are administered in Grade 5 and used for admission into secondary level schooling. We would recommend taking a number of measures to mitigate the problem of lack of motivation and its potential effect on student performance on tests; we have for example in some international locations used public recognition of highest performing schools to increase participants' motivation to perform as well as they can on external achievement tests; we also would highlight the need to provide participating schools with reports of their performance, compared with that of the district, region, or country, and in as timely a manner as possible – this has the effect of showing schools that external assessments of this type are serious and can provide useful feedback to schools.
- There are competing interests between the Grade 5 terminal examinations (the PECE), which focus more on measuring the acquisition of knowledge (memorization and recall skills), and the NSA which focuses on measuring a range of cognitive skills from knowledge to understanding to application and other higher order skills.

General Highlights in the Results of the NSA 2015

The following highlights in the results of the NSA 2015 demonstrate the continued achievements made in student learning outcomes. As we have suggested, we would expect to see more significant gains in the coming years as implementation efforts related to the revised national curriculum take hold and their impact begins to be felt in the instructional behavior of teachers and learning outcomes of students.

- The NSA 2015 scores show that Grade 5 achievement was significantly higher than Grade 3 – this is consistent with growth shown in NSA 2011 and NSA 2013, and indicates sustained and consistent growth in achievement from Grade 3 to Grade 5 in both Bangla Language and Mathematics. Although this growth looks impressive at first sight, it will be more important to evaluate how students achieved in relation to the performance objectives and expectations of their respective grade. From that perspective, the process of setting grade level performance standards is a highly recommended step in defining an evaluation framework for educational attainment in Bangladesh.
- In all subjects and at both grades, KG schools have consistently been top performers among all school types, in both 2013 and 2015(data are not available for KG schools in the 2011 administration). KG schools in Bangladesh are privately owned and run and offer resources that may account for the sustained successful performance of these schools.
- The NSA 2015 scores show clear evidence of gender parity in both Grades 3 and 5 in Bangla and Mathematics – this is not typical of gender-differentiated performance in many other countries in this region and represents an important achievement of the Primary Education Development 3 program (PEDP3), in addition to a goal that has already been achieved in terms of primary education enrolment (i.e., access).
- The NSA 2015 scores of Grade 3 students are largely within the range of Bands 3 and above (as defined in the ACER technical report of the NSA 2011 and which we refer to as the "legacy bands").A low percentage of Grade 5 student scores, however, fall within legacy Band 5. (Although legacy Bands 1-5 bear serious limitations, we are using the bands in this report for descriptive purposes and to provide a link with the 2011 and 2013 data which were interpreted in terms of the same bands. We recommend conducting in-grade standard-setting to overcome the limitations of the legacy bands.)
- The Rajshahi division was the highest performing region of all regions in Bangladesh, in both Bangla and Mathematics and at both grades – this has also been observed in the Learning Assessment in SEQAEP Institutions (LASI) 2015 Grades 6 and 8 Bangla Language, English Language and Mathematics.
- The Sylhet division was consistently the lowest performing region in Bangladesh in both primary (NSA 2015) and secondary (LASI 2015) assessments.

Highlights in the Bangla Language Results of the NSA 2015

The following results in student scores on the Bangla Language assessment stand out as important highlights:

- 65% of Grade 3 students scored at Band 3 performance level or higher on Grade 3 Bangla.
- 91% of Grade 5 students scored at Band 3 or higher on Grade 5 Bangla Language; this would suggest considerable growth across Grades 3 – 5 (see Table5 below).
- By contrast, only 23% of Grade 3 and Grade 5 students combined scored at Band 5 on Grade 5 Bangla, which, given other indicators of performance on Bangla Grade 5 suggests that the legacy bands may not be suitable for the evaluation of student achievement relative to grade level expectations.

Table 5. Bangla Language Performance Level son the NSA 2015

NSA 2015 Bangla Language Performance Levels

Students Attaining (in %):	Band 1	Band 2	Band 3	Band 4	Band 5
Overall Grade 3 Bangla	8	27	40	23	2
Overall Grade 5 Bangla	1	8	26	42	23

Note: The Band range 1-5 describes skills and knowledge measured on both Grade 3 and 5 tests, with the more sophisticated skills and knowledge at the Band 3-5 range. According to the initial interpretation of the 2015 NSA results, performance between Bands 3-5 indicates that students are working at Grade 3, above Grade 3 level, and at Grade 5 level, while at Bands 1 and 2 students are working below Grade 3 level.

- For both Grades 3 and 5, the vocabulary tasks were the easiest while the Reading Comprehension tasks were the most challenging.
- Students answered larger proportions of Knowledge and Understanding questions correctly than Application and Higher Order Thinking questions for both grades.
- Students answered a larger proportion of selected response questions correctly than the constructed response questions.
- In regard to mean scores by school type, for Grade 3 in both 2013 and 2015, KG schools had the highest mean scores, a full 9.8 points higher than the lowest performing category of school type in 2015.
- Schools with High Schools Attached (HSA) scored in the top three in both 2013 and 2015.
- For Grade 5, KG and HSA schools were consistently high scoring school types in 2013 and 2015.
- Madrasah and Reaching Out-of-School Children (ROSC) schools were the lowest scoring in both assessment years.
- There is more differentiation and less mean score bunching by school type on Grade 5 than for Grade 3.
- In regard to geographical division, for Grade 3, the Rajshahi mean score was the highest in 2015 and second highest in 2013. The lowest scoring division, Sylhet, was 9.4 mean points below Rajshahi in 2015, almost one standard deviation in score difference.
- Rajshahi and Rangpur were both in the top four divisions for both assessment years.
- Dhaka's rank order moved from seventh (2013) to second (2015).
- For Grade 5, Rajshahi was also the highest scoring division while Sylhet was again the lowest scoring division. The relationship was similar in 2013 but with Barisal attaining the highest mean score in 2013.
- Between NSA 2013 and 2015, there was is a slight decrease in NSA scores for Grade 3 Bangla Language. The total mean score difference(104.2 to 100.8)was associated with a relatively small effect size for this difference (0.33),[3]making Bangla Grade 3 average on NSA 2015 almost the same as on the 2011 (100.8 vs. 100.2).
- For Grade 5 Bangla Language, the total mean scores how negligible differences between NSA 2011, 2013, and 2015 (116.2; 115.3; and 114.1 respectively), with very small effect sizes.

---

[3]As large sample sizes often return statistically significant results, effect size estimations were calculated for statistical tests using Cohen's D. Values at 0.2 are considered small, at 0.5 are considered moderate, and at 0.8, large.

- For Grade 3, Government Primary Schools (GPS) and Newly Nationalized Primary Schools (NNPS) both decreased by 4-5 points from 2013 to 2015. Together, these two school types contain about 77% of the sampled population.

Highlights of the Mathematics Results of the NSA 2015

The following results in student scores on the Bangla Language assessment stand out as important highlights:

- 41% of Grade 3 students scored at Band 3 or higher on Grade 3 Mathematics
- 81% of Grade 5 students scored at Band 3 or higher on Grade 5 Mathematics which suggests considerable growth across these grades (see Table 6 below).
- By contrast, only 10% of Grade 3 and Grade 5 students combined scored at Band 5 level, which, given other indicators of performance on Mathematics Grade 5, suggests that legacy bands may not be an accurate framework for the evaluation of student achievement relative to grade level expectations.

Table 6. Mathematics Performance Level son the NSA 2015

NSA 2015 Mathematics Performance Levels					
Students Attaining (in %):	Band 1	Band 2	Band 3	Band 4	Band 5
Overall Grade 3 Mathematics	23	36	32	7	2
Overall Grade 5 Mathematics	2	17	42	29	10

Note: The Band range 1-5 describes skills and knowledge measured on both Grade 3 and 5 tests, with the more sophisticated skills and knowledge at the Band 3-5 range. According to the initial interpretation of the 2015 NSA results, performance between Bands 3-5 indicates that students are working at Grade 3, above Grade 3 level, and at Grade 5 level, while at Bands 1 and 2 students are working below Grade 3 level.

- For Mathematics Grade 3, students scored highest on Shape and Space but for Grade 5 they scored the highest on Measurement and Units. For both grades, students scored higher on items assessing Understanding and Knowledge than Application and Higher Order Thinking Skills.
- For Grade 3 by school type, KG schools scored the highest, at a statistically significant difference level from all other schools in the group. The lowest scoring school mean in 2015 were ROSC schools, 8.3 points less than the top mean score.
- For Grade 5, the same two school types (KG schools and GPS) were the top performers in both 2013 and 2015, though their relative positions were reversed with KG's performance improving from 2013 to 2015.
- In 2015, Madrasah schools were the lowest scoring school type, with a mean score exactly ten points below the highest scoring school type. ROSC schools were at or near the bottom in both 2013 and 2015.
- For Grade 3, the Rajshahi division scored the highest, at a statistically significant level above the rest of the group in 2013 and 2015. The lowest scoring division mean, Sylhet, was 9.3 points below Rajshahi in 2015.
- For Grade 5 by division, Barisal fell from 2013 and Dhaka moved up in 2015 in relative ranking, scoring at the same level as Rajshahi. Sylhet remained at the bottom and Rangpur, Chittagong, and Khulna remained in the middle in both assessment years.

- For Grade 3 Mathematics, the total mean scores show irregular variation across the 3 NSA years (2011, 2013, and 2015) from 100.8 to 103.7 to 98.2 respectively. The effect size between 2013 and 2015 was 0.45 whereas between 2011 and 2015 it was just 0.21.
- For Grade 5, total mean scores show a steady decrease across the 3 NSA years (2011, 2013, and 2015) from 118.6 to 115.8 to 110.2 respectively. Between 2013 and 2015, the effect size was 0.49; however, the decrease from 2011 to 2015 shows and effect size of 0.76. We have provided a number of explanations at the beginning of this chapter as to why we think this may be occurring, chief among them the curriculum revision efforts that were started in 2013 and have still to be consolidated at the classroom level.

CHAPTER2. THE 2015 BANGLA LANGUAGE ASSESSMENT

The NSA 2015 Bangla Language assessment for Grades 3 and 5 was designed based on a 2015 revision of the Bangla assessment framework. An assessment framework defines the organizing structure for the construction of tests. It defines the content to be assessed and guides the development of the assessment instrument. Frameworks capture a range of subject and grade-specific content and cognitive skills and are defined by curriculum documents and professional best practice. The framework prescribes curriculum balance and the range and type of test items that are to be used.

Curriculum Objectives and Content Expectations

As the mother tongue for the vast majority of citizens of Bangladesh, Bangla should enable students to develop creative thinking, imagination, and artistic awareness. It is also the medium to understand other subjects on the curriculum, and to attain knowledge and progress in life. So that students may use the language effectively for these purposes the curriculum aims to enable them to read, write and speak creatively as well as correctly. Appendix 1 provides a list of the key reading skills, organized from key area (domain) to strands to learning outcomes, that students are expected to develop in Grades 3 and 5.

The content standards for both Grades 3 and 5 Bangla Language are in many respects the same; however, the content differs in terms of text appropriateness for grade and age, as well as grammatical and lexical complexity. While Bangla Language content measured in earlier iterations of the NSA compared with NSA 2015 was, in respect of reading comprehension, essentially the same (i.e., the measurement of ideas communicated in a text) and aligned to the curriculum, the content for NSA 2015 was organized to facilitate more logical sub-score analyses.

Bangla Assessment and Content Expectations
Reading

In the Grade 3 and Grade 5 reading assessments, students were required to identify, interpret, infer and synthesize information focusing on:

- reading for meaning in literary texts
- reading for meaning in factual texts
- textual devices (e.g. spelling, punctuation, word construction)
- syntax
- vocabulary

Text types

The Bangla assessment included three broad categories of texts: Imaginative texts, information texts and argumentative (or persuasive) texts. Texts were between 50 – 150 words in Grade 3 and 80 – 200 words in Grade 5.

- Imaginative texts: texts that involve the use of language to represent, recreate, shape and explore human experiences in real and imagined worlds. They include, for example, fables, short stories, novels and plays. Included in imaginative texts are narrative and descriptive fictional text types.
- Informative/descriptive texts: non-fictional texts that involve the use of language to represent ideas and information related to people, places, events,

things, concepts and issues. They include, for example, reports, descriptions, biographies, explanations, news articles.

- Argumentative/persuasive texts: texts that systematically present a point of view and seek to persuade or change the behavior or attitude of the reader. They include, for example, formal essays, letters, advertisements, interviews and reviews.

The assessment provided a measure of reading performance that reflected students' typical reading experiences. Contexts were relevant to students and grade/age appropriate. Texts were self-contained and did not depend on prior knowledge or knowledge of other texts. The following are key administration features of the Bangla Language tests:

Table **7.** General Features of the Bangla Language Tests

Feature	Detail
Grades	3 and 5
Number of test sessions	N = 1, November 14, 2015
Test time	60 minutes + 15 minutes (excluding preliminary organization time)
Number of test forms (booklets)	2 per grade
Number of texts	5-7 per form
Length of texts	Grade 3: maximum 150 words Grade 5: maximum 200 words
Balance of texts by text type (both grades)	Imaginative texts: 2-3 of varying difficulty Information texts: 2-3 of varying difficulty Persuasive texts: 2-3 of varying difficulty
Number of scored items	Grade 3: 36 Grade 5: 40
Item types	Multiple choice: Grade 3: 31 items; Grade 5: 35 items Constructed response: Grade 3: 5 items; Grade 5: 5 items

Bangla Language Results and Reliability Estimations

Reliability measures for Bangla Language in both grades were estimated using Cronbach's alpha, a coefficient of scale reliability, and the Spearman-Brown Split half method. A reliability coefficient is an estimation of the internal consistency of test items. Internal consistency refers to the extent to which the items in the test are consistently measuring the same construct. As the alpha coefficient increases, the portion of a score that can be attributed to error will decrease; hence, higher values are desirable (generally above 0.80). A first analysis of the Bangla data for both grades revealed satisfactory reliability for both grade levels (see Table 8below).

Table 8: Bangla Language Reliability Coefficients

Test	Number of items	Cronbach-alpha	Spearman-Brown Split-half
Bangla Grade 3	36	0.89	0.91
Bangla Grade 5	40	0.85	0.87

Bangla Performance Level Standards

One meaningful way to report NSA scores is to present results in terms of percentages of students attaining specific performance levels. In order to report performance levels, it is essential to have clear performance standards for each subject and grade. Setting performance standards is a process of developing conceptual and operational definitions of knowledge and skills that students should acquire in order to be considered as achieving at a specific mastery level. Content standards are specifications of "what" students are expected to learn in a given subject and grade, whereas performance standards refer to "how much" students are expected to know and be able to do to be placed in a particular performance level.

The procedure for setting performance levels typically employs a focus group method to solicit the judgments of field experts (teachers, supervisors, curriculum experts). The current performance ("legacy band") levels originate from the NSA 2011 (ACER, 2012) and were intended to capture achievement of Grade 3 and 5 students across Bangla Language and Mathematics content from both Grade 3 and 5. The bands were developed based on the empirical analysis of assessment data and mapping item difficulties across the scale. The authors (ACER, 2012) indicate they were aligned with the Aims and Objectives of Primary Education as a guide. The 2011 report explains the following:

> "Using the *Aims and Objectives of Primary Education* as a guide, five bands of achievement have been identified from the assessment data. The bands are broad descriptions of skills summarized from the detail of all the questions used to test pupils at both grades. They provide a more generalized picture of development in each subject and are useful as a frame of reference for monitoring growth over the grades of schooling." (ACER, 2012).

Table 9.Performance Band Descriptors and Cut Scores for Bangla Language

BSS	BAND DESCRIPTORS FOR BANGLA LANGUAGE (from ACER, 2012) (with cut scores on Bangla Language scale)
138	
	Pupils working in Band 5 • Read a range of short, more challenging texts, including poems • Interpret figurative language • Identify literal and implied meaning • Connect ideas in different parts of a text
122	• Show detailed knowledge of the rules of punctuation
	Pupils working in Band 4 • Read a range of short texts with more complex ideas • Identify main ideas, literal meaning • Make inferences • Understand the sequence of events in imaginative texts • Identify text types based on format • Identify meanings of familiar words in new contexts
108	• Know how to punctuate direct speech
	Pupils working in Band 3 • Read short, simple texts of different types with some unfamiliar vocabulary • Make use of simple clues to make simple inferences and identify main ideas • Deduce simple word meanings
96	• Show knowledge of word formation
	Pupils working in Band 2 • Read short, simple, mostly imaginative texts
85	• Locate and interpret directly stated information

	• Identify correct word orders of simple sentences • Identify the meaning and correct spelling of high frequency words • Recognise correct use of some punctuation
< 85	**Pupils working in Band 1** • Read simple, highly familiar texts, such as signs that contain strong visual support to interpret and locate information

Note: BSS = Bangla Scale Score

The NSA 2011, 2013 and 2015 results are presented below in Figure3 in terms of the legacy bands and cut scores developed in 2011.[4] For convenience sake, Bands 3 and above are presented above the reference line, and Bands 1 and 2 below the reference line.

Figure3. Overall Bangla Results by Performance Bands(2011, 2013,and 2015)
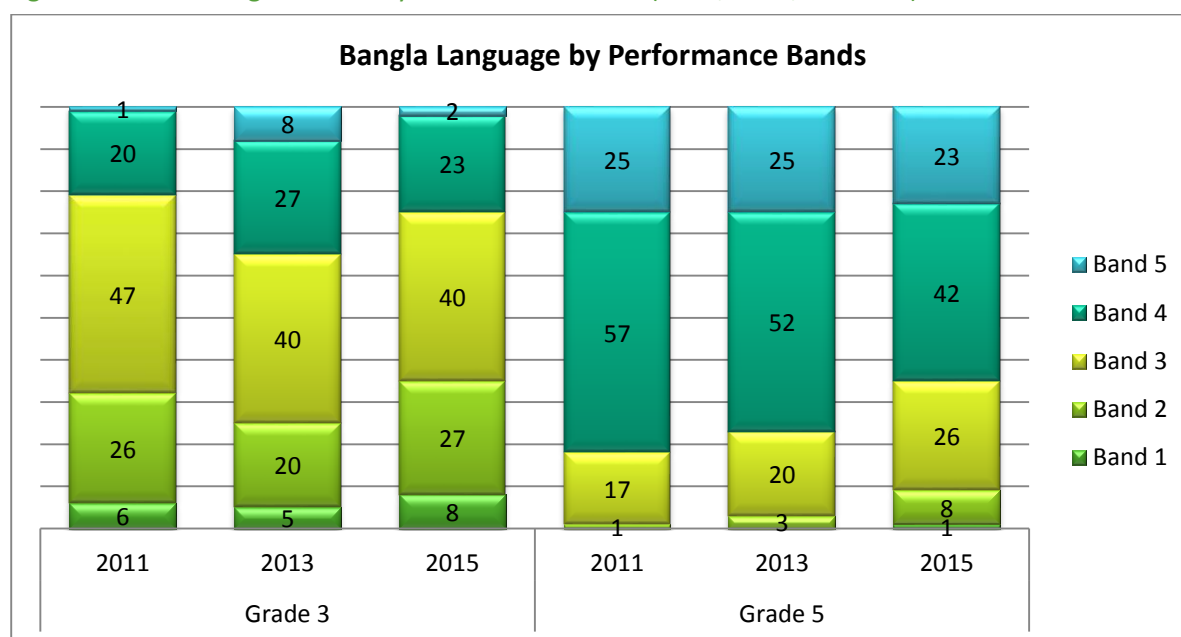
Comparability between the three assessments (2011, 2013, and 2015) in terms of performance bands is enabled through the application of IRT-based horizontal equating procedures. As can be seen in Figure3 above, there is growth in the proportion of students scoring in the three highest bands from Grade 3 to Grade 5; this holds true for the three NSA administrations in2011, 2013 and 2015. In the period between 2011 and 2015, the percentage of Grade 3 students scoring at Band 3 and above was 65% -75%, which suggests that over 2/3 of Grade 3 students perform at or above the grade level expectations set by the legacy bands. In the same period the percentage of Grade 5 students scoring at the same bands (3 and above) was 91% -99%, which shows a considerable learning growth from Grade 3 to Grade 5. However, considering that the percentage of Grade 5 students scoring at Band 5 (i.e., achieving at Grade 5 level expectation) is only 23% - 25%, the viability of interpretations of legacy bands in terms of grade level expectations is dubious since approximately 75% of Grade 5 students are only achieving at below Grade 5 level. It would not be reasonable to say that over 2/3 of students

---

[4] AIR has proposed a plan for conducting standard-setting including the setting up of grade specific performance scales by subject area and defining cut-points on each scale using data from the next iteration of the NSA.

in Grade 3 perform at or above grade expectations, while less than 1/4 of students in Grade 5 perform at grade expectations. This evidence strongly suggests that legacy bands may not be a suitable framework for the evaluation of student performance against grade specific curriculum expectations.

Grade 3 Bangla Performance by Band Level and Gender

As can be seen from the results of Bangla Language in Grade 3 disaggregated by gender(Figure4 above), approximately equal percentages of boys and girls are achieving at Bands 3 and higher in all three administration years. In NSA 2011 and 2013 the percentage of girls in Grade 3 achieving at Band 3 and above is just 1% more than the percentage of boys, whereas in NSA 2015 the percentage of girls achieving at Band 3 and higher is 4% better than for boys (66% and 70% respectively).

Figure5. Bangla Performance Bands by Gender

Grade 5 Bangla Performance by Band Level and Gender
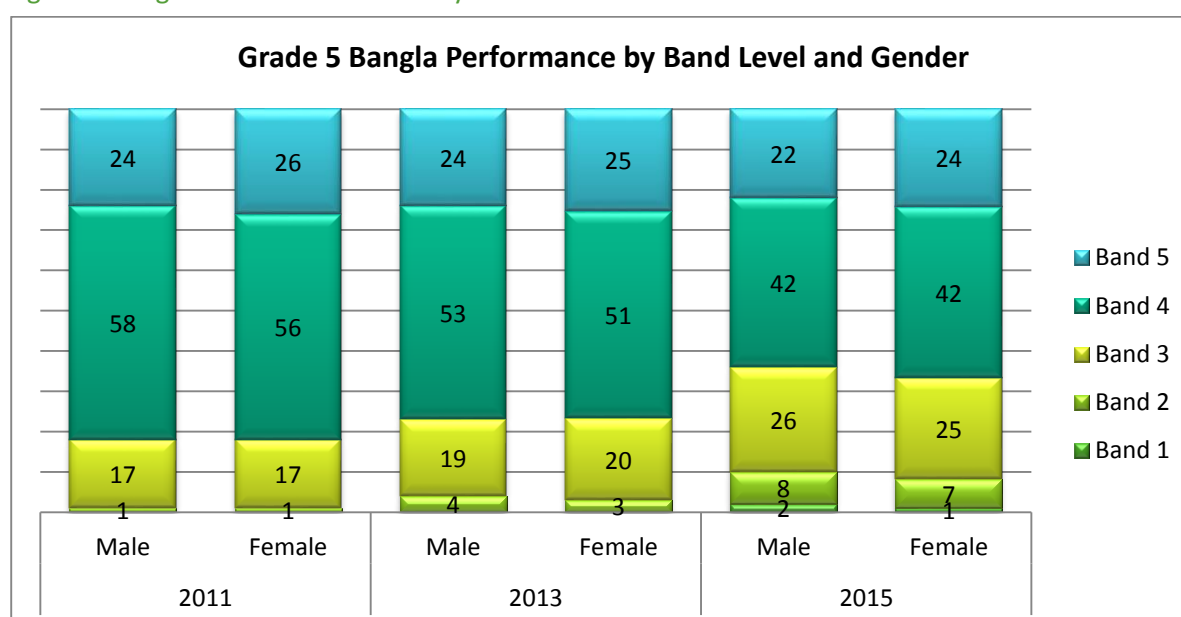
The results of Bangla Language in Grade 5 disaggregated by gender (see Figure5 above) are congruent with the general trend between grade levels in all three NSA administration years; the percentage of boys and girls achieving at Band levels 3, 4, and 5 rises significantly at Grade 5 compared to Grade 3. However, the difference between genders in Grade 5 remains negligible. These results speak in favor of gender equity in Bangladesh.

NSA Mean Scores

NSA 2015 mean Bangla Scale Scores (BSS) were 100.8 for Grade 3 and 114.1 for Grade 5. As Figure6 demonstrates, in the period between NSA 2011 and 2015 there was a small variation in NSA mean BSS for both grades. For Grade 3 the average BSS decreased from 104.2 in NSA 2013 to 100.8 in NSA 2015, but the effect size for this difference was small, at 0.29. For Grade 5 Bangla, the difference between average scale scores in NSA 2013 and 2015 was almost non-existent (115.2 vs. 114.1), with negligible effect size of 0.09.

Figure6. Overall Bangla Scale Scores(2011, 2013,and 2015)
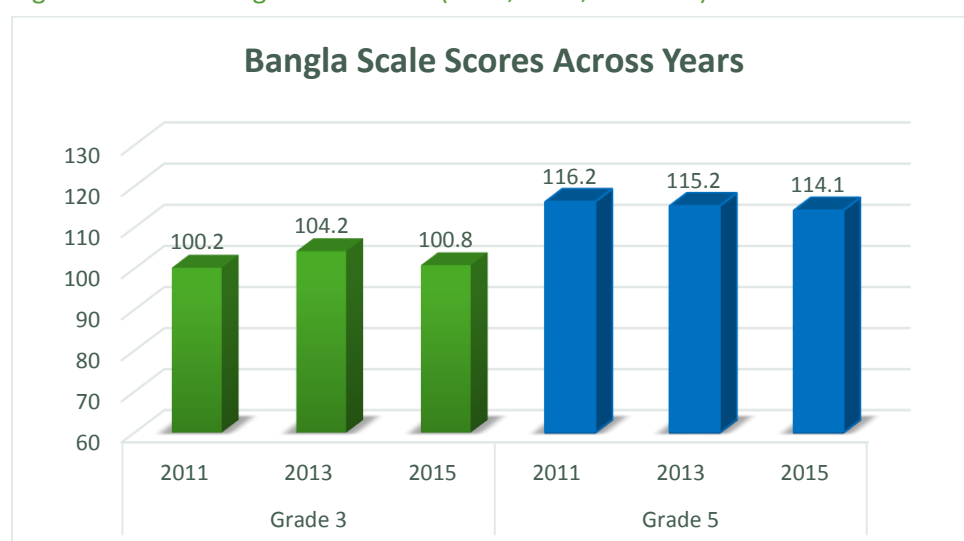
Figures7 and 8 present Bangla mean scores for Grades 3 and 5 by the two most prominent school types, Government Primary Schools (GPS) and Newly Nationalized Primary Schools (NNPS) for the

Figure7. Bangla Language Means by GPS and NNPS (Grade 3)
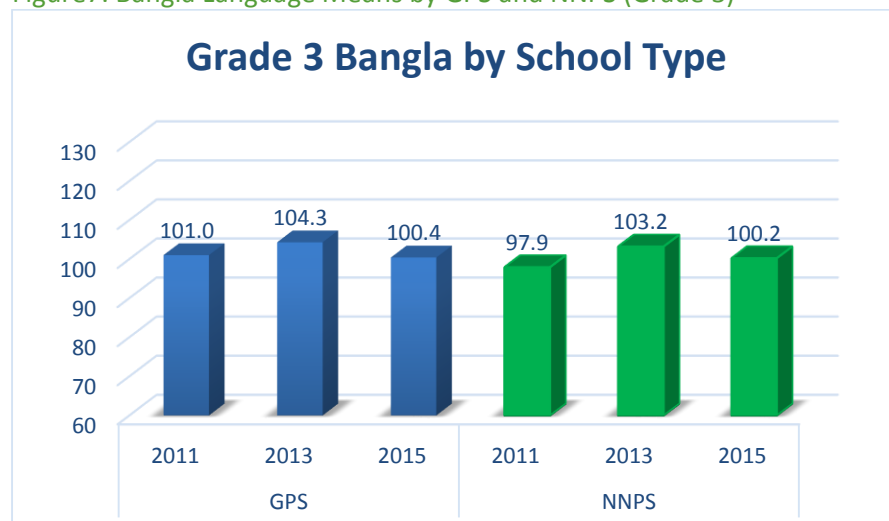
period of the three NSA administrations (2011, 2013, and 2015). Although the variation between administration years is relatively small, it should be noted that in both school types Bangla average scale scores dropped by 2-4 points from 2013 to 2015. We provide more detailed information regarding performance of all school types in a section toward the end of this chapter.

Figure8.Bangla Language Means by GPS and NNPS (Grade 5)

Grade 5 Bangla by School Type

GPS: 2011 = 117.1, 2013 = 116.3, 2015 = 114.6
NNPS: 2011 = 113.3, 2013 = 113.1, 2015 = 110.7
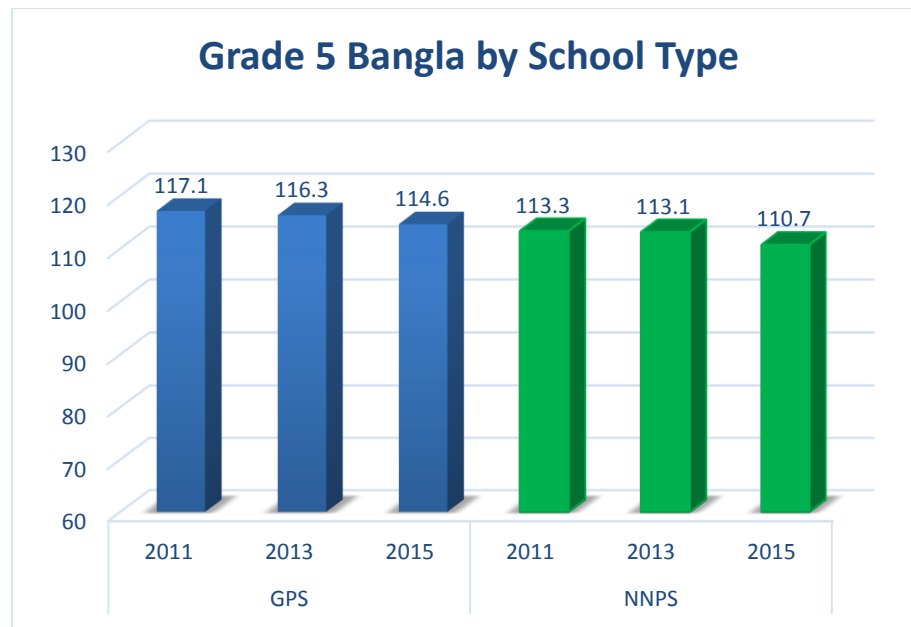
Bangla Language and Gender Mean Scores

In terms of scale scores, girls tended to outperform boys by around one point on Bangla Language at both grades in 2015 (see Figure 9 below). These differences were, however, of negligible effect sizes. Considering the previous two NSA administration years, the differences between boys and girls are even smaller, which all together strongly suggests that a relative gender parity exists in Bangladesh in terms of Bangla Language achievement and is consistent across grades and administration years and, as we indicated above, was also borne out in interpretation of scores by reference to performance levels.

Figure9.Bangla Language Means by Gender

Bangla Scale Scores by Gender

Grade 3 — 2011: Boys 100.0, Girls 100.5; 2013: Boys 103.7, Girls 104.7; 2015: Boys 100.1, Girls 101.5
Grade 5 — 2011: Boys 116.2, Girls 116.3; 2013: Boys 115.2, Girls 115.2; 2015: Boys 113.6, Girls 114.6

Boys   Girls
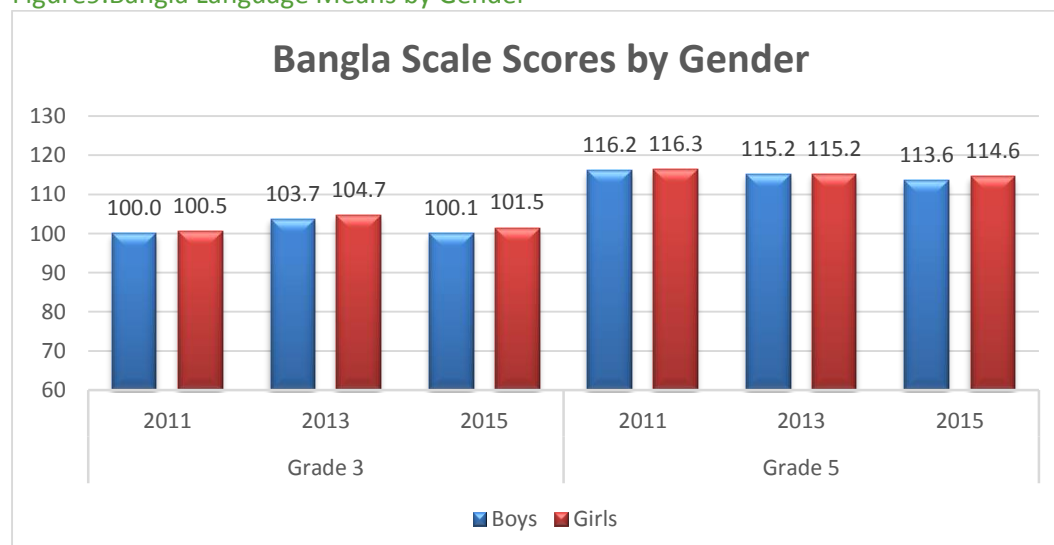
Bangla Language Results by Content Domain

The major domain of interest on the Bangla Language assessment was Reading Comprehension. Two additional domains–Grammar and Vocabulary –closely connect to and support the acquisition of Reading Comprehension.[5] The 2015 results for these content areas are presented below. As can be seen from the data in Figure10, for Grade 3 the Vocabulary tasks were the easiest while the Reading Comprehension tasks (presented here as % correct) were more challenging.

For Grade 3, girls scored numerically higher than boys in each of the three content domains and all these differences were statistically significant. However, recalling that the probability of a statistically significant result increases with large sample sizes, the data were also analyzed with an effect size measure. The results indicated that only in the case of Reading Comprehension was the effect of the difference notable, and very small at 0.1.

Figure10. Bangla Results by Content Domains (Grade 3)
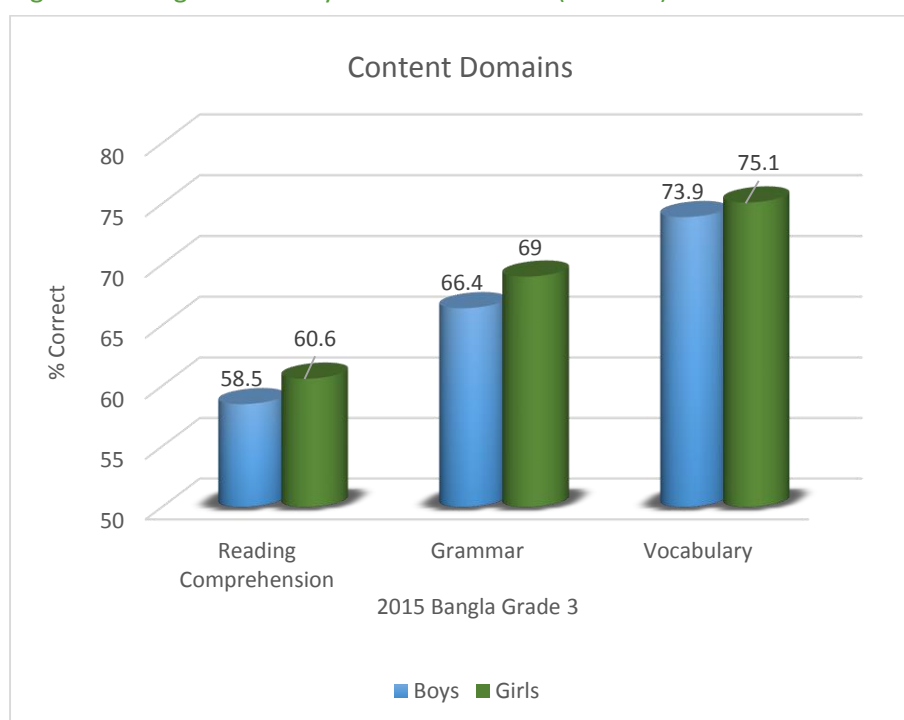
For Grade 5, while girls again scored at statistically significant higher levels than boys in all three content areas, the effect size for each comparison was negligible in all cases. Note that the overall trend is different for Grade 5 than for Grade 3. Students in Grade 5 scored higher on the Reading Comprehension section than on Grammar.

---

[5]The 2015 NSA did not measure oral skills (listening and speaking) because of the time and costs associated with this type of assessment. In 2015 it was also agreed that writing would not be measured given the limited amount of testing time available.

Figure1. Bangla Results by Content Domains (Grade 5)

Content Domains

2015 Bangla Grade 5

Boys ■ Girls
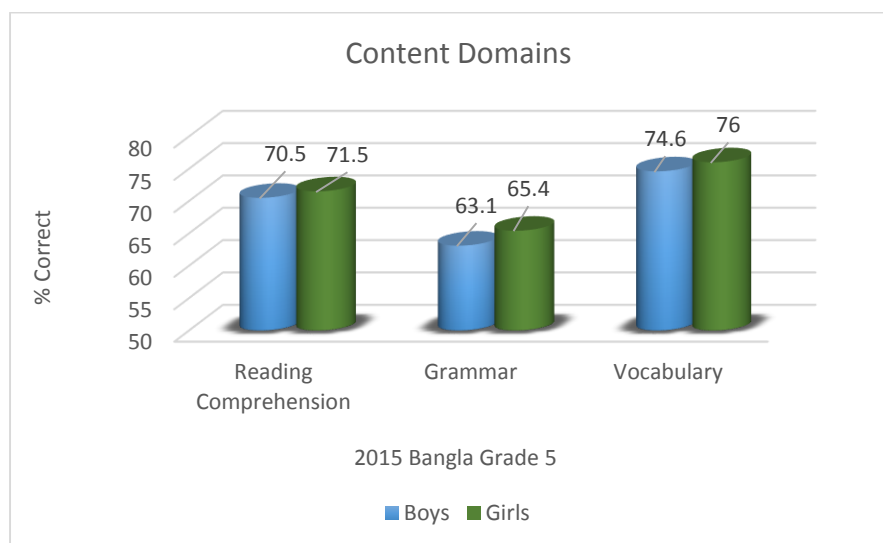
Bangla Results by Cognitive Processing Level

In the Bangla test design, items were arranged within the test in a logical order, presenting to students a cognitive flow logically related to the curriculum. Regarding difficulty, items were sequenced within the test from easiest to more difficult, with a peak of difficulty somewhere around the middle of the second half of the test. Items were classified into 1 of 4 categories defined by the cognitive level measured by the item and in the following approximate proportions:

Table 10. Cognitive Processing Levels for Bangla Language Tests

Cognitive processing level	% weight
Knowledge	30%
Understanding	40%
Application	20%
Higher Order Thinking Skills	10%

The methodology used for assigning cognitive processing level to the design of a test item was Bloom's taxonomy (see Appendix 2). For the purposes of item development on the NSA 2015, the first 3 levels of Bloom's taxonomy were preserved: 1. Knowledge (giving rise to items that measure recall or location of information); 2. Comprehension (giving rise to items that measure understanding of concepts); 3. Application (giving rise to items that measured the use of knowledge and concepts for the resolution of problems). The fourth level used on the NSA 2015 was a combination of the top three levels of Bloom's taxonomy, combining Analysis, Synthesis, and Evaluation (and other related skills) into a single level defined as Higher Order Thinking Skills.

The overall Grade 3 trend is clear in the data presented in Figure12 below. Students answered larger proportions of Knowledge and Understanding questions correctly than Application and Higher Order Thinking questions. It should be noted however, that the items measuring Knowledge and Understanding comprised a full 70% of the total number of items. Nonetheless, the pattern is worth noting and it is the same for both Grades 3 and 5.

Figure12.Bangla Cognitive Processing Levels (Grade 3)

Figure12.Bangla Cognitive Processing Levels (Grade 3)
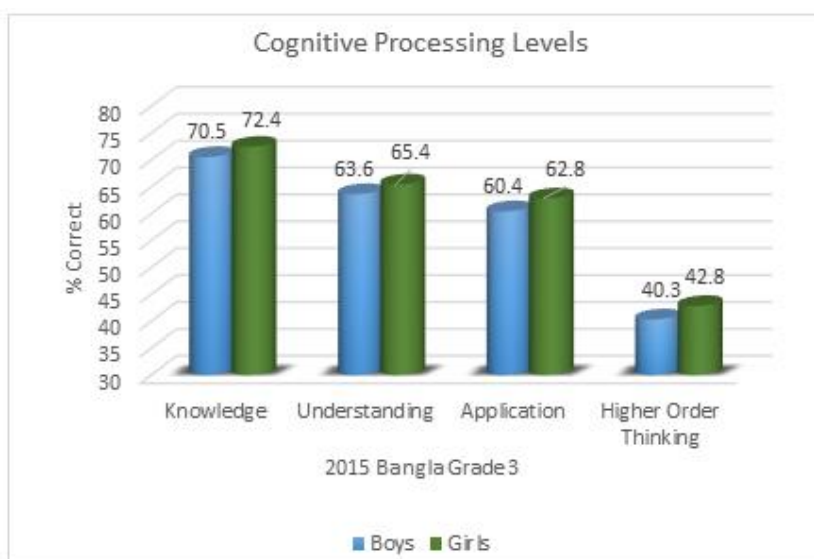
For Grades 3 and 5, girls again scored at statistically significant higher levels in all four of the cognitive processing categories as can be seen in Figures12 and 13. As in the content domain scores, however, effect sizes were negligible for three out of four of the processing levels. For Grade 3, girls scored on Higher Order Thinking with a very small effect size of around 0.10, and at 0.11 for Grade 5.

Figure2.Bangla Cognitive Processing Levels(Grade 5)

Figure2.Bangla Cognitive Processing Levels(Grade 5)
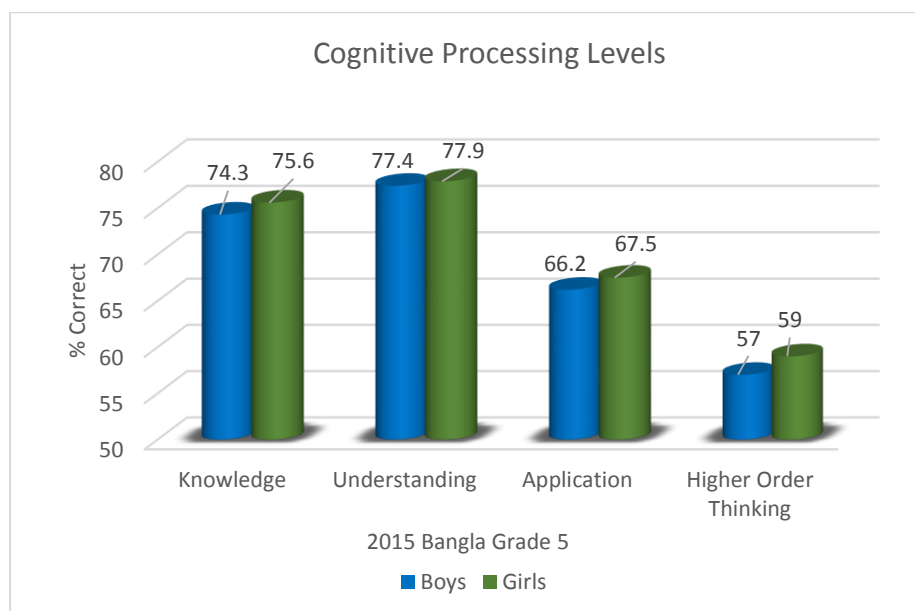
Bangla Results by Item Type

The reading test contained two types of items, selected response or multiple choice (31 items) and constructed response (5 items) for Grade 3, and 35 and 5 items for Grade 5, respectively. The constructed response items were used when the skill, knowledge or understanding – such as providing

an explanation or evidence to support an answer – could not be assessed in selected response format. Constructed response items were also used when there were insufficient plausible

distractors for a text. During the item development phase, each constructed response item was constructed along with its scoring rubric, and both elements were pilot tested. As can be seen in Figure 14above, students answered selected response items correctly in greater percentages than the constructed response items. However, by Grade 5 students are performing better in constructed response than they were at Grade 3. Girls scored higher than boys in this regard in both grades, though the effect size for selected response difference was negligible in both grades.

Bangla Scores by School Type

Table 11 below presents Bangla mean scores, standard deviations, and numbers of students sampled for Grade 3 by school type, listed from highest to lowest scoring school types. The orange shaded area represents mean scores that were above the national average of 100.2 in 2015. In both 2013 and 2015, Kindergarten Schools had the highest mean scores, a full 9.8 points higher than the Reaching Out-of-School Children (ROSC) schools, the lowest performing category of school type in 2015. Schools with High Schools Attached (HSA)scored in the top three in both 2013 and 2015. Note that scores for both Government Primary Schools (GPS) and Newly Nationalized Primary Schools (NNPS) both fell by 4-5 points from 2013 to 2015. However, together these two school types contain about 77% of the sampled population. There were other important changes across the comparison years. For example, the ROSC schools dropped considerably in the overall rank between the two years.

Table 11. Bangla Achievement by School Type (Grade 3)

2015				2013				2011			
School Type	Mean BSS	Std. Dev	N	School Type	Mean BSS	Std. Dev	N	School Type	Mean BSS	Std. Dev	N
KG	106.8	10.4	1,723	KG	107.1	10.5	1,485	GPS	101.0	9.8	13,348
HSA	103.4	11.8	632	ROSC	105.5	12.7	1,082	RNGPS	97.9	9.4	4,267
BRAC	102.2	10.1	588	HSA	105.2	9.8	869				
GPS	100.4	12.2	13,531	GPS	104.3	12.5	13,322				
NNPS	100.2	11.7	4,042	Madrasah	103.5	12.2	1,078				
Madrasah	99.8	13.3	1,018	RNGPS	103.2	11.5	4,619				
ROSC	97.3	12.0	1,355	BRAC	98.7	7.8	414				
TOTAL			22,889	TOTAL			22,869	TOTAL			13,615

Key: KG = Kindergarten; HSA = High School Attached; BRAC = Bangladesh Rural Advancement Committee; GPS = Government Primary School; ROSC = Reaching Out-of-School Children; ; NNPS = Newly Nationalized Primary Schools

Table 12 illuminates how the mean score differences relate to each other in terms of whether or not group differences were statistically significant. Recall that for dichotomous mean score comparisons such as gender, an independent samples t-test was run in SPSS to test for statistical significance of the difference. For comparisons across several groups, such as in school type or division, an ANOVA was used to determine whether the mean score differences were statistically significant.[6] When significant, a Cohen's d effect size measure was applied to determine the level of significance of statistically significant differences. Recall the interpretation of Cohen's d = 0.20 (small effect), 0.50 (moderate effect), and 0.80 (large effect).

The way to interpret the data in Table11 (and similar tables that follow) is that scores that fall under the same homogeneous group (1, 2, 3, or 4 in the table below) do not have mean score differences that are statistically significant. For example, the differences in mean scores between all three school types in group 2 (Madrasah, NNPS, and GPS) were not statistically significant. The mean score difference between ROSC and KG was statistically significant and the large effect size of 0.87 indicates a practical difference. By comparison, the difference between High School Attached and KG was also statistically significant but the effect size was more moderate to small at 0.41.

Table 12. 2015 Score Differences Grouped by Statistical Significance (Bangla Grade 3)

School Type	Student N	Homogeneous Groups			
		1	2	3	4
ROSC	1,355	97.3			
Madrasah	1,018		99.8		
NNPS	4,042		100.2		
GPS	13,531		100.4		
BRAC	588			102.2	
HS Attached	632			103.4	
Kindergarten	1,723				106.8
TOTAL	**22,889**				

---

[6] One note of caution is that comparing means with statistical tests is not 100% accurate when the difference in sample sizes is large (as is sometimes the case with the NSA data).

Bangla Grade 5 mean scores are presented by school type in Table 13 below. The table contains the same basic data for Bangla Grade 5 as presented above in Table 11 for Bangla Grade 3. The orange shaded area represents mean scores that were above the Bangla Grade 5 national average of 114.1 in 2015. KG schools and HSA were again consistently high scoring school types in 2013 and 2015. The difference between top scoring and low scoring was over ten points in 2015. Madrasahs and ROSC were the lowest scoring in both assessment years.

Table13. Bangla Scale Scores by School Type (Grade 5)

2015				2013				2011			
School Type	Mean BSS	Std. Dev	N	School Type	Mean BSS	Std. Dev	N	School Type	Mean BSS	Std. Dev	N
KG	121.6	16.7	1,491	KG	118.2	10.2	1187	GPS	117.1	8.6	10,695
HSA	118.3	16.7	560	GPS	116.3	11.1	10,633	RNGPS	113.3	8.2	3,159
BRAC	115.5	13.8	1,060	HSA	114.1	11.8	710				
GPS	114.6	13.8	11,526	RNGPS	113.1	10.3	3,419				
NNPS	110.7	13.1	3,359	BRAC	112.4	8.2	994				
ROSC	108.1	13.8	447	Madrasah	110.4	12.0	935				
Madrasah	108.0	14.7	945	ROSC	--	--	--				
TOTAL			19,388	TOTAL			17,878	TOTAL			13,854
Key: KG = Kindergarten; HSA = High School Attached; BRAC = Bangladesh Rural Advancement Committee; GPS = Government Primary School; ROSC = Reaching Out-of-School Children; NNPS = Newly Nationalized Primary Schools											

As above for Grade 3, Table 14 presents the school type data for Grade 5 by homogeneous groups according to statistical significance levels. Note there is more differentiation and less bunching by groups for Grade 5 than for Grade 3. The top scoring Kindergarten mean score is again statistically significant from all other scores. The effect size of the difference between low scoring Madrasahs and KG schools is quite large at 0.85, almost one full standard deviation in mean score average.

The mean score differences between Madrasahs and ROSC and between GPS and BRAC were not statistically significant. Note that the effect size of the difference between BRAC (115.5, group 3) and HSA (118.3, group 4) was small at 0.19. This underscores the point that the demarcation of statistical significance by groups is not always demonstrative of large practical differences between distinct groups: the meaning of the boundaries between groups needs to be interpreted carefully. The effect size for the difference between NNPS (110.7, group 2) and BRAC (115.5, group 3) is small to moderate at 0.36. While the effect size values for each possible combination of pairs is not presented in this report, the DPE has this data and it can be made available for further study.

Table 14. 2015 Score Differences Grouped by Statistical Significance (Bangla Grade 5)

School type	Student N	Homogeneous Groups				
		1	2	3	4	5
Madrasah	945	108.0				
ROSC	447	108.1				
NNPS	3,359		110.7			
GPS	11,526			114.6		
BRAC	1,060			115.5		
HS Attached	560				118.3	
Kindergarten	1,491					121.6
TOTAL	19,388					

Bangla Language Results by Division

Total mean scores for Bangla Grade 3 by geographic division are presented below in Table 15, ordered from highest to lowest scores for both 2013 and 2015. The orange shaded area represents mean scores that were above the national average in 2015. We see that the Rajshahi division mean score was the highest in 2015 and second highest in 2013. The lowest scoring division, Sylhet, was 9.4 mean points below Rajshahi in 2015. Rajshahi and Rangpur were both among the top four divisions for both assessment years but only Sylhet was in the bottom three in both years. Note that Dhaka's rank order moved from seventh (2013) to second (2015).

Table 15. Bangla Grade 3 Scores by Geographic Division (2011, 2013, and 2015)

2015				2013				2011			
Division	Mean BSS	Std. Dev	N	Division	Mean BSS	Std. Dev	N	Division	Mean BSS	Std. Dev	N
Rajshahi	104.0	11.8	2,901	Barisal	108.5	11.9	1,278	Dhaka	101.7	9.3	5,186
Dhaka	101.6	11.6	6,940	Rajshahi	106.9	12.8	2,782	Chittagong	101.2	10.2	3,715
Khulna	101.3	12.9	2,336	Chittagong	105.7	12.1	4,962	Khulna	99.6	9.6	1,901
Rangpur	100.8	10.9	2,845	Rangpur	105.5	10.7	2,606	Barisal	99.4	9.1	1,259
Chittagong	100.1	12.3	4,474	Khulna	103.7	11.2	2,430	Rajshahi	99.0	9.8	2,186
Barisal	99.6	11.7	1,425	Dhaka	101.9	11.7	6,883	Rangpur	98.7	10.0	2,028
Sylhet	94.7	12.6	1,968	Sylhet	100.9	12.7	1,928	Sylhet	97.9	10.5	1,340
TOTAL			22,889	TOTAL			22,869	TOTAL			17,615

Statistical testing was employed to compare the mean scores across geographic divisions (Table 16 below). The differences in mean scores for all three divisions in group 4 (Rangpur, Khulna, and Dhaka divisions) were not statistically significant. The mean score difference between Barisal and Chittagong divisions was also not statistically significant. Note that the highest mean score (Rajshahi) and lowest mean score (Sylhet) divisions scored at statistically significant levels from all other divisions. Also note that the difference between Sylhet (group 1) and Barisal (group 2) has a moderate/small effect size at 0.40, while the difference between Chittagong (group 2) and Rangpur (group 3) had a negligible effect size at 0.07. While Rajshahi scored 2.4 points higher than Dhaka, the effect size of this statistically significant difference is very small at 0.20.

Table 16. 2015 Score Differences Grouped by Statistical Significance (Bangla Grade 3)

Division	Student N	Homogeneous Groups				
		1	2	3	4	5
Sylhet	1,968	94.7				
Barisal	1,425		99.6			
Chittagong	4,474		100.1	100.1		
Rangpur	2,845			100.8	100.8	
Khulna	2,336				101.3	
Dhaka	6,940				101.6	
Rajshahi	2,901					104.0
TOTAL	22,889					

It is interesting to observe how differences among divisions in Bangla Language performance increase across the 3 administrations of the NSA. For Bangla Grade 5, the range of differences among divisions

in 2011 was small at less than 2 points; however, it increased to 7points in 2013, and to over 10 points in 2015. Rajshahi was again the highest scoring division in 2015, while Sylhet was again the lowest scoring division, with a 10.3 mean score difference between them. The relationship was similar in 2013 but with Barisal attaining the highest mean score in 2013. Dhaka was relatively higher in rank order in 2015 and 2011 than it was in 2013. The orange shaded area represents mean scores that were above the national average in 2015.

Table 17. Bangla Grade 5 Scores by Division (2011, 2013, and 2015)

2015				2013				2011			
Division	Mean BSS	Std. Dev	N	Division	Mean BSS	Std. Dev	N	Division	Mean BSS	Std. Dev	N
Rajshahi	117.6	14.8	2,530	Barisal	118.2	10.8	1,115	Chittagong	116.8	8.9	2,977
Dhaka	116.1	15.4	5,727	Rajshahi	117.6	10.8	2,171	Dhaka	116.5	8.7	4,059
Khulna	115.9	13.8	1,977	Rangpur	116.4	11.0	2,054	Rangpur	116.4	8.2	1,564
Barisal	112.6	15.1	1,254	Chittagong	115.3	11.5	3,919	Khulna	116.1	9.0	1,536
Rangpur	112.5	13.8	2,138	Dhaka	114.7	10.6	5,145	Barisal	115.9	8.4	937
Chittagong	112.2	12.5	4,177	Khulna	113.9	9.2	2,038	Rajshahi	115.5	8.3	1,739
Sylhet	107.3	13.4	1,603	Sylhet	111.2	12.2	1,386	Sylhet	115.0	8.9	1,041
TOTAL			19,406	TOTAL			17,828	TOTAL			13,853

In terms of statistical significance (Table 18 below), the effect size of the difference between the top (Rajshahi) and bottom (Sylhet) divisions was large at 0.72. While the top and bottom scorers had means that were statistically significant from all others, there was much clustering in groups 2 and 3 as can be seen below. Further, the statistically significant difference between Chittagong (group 2) and Khulna (group 3) was small at 0.29. And, the effect size of the difference between the top two scorers Rajshahi and Dhaka was also negligible at 0.10.

Table 18. 2015 Score Differences Grouped by Statistical Significance (Bangla Grade 5)

Division	N	Homogeneous Groups			
		1	2	3	4
Sylhet	1,603	107.3			
Chittagong	4,177		112.2		
Rangpur	2,138		112.5		
Barisal	1,254		112.6		
Khulna	1,977			115.9	
Dhaka	5,727			116.1	
Rajshahi	2,530				117.6
TOTAL	19,406				

CHAPTER 3. THE 2015 MATHEMATICS ASSESSMENT

The objective of Mathematics instruction in the targeted grades was to acquaint learners with arithmetical logic, methods, and skills so that students become imaginative, curious, creative and intellectual learners; and to increase students' abilities to apply such knowledge and skills for problem solving in real world contexts and activities. Appendices 3 and 4provide a complete description of the key Mathematics skills that students were expected to develop in Grades 3 and 5.

The framework for Mathematics was written with a consistent focus on collecting information on student performance in four key areas of mathematical content:

- Number Properties and Operations (including computation and understanding of number concepts);
- Measurement and Units of Measurement (scale of measurement; principles of measurement; metric system of measurement; application of processes and concepts of area; differentiate between and carry out operations);
- Shape and Space (understand concepts and use instruments);
- Data (graphical representations, relationships, and central tendency of data).

Table 19.General Features of the Mathematics Tests

Key areas	Grade 3	Grade 5
Number Properties and Operations	50%	40%
Measurement and Units of Measurement	25%	30%
Shape and Space	25%	20%
Feature	**Detail**	
Number of test sessions	1	
Test time	60 minutes + 15 minutes (excluding preliminary organization time)	
Number of scored items	Grade 3: 35 Grade 5: 40	
Item types	Selected Response: Grade 3: 30 items; Grade 5: 35 items Constructed response: Grade 3: 5 items; Grade 5: 5 items	

Mathematics Reliability Estimations

Reliability coefficients for Mathematics in both grades were estimated using Cronbach's alpha, a coefficient of scale reliability, and the Spearman-Brown, Split half method. The reliability coefficient is an estimation of the internal consistency of the items. Internal consistency refers to the extent to which the items in the test are consistently measuring the same construct. As the alpha coefficient increases, the portion of a score that can be attributed to error will decrease: Hence higher values are desirable (generally above 0.80). A first analysis of the Mathematics data for both grades revealed satisfactory reliability for both grade levels (see Table 20 below).

Table 20.Reliability Coefficients for Mathematics Assessments

Test	Number of items	Cronbach-alpha	Spearman-Brown Split-half
Mathematics Grade 3	35	0.89	0.90
Mathematics Grade 5	40	0.89	0.90

Mathematics Results by Performance Levels

NSA Mathematics results for Grades 3 and 5 are presented below. Comparability across the three NSA years (2011, 2013, and 2015) is enabled through the application of IRT-based horizontal equating procedures. Similarly, using vertical equating procedures, the Grade 3 and Grade 5 NSA scores were placed on the same vertical scale, so that comparison across grade levels is possible. Comparisons are provided in terms of performance bands and Mathematics Scale Scores (MSS) at the national level, and disaggregated by gender, division (region), and school type.

As with the Bangla results, results are first presented in terms of percentages of students attaining specific performance levels, or bands (see the section on performance levels for Bangla Language for more on how the bands were created). Finally, analyses of results by content domain, cognitive processing level, and item type are also presented.

Table 21. Performance Band Descriptors and Cut Scores for Mathematics

BAND DESCRIPTORS FOR MATHEMATICS (with cut scores on Mathematics scale)	
MSS	
145	
124	**Pupils working in Band 5** • apply strategies to simplify numerical expressions and solve word problems on percentages and unitary method • apply geometric properties and relations in solving simple problems on angles • calculate the perimeter of simple geometric shapes in real context
113	**Pupils working in Band 4** • apply strategies to solve word problems including money transactions using skills of addition, subtraction, multiplication and division of whole numbers, add/ subtract and simplify decimals, find the Highest Common Factor of small numbers, identify and represent fractions, multiply and divide whole numbers by fractions, solve word problems related to addition and subtraction • set up a mathematical expression(equation) for a given situation, find the value of an unknown in a given simple mathematical expression • convert different units of length measure (cm/mm to cm/m, kg to gm) and area measure (square metres to hectares) , calculate area of a triangle from given dimensions • identify the distinguishing properties of 2D objects • calculate averages from data presented pictorially
101	**Pupils working in Band 3** • add and subtract 6-digit numbers (negative numbers excluded) identify the remainder on division by 100, find Lowest Common Multiple of given numbers, uses addition/subtraction and multiplication to solve 2 stage word problems, can convert fractions to mixed fractions, percentages and decimals, add, subtract and multiply like fractions including decimal fractions by whole numbers, identify equivalent mathematical processes form simplification, find the unit price of an item using unitary method • calculate elapsed time and read a 24hour clock format • measure the volume of a liquid shown in a graduated cylinder and calculate the area of a rectangle

	• identify 3D shapes and classify triangles • use tally charts and frequency tables
90	**Pupils working in Band 2** • identify place value in numbers up to 4-digit numbers, orders 2-digit numbers, compare two numerical expressions • add and subtract numbers up to 4-digits (without carry over) divide a 3-digit number by a 1-digit number, use addition, subtraction and multiplication to solve two stage problems, recognise, order and find equivalent simple fractions • recognise and name currency in words and figures • read time an analogue clock to the nearest quarter hour, convert hours to days • identify appropriate unit of measurement, convert metres and centimetres to metres, calculate area of a rectangle
< 90	**Pupils working in Band 1** • identify, count and compare numbers up to 3-digits, add and subtract numbers up to 4-digits (without carry over), identify even and odd numbers • read date and day on a calendar • read simple graphs • recognise and draw simple 2D shapes and identify types of surfaces(plane surface)

The results of the three NSA years (2011, 2013, and 2015) expressed in terms of performance bands are presented in Figure15 below. For convenience sake, the portions of the bars representing the top three bands (3, 4, and 5) are placed above the reference line, and portions representing Bands 1 and 2 are shown below the reference line. Note that there is growth in the percentages of students scoring in the three top bands from Grade 3 to Grade 5; this holds true for all three NSA years although this is hardly a significant fact given that growth is expected to occur from one grade to the next. When looking across administration years, it can be seen that Mathematics scores in NSA 2015 are lower than in previous NSA years.

In the NSA 2013, 57% of Grade 3 students achieved Mathematics scores at Band Levels 3 and above (i.e., at Grade 3 level and above), while 89% scored at these band levels by Grade 5. However, in terms of percentages of scores at Band 5 (i.e., scoring at Grade 5 level), only 25% of Grade 5 students achieved Band 5 level in the NSA 2013. A similar pattern holds for NSA 2015 where 41% of Grade 3 students were at Bands 3 and above, and 81% of Grade 5 students were at the same Bands 3 and above; however, only 10% of Grade 5 students reached Band 5 level.

As mentioned in the section regarding Bangla performance bands, the viability of using legacy bands as an evaluation framework for the performance of students against grade specific standards and expectations is seriously challenged by the reasonableness check – it may not be likely that the percentage of students reaching grade level curriculum objectives is so different between Grade 3 and Grade 5 (41% and 10%, respectively). This evidence suggests a need for establishing a system of performance standards that will serve as a suitable framework for evaluating students' academic achievement in the context of grade specific curriculum objectives.

Figure15.MathematicsPerformance Levels by Bands(2011, 2013,and 2015)

Mathematics by Performance Bands
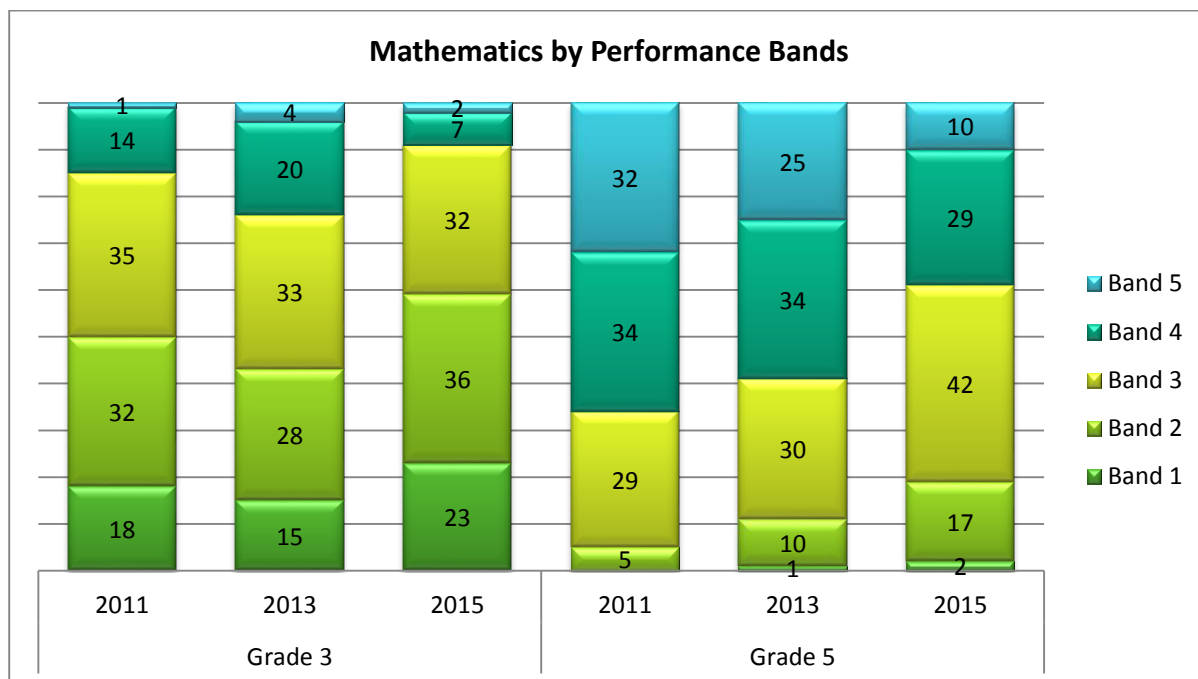
As can be seen from the data in Figures16 and 17 below, approximately equal proportions of boys and girls achieve at Bands 3 and higher in all three NSA years (2011, 2013, and 2015).

Figure16.MathematicsBand Levels by Gender (Grade 3)

Grade 3 Mathematics Performance by Band Level and Gender
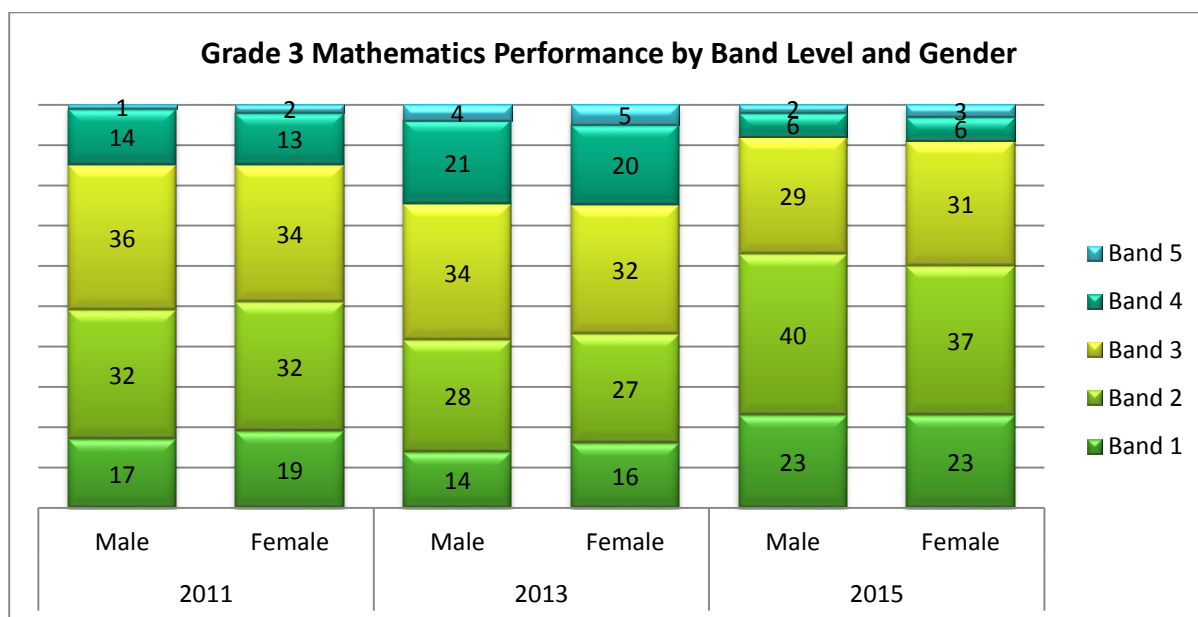
Figure17.Mathematics Band Levels by Gender (Grade 5)

Grade 5 Mathematics Performance by Band Level and Gender

Band 5	1	2	4	5	2	3
Band 4	14	13	21	20	6	2/6
Band 3	36	34	34	32	29	31
Band 2	32	32	28	27	40	37
Band 1	17	19	14	16	23	23
	Male	Female	Male	Female	Male	Female
	2011		2013		2015	
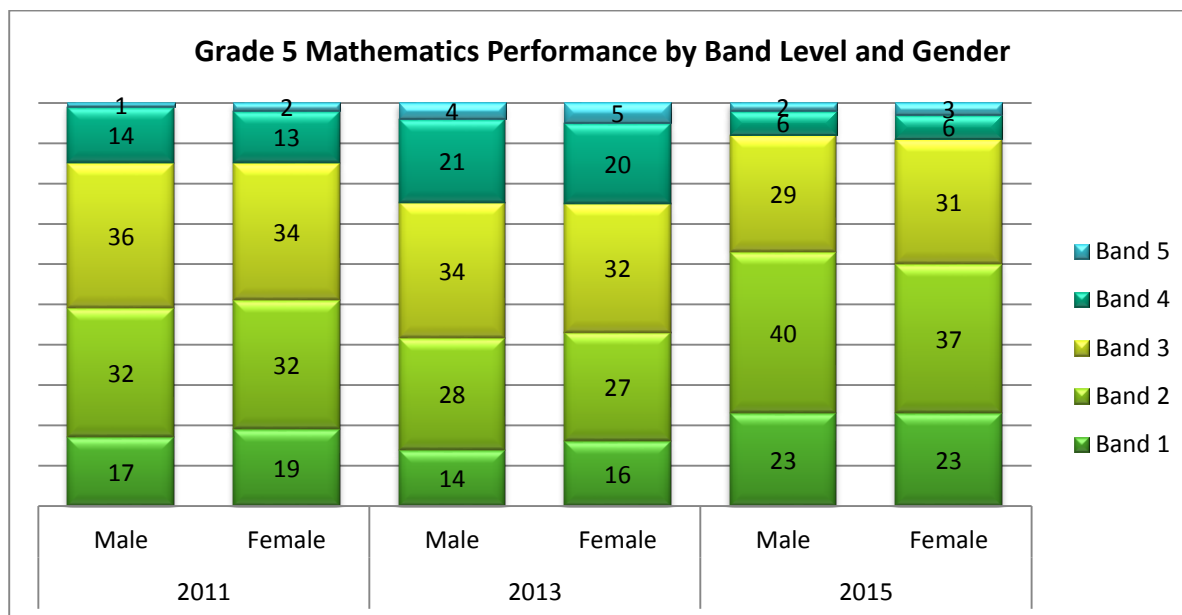
Mathematics Scale Scores

The overall mean scale scores in Mathematics across all three NSA years are presented in Figure18 below. Overall 2015 Mathematics mean scores for Grades 3 and 5 were clearly down from the 2013 NSA. The overall mean scores fell from 103.7 to 98.4 for Grade 3 and from 115.8 to 110.2 for Grade 5. These mean score differences were statistically significant with moderate effect sizes (0.44 for Grade 3 and 0.48 for Grade 5). Looking across all three NSA years it can be observed that Mathematics scale scores in Grade 5 consistently decreased, whereas in Grade 3 the variation is less systematic showing ups and downs.

Figure18. Overall Mathematics Mean Scores by Grade and Year (2011, 2013, and 2015)

Mathematics Scale Scores Across Years

	Grade 3			Grade 5		
Year	2011	2013	2015	2011	2013	2015
Score	100.8	103.7	98.4	118.6	115.8	110.2
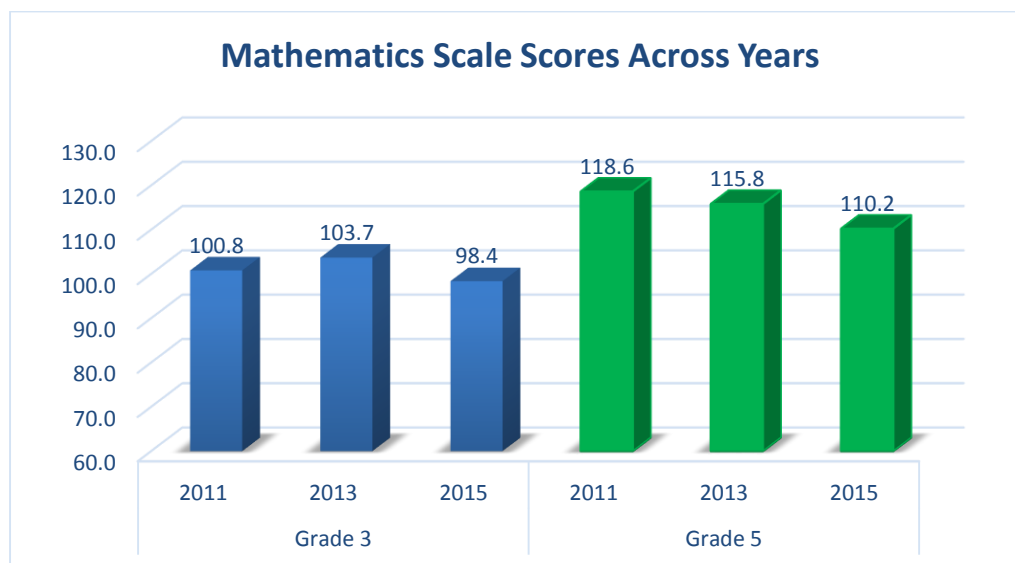
NSA Mathematics mean score averages across all three assessment years for the two most prominent school types, GPS and NNPS (RNGPS in 2011 and 2013), are shown in Figures19 and 20. Both GPS and NNPS mean scores were significantly down in 2015 from 2013 in both grade levels.

Figure19. Mathematics Means by Main School Types (Grade 3)

Grade 3 Mathematics by School Type

	GPS			NNPS		
	2011	2013	2015	2011	2013	2015
	101.3	104.1	98.0	99.1	102.3	99.2
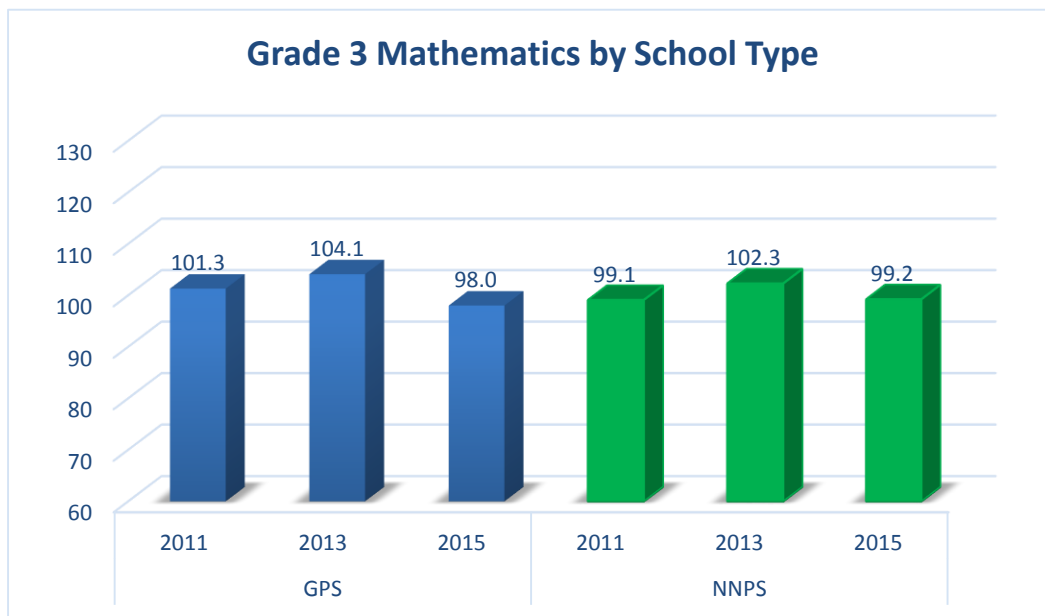
Figure20. Mathematics Means by Main School Types (Grade 5)

Grade 5 Mathematics by School Type

	GPS			NNPS		
	2011	2013	2015	2011	2013	2015
	119.2	117.2	111.0	116.5	113.9	108.9
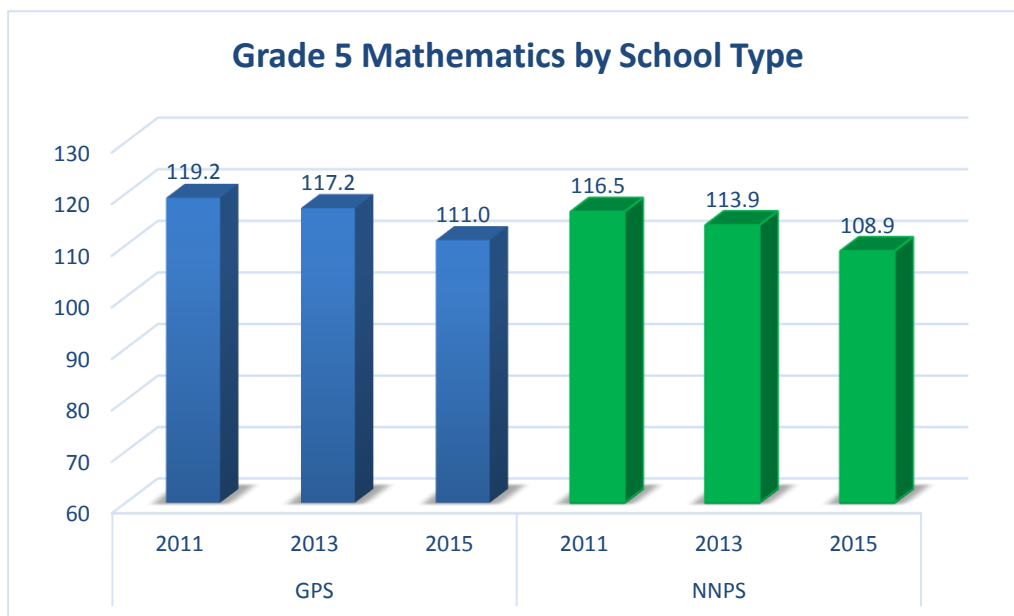
In Figure21, gender parity in Mathematics achievement is again evident across grade levels and assessment years.

Figure21.MathematicsMeans by Gender

Mathematics Scale Scores by Gender
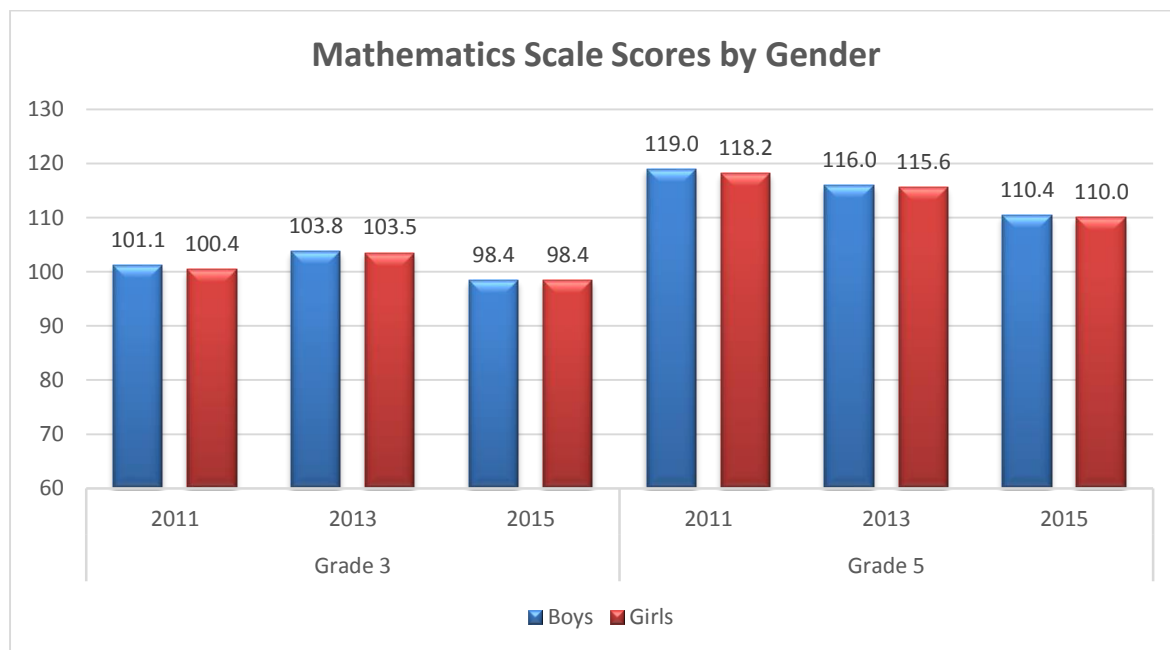
Mathematics Results by Content Domain

Unlike the Bangla assessments, girls did not score numerically higher across the Mathematics content domains. In Grade 3 they scored only slightly higher on Number and Operations and Shape and Space –but neither mean score difference was statistically significant. On Measurement and Units, boys scored at a statistically significant higher level but with a negligible effect size of 0.12. In Grade 5 the two content areas where there were statistically significant differences (Numbers and Operations, Measurement) the effect size levels were also negligible. In other words, there were no practical differences between boys and girls on these three content domains for Grade 5.

Figure22.Mathematics Achievement by Content Domain (Grade 3)
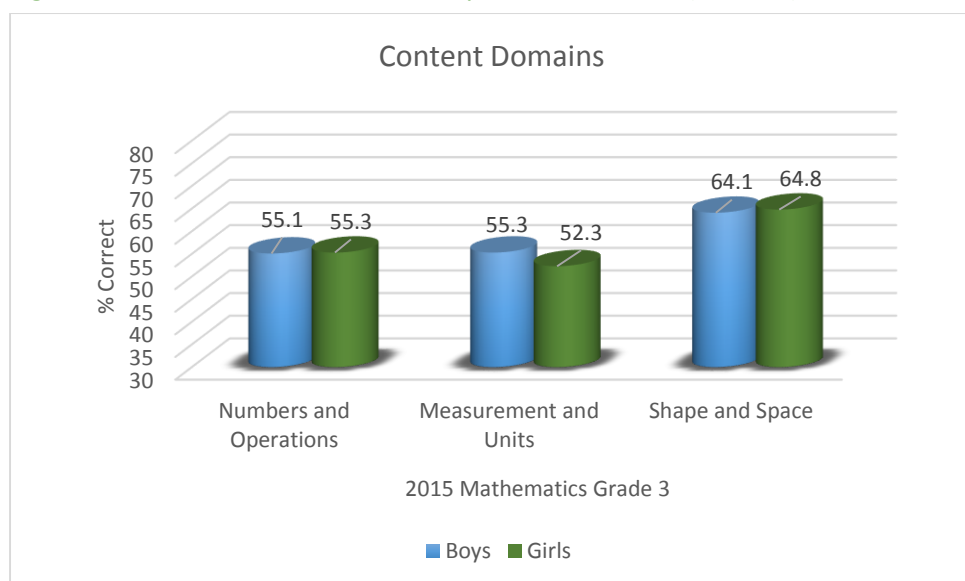
Content Domains

Figure23.Mathematics Achievement by Content Domain (Grade 5)

Content Domains

2015 Mathematics Grade 5

Boys ■ Girls
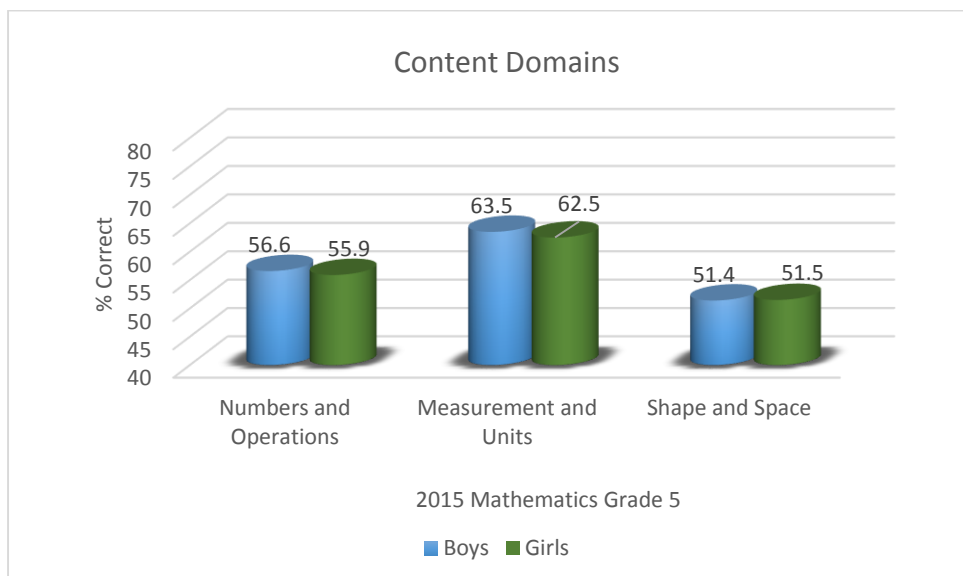
Mathematics Results by Cognitive Processing Level

Items were classified into one of 4 categories defined by the cognitive level measured by the item in the following approximate proportions:

Table 22.Cognitive Processing Levels for Mathematics Tests

Cognitive processing level	% weight
Knowledge	30%
Understanding	40%
Application	20%
Higher Order Thinking Skills	10%

Figure24. Cognitive Processing Levels (Grade 3)

Cognitive Processing Levels

2015 Mathematics Grade 3

Boys ■ Girls
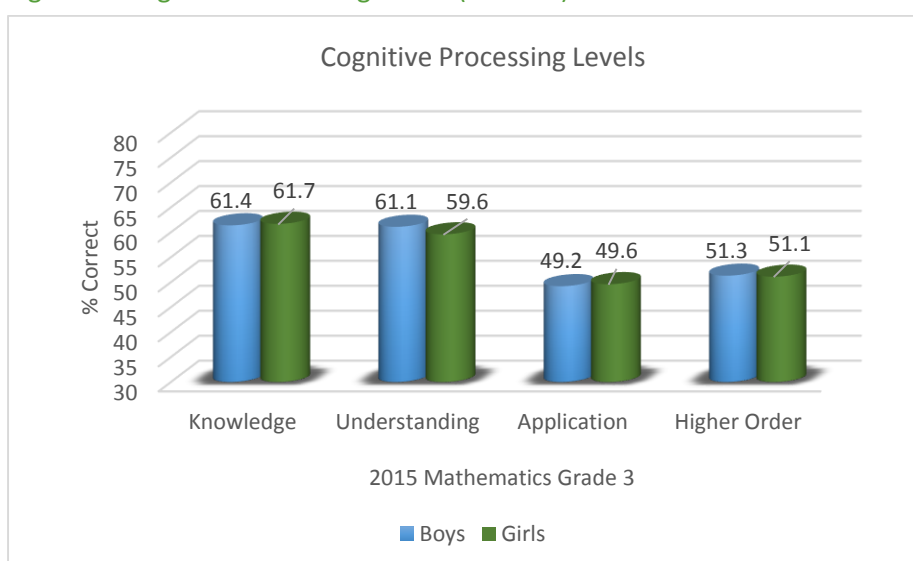
As with the Bangla assessments, a slight downward trend is evident for Mathematics as the cognitive demands increase. Note in Figures24 and 25 that students answered larger proportions of Knowledge and Understanding questions than those questions requiring Application and Higher Order Thinking skills. For Grade 3 there were no significant differences between boys and girls on three of the four levels. On Understanding, however, boys scored at a statistically significant higher level but the effect size for this difference was negligible.

Figure25.Cognitive Processing Levels (Grade 5)
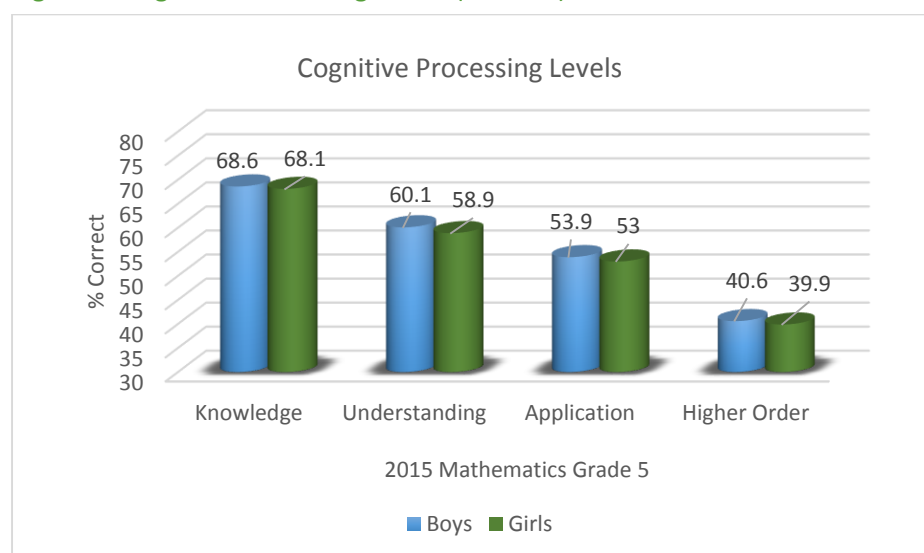
For Grade 5, there were no statistically significant differences by gender on the Knowledge and Higher Order Thinking processing levels. For the two levels where there were statistically significant differences in favor of boys (Understanding, Application), the effect size levels were negligible.

Mathematics Results by Item Type

The Mathematics test contained two types of items, selected response and constructed response. Constructed response items were distributed throughout the test with the following weights by grade:

Table 23. Mathematics Item Types and Weights by Grade

	Grade 3	Grade 5
Selected Response	30 items	35 items
Constructed Response	5 items	5 items

As on the Bangla assessments, students performed proportionately better on the selected response items than on the constructed response items. For Grade 3, there were no statistically significant differences by gender for performance by type of item. For Grade 5, boys were favored at statistically significant levels but effect sizes were negligible for both item types.

Figure26.Mathematics Item Types
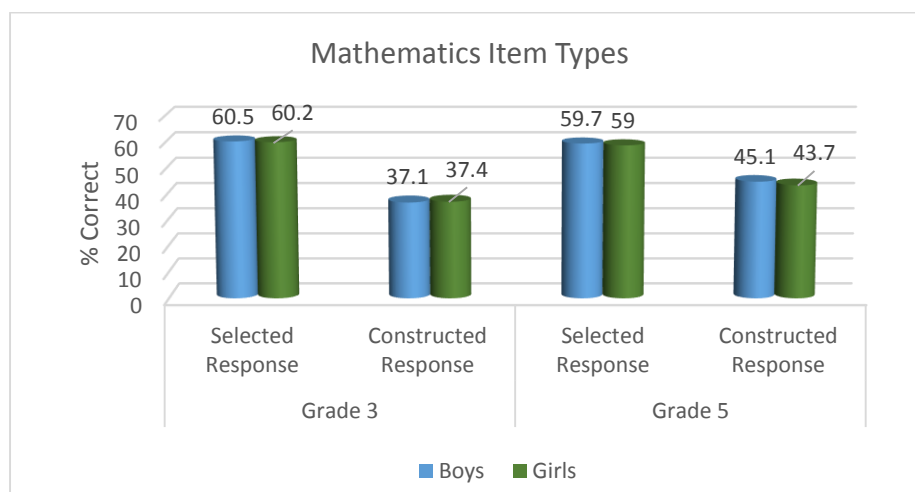
Table 24 presents mean scores for Grade 3 Mathematics by school type in all three NSA administration years, listed from highest to lowest scorers. The orange shaded area represents mean scores that were above the national mean score in 2015. Looking at Grade 3 by school type, we see that KG schools scored at a statistically significant higher level than all other school types in the group. The lowest scoring school mean in 2015 was ROSC, 8.1 points less than the top mean score achieved by KG schools. Interestingly, ROSC was the second highest scorer in 2013 but fell in 2015 with an almost ten point mean score decrease.

Table 24. Mathematics Results by School Type (Grade 3)

2015				2013				2011			
School Type	Mean MSS	Std. Dev	N	School Type	Mean MSS	Std. Dev	N	School Type	Mean MSS	Std. Dev	N
KG	103.3	11.6	1,729	KG	105.0	9.9	1,486	GPS	101.3	11.7	13,356
NNPS	99.2	11.2	4,051	ROSC	104.6	12.5	1,079	RNGPS	99.1	11.3	4,270
BRAC	98.4	9.6	596	Madrasah	104.2	14.5	1,078				
GPS	98.0	11.4	13,575	GPS	104.1	13.6	13,454				
HSA	97.9	9.7	632	HSAPS	103.4	11.4	891				
Madrasah	96.0	11.4	1,006	RNGPS	102.3	12.0	4,662				
ROSC	95.2	10.3	1,365	BRAC	97.5	8.8	414				
TOTAL			22,889	TOTAL			23,064	TOTAL			17,626
Key: KG = Kindergarten; HSA = High School Attached; BRAC = Bangladesh Rural Advancement Committee; GPS = Government Primary School; ROSC = Reaching Out-of-School Children; NNPS = Newly Nationalized Primary Schools											

Table 25 shows which Mathematics mean score differences by school type were statistically significant. As per the Bangla Language school types and divisions (presented in the previous section), for comparisons across groups an ANOVA was used to determine whether the mean score differences under study were statistically significant. Scores that fell under the same group number (groups 1, 2, or 3 in the table below) did not have mean score differences that were statistically significant.

The difference between the means of the top and bottom scoring school types (KG and ROSC) was statistically significant with a large effect size, 0.73. While the differences in scores between the

school types in different groups were statistically significant, note that the effect size for Madrasah and GPS was small at 0.18 and the effect size for NNPS and KG was small to moderate at 0.36.

Table 25.2015 Score Differences Grouped by Statistical Significance (Mathematics Grade 3)

School Type	Student N	Homogeneous Groups		
		1	2	3
ROSC	1,365	95.2		
Madrasah	1,006	96.0		
HS Attached	632		97.9	
GPS	13,575		98.0	
BRAC	596		98.4	
NNPS	4,051		99.2	
KG	1,729			103.3
TOTAL	22,954			

For Mathematics Grade 5, the same two school types (KG and GPS) were the top performers in both 2013 and 2015, though their relative positions were reversed (i.e., KG moved from first to second from 2013 to 2015, with GPS in the reverse order). In 2015, Madrasah schools were the lowest scoring school type, with a mean score almost ten points below the highest scoring school type. ROSC was at or near the bottom in both 2013 and 2015.

Table 26. Mathematics Results by School Type (Grade 5)

2015				2013				2011			
School Type	Mean MSS	Std. Dev	Student N	School Type	Mean MSS	Std. Dev	Student N	School Type	Mean MSS	Std. Dev	N
KG	113.8	11.1	1,490	GPS	117.2	13.1	10,620	GPS	119.2	11.3	10,667
GPS	111.0	11.2	11,513	KG	116.8	11.1	1,187	RNGPS	116.5	10.3	3,160
HSA	110.1	10.9	567	RNGPS	113.9	11.3	3,413				
NNPS	108.9	10.0	3,363	H/A PS	112.8	11.7	710				
BRAC	107.6	8.5	1,068	Madrasah	112.7	11.2	932				
ROSC	107.3	8.8	443	BRAC	110.2	7.8	944				
Madrasah	104.0	10.6	939	ROSC	-	-	-				
TOTAL			19,388	TOTAL			17,806	TOTAL			13,827
Key: KG = Kindergarten; HSA = High School Attached; BRAC = Bangladesh Rural Advancement Center; GPS = Government Primary School; ROSC = Reaching Out-of-School Children; NNPS = Newly Nationalized Primary Schools											

For Grade 5, there were six distinct groups with statistically significant differences across mean scores (Table 27 below). Note that several school types are clustered in more than one group depending on the relationship of the statistical significance. High scoring KG schools and low scoring Madrasah were in groups by themselves but other school types were located in various pair combinations. The effect size for the KG and Madrasah difference was large, at 0.90, almost a full 1 standard deviation mean score difference. The difference between ROSC and BRAC was not statistically significant, nor was the difference between BRAC and NNPS. However, the mean score difference between BRAC and HSA was statistically significant, though the effect size was small at 0.26. The effect size for the Madrasah and HSA mean score difference was moderate at 0.58.

Table 27.NSA 2015 Score Differences Grouped by Statistical Significance (Mathematics Grade 5)

School Type	Student N	Homogeneous Groups					
		1	2	3	4	5	6
Madrasah	939	104.0					
ROSC	443		107.3				
BRAC	1,068		107.6	107.6			
NNPS	3,363			108.9	108.9		
HS Attached	567				110.1	110.1	
GPS	11,513					111.0	
Kindergarten	1,490						113.8
TOTAL	19,383						

Mathematics Results by Geographic Division

The divisions that scored above the national average are highlighted in orange in Table 27. Looking at Mathematics Grade 3 by division, we see that Rajshahi scored the highest, at a statistically significant level above the rest of the group. Rajshahi was also the highest scorer in 2013. The lowest scoring division mean, Sylhet, was 9.3 points below Rajshahi in 2015, but it was second lowest in 2011. Rangpur was in the upper range for both years, while Dhaka moved up in 2015. Interestingly, Barisal dropped from second highest in 2013 to second lowest in 2015. However, before making inferences about the meaning of these rank scores, it is important to identify which mean score differences were statistically significant, as presented in Table 29.

Table 28.NSA Mathematics Achievement by Geographic Divisions (Grade 3)

2015				2013				2011			
Division	Mean MSS	Std. Dev	Student N	Division	Mean MSS	Std. Dev	Student N	Division	Mean MSS	Std. Dev	Student N
Rajshahi	101.3	11.3	2,906	Rajshahi	107.9	14.5	2,791	Dhaka	102.2	11.0	5,190
Rangpur	99.7	10.3	2,866	Barisal	106.4	12.8	1,458	Chittagong	102.2	12.0	3,719
Dhaka	98.9	11.5	6,919	Rangpur	105.9	11.5	2,607	Barisal	100.6	10.9	1,258
Khulna	98.8	11.3	2,336	Chittagong	105.5	12.6	4,968	Khulna	100.2	11.3	1,904
Chittagong	97.8	11.2	4,525	Khulna	102.7	11.8	2,432	Rangpur	100.1	12.1	2,033
Barisal	96.8	10.2	1,433	Dhaka	100.9	12.3	6,881	Rajshahi	98.5	11.7	2,181
Sylhet	92.4	11.6	1,969	Sylhet	98.4	13.2	1,927	Sylhet	96.9	11.6	1,340
TOTAL			22,954	TOTAL			23,064	TOTAL			17,625

Table 29. 2015 Score Differences Grouped by Statistical Significance (Mathematics Grade 3)

Division	N	Homogeneous Groups					
		1	2	3	4	5	6
Sylhet	1,969	92.37					
Barisal	1,433		96.83				
Chittagong	4,525			97.79			
Khulna	2,336				98.75		
Dhaka	6,919				98.87	98.87	
Rangpur	2,866					99.72	
Rajshahi	2,906						101.34
TOTAL	22,954						

Score differences for high scoring Rajshahi and low scoring Sylhet were statistically significant formal other divisions. Note that while there was a numerical difference between Dhaka and Khulna, this difference was not statistically significant. There was also a statistically significant difference in mean score between Chittagong and Barisal. The mean score for Rangpur was statistically significant from the groups above and below Rangpur.

Looking at Mathematics Grade 5 by division across years (Table 30), Rajshahi scored the highest in 2015 and second highest in 2013, however, it was second lowest in 2011. Barisal fell from the leading position in 2013 to second lowest in 2015, and Dhaka moved up in 2015 in relative ranking from second lowest to second highest, scoring basically at the same level as Rajshahi. Sylhet remained at the bottom in all three NSA years, and Khulna remained in the middle in all three assessment years. Except for Rajshahi being in the upper echelon and Sylhet being in the bottom for both NSA 2013 and 2015, there is considerable changing of places in the rank order of regions.

Table 30.NSA Mathematics Achievement by Geographic Division (Grade 5)

2015				2013				2011			
Division	Mean MSS	Std. Dev	Student N	Division	Mean MSS	Std. Dev	Student N	Division	Mean MSS	Std. Dev	Student N
Rajshahi	111.9	11.6	2,536	Barisal	119.6	12.6	1,115	Chittagong	121.4	11.5	2,977
Dhaka	111.6	11.1	5,706	Rajshahi	118.3	12.7	2,171	Rangpur	119.1	11.2	1,565
Khulna	110.8	11.0	1,983	Chittagong	117.2	12.4	3,927	Dhaka	118.9	10.9	4,052
Rangpur	110.5	10.9	2,147	Khulna	115.5	11.3	2,030	Khulna	118.3	11.1	1,530
Chittagong	109.5	10.2	4,163	Rangpur	115.4	12.4	2,054	Barisal	117.7	10.8	940
Barisal	108.9	10.8	1,253	Dhaka	114.4	11.9	5,123	Rajshahi	116.6	10.3	1,724
Sylhet	104.4	9.5	1,595	Sylhet	111.0	13.1	1,386	Sylhet	112.8	9.6	1,039
TOTAL			19,383	TOTAL			17,806	TOTAL			13,827

However, before making inferences about the meaning of these rank scores, it is important to identify which mean score differences were statistically significant, as presented in Table 31. In Table 30 we see that the mean score difference between Rajshahi and Dhaka was not statistically significant. The lowest mean score for Sylhet was different at a statistically significant level from all other scores. The differences between Rangpur and Khulna and Barisal and Chittagong were also not statistically significant.

Table 31.2015 Score Differences Grouped by Statistical Significance (Mathematics Grade 5)

Division	Student N	Homogeneous Groups				
		1	2	3	4	5
Sylhet	1,595	104.44				
Barisal	1,253		108.86			
Chittagong	4,163		109.48			
Rangpur	2,147			110.45		
Khulna	1,983			110.83	110.83	
Dhaka	5,706				111.60	111.60
Rajshahi	2,536					111.93
TOTAL	19,383					

CHAPTER 4. FACTORS ASSOCIATED WITH ACHIEVEMENT IN BANGLADESH

In 2015 three surveys were administered to students, teachers, and head teachers, to collect information about their backgrounds and demographic status. This information was used to analyze what factors were associated with student achievement as highlighted in this report. Some of those findings are presented here in Chapter 4. More rigorous study and analyses can be conducted by the DPE with the data now available.

Differences in Achievement between Schools

In the 2013 NSA Report, it was noted that between-school factors accounted for 65% of the total variance for Bangla Language and 66% for Mathematics (p. 42). In 2015, the between-school variance accounted for 59% of the total variance for Bangla Grade 3, less than in 2013 but still a relatively high proportion. In 2013, the large proportion of variance explained by between-school factors held at both Grades 3 and 5, was found in both Mathematics and Bangla Language, and was evident across all divisions in the country. The implication from both 2013 and 2015 is that there are tremendous differences in achievement across schools, with some schools in the sample having mean score differences that were 3 and even 4 standard deviations apart, very large gaps by any standard. Further, some of these large differences were within the same school type category.

The meaning of these numbers is important because they are indicative of very large achievement gaps. In 2013, several plausible explanations for this state of affairs were postulated: differences in human and material resources, impediments to implementing improvement programs being much greater in some schools than others, the concentration of some schools in poor regions, and the increases in enrollment of students (perhaps many unprepared for study) who were never "in the system" in previous years. The 2013 report notes that this level of between-school variance in Bangladesh is very high by international standards and that levels of between-school variance tends to decrease with increased per capita income levels of a given country.

Educational Attainment of the Mother

According to the data collected, 64.6% of students had mothers who were either illiterate (23.3%) or had attained Grade 5 education (41.3%). While the educational background of the mother explained a low proportion of the overall variance (less than 3%), there were nonetheless statistically significant relationships between mothers' educational attainment levels and student performance. Estimating a linear model, education levels attained by mothers was controlled for at several levels. The comparison group was "illiterate" and Table32 below presents the scaled score changes associated with the attainment of additional levels of education, in comparison with mothers being illiterate.

For example, having completed Grade 8 is associated with a 3.67 point increase to the Bangla Grade 3 test scores. Having completed secondary education is associated with a 5.57 point increase to the same scores. Perhaps not surprisingly, the trend continues upward: increases in the educational attainment levels of the mother are associated with additional points gained on the Bangla test.[7]

---

[7] Note that the standard error increases as education levels rise. This is plausibly due to the increasingly smaller proportions of these groups in the sample, i.e. only 0.9% of the sample had mothers who had attained the level of "post graduate."

Table 32: Effect of Mothers' Educational Attainment on Test Scores (Bangla 3)

Mothers Educational Attainment	Test Score Change	Std. Error	% in NSA Sample
Illiterate	--	--	23.3%
Primary Grade 5	+ 1.99	0.220	41.3%
Primary Grade 8	+ 3.67	0.262	19.1%
Secondary School Certificate	+ 5.57	0.333	9.0%
Higher Secondary Certificate	+ 5.03	0.469	3.8%
Higher Education (Bachelors)	+ 6.94	0.563	2.5%
Post Graduate (Masters)	+ 7.22	0.924	0.9%

Educational Attainment of the Father

While attention in the education literature is frequently given to the impact of the mother's educational attainment on student achievement, the educational attainment of the father can also plausibly be associated with varying achievement levels. Table 33 presents the scaled score changes associated with the attainment of additional levels of education, in comparison with a baseline of being illiterate. As with the mother, mean score gains are associated with increases in fathers' educational attainment.

Note however, that in comparison with change by mothers' level of attainment, the gains are just slightly smaller for the lower levels of educational attainment. For example, instead of Grade 5 attainment being associated with a 3.67 point increase to the Bangla Grade 3 test scores, it is associated with only a 1.7 increase when the father has that level of educational attainment.[8] Having completed secondary education is associated with a 5.57 point increase for mothers while 4.07 for fathers. Then, for the two highest levels of educational attainment, the gains are just slightly higher for fathers than for mothers. The overall trend, however, does continue upward: increases in the educational attainment levels of the father are also associated with additional score gains on the Bangla test, when the impact of mother and father educational attainment is analyzed independently.

Table 33: Effect of Fathers' Educational Attainment on Test Scores (Bangla 3)

Fathers Educational Attainment	Test Score Change	Std. Error	% in NSA Sample
Illiterate	--	--	23.2%
Primary Grade 5	+ 1.70	0.227	35.3%
Primary Grade 8	+ 3.07	0.257	20.6%
Secondary School Certificate	+ 4.07	0.325	9.6%
Higher Secondary Certificate	+ 4.70	0.404	5.5%
Higher Education (Bachelors)	+ 6.95	0.451	4.2%
Post Graduate (Masters)	+ 7.85	0.692	1.6%

Economic Status and Educational Achievement

Information about student living conditions and availability of home amenities was collected through a student survey in order to establish relationships between achievement outcomes and economic status of students in the sample. In order to determine how economic status was associated with achievement outcomes, a series of proxies for economic status was developed (access to food, access

---

[8] Interaction effects, and an interpretation of the effect of both parents (in some combined measure), were not accounted for in this analysis.

to Safe Drinking water, access to a sanitary latrine, possession of a T.V., bicycle, mobile phone, motor cycle, and a separate room in the home for studying). Respondents were asked to select "yes" or "no" if these amenities were available in their dwellings. The table below presents the percentage of students for Grade 3 who marked having these possessions in their households. Of course, there are some limitations in accuracy when collecting data from young students.

Table 34: Percentages of Students with Access to Amenities in their Dwellings

Percent Reporting:	% Access to:
Sufficient Food	85.0%
Separate Room for Study	49.7%
Safe Drinking Water	87.5%
Sanitary Latrine	72.3%
Electricity	73.8%
Cell phone	87.0%
Bicycle	43.9%
Motorcycle	17.3%

In order to understand the proportion of variance attributed to these factors and to understand which – if any– of these economic indicators were related to achievement outcomes, a stepwise linear regression analysis in SPSS was employed using the indicators as categorical variables. The overall proportion of variance explained by these factors was very low, less than 2% for even the highest factor. However, there were statistically significant relationships between the presence of these amenities and scaled scores for Bangla Grade 3. These impacts, however, were modest. For example, having access to water or a sanitary latrine, the highest coefficients, were associated with 1.15 and 1.36 increase in Bangla scores, respectively. Having a bicycle and mobile phone were also associated with approximately a 1.2 increase in Bangla score. Interestingly, access to food was associated with a score decrease for Bangla 3. This could be explained as being caused by a statistical artifact.

Table 35: Economic Indicators (coefficients) [a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	96.984	0.332		291.749	0.000		
Sanitary Latrine	1.360	0.215	0.049	6.335	0.000	0.787	1.270
Bicycle	1.220	0.176	0.049	6.944	0.000	0.955	1.047
TV	0.935	0.192	0.037	4.863	0.000	0.817	1.224
Mobile	1.290	0.264	0.035	4.895	0.000	0.923	1.084
Safe Drinking Water	1.152	0.280	0.031	4.119	0.000	0.848	1.180
Access to Food	-1.141	0.255	0-.033	-4.475	0.000	0.879	1.138
Separate Room	0.681	0.185	0.027	3.670	0.000	0.844	1.185

a. Dependent Variable: Scale Score Bangla Grade 3

Economic Status by Dwelling

Another way to examine economic status was to analyze the data based on student dwelling. "Puccas" are concrete dwellings, like urban apartment complexes. "Semi-puccas" are generally considered more modest, brick wall homes, usually with tin roofs. Finally, "kucchas" are marked by the use of bamboo, wood, or other natural materials and can be found in urban, semi-urban, and rural areas.

The difference in Bangla scores between those students living in Pucca's or Semi-Pucca's was not statistically significant. However, the difference between students living in Kuccha's and Pucca's and Semi-Pucca's was statistically significant. In a linear regression analysis, living in Kuccha's was associated with -1.8 score (i.e. decrease in scores) in comparison to those living in Puccas.

Table 36: Housing as an Economic Indicator

Housing Type	Number	%	Mean Score	St. Dev.
Pucca (cement building)	4,160	20.0%	101.56	12.398
Semi-Pucca (brick walls, tin roof)	5,207	25.0%	101.47	11.818
Kuccha (Bamboo, wood)	11,419	54.9%	99.78	12.561

CHAPTER 5. HEAD TEACHER and TEACHER SURVEYS

Head Teacher and Teacher Demographics

Drawing on data collected from the Head Teacher Surveys from both grades and subjects, the average age of Head Teachers was 43 (median 44, mode 45), while the average number of years Head Teachers had served in that capacity was 11.9 years - 18 mean years of total service to the school overall. In terms of educational attainment, a total of 33.1% of Head Teachers had a post graduate degree, 36.4% had attained a bachelor's degree, 19.1% a completed high school certificate, 7.7% a lower secondary certificate, while 2.5% had not attained a secondary certificate. In terms of Head Teacher professional degrees, the data indicate that 57.8% had a Certificate in Education, 16.4% a Bachelor's in Education, 1.1% a Degree in Pedagogy, and 1.8% a Master's Degree in Education.

A total of 3,278 total teachers (both grades and subjects) completed the teacher survey. 61.6% of teachers surveyed were from Government Primary Schools and 61.9% were female. The mean length of teacher service was 12.8 years, with a median of 10 years and a mode of 5 years. The average age of the teachers surveyed was 36, with a standard deviation of 9.7. The oldest teacher surveyed was 78 years old. In terms of educational attainment, a total of 23.3% had some type of graduate degree, 33.3% were graduates of higher education institutions, while 24.7% had completed secondary education. 13.2% had lower secondary education only and 3.9% had attained below lower secondary education. In regards to professional qualifications, 60.5% had a teaching certificate in education, while 7.4% had a Bachelors in Education.

Interestingly for the question about professional degree, data was missing for 29.1% of the teachers surveyed. Of the 84% of teachers who did respond to the question about major concentrations of study (for last degree obtained), a full 53% had major concentrations in areas other than Bangla (37.7%) or Mathematics(9%). While 37% of the sample had missing data for the question about what subjects they were currently teaching, from those who did respond for Grade 3, 54.1% were teaching Bangla while 45% were teaching Mathematics. For Grade 5, there was also a large proportion of missing data (37.8%). For those who did respond, 51.7% were teaching Bangla for Grade 5 and 45.1% were teaching Grade 5 Mathematics.

In regard to professional support (by way of classroom observation), the mean number of classroom observations of teachers was 2.5 by Assistant Upazila Education Officers (AUEO) (median 2, mode 2) and 1.5 by Upazila Resource Centre (URC) Instructors (median 1, mode 1). The minimum number of observations was 1 for both categories and the maximum was 10 for AUEO and 9 for URC. 23% of teachers did not respond to the question about which subject they received their last training in. Of the 77% that did respond, the percentages were 32.2% (Bangla), 32.3% (Math), and 34.8% (other subject).

The Head Teacher and Teacher surveys provided other interesting information about the schools surveyed. According to the data provided by the Head Teachers, the mean number of boys per school was 123.5 (Grades 1-5) and the mean number of girls per school was 129.3 (Grades 1-5). An interesting source of variation across schools was the length (in minutes) of the typical lesson in these two primary subjects. For example, 35.7% of teachers reported holding 50 minute lessons, while 29.6% report holding 35 minute lessons: a large time difference when the cumulative effect of time on task is considered over the course of a year. 11% of teachers report having 40 minute lessons while 9.1% report having 45 minute lessons. We also learn from the Head Teachers survey that the average

number of male teachers per school was 3.5 (median 2, mode 2), while the average number of female teachers per school was 4.3 (median 4, mode 2).

Teacher Beliefs about Themselves and the Profession

The surveys elicited teachers' opinions and beliefs about a variety of personal, professional, and pedagogical issues. In Table 37, we see that the majority of respondents intend to stay in teaching (59.1%), while another 23.6% hope to become Head Teachers. It would seem encouraging that such a proportion would desire to stay in the profession. While the survey is based on self-reporting, the data nonetheless provide an interesting picture of teachers' attitudes, beliefs, and dispositions. After presenting descriptive data on teacher beliefs (below) on self-efficacy and other key questions, in the next section we present findings on what teacher-related factors were associated with student outcomes.

Table 37. Future Ambitions of Teachers

Future ambitions	% reporting
Continue working in the teaching profession	59.1%
Be a head teacher	23.6%
Be an educational administrator	8.6%
None of these plans	4.2%
Leave teaching profession to do something different	1.8%
Leave the teaching professional immediately	.5%

As can be seen in Figure27, over 80% of teachers believe that their work is contributing to building good citizens. A majority also fully agree or agree that the working environment is good and that teaching is a stable profession. Almost all respondents agreed that they enjoyed working with children. Note, however, that a majority do not agree that remuneration in the profession is high.

Figure27. Teacher Beliefs about the Conditions of Teaching and Their Motivations
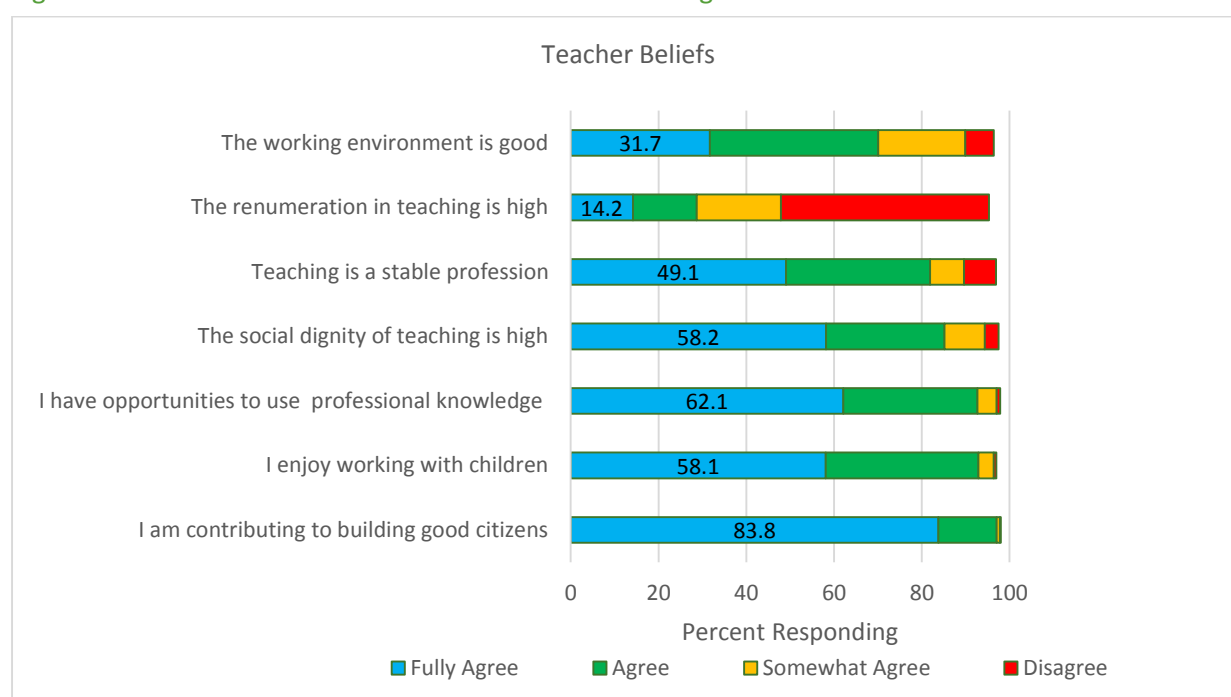
In regard to problems faced by teachers (Figure28), helping students learn who are not eager or able to understand is a major problem faced by teachers, with just over 50% either fully agreeing or agreeing. Finding funds to support instructional activities and lack of time also seems to be an issue for around half of all teachers surveyed. However, large majorities disagreed that there is a lack of cooperation from the Education Office or from the School Management Committee. More than half also do not agree that there is a lack of subject matter help from colleagues. According to the survey, there is some support for the idea that the majority of teachers feel supported by their peers.

Figure28. Teacher Beliefs about Problems in the Profession
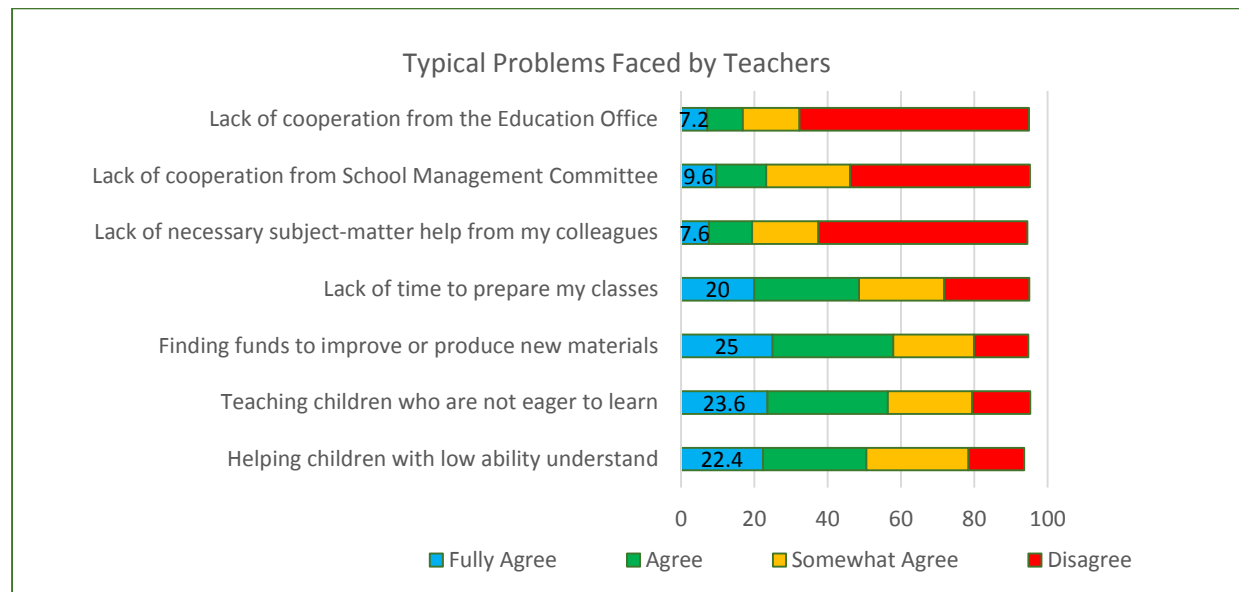
Figure29. Teacher Beliefs about Self-Efficacy (1)
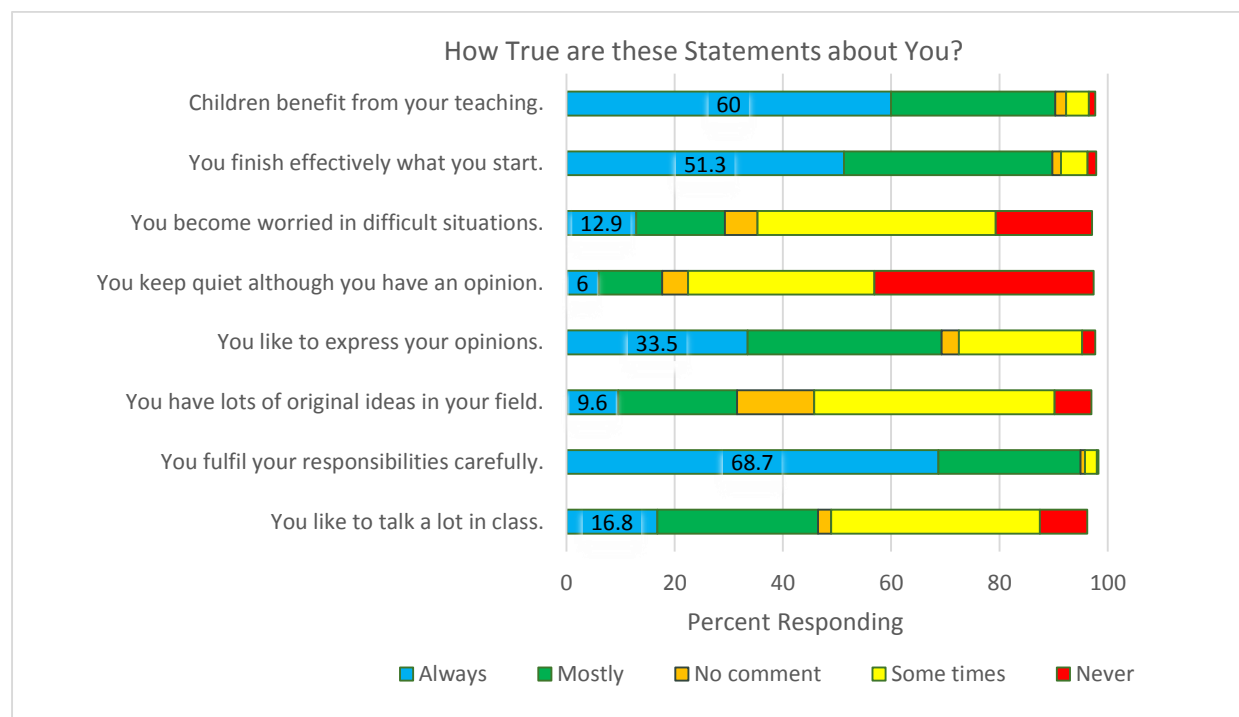
In regard to beliefs about self-efficacy (Figure29 above), an overwhelming majority believes that children benefit from their teaching. Teachers also report that they fulfil their responsibilities carefully. Almost half believe that they are not silent when they have an opinion on something.

Figure30. Teacher Beliefs about Self-Efficacy (2)

How True are these Statements about You?

Statement	Always
You are very submissive.	51.1
You think of the consequences before you start work.	51.5
You enjoy things like nature, art and music.	37.2
You prefer longer tasks than shorter tasks.	14.6
You need a lot of rest to be able to work.	10.2
You are eager to learn new things.	62.2
You often get worried in the classroom.	13.8
You forgive others easily.	42.2
You go on working when others tend to give up.	15.5

Legend: Always, Mostly, No comment, Some times, Never

Figure31. Teacher Beliefs about Self-Efficacy (3)

How True are these Statements about You?

Statement	Always
The work that you do influences others.	26
If you do not understand, you seek help from others.	56.4
Students consider you an ally.	76.3
You think carefully before taking any important decision.	65.7
You make friends easily.	54.2
You are generous to others with your time and money.	30.2
You become nervous easily.	14.2
You complete work accurately and on time.	35.2

Legend: Always, Mostly, No comment, Some times, Never

Though teachers noted they are not quiet when they have opinions, nonetheless over 80% consider themselves to be "submissive" and that they think of the consequences before they start work (Figure30 above). Around 50% of teachers noted that they never get worried in the classroom and that they do not need a lot of rest in order to be able to work.

Finally, over 90% of teachers believe that they think before they make an important decision and that students would consider them to be allies.

Teacher Time on Task

All teachers were asked by the survey to indicate how much time they spent engaged in certain classroom activities during their last lesson taught. They were not provided with a list of answer options but instead were required to fill in their own numbers, thus the numbers do not add up to 100%. A summary of the response data is presented below in Table 38. The five categories of time presented were selected somewhat arbitrarily but represent the most commonly noted number of minutes. So, for example, 44% of the teachers spent 5 minutes on homework while 44% also spent 5 minutes on question-answer or quizzes.

Table 38. Teacher Time on Task in the Classroom (% selecting)

Task	2 min	3 min	5 min	10 min	15 min
Check homework?	3	5	44	23	4
Lecture method?	5	7	37	17	0
Reading silently or aloud?	4	5	42	17	2
Helping students solve problems?	2	6	42	22	4
Students solve own problems (without help)?	7	9	37	12	2
Student group work?	3	4	38	23	4
Question-Answer or Quizzes?	12	12	44	5	1
Class management besides giving lessons?	30	16	24	1	0
Other?	17	14	22	2	0

Teacher Factors Associated with Student Outcomes

In order to understand which teacher background factors were associated with student performance, regression analyses were conducted with the data for both Bangla Language and Mathematics, at both Grades 3 and 5. The variables analyzed for both grades and subjects included teacher qualifications (level of educational attainment), professional degree (focus on education or other subject), the teachers' major concentration of study in higher education, and the last type of training (Bangla, Mathematics, other) the teacher received. Few statistically significant relationships between these four factors and student outcomes were established with two exceptions.

First, for Bangla Grade 3 data, teachers who received their last professional training in Bangla Language did have students who scored at statistically significant higher levels on the Bangla assessment than those students whose teachers completed other types of training. Second, from all four groups, teacher educational attainment was positively associated with increased student outcomes for Mathematics Grade 5. Table 39 below presents the output of the linear regression analysis using categorical data (attainment levels).

The results are interpreted in relationship to the teachers who had not completed any level of secondary education (left out of the model). For example, the completion of higher (complete)

secondary school is associated with a 3.54 point higher test score. Having a teacher with a university degree was associated with a 3.91 point higher score; with a graduate degree, 5.12 point higher test scores. It is plausible that this educational attainment effect is prevalent at the upper grades (Grade 5 not Grade 3) as mathematical subject matter becomes more complex and places greater knowledge demands on teachers. For no other subject or grade was there a significant association between teacher educational attainment and student outcomes.

Table 39. Effect of Teacher Educational Attainment on Student Outcomes (Mathematics Grade 5)[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	105.871	1.420		74.536	0.000
Lower Secondary School	3.511	1.776	0.106	1.977	0.048
Complete Secondary	3.548	1.589	0.151	2.233	0.026
University Graduate	3.910	1.500	0.212	2.607	0.009
Post Graduate Work	5.126	1.525	0.260	3.360	0.001

a. Dependent Variable: ss100_mean

Teacher Attitudes towards Teaching and Student Outcomes

In order to understand any possible relationships between teacher attitudes and dispositions and student outcomes, we created two "attitudinal scales" with the data collected. Two constructs – "Job Positivity" and "Job Negativity" –were composed of 7 items each. Under "Job Positivity" teachers responded to the statements under the heading "I like teaching because of _____." For "Job Negativity" the statement reads "I have some problems such as _____."Responses to the items required a "level of agreement rating" from the respondent where the score range was *Fully Agree* (4 points), *Agree* (3 points), *Somewhat Agree* (2 points), and *Disagree* (1 point). Scores were averaged for both constructs. Thus, high "Job Positivity" reflected strong agreement with positive statements about teaching, while high levels of "Job Negativity" reflected strong agreement with negative statements about teaching. For the complete list of 7 statements for each construct, please see the Teacher Survey, Tasks O and P.

The results of linear regression analyses indicated no association between the measure of teacher "Job Positivity" and student scores for Bangla or Mathematics. However, interestingly, "Job Negativity" was associated (p-value 0.009, Grade 3) with decreases in scores at statistically significant levels for both Bangla Grade 3 (-1.1) and Grade 5 (-1.0), but not for Mathematics. In other words, an increase in one unit of teacher "Job Negativity" is associated with a decrease of student score at the levels noted above.

Classroom Management, Mentoring, and Homework

Table 40 presents descriptive summary data in regard to classroom management, use of mentoring, and homework. 69.4% of teachers sometimes experienced problems managing their classrooms, while only around 13.5% experience management problems "all the time" or "often." In terms of collaborating with mentors, 47.9% note that they received advice from the Head Teacher or other supervisory unites "all the time."

Table 40. Teacher Experience with Classroom Management andAdvice Seeking

	All the time	Often	Sometimes	Never
Problems managing the classroom	6.3%	7.2%	69.4%	14.8%
Receive advice from head teacher, AUEO, or other colleagues in class management or teaching	47.9%	23.6%	22.2%	2.6%

For Bangla Grade 3, Mathematics Grade 3, and Mathematics Grade 5, there were no associations between teachers reporting classroom management problems and student outcomes. However, interestingly, for Bangla Grade 5, there were statistically significant increases in test scores associated with teachers who acknowledged management problems –"sometimes" (+1.8, p-value 0.03) and "often" (+2.95, p-value, 0.03) in comparison to teachers who "never" had classroom management problems. Intuitively, teachers with classroom management problems might seem less likely to maintain environments conducive to learning. On the other hand, there may be traits associated with teacher honesty and transparency that are also strongly correlated with quality teaching. In other words, "admitting challenges in classroom management" may be associated with stronger teachers, not weaker ones. This result underscores the need to be cautious about inferences based only on regression analyses. No statistically significant associations were discovered between teacher advice seeking and student outcomes.

41.9% of teachers give between 0 and 30 minutes a day while only 1.1% give more than 60 minutes. An analysis of the relationship between amount of homework given and student outcomes revealed no relationship for Bangla Grade 3, Bangla Grade 5, and Mathematics Grade 3. However, for Grade 5 Mathematics, "some homework" (30-60 min) is associated with declines in test scores (-2.6 point, p-value = 0.007). More investigation is needed to understand what might be going on with Grade 5 Mathematics homework.

Table 41. Amount of Daily Homework

How much homework per day?	0 – 30 min	30- 60 min	>60	Do not give
	41.9%	28.9%	15.1%	7.4%
How do you check homework?	Check each	Tell them answers, self-check	Ask, random check	Do not give
	58.3%	16.5%	12.6%	6.1%

Finally, teachers were asked to select "3 items that are the most important for you in your teaching approach, and 3 items that are the least important for you." From 11 possible selections, the teachers' most popular choices can be seen below in Table 42. Note that "following curriculum guidelines", "using appropriate materials", and "guiding students to memorize subject matter from the textbook" were the three most popular first choices. "Using appropriate materials" was also the most popular second choice, followed by "involving students in group work." On the one hand, these findings seem to indicate that teachers have strong predilections towards following standard books and materials in their teaching. On the other hand, as can be seen in Table 42, while "guiding students to memorize" was the third most selected "most important thing" it was also selected as the number one "least

important thing" at 37.6%. This would indicate considerable diversity of opinion within the teaching ranks.

Table 42. Most Popular First and Second Choices: "What is *Most* Important?" (% selecting)

First Choice	
Follow Curriculum Guidelines /Teaching Guide	**36.2%**
Use Appropriate Materials during teaching learning process	**29.1%**
Guide students to memorize subject matter from the textbook	**11.1%**
Second Choice	
Use Appropriate Materials during teaching learning process	**30.9%**
Involve students in group work	**29.6%**

Table 43. Most Popular First and Second Choices: "What is *Least* Important?" (% selecting)

First Choice	
Guide students to memorize subject matter from the textbook	**37.6%**
Put emphasis on lecture methods	**13.7%**
Follow Curriculum Guidelines and Teaching Guide	**13.3%**
Second Choice	
Put emphasis on lecture methods	**26.9%**
Organize quizzes and small tests	**13.7%**
Support Individual student to do their work	**12.5%**

CHAPTER 6. DISCUSSION AND CONCLUDING REMARKS

Understanding the state of Bangla Language and Mathematics instruction is a first step towards improving instruction and achievement outcomes. It is essential to understand what is happening in these content areas and to monitor progress at all levels in order to adjust interventions and supports as necessary. It is also important that any proposed initiatives or reforms to the system be tightly aligned with other parts of the system. The purpose of this section is not to suggest a list of policy or program interventions but rather to raise pertinent questions related to 2015 NSA results and how they can be studied, analyzed, and employed effectively to improve instruction and achievement outcomes.

Discussion for Policy Makers

Content Coverage on the NSA

For Bangla Language Grades 3 and 5, the vocabulary tasks were the easiest while the reading comprehension tasks were the most challenging. Students correctly answered a larger proportion of selected response questions than constructed response questions. For Mathematics Grade 3, students scored highest on Shape and Space but for Grade 5 they scored the highest on Measurement and Units. The DPE will seek to engage with MoPME system-level counterparts as well as teacher training institutes in structured conversations around the following core questions:

- In addition to this National Report, what are the other mechanisms through which the Government of Bangladesh (GOB) will ensure that lessons are learned from NSA results system-wide and that findings are disseminated widely and acted upon all the way down to the classroom level?
- For example, as reading comprehension and higher order processing skills need more focused attention, what is the mechanism through which this information becomes available at the school level?
- Do teachers and school administrators currently have the necessary resources and support to improve students' knowledge in the needed areas? What are the barriers to improvement? Are these barriers related to materials, teacher knowledge, or other constraints?
- How can NSA reporting provide more focused strand and item level analyses? How can NSA reporting with strand and item level analyses be produced and distributed efficiently so that teachers receive adequate information about areas of student weaknesses?
- What are some of the barriers to enhancing the utility of the NSA so that NSA results reach the classroom level and impact instruction? How can these barriers be overcome?
- Are instructional priorities and materials at the national, divisional and upazila level(s) tightly aligned with the content covered on the NSA? If not, how can that alignment be improved?

Cognitive Processing Levels on the NSA

NSA results for Bangla Language and Mathematics indicate that students at both Grades 3 and 5 answered questions correctly requiring lower level cognitive processing (Knowledge and Understanding) in greater proportions than they answered questions requiring application and higher order thinking. Acknowledging the challenges of employing instructional approaches that seek to be more rigorous in cognitive demand, more focus and investment could be directed towards this endeavor. The DPE will seek to engage with MoPME system-level counterparts as well as teacher training institutes in structured conversations around the following core questions.

- At what level(s) of analysis do teachers need information or feedback from the NSA results in order to realistically adjust instructional approaches based on NSA findings related to cognitive processing levels?
- Do the current curricula and mandated cognitive processing standards align tightly with how teachers are trained in regard to instructional best practices and methods? If not, what can the DPE or other agencies do to focus more attention on this issue?
- Is more training necessary through pedagogical or other institutes in order to improve the alignment between NSA results and instructional practice?
- Are there currently incentives for teachers to embrace the additional work required in preparation time to develop lessons that demand more rigor from students? If not, what are the barriers to incentivizing teachers to make changes? Who will be responsible for creating the incentives necessary for change in instructional practice?

Quality Gaps by Division, between Schools, and by School Types

While the evidence from the 2013 and 2015 NSA indicates no large gaps in achievement by gender, there are achievement gaps by school type and division. The large amount of variance between schools also indicates that within divisions and upazilas there are large quality gaps across schools. As the 2013 NSA report noted, DPE and other stakeholders can analyze NSA results at the upazila and school levels, identify those schools performing poorly, and consider providing more targeted support and outreach to those schools most in need.

In regard to mean scores by division, we see that the Rajshahi has retained a relatively higher position in comparison to other divisions across grades and subjects. Sylhet Division has consistently been in the lower ranks at both grade levels and subjects. The achievement gap between Sylhet and the highest scorers is large in some cases, almost one standard deviation in score difference. Rangpur has also been a top four scorer for both assessment years. Khulna and Dhaka were relatively higher in position in 2015 than in 2013, with Dhaka moving from seventh (2013) to second (2015) on one of the Bangla years while Barisal dropped in Grade 3 Mathematics from second highest in 2013 to second lowest in 2015. Sylhet remained at the bottom for Grade 5 Mathematics and Rangpur, Chittagong, and Khulna remained in the middle in both assessment years. How can resources be better focused on those divisions that scored the lowest on the NSA?

KG Schools and High attached primary schools performed at the highest levels in both grades in both 2013 and 2015. Madrasah and ROSC schools scored relatively lower in both assessment years. For Bangla Grades 3 and 5, Government Primary Schools (GPS) and Newly Nationalized Primary Schools (NNPS) both fell by 4-5 points from 2013 to 2015. Together, these two school types contain about 77% of the sampled population. For Grade 5 Mathematics, the same two school types were the top performers in both 2013 and 2015 – KG and GPS, though their relative positions were reversed. Madrasah and ROSC were also near the bottom in both 2013 and 2015. The differences between the top scoring school types and the bottom scoring school types were large, close to 1 standard deviation in some cases. How can resources be better focused on those school types that scored the lowest on the NSA?

The DPE will seek to engage with MoPME system-level counterparts as well as teacher training institutes in structured conversations around the following core questions:

- In particular, what factors might be contributing to the overall decline of GPS and NNPS scores from 2013 to 2015? Are there systemic issues that might explain the decline in performance for these two school types?

- Beyond the obvious economic and background factors that may explain achievement gaps by school type or division, are there approaches or methods that the higher achieving school types and regions are doing that lower achieving school types or areas are not doing?
- If yes, differences need to be diagnosed (in terms of curricula and assessment alignment, investment, resource allocation, teacher attraction, teacher retention, or pedagogical training) and analyzed. What means does DPE have to identify gaps in these areas? Which institution or agency will lead this initiative?
- How could more resources be focused on closing achievement gaps by school type and between school differences within the same school type but where achievement gaps are large?
- Is the issue of directing resources towards improvement related primarily to limitations on funding or are there other contributing factors such as bureaucratic obstacles or teacher incentives?

Recommendations for Improvements in the NSA 2017

The NSA 2015 has introduced into the procedures for test development a number of modifications designed to improve the quality of the NSA. These changes include: integrating into the operational test forms a small number of new items that will be piloted during administration; redesign of the test blueprints to align with the newly revised 2011 national curriculum; organization of all content standards measured on the tests in terms of a horizontal structure; use of item cards containing all piloted items with content and psychometric information to facilitate item review and test assembly, among other changes.

The design, development, and implementation of the NSA 2015 has provided all of the participating stakeholders with extensive opportunities for reflection on the types of changes that might be suggested for improving procedures for the next iteration of the NSA. These suggestions are presented below:

1. We recommend that the NSA 2017 introduce an independent test of the writing domain (i.e., creative writing) as part of the Bangla Language test. Writing, which is an important subject on the curriculum, is complex to assess, mostly related to the need for objective procedures for scoring via rubrics. The DPE and NCTB have some experience handling open-ended test items in both Mathematics and Bangla Language (in the Reading Comprehension domain). We would recommend conducting an assessment of writing through a sub-sample of the main sample.

2. The report of the NSA 2015 administration points out that the "legacy bands" established in 2011 and used for providing performance levels defined by the content assessed on the tests has certain limitations. These limitations are in part due to the fact that the performance scale cuts across the grades assessed (viz. Grades 3 and 5). Data from the NSA 2015 show that there is a large difference in the percentage of students in Grade 5 who achieve grade level performance (between 11%-32% depending on the subject) compared to the percentage of students in Grade 3 who perform at Grade 3 level (41%-75% depending on the subject). We would recommend an alternative methodology based on setting up in-grade performance scales by subject area and establishing cut-points on each scale using the 2017 operational test data. Results on the NSA 2017 can be reported by means of these grade-specific performance scales. The scales can also be used retrospectively to plot 2011, 2013, and 2015 scores. These performance scales have a number of advantages, including allowing teachers, schools, and districts to set annual achievement targets by level on the scale and then monitoring achievement by reference to the performance scale.

3. We suggest Intensifying MoPME and DPE capacity-building around key technical areas in test development, administration, analysis and reporting in order to consolidate and improve the skills already acquired by technical staff. Long-term sustainability of the assessment system in the hands of local experts must continue to be a major goal of the NSA.

4. One of the limitations of the NSA 2015 is that it is not possible to measure the link between student achievement, as measured by the NSA, and program support provided under the PEDP3. While it is possible to informally suggest the effects of PEDP on student learning outcomes, this can't be done through any formal methodology. We recommend identifying implementation indicators from the post-PEDP3 program implementation plans and measure them concurrently with the NSA so that achievement results on the NSA can be correlated with school support initiatives.
5. More preparation and monitoring of the administration of the NSA in the field as well as data cleaning and scoring needs to take place in order to ensure availability of high quality data.
6. Greater focus on formative uses of the results of the NSA should be integrated into the study; teachers, schools and districts can benefit greatly from the availability of data and school reports for instructional planning purposes.
7. We recommend that the NSA 2017 be considered as a baseline for the implementation of the newly modified national curriculum as well as a baseline for post-PEDP3 programming.
8. Given the somewhat low levels of motivation of students and schools reported in the administration of the NSA 2015, we would recommend greater marketing of the NSA highlighting the differences between the PECE and the NSA, what the different goals of both testing programs are, and why it is necessary for students and schools to take the NSA seriously.
9. We would recommend collapsing the 2 higher cognitive levels (combining application and higher order thinking) into one single level – distinguishing between these two cognitive levels when designing test items for primary levels, especially at Grade 3, is not easy, although one begins to see greater differentiation of higher level cognitive processing in the upper grades.

APPENDIX 1. Grade 3 and Grade 5 Bangla Language Objectives and Expectations

Key areas	Strands	Learning outcomes
Reading comprehension	Read for meaning in grade-level appropriate **imaginative** texts	Understand the **main ideas, secondary ideas, global ideas,** and **inferential ideas** of grade-level appropriate imaginative texts
	Read for meaning in grade-level appropriate **informational** texts	Understand the main ideas, secondary ideas, global ideas, and inferential ideas of grade-level appropriate informational texts
	Reading for meaning in grade-level appropriate **persuasive** texts	Understand the main ideas, secondary ideas, global ideas, and inferential ideas of grade-level appropriate persuasive texts
Grammar	Word grammar	Identify and use key elements in the construction of a word in Bangla Language to support comprehension
	Sentence grammar	Identify and use key elements in the **construction of a sentence** in Bangla Language to support comprehension
Vocabulary	Known words	Identify the **meaning of words in Bangla that should be known** and are grade appropriate
	New (i.e., above grade-level) words	Identify the **meaning of new words in Bangla** (appropriate for a maximum of 2 grades above the targeted grade) through the use of grammatical and semantic contextual information
(see: *NSA Assessment Frameworks_NSA06022013_Final.pdf*)		

APPENDIX 2. Bloom's Taxonomy[9]

EVALUATION
Assessing theories; Comparison of ideas; Evaluating outcomes; Solving; Judging; Recommending; Rating

SYNTHESIS
Using old concepts to create new ideas; Design and Invention; Composing; Imagining; Inferring; Modifying; Predicting; Combining

ANALYSIS
Identifying and analyzing patterns; Organisation of ideas; recognizing trends

APPLICATION
Using and applying knowledge; Using problem solving methods; Manipulating; Designing; Experimenting

COMPREHENSION
Understanding; Translating; Summarising; Demonstrating; Discussing

KNOWLEDGE
Recall of information; Discovery; Observation; Listing; Locating; Naming

[9]https://juliaec.wordpress.com/2011/03/23/blooms-taxonomy-encouraging-higher-cognitive-thinking-in-primary-school-classrooms/

APPENDIX 3. NSA 2015 Grade 3 Mathematics Objectives and Expectations

Key areas	Strands	Learning outcomes
Numbers and operations	Counting and ordering numbers	Count, read, write and order numbers up to 10,000 in various ways
	Basic operations	Add and subtract numbers without and with carry over, multiply by 2-digit numbers and divide by 1-digit numbers using place value
	Problem solving using basic operations	Solve 2-stage problems on addition or subtraction and multiplication or division involving 2-digit numbers (multiplier to be 2-digit number at most and in case of division the divisor to be 1-digit number)
	Fractions	Identify fractions, equivalent fractions, and compare fractions, add and subtract like fractions (denominator being 1-digit number only)
Measurement and units	Currency	Identify, read and write Bangladeshi currency, convert Taka to paisa and vice versa and use it to solve problems on money transactions
	Length	Use measurements of length (m, cm, mm) to measure and draw line segments of given length
	Weight	Convert between different units of measurement of length, mass and capacity
	Time	Read time to the minutes and find duration of short activities
Shape and space	Lines, points, and planes	Differentiate between a point, line and plane
	Angles	Identify and draw acute angle, right angle and obtuse angle
	Quadrilaterals	Identify quadrilaterals and differentiate between square and rectangle
	Circles	Identify and draw a circle

APPENDIX 4. NSA 2015 Grade 5 Mathematics Objectives and Expectations

Key areas	Strands	Learning outcomes

Numbers and operations	Operations	Do operations of addition, subtraction, multiplication and division and apply to solve three-stage problems
	Brackets and expressions	Simplify numerical expressions involving brackets
	Averages	Find averages and apply in daily life situations
	GCD and LCM	Find the GCD and LCM of numbers and apply them
	Literals	Use literals for unknown quantities and find the value of a literal from given information
	Fractions	Find equivalent fractions, compare proper fractions, convert between improper and mixed fractions
		Add, subtract, multiply and divide fractions and apply them
	Decimals	Add, subtract, multiply and divide decimals and apply them
	Percentages	Convert between fractions and percentages and apply percentages to solve problems
Measurement and units	Length, mass and capacity	Use metric system of measurement to add and subtract length, mass and capacity
	Conversion of units of measurement	Convert between different units of measurement
	Area	Calculate the area of a square, rectangle, triangle and apply it
	Time	Convert between different units of time and convert time from 12 hour format to 24 hour format and vice versa
Data	Data use	Understand data, read, draw and interpret from bar graphs and apply to situations involving population
Shape and space	Quadrilaterals	Differentiate between different types of quadrilaterals (parallelogram, rhombus, square and rectangle) and draw them
	Circles	Identify parts of a circle and draw a circle
Calculator and computer	Technology	Understand the usage of basic calculator and computer peripherals in daily life

APPENDIX 5. Setting Performance Standards for Bangladesh

Why Performance Standards are Important for the Bangladesh Education System

The principal reason why Performance Standards are important for a national education system is that they significantly improve the interpretability of test results like those of the Bangladesh NSA. When there is improved interpretability, there is the potential for a much greater and more effective impact on instruction

and achievement. Thus, instead of reporting a mean percentage score, based on raw test scores, representing how well students/schools/districts/the country did on a test (e.g., "the mean score on the Grade 3 Mathematics test at the national level was 62.5%"), we are able to report student performance by reference to test scores that are mapped onto a performance scale. The scale is developed to discriminate between 3 or 4 different levels of performance and each level is defined in terms of what students have achieved relative to the content standards measured on the test.

Thus, with Performance Standards, one can now report for example that "25% of Grade 3 Mathematics students are performing at the **insufficient level**, 24% at the **basic level**, 42% are performing at the **proficient level**, and 9% at the **Advanced level**". This is significantly richer information, not only because we can now distinguish between different levels of performance ("still one quarter of all students in Grade 3 is failing Mathematics, while another quarter are only achieving the minimum"), but also because we know exactly what students know and can do (or don't know and can't do) at their level in terms of the content measured: e.g., at the **insufficient level** students can't convert taka to paisa; they can't read the time, and can't calculate the duration of an activity; they can't compare equivalent fractions, etc. This type of information provides a script to teachers/school directors/district supervisors for setting targets for schools to reach ("next academic year we need to lower the percentage of students in the **insufficient level** by 10%", which means providing students with more effective instruction in converting currency/identifying the time/calculating duration/comparing fractions, etc., and probably improving teachers' ability to provide relevant instruction on these topics.)

Through the use of performance standards, comparisons of student performance across different years is far more pertinent and, given the right application of test equating and scaling methodologies, the use of a performance scale can facilitate comparisons between different grades, and even different subject areas, e.g., "25% of Grade 3 Mathematics students are performing at the **insufficient level**, while in Bangla Language 37% of Grade 3 students are performing at that level; by Grade 5, the percentage at that level has dropped to 18%". It is important to note that under the traditional score reporting approach ("the mean score on the Grade 3 Mathematics test at the national level was 62.5%") we know very little about what 62.5% actually means and therefore cannot do much to improve the situation.

How AIR proposes to Develop Performance Standards with the DPE

AIR proposes to develop 4 different performance scales, one scale in each of the following NSA focused subjects/grades: Mathematics and Bangla Language Grades 3 and 5. We require an independent scale for each subject/grade tested so that we can interpret the results of the test in the rich and informative way indicated above. Each performance scale will have 4 levels (in discussions with the DPE stakeholders it may be decided that 3 levels, or even 5 levels, is more appropriate to meet local needs, although AIR would recommend 4 levels). Each level on each of the scales would be defined in 2 ways: 1) with a general description which would be common across all grades and all subject areas; and 2) with a specific description of the expected performance relative to content standards and cognitive skills measured on the test. The sequence of activities involved in developing performance standards (the scale + the descriptions) will be the following:

1. Decide how many levels the scale will be made up of (this decision will govern all of the performance scales for the country);
2. Decide what labels to use to define each level on the scale;
3. Develop the general descriptions of each level on the scale in such a way that the descriptions are coherent across grades and across subject areas;
4. Develop the specific descriptions of each level on the scale using the content standards and cognitive skills expected of students and measured by the test;
5. Using real data from an operational administration of the test as well as the content measured by the test items, map test scores onto the scale and determine the statistical range of each level on the scale.

These activities will be guided by AIR technical experts working with DPE content and assessment specialists, together with local expert subject teachers by targeted grades. The complete activity will take a total of 2 intensive weeks.

What the DPE will be able to do with Performance Standards

When the Performance Standards have been developed, the DPE will be able to carry out the following:

1. Interpret the results of the NSA test administration in ways described above;
2. Design strategies to help improve instruction and student achievement for the upcoming academic year including setting targets for teachers, schools, and districts to aim to achieve;
3. Retrospectively re-interpret the test results from the NSA 2011 and 2013 by reference to the performance standards and plot trends from 2011 to the current administration also by reference to the performance standards;
4. Use the performance standards for all future administrations of the NSA, to look at both horizontal change (from year to year) and vertical change (from grade to grade);
5. Provide the teaching profession/education system with a rich array of formative information that is empirically derived from high quality testing and that involves ensuring that test forms from one administration are appropriately equated and test results can be mapped onto the performance scale;
6. Use the performance standards to hold schools/districts accountable for developing and achieving improvement targets;
7. If the LASI tests similarly develop performance standards for their targeted subjects/grades, then comparisons, both horizontal and vertical, can be made for the complete range of grades covering primary and middle school education in Bangladesh. AIR can modify its proposal to support the development of the 4 performance scales for LASI together with those of NSA.

APPENDIX 6. Statistical Tables for Key NSA Data

Bangla Grade 3: All Students					
Year	**N**	**mean**	**St dev**	**Min**	**Max**
2015	23,040	100.2	11.9	60	140
2013	22,869	104.2	12.1	49	140

2011	17,626	100.8	11.6		

Bangla Grade 5: All Students

Year	N	Mean	St dev	Min	Max
2015	19,406	112.0	12.9	60	154
2013	17,828	115.2	11.0	67	151
2011	13,827	118.6	11.1		

Mathematics Grade 3: All Students

Year	N	Mean	St dev	Min	Max
2015	23,034	98.2	11.4	60	140
2013	23,064	103.7	13.0	45	142
2011	17,615	100.2	9.8		

Mathematics Grade 5: All Students

Year	N	Mean	St dev	Min	Max
2015	19,395	110.2	10.9	60	150
2013	17,806	115.8	12.4	45	154
2011	13,854	116.2	8.7		

Bangla Grade 3: By Gender

	Boys					Girls				
Year	N	Mean	St dev	Min	Max	N	Mean	St dev	Min	Max
2015	11,347	99.5	11.6	60	140	11,693	100.8	12.1	60	140
2013	10,862	103.7	11.9			12,007	104.7	12.2		

Bangla Grade 5: By Gender

	Boys					Girls				
Year	N	Mean	St dev	Min	Max	N	Mean	St dev	Min	Max
2015	9,096	111.5	12.9	60	154	10,310	112.4	12.8	60	154
2013	8,032	115.2	10.8			9,796	115.2	11.1		

Mathematics Grade 3: By Gender

	Boys					Girls				
Year	N	Mean	St dev	Min	Max	N	Mean	St dev	Min	Max
2015	11,364	98.3	11.2	60	140	11,670	98.2	11.6	60	140
2013	10,965	103.8	12.6			12,099	103.5	13.3		

Mathematics Grade 5: By Gender

	Boys					Girls				
Year	N	Mean	St dev	Min	Max	N	Mean	St dev	Min	Max
2015	9,130	110.4	10.8	60	150	10,265	110.0	11.1	60	150
2013	8.030	116.0	12.2			9,776	115.6	12.7		

Bangla Grade 3: By Main School Types

	Government Primary Schools					Newly Nationalized Schools Primary Schools				
Year	N	Mean	St dev	Min	Max	N	Mean	St dev	Min	Max
2015	13,617	**99.75**	11.8	60	140	4,058	**99.9**	11.5	60	133
2013	13,322	**104.3**	12.5	71	144	4,619	**103.2**	11.5	68	143

Bangla Grade 5: By Main School Types										
	Government Primary Schools					**Newly Nationalized Primary Schools**				
Year	N	Mean	St dev	Min	Max	N	Mean	St dev	Min	Max
2015	11,539	**112.4**	12.39	60	154	3,360	**108.9**	12.03	60	154
2013	10,633	**116.3**	11.1	82	155	3,419	**113.0**	10.3	83	154

Mathematics Grade 3: By School Type										
	Government Primary Schools					**Newly Nationalized Primary Schools**				
Year	N	Mean	St dev	Min	Max	N	Mean	St dev	Min	Max
2015	13,615	**97.9**	11.37	60	140	4,058	**99.2**	11.32	60	140
2013	13,454	**104.1**	13.6			4,662	**102.3**	12.0		

Mathematics Grade 5: By School Type										
	Government Primary Schools					**Newly Nationalized Primary Schools**				
Year	N	Mean	St dev	Min	Max	N	Mean	St dev	Min	Max
2015	11,520	**110.9**	11.15	60	150	3,363	**108.9**	10.01	67	145
2013	10,620	**117.2**	13.1			3,413	**113.9**	11.3		

REFERENCES

ACER (The Australian Council for Educational Research)(2012). *2011 National Student Assessment Grades 3 & 5. Revised Report. June 2012. ACER: Melbourne.*

ACER (The Australian Council for Educational Research). 2014. *2013 National Student Assessment for Grades 3 & 5: Public Report*. ACER. Melbourne.

Cohen, J. (1977). *Statistical Power Analysis for Behavioral Sciences* (revised ed.). New York: Academic Press.

DPE (Directorate of Primary Education) (2013). *Bangladesh Primary Education Annual Sector Performance Report* (ASPR) 2013. Ministry of Primary and Mass Education, Dhaka.

Monitoring and Evaluation Division). *National Student Assessment 2011 & 2013 for Grades 3 and 5.* Directorate of Primary Education, Mirpur-2, Dhaka.

World Bank (2013). *Bangladesh Education Sector Review: Seeding Fertile Ground Education what Works for Bangladesh.* World Bank: Washington, D.C.