

# An Analysis of 2015 HSC Examination Results for Selected Subjects of Three Examination Boards

By the BEDU team led by:  
Robiul Kabir Chowdhury, Jack Holbrook, Chowdhury Mufad Ahmed,  
Md. Ali Ahsan, Ahmed Obaidus Sattar Bhuiya\*

## Abstract

*The results from the 2015 Higher Secondary Certificate (HSC) examinations give some indication of how the new Creative Questions (CQ) and the accompanying modified Multiple Choice Questions (MCQ) are being treated across the various Boards of Intermediate and Secondary Education (BISE). This analysis is based on data on 8 HSC subjects on which tests were conducted by three BISEs. It reveals the discrepancies in test results for different subjects within and among Boards which raise questions about validity and reliability of the tests. It is recommended that, at the least, grade point (GP) and grade point average (GPA) should be calculated on the basis of standardised scores rather than raw scores. Continuing evaluation and analysis of the tests are needed; particularly, regular item analysis of questions in order to improve validity and reliability of tests.*

**Key Words.** *Secondary Education Public Examinations, Standardisation, Multiple Choice Questions, Creative Questions, Bangladesh Secondary Education.*

## 1. Background

Examinations in Bangladesh at the higher secondary certificate (HSC) level are conducted by 8 examination boards (BISE). Each board covers the same range of subjects, but currently sets its own question papers. In most subjects there are two papers or parts – one described as Creative Questions (CQ) for which each question has 4 sub-parts leading to a maximum of 10 marks. There is a choice with more questions than students need to answer. Marks are awarded to the 4 sub-parts based on a hierarchical model derived from Bloom’s taxonomy of learning domains designed to encourage higher order learning of analysis, evaluation and application of knowledge, rather than only knowing and recalling facts. The CQ papers are marked by examiners who are expected to follow a marking scheme created by the question setters. The other paper is a collection of multiple choice questions (MCQ), also based on the same hierarchical model as for the CQ paper, but seeking to have a wider content coverage. The MCQ papers are marked by machine, the Optical Marking Reader (OMR).

---

\* In addition to the above named, the Bangladesh Examination Development Unit (BEDU) team contributing to the study included: Shamima Akhter, Lipika Rani Saha, Ranjit Kuma Sarkar, Salma Akter, Md. Khalid Hossain., Muhammad Saiful Islam, Muhammad Aslam Khaled, Md. Shamsul Huda, and Mohammad Nasir Uddin.

The marks obtained in each paper are combined (proportionately for a total of 100 for a particular subject). This is then converted to grades based on the following formula: A+ => 80 marks; A = 70-79; A- = 60-69; B = 50-59; C = 40-49; D = 33-39; F = <33. These are recorded as grade points (GP) with A+ =5 and so on, and F=0. Each Board follows the same formula and declares for each student a combined result as a grade point average (GPA).

BEDU, the Bangladesh Examinations Development Unit, as part of its job of improving the examination system and engaging in needed research, has undertaken this analysis. The main purpose of this study is to examine the validity and reliability of the examinations, recognising that examinees' marks for each subject are indicated in terms of a grade point (GP) based on raw scores for each paper without taking account of variability of papers for different subjects set by the various BISE and the reliability of marking by the examiners appointed by each BISE (Ministry of Education, n.d.)

## 2. Study Questions

The report is in three parts which deal with: (a) comparison of scores of the CQ and MCQ papers, (b) effect of applying a standardisation formula on scores, and (c) comparison of percentages of high GPA for selected subjects in three Boards (NFER, n. d.).

This study, therefore, addresses the following research questions:

- i. How suitable is the marking of question papers undertaken by examiners compared to marks obtained by a computerised optical marking record (OMR) across the various BISEs?
- ii. Are the results at HSC level, by each BISE, equivalent? If not, is it possible to ensure equivalency of results from all BISEs?
- iii. Can meaningful grade point and grade point averages be established to ensure an equivalent standard of examination results from all BISEs?

## 3. Comparison of CQ and MCQ Scores

This section compares:

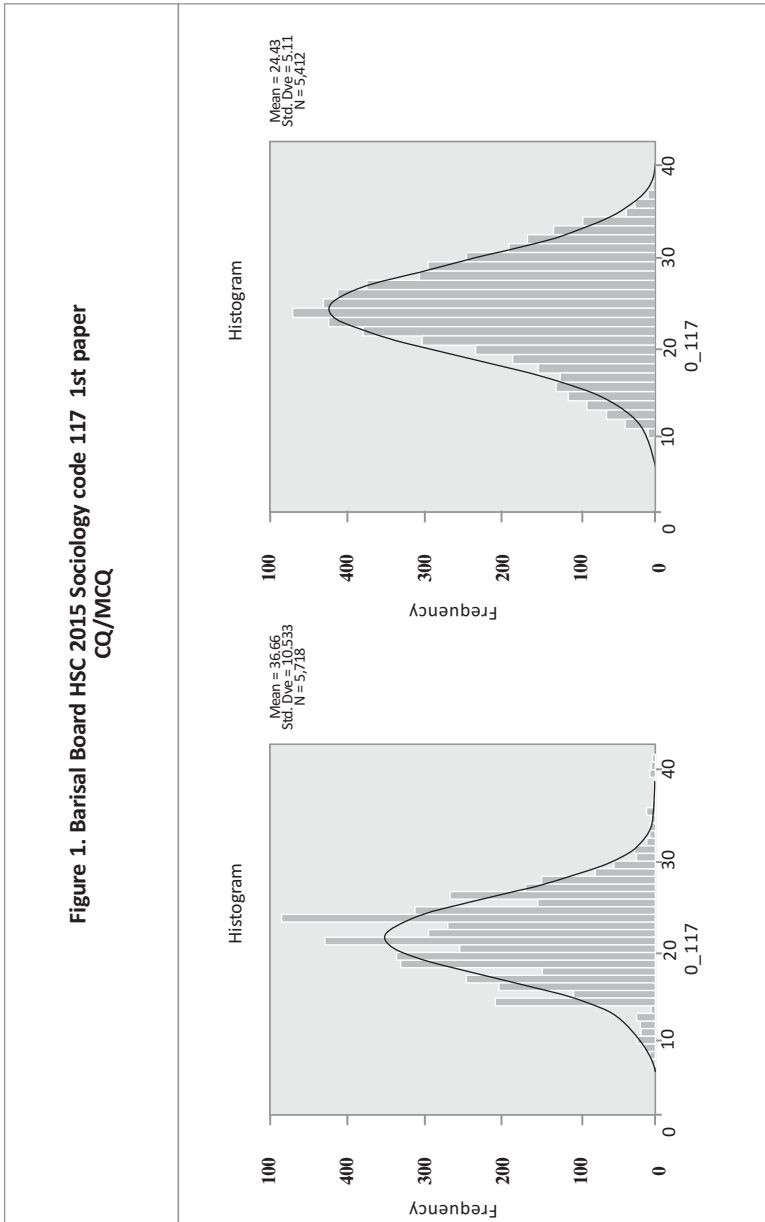
- Mark distributions for a selection of HSC subject papers from three BISEs
- Mean scores per subject based on marks awarded for CQ and MCQ papers, and
- Practical examination marks.

### Discussion of Findings

Marks obtained by students vary greatly with distinct troughs and peaks for CQ papers. The peaks are seen at the point of boundaries for grade point, such as a score of 80 for GP5 or letter grade A. However, marks for MCQ papers are distributed more evenly. MCQ items are machine marked; there is no examiner influence in this case. It can be reasonably argued that machine marking is more objective and, therefore, has eliminated peaks at the grade point boundaries.

Comparison of CQ and MCQ results

The difference can be clearly seen in graphical representation (Figure 1). The graph for CQ and MCQ marks distribution for Sociology papers set by Barisal BISE illustrates the point.



The difference in CQ and MCQ marks distribution points to undue and damaging examiner interference that appears to ignore the marking scheme. The effect is somewhat mitigated when the CQ marks and MCQ marks are combined (since there is no peaks and troughs in MCQ marks) to arrive at a total that is converted to grades.

There is a wide range of mean scores across Boards for the same subject, as well as across subjects for both CQ and MCQ papers.

The following table shows the range of CQ mean scores per subject across Boards (expressed as a percentage), plus the range of mean scores across subjects and across Boards.

**Table 1: Range of mean scores for CQ papers of 8 subjects in three BISEs**

| Subject Code       | Subject       | Barisal (B)<br>Mean %    | Comilla (C)<br>Mean %    | Sylhet (S)<br>Mean %     | Range of<br>Mean Scores |
|--------------------|---------------|--------------------------|--------------------------|--------------------------|-------------------------|
| 117                | Sociology I   | 61.1                     | 46.4                     | 55.7                     | 46.4 – 61.1<br>C B      |
| 118                | Sociology II  | 68.7                     | 52.7                     | 58.6                     | 52.7 – 68.7<br>C B      |
| 123                | Psychology I  | 69.9                     | 64.2                     | 41.5                     | 41.5 – 69.9<br>S B      |
| 124                | Psychology II | 64.5                     | 60.3                     | 34.8                     | 34.8 – 64.5<br>S B      |
| 176                | Chemistry I   | 57.0                     | 51.7                     | 60.2                     | 51.7 – 60.2<br>C S      |
| 177                | Chemistry II  | 55.5                     | 53.9                     | 55.8                     | 53.9 – 55.8<br>C S      |
| 178                | Biology I     | 54.7                     | 52.8                     | 64.0                     | 52.8 – 64.0<br>C S      |
| 179                | Biology II    | 62.4                     | 52.5                     | 65.9                     | 52.5 – 65.9<br>C S      |
| 253                | Accounting I  | 48.2                     | 50.4                     | 48.6                     | 48.2 – 50.4<br>B C      |
| 254                | Accounting II | 62.7                     | 47.6                     | 50.9                     | 47.6 – 62.7<br>C B      |
| 265                | Maths I       | 66.9                     | 59.7                     | 51.5                     | 51.7 – 66.9<br>S B      |
| 266                | Maths II      | 68.9                     | 54.9                     | 58.3                     | 54.9 – 68.9<br>C B      |
| 269                | Civics I      | 60.3                     | 49.5                     | 59.3                     | 49.5 – 60.3<br>C B      |
| 270                | Civics II     | 51.3                     | 50.3                     | 54.9                     | 50.3 – 54.9<br>C S      |
| 304                | History I     | 58.1                     | 49.5                     | 55.2                     | 49.5 – 58.1<br>C B      |
| 305                | History 2     | 63.0                     | 53.0                     | 53.4                     | 53.0 – 63.0<br>C B      |
| Range<br>(subject) |               | 48.2 – 68.9<br>(253 266) | 47.6 – 59.7<br>(254 265) | 48.6 – 65.9<br>(253 179) |                         |

The table shows that there is a wide discrepancy for most subjects for mean scores within each Board and between different Boards for the same subject. In some cases the mean percentage score is quite high indicating an easy paper (leading to poor discrimination between the very able and not so able examinees).

It also points to poor validity of the question papers in maintaining a meaningful and uniform standard of results and in setting suitable questions, based on the CQ model (as indicated by the low standard deviation of mark distribution). It also suggests poor reliability of the marking, raising marks for some candidates while penalising others.

#### Comparison of Mean Scores

Table 2 considers the mean percentage scores per Board per subject for the MCQ papers, expressed in percentages of maximum possible mark.

**Table 2: Range of mean scores for MCQ papers of 8 subjects in three BISEs**

| Subject Code | Subject       | Barisal (B) Mean %    | Comilla (C) Mean %    | Sylhet (S) Mean %     | Range of Mean Scores |
|--------------|---------------|-----------------------|-----------------------|-----------------------|----------------------|
| 117          | Sociology I   | 61.1                  | 49.2                  | 51.4                  | 49.2- 61.1<br>C B    |
| 118          | Sociology II  | 65.8                  | 51.8                  | 47.8                  | 47.8- 65.8<br>S B    |
| 176          | Chemistry I   | 74.9                  | 62.9                  | 52.9                  | 52.9- 74.9<br>S B    |
| 177          | Chemistry II  | 73.6                  | 63.5                  | 56.5                  | 56.5- 73.6<br>S B    |
| 178          | Biology I     | 66.2                  | 66.0                  | 56.7                  | 56.7- 66.2<br>S B    |
| 179          | Biology II    | 72.3                  | 63.3                  | 58.8                  | 58.8- 72.3<br>S B    |
| 253          | Accounting I  | 71.1                  | 50.6                  | 55.9                  | 55.9- 71.1<br>S B    |
| 254          | Accounting II | 70.5                  | 51.7                  | 47.7                  | 47.7- 80.6<br>S B    |
| 269          | Civics I      | 58.9                  | 58.3                  | 59.6                  | 58.3 59.6<br>C S     |
| 270          | Civics II     | 66.2                  | 71.1                  | 74.3                  | 66.2- 74.3<br>B S    |
| 304          | History I     | 71.2                  | 64.5                  | 66.4                  | 64.5- 71.2<br>C B    |
| 305          | History II    | 59.1                  | 53.3                  | 56.2                  | 53.3- 59.1<br>C B    |
| Range        |               | 58.9- 74.9<br>269 176 | 49.2- 71.1<br>117 270 | 47.7- 74.3<br>254 270 |                      |

Table 2 illustrates, similar to CQ papers, that there is also wide discrepancy between mean scores per Board for the same subject and between subjects. The results again point to poor validity of the test items in maintaining a meaningful standard, based on the model intended, as indicated by the low standard deviation for scores..

A comparison of means for CQ and MCQ papers may show the level of difficulty of the two types of questions. However, results show no general pattern across subjects and Boards. .

A more detailed comparison of MCQ and CQ scores means is shown in Table 3.

**Table 3: Comparison of the range of mean scores for CQ and MCQ papers**

| Subject Code | Subject       | Range of CQ Mean Scores | Range of MCQ Mean scores |
|--------------|---------------|-------------------------|--------------------------|
| 117          | Sociology I   | 46.4 – 61.1<br>C B      | 49.2 -61.1<br>C B        |
| 118          | Sociology II  | 52.7 – 68.7<br>C B      | 47.8 - 65.8<br>S B       |
| 123          | Psychology I  | 41.5 – 69.9<br>S B      | NA                       |
| 124          | Psychology II | 34.8 – 64.5<br>S B      | NA                       |
| 176          | Chemistry I   | 51.7 – 60.2<br>C S      | 52.9 – 74.9<br>S B       |
| 177          | Chemistry II  | 53.9 – 55.8<br>C S      | 56.5 – 73.6<br>S B       |
| 178          | Biology I     | 52.8 – 64.0<br>C S      | 56.7 – 66.2<br>S B       |
| 179          | Biology II    | 52.5 - 65.9<br>C S      | 58.8 – 72.3<br>S B       |
| 253          | Accounting I  | 48.2 – 50.4<br>B C      | 55.9 – 71.1<br>S B       |
| 254          | Accounting II | 47.6 – 62.7<br>C B      | 47.7 – 80.6<br>S B       |
| 265          | Maths I       | 51.7 – 66.9<br>S B      | NA                       |
| 266          | Maths II      | 54.9 – 68.9<br>C B      | NA                       |
| 269          | Civics I      | 49.5 – 60.3<br>C B      | 58.3 – 59.6<br>C S       |
| 270          | Civics II     | 50.3 – 54.9<br>C S      | 66.2 – 74.3<br>B S       |
| 304          | History I     | 49.5 – 58.1<br>C B      | 64.5 – 71.2<br>C B       |
| 305          | History 2     | 53.0 – 63.0<br>C B      | 53.3 – 59.1<br>C B       |

*Note: NA – No MCQ tests used.*

The MCQ mean score is often higher than the CQ score, as seen in Table 3. The MCQ mean score is excessively high for some subjects in some Boards (This is in contrast with the SSC results where the MCQ mean score is more often lower than the CQ mean score, not shown in this paper.) Especially high mean scores suggest the MCQ papers may be poorly set and not following a suitable assessment model.

It is necessary to undertake item analysis for each paper to have a better understanding of validity and reliability across subjects and across Boards. This is a task to which BEDU needs to give attention and develop its capacity for this purpose.

### Practical examination Scores

Practical examinations are conducted for certain subjects as part of the public examination, where applicable. The practical examination scores all cluster around the maximum possible mark. The practical examination results, therefore, are totally unreliable as a measure of learning achievement. All the practical mark does is to boost the mean overall score for each student taking that subject, irrespective of the Board. The mean scores are shown in Table 4.

**Table 4: Mean scores for practical examinations in 8 papers for 3 BISEs.**

| Subject Code | Examination Board |                  |                 |
|--------------|-------------------|------------------|-----------------|
|              | Barisal Mean (%)  | Comilla Mean (%) | Sylhet Mean (%) |
| 123          | 24.74 (99.0 %)    | 24.59 (98.4 %)   | 23.33 (93.3 %)  |
| 124          | 24.74 (99.0 %)    | 24.61 (98.4 %)   | 23.73 (94.9 %)  |
| 176          | 23.75 (95.0 %)    | 23.77 (95.1 %)   | 22.82 (91.3 %)  |
| 177          | 23.79 (95.2 %)    | 23.99 (96.0 %)   | 23.02 (92.1 %)  |
| 178          | 24.12 (96.5 %)    | 24.02 (96.1 %)   | 22.92 (91.7 %)  |
| 179          | 24.12 (96.5 %)    | 24.17 (96.7 %)   | 23.13 (92.5 %)  |
| 265          | 24.40 (97.6 %)    | 24.17 (96.7 %)   | 23.25 (93.0 %)  |
| 266          | 24.41 (97.6 %)    | 24.31 (97.2 %)   | 23.31 (93.2 %)  |

The mean scores of all practical examinations are exceptionally high, having a very low standard deviation. There is virtually no discrimination seen among examinees. It denotes that teachers are seemingly more interested in the examination results than in educating their students in meaningful ways. The practical scores boosting the overall scores give an undue advantage to the students having practical subjects, especially, the science stream students.

### 4. Standardization of examination results

One of the objectives of the study is to ascertain whether the results at HSC level from the different BISEs maintain a comparable standard across Boards and subjects. This study, conducted by BEDU, uses relevant data from Barisal, Comilla and Sylhet Boards.

It is noteworthy that the results at HSC level are published on the basis of the raw score overlooking the nature and pattern of the scores of the respective subjects. This leads to non-comparable results across boards, subjects and years. The average marks obtained in the subjects of the Science stream having the advantage of practical marks are higher than that of Commerce and Humanities subjects. As a consequence, fewer numbers of students from the latter streams obtain GPA 5 compared to the science students. The commerce and humanities students are thus placed at a disadvantaged position when their performance is assessed for higher education admission, job markets and other purposes (NFER, nd.)

This section focuses on publishing the results which would be comparable across boards, years and subjects. For this purpose, two steps were necessary:

- Collecting raw scores from HSC examination subjects in order to use these to create standardised scores.
- Comparing raw scores and standardized scores for selected subjects from three Boards.

The research question for this section, therefore, is: Are the results at HSC level of different BISEs of an equivalent standard? If not, is it possible to establish comparable examination results from all BISEs?

### **Methodology**

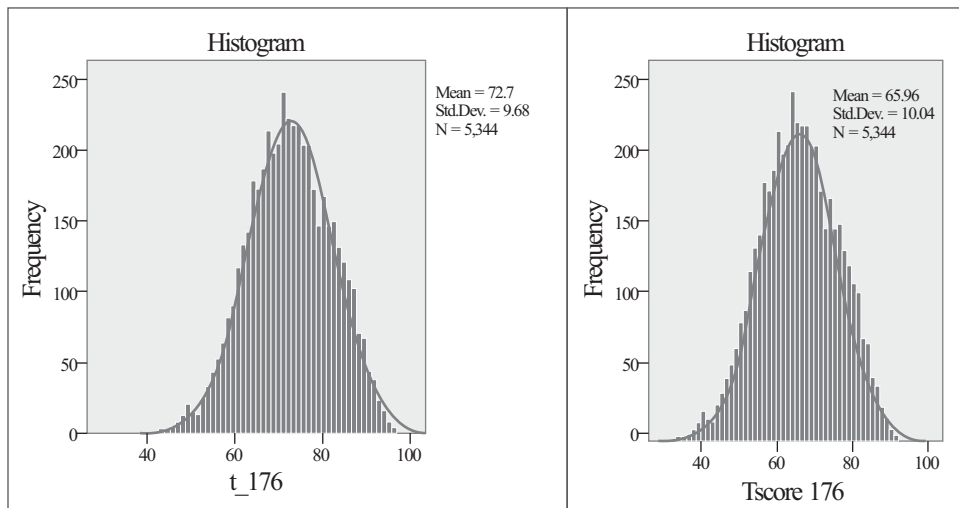
- HSC examination results for 2015 were obtained from three BISEs.
- The results were analysed to determine the distribution of examinees' marks in the different subject papers.
- The distribution of marks from total raw score, adding CQ and MCQ scores, and the distribution of marks from standardized scores were derived.
- The distribution of marks based on raw scores and standardised scores were checked against a normal distribution curve, which is predicted to represent the expected mark distribution for the large numbers taking each examination.
- The resulting graphs were analysed in order to derive answers to the research question posed.

### **Discussion of Findings -- Comparison of raw score and standardized score**

Graphs of raw score and standardized score of Barisal in Chemistry (Figure 2.1) show an almost normal distribution in contrast to what was seen above (Figure 1) for Sociology in the same Board. This suggests a difference in scores and marking pattern for science and other subjects. There may also be differences in standard and criteria in constructing questions.



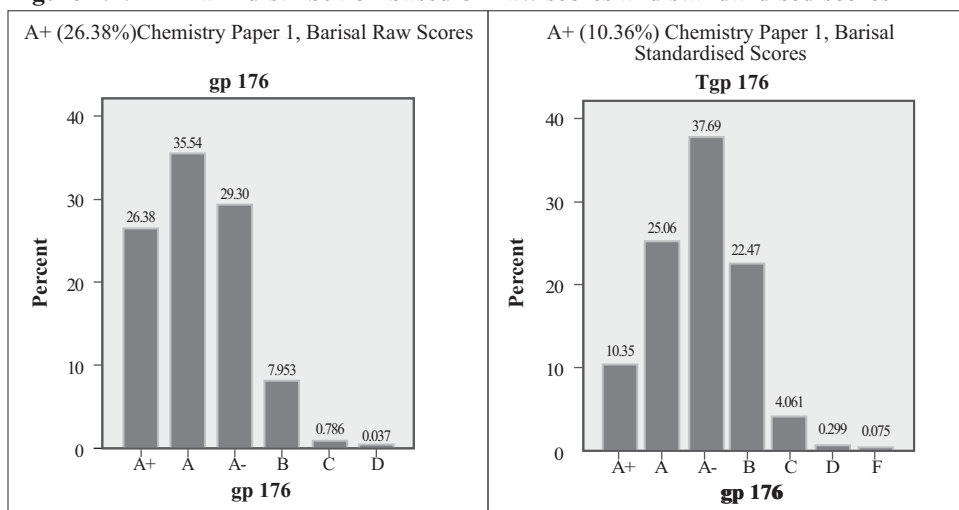
**Figure 2.1: Barisal Board HSC 2015 Chemistry Code 177 2<sup>nd</sup> paper.**



Note the graphs for the standardized scores are similar in appearance to those from the unstandardized (raw) scores. This is not so for all subjects and all Boards, which suggests problems in the quality of setting questions and marking standard.

The percentages of GP A+ across boards are different for different subjects and for different Boards as determined from the raw scores (not shown). On the other hand, the percentage of GP A+ across subjects and Boards are likely to be similar for standardized score (approximately about 10%). (Figure -2.2 for Chemistry paper, Barisal)

**Figure 2.2: A+ mark distribution based on raw scores and standardised scores**

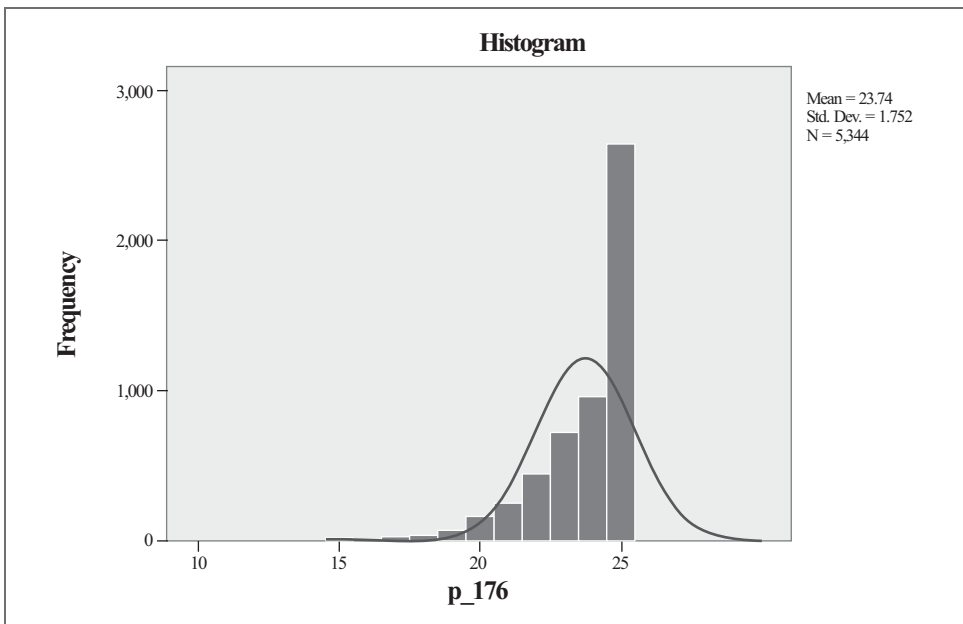


In some cases, the total raw score graphs show peaks rather than a smooth distribution of marks. The peaks occur from marking irregularities by examiners, which are not fully eliminated when marks from the CQ and MCQ papers are combined.

When both the graphs of raw and standardized scores do not follow a normal distribution, this indicates weaknesses in setting the question paper and perhaps marking the papers by examiners.

Figure 2.3 showing practical examination results peaking at the point of maximum score indicates faulty and rather useless testing and grading.

**Figure 2.3: Chemistry, Code 176, Practical Test Scores, Barisal Board**



## 5. Making Grade Point Averages (GPA) equivalent across Boards

This section explores whether grade points and grade point averages for different Boards can be equivalent.

### Methodology

- HSC examination results for 2015 were obtained from different BISEs.
- The results were analysed to determine the distribution of GP for student scores in different subject papers. Graphs (histograms) were plotted for the percentages of GP (Grade Point) students attained in each examination paper for total raw score and for standardized score.

- The distribution of GP based on raw scores and standardized score are compared.
- The distribution of GPs were checked against a normal distribution, predicted to represent the expected GP distribution for the large student numbers taking each examination.

### Findings

Table 5 shows the combined raw scores for CQ and MCQ papers per subject across boards, plus the range of mean scores by subject and by Board.

**Table 5: Combined CQ and MCQ raw score means and their range**

| Subject Code                                  | Examination Board        |                          |                          |                         |
|---|--------------------------|--------------------------|--------------------------|-------------------------|
|   | Barisal (B)<br>Mean %    | Comilla (C)<br>Mean %    | Sylhet (S)<br>Mean %     | Range of<br>Mean Scores |
| 117   | 59.8                     | 46.9                     | 53.5                     | 46.9 – 59.8<br>C B      |
| 118   | 66.1                     | 51.4                     | 54.4                     | 51.4 – 66.1<br>C B      |
| 176   | 72.7                     | 66.4                     | 65.3                     | 65.3 – 72.7<br>S B      |
| 177   | 71.6                     | 67.6                     | 65.0                     | 65.0 – 71.6<br>S B      |
| 178   | 68.4                     | 67.0                     | 67.5                     | 67.0 – 68.4<br>C B      |
| 179   | 73.5                     | 66.6                     | 69.1                     | 66.6 – 73.5<br>C B      |
| 253   | 56.3                     | 49.4                     | 51.1                     | 49.4 – 56.3<br>C B      |
| 254   | 64.8                     | 48.1                     | 49.3                     | 48.1 – 64.8<br>C B      |
| 269   | 59.7                     | 53.1                     | 59.4                     | 53.1 – 59.7<br>C B      |
| 270   | 57.3                     | 58.6                     | 63.0                     | 57.3 – 63.0<br>B S      |
| 304   | 63.4                     | 55.5                     | 59.7                     | 55.5- 63.4<br>C B       |
| 305   | 61.5                     | 53.1                     | 54.5                     | 53.1- 61.5<br>C B       |
| Range of<br>mean scores<br>across<br>subjects | 56.3 – 73.5<br>(253 179) | 46.9 – 67.6<br>(117 177) | 49.3 – 69.1<br>(254 179) |                         |

Table 5 shows:

1. There is a wide discrepancy between mean scores based on raw student marks for different Boards for the same subject in many cases.
2. There is a wide discrepancy between mean scores based on raw marks per subject for the same board in many cases.
3. The lower mean scores were for humanities (e.g., Sociology, 1st paper), while science mean scores were higher (e.g., Biology 2nd paper); the mean of all science subjects are higher than for other streams.
4. Mean scores for the Comilla Board were lower on average, while mean scores for Barisal Board were higher, showing another variability of mean student marks across Boards.

The results point to poor validity of the question papers in terms of maintaining a meaningful and uniform standard in setting suitable questions.

Table 6 shows the standardized scores per subject across Boards, plus the range of mean scores by subjects and by Boards.

Table 6 illustrates the effect of standardization of scores. The results based on standardised marks point out that through standardization the variation of mean scores across the subjects and across the Boards can be minimised.

The findings presented in Table 7 show the percentage of A+ distribution for raw GP and standardized GP per subject across Boards. (Also see Figure 2.2 above.)

Analysis of Table 7 data indicates that on applying standardization, the percentage of A+ is similar and comes close to 10% for each subject across each Board. It also shows:

1. For raw scores, there is a very high discrepancy in the percentage of A+ within the same subject across Boards.
2. For raw scores, there is a very high discrepancy in the percentage of A+ within the same Board across subjects.
3. The percentage of A+ tends to be lower for humanities subjects but much higher for science subjects.

### **Discussion - Comparison of raw score and standardized score distribution**

The results highlight a lack of common standards among the Examination Boards arising from setting questions and marking papers at different standards by each Board. Applying standardisation to raw scores and calculating and publishing results based on standardization of scores can contribute to establishing common standards in results across academic streams and Boards.

It is important to note that the benchmark of 10% for A+ was an assumption made by BEDU and was not derived from any psychometric analysis of measurement. It was chosen to represent a meaningful standard, which could be used to equate grade points across Boards, based on a normal distribution assumption. In practice, however, exactly 10% was not

**Table 6: Mean standardised scores for subjects across Boards and their range.**

| Subject Code                         | Examination Board |                |               |                      |
|--------------------------------------|-------------------|----------------|---------------|----------------------|
|                                      | Barisal Mean %    | Comilla Mean % | Sylhet Mean % | Range of Mean Scores |
| 117                                  | 64.0              | 63.8           | 64.0          | 63.8 – 64.0<br>C B/S |
| 118                                  | 67.0              | 62.9           | 65.0          | 62.9 – 67.0<br>C B   |
| 176                                  | 66.0              | 65.0           | 64.0          | 64.0 – 66.0<br>S B   |
| 177                                  | 66.0              | 64.0           | 62.5          | 62.5 – 66.0<br>S B   |
| 178                                  | 64.0              | 65.0           | 65.0          | 64.0 – 65.0<br>B C/S |
| 179                                  | 64.0              | 64.0           | 63.5          | 63.5 – 64.0<br>S B/C |
| 253                                  | 60.0              | 61.5           | 58.0          | 58.0 – 61.5<br>S C   |
| 254                                  | 60.0              | 60.0           | 59.0          | 59.0 – 60.0<br>S B/C |
| 269                                  | 65.8              | 63.0           | 64.0          | 63.0 – 65.8<br>C B   |
| 270                                  | 63.0              | 61.9           | 63.0          | 61.9 – 63.0<br>C B/S |
| 304                                  | 59.0              | 58.0           | 59.7          | 58.0 – 59.7<br>C S   |
| 305                                  | 65.0              | 62.1           | 64.0          | 62.1 – 65.0<br>C B   |
| Range of mean scores across subjects | 59.0 – 67.3       | 50.0 – 68.9    | 56.0 – 68.0   |                      |

always possible to achieve. As the 10% benchmark was arbitrary, the selected A+ percentage could be modified to relate to other grade point average cut-off still approximating a normal distribution. The A+ % thus represents a selected standard, but once chosen it needs to be applied across Boards and across examination years without changing it frequently. Thus, comparison could be made per subject, across different cohorts of examinees and over time. One important criterion of judging the quality of the examination would be approximation of results to a normal distribution.

**Table 7: Percentage of A+ marks for Boards based on raw and standardised scores**

| Subject Code        | A+ %           |                     | A+ %           |                     | A+ %           |                     | Range (Board to Board)                  |
|---------------------|----------------|---------------------|----------------|---------------------|----------------|---------------------|---|
|                     | Barisal raw    | Barisal standard    | Comilla raw    | Comilla standard    | Sylhet raw     | Sylhet standard     |   |
| 117                 | 4.8            | 9.9                 | 0.3            | 10.0                | 1.3            | 9.8                 | 0.3–4.8 (raw)<br>9.8–10.0 (standard)    |
| 118                 | 8.6            | 11.2                | 0.9            | 10.5                | 0.7            | 11.3                | 0.7–8.6 (raw)<br>10.5–11.3(standard)    |
| 123                 | 56.0           | 11.4                | 33.1           | 9.9                 | 2.3            | 9.8                 | 2.3–56.0 (raw)<br>9.8–11.4(standard)    |
| 124                 | 36.9           | 10.6                | 28.2           | 11.5                | 2.6            | 10.0                | 2.6–36.9 (raw)<br>10.0–11.5(standard)   |
| 176                 | 26.4           | 10.4                | 14.4           | 10.5                | 12.2           | 10.4                | 12.2–26.4 (raw)<br>10.4–10.5(standard)  |
| 177                 | 25.4           | 10.0                | 21.2           | 11.0                | 13.9           | 10.8                | 13.9 –25.4 (raw)<br>10.0–11.0(standard) |
| 178                 | 14.0           | 10.50               | 18.4           | 11.9                | 18.5           | 11.2                | 14.0–18.5 (raw)<br>10.5–11.9(standard)  |
| 179                 | 29.4           | 10.1                | 17.2           | 11.6                | 21.9           | 9.9                 | 17.2–29.4(raw)<br>9.9–11.6(standard)    |
| 253                 | 2.9            | 8.3                 | 2.5            | 11.9                | 4.8            | 9.4                 | 2.5–4.8 (raw)<br>8.3 –11.9(standard)    |
| 254                 | 17.8           | 9.8                 | 2.1            | 10.4                | 3.9            | 10.4                | 2.1–17.8 (raw)<br>9.8–10.4(standard)    |
| 269                 | 3.0            | 10.8                | 1.4            | 9.5                 | 3.5            | 10.1                | 1.4 – 3.5 (raw)<br>9.5 – 10.8(standard) |
| 270                 | 3.8            | 9.7                 | 5.3            | 9.9                 | 9.6            | 10.1                | 3.8 – 9.9 (raw)<br>9.6 –10.0(standard)  |
| 304                 | 12.1           | 9.7                 | 6.9            | 9.6                 | 10.3           | 10.3                | 6.9 -12.1(raw)<br>9.6 -10.3(standard)   |
| 305                 | 3.2            | 9.0                 | 1.5            | 9.7                 | 1.5            | 9.9                 | 1.5 -3,2(raw)<br>9.0-9.9(standard)      |
| Range (sub.to sub.) | 3.0-56.0 (raw) | 8.3-11.4 (standard) | 0.3-33.1 (raw) | 9.5-11.9 (standard) | 0.7-21.9 (raw) | 9.4-11.3 (standard) |   |

## 6. Conclusions

The analysis of findings regarding the research questions above leads to several conclusions. Parallels can be seen in these conclusions to other developing countries in the region which are also addressing similar issues (Hill, 2013).

- i. Marking by examiners of CQ papers clearly shows subjectivity which puts reliability of the tests under question. Whether and how the marking scheme is being followed

by examiners and how it is understood by examiners need to be given attention and assessed.

- ii. Marking by machine (OMR) is superior in terms of reliability. However, the validity of MCQ papers in respect of measuring competence and discriminating student performance need measures to ensure appropriate content and quality of the test items.
- iii. There is no inherent superiority of CQ over MCQ or vice versa; they serve different test objectives. Greater reliability in marking and possibility of covering a broader range of content exist with MCQ; whereas, higher order learning domain may be more easily tested with CQ. However, all this depends on ensuring the quality of question setting of both kinds and minimising subjectivity in marking CQ papers.
- iv. Marking of practical examination is seriously flawed and does not add any value to measurement of student learning. In fact, it distorts results by giving undue advantage to students of the science stream.
- v. Rigorous individual item analysis is necessary to pass definitive judgement on validity and reliability of tests or improve their effectiveness as measurement tools.
- vi. A method of standardization is needed to establish equivalence of marks across subjects and Boards and over time. Without standardization students are rewarded or penalized depending on subjects for which they take the test, and the Board that gives the test, rather than on the basis of students' real performance.
- vii. By means of standardization it is possible to establish meaningful grade point distributions which can be similar across subjects and across Boards. By such means, the very able students are not penalized, which is the case if the percentage of examinees obtaining the grade point A+ is allowed to become excessive.

## 7. Recommendations

The following recommendations are offered on the basis of the findings and conclusions.

- i. Standardised marks should be used for calculating grade points and grade points averages to determine student results. A further logical step would be to have common question papers for all boards and establishing and following common standards and process for setting questions and marking tests.
- ii. Given the comparative advantages of both CQ and MCQ, it would be reasonable to give equal weight in tests to both, with due attention to maintaining quality in setting and marking questions. Recently, a decision has been taken to reduce the proportion of MCQ items and increase CQ items in SSC and HSC examinations, because of dissatisfaction expressed about the quality of MCQ items.
- iii. Preparation of test items (both CQ and MCQ) for large scale public examination is a highly technical job; the necessary capacity has to be built jointly by the examination

boards, a large repository of questions of both MCQ and CQ type have to be built up for all subjects, and regular item analysis should be undertaken to maintain and improve quality, validity and reliability of questions.

- iv. Options should be considered for dysfunctional practical parts of the examinations, such as, describing processes of experiments in writing or gradually introducing computer-based simulation of experiments.
- v. Apart from attending to logistical and procedural issues of public examinations and ensuring security and confidentiality of tests, greater attention should be given to reducing pressure and burden on students, by reducing total duration of tests, length of each test, and the possibility of what can be taken out of the domain of public examinations and shifted to school-based evaluation.

### References

- Ministry of Education, Education Board Computer Center (n. d).  
<http://www.educationboard.gov.bd/computer/>
- NFER (n.d). "What do test scores mean – Standardised Scores and Percentile Ranks." National Foundation for Educational Research in England and Wales, Centre for Assessment. <https://www.nfer.ac.uk/research/>
- Hill, P. (2013). Examination Systems. Asia Pacific Secondary Education System Review Series 1. Bagkok: UNESCO Asia and Pacific Regional Bureau of Education.