A Web-based System for Assurance of Assessment Quality and Diagnosis of Student Performance

Fung Tze-ho, Eric
thfung@hkeaa.edu.hk
Hong Kong Examinations and Assessment Authority

Abstract

After marking all student responses in an exam conducted in a school, different statistics and graphs could be compiled to diagnose overall student performance, and item and paper quality. However, this involves various formulae, algorithms and complicated compilations. To facilitate the preparation of various statistics and graphs for such a diagnosis, Hong Kong Examinations and Assessment Authority has developed a web-based system, called Assessment Quality-assurance Platform (AQP) for schools. It is expected that AQP could inform school teachers about achievements of students and quality of different items and papers. Accordingly, AQP provides three perspectives of analysis; namely (i) Paper; (ii) Item and (iii) Student. With these analysis results, teachers could have a direction to improve the paper and item design, and strengthen the weakness of their students.

Besides, AQP provides utilities for school teachers to form various teams. Being members of a team, they could study results of an analysis conducted by the other member and share their own views. Moreover, school teachers could store quality items in the system as an item bank. A test could then be automatically generated by drawing items from item banks according to a set of criteria given by school teachers.

Keywords: Web-based system, Assessment, Quality

Introduction

The Assessment Quality-assurance Platform (AQP) for schools is a web-based computer system designed to enable primary and secondary school teachers to perform quality checking of their own examination papers. The goal of developing the AQP for schools is to provide a software tool for the education community in order to enhance the quality of assessment and at the same time encourage the practice of assessment for learning. The AQP for schools provides teachers with various statistical indicators that can help them gain a better understanding of their students' performance, item and paper quality. With these pieces of information, the system helps them improve their assessment design and deliver appropriate and just in-time guidance to students.

AQP provides three perspectives of analysis; namely (i) Paper; (ii) Item and (iii) Student. With these analysis results, teachers could have a direction to improve the paper and item design, and strengthen the weakness of their students. Besides, AQP provides utilities for school teachers to form various teams. Being members of a team, they could view and study results of an analysis conducted by other members, and share their own views and opinions.

Moreover, school teachers could store quality items in an item bank via utilities provided by the system. A test could then be automatically generated by drawing items from the relevant item banks according to a set of criteria given by school teachers. In the following, we briefly outline the functions available in AQP, together with some relevant screen captures.

Paper Level Analysis

In the paper level analysis, the statistics are mainly divided into five sections, namely:

A. *Figures of basic statistics*: Basic paper statistics such as mean, standard deviation of paper marks, the corresponding 25%, 50%, 75% percentiles and paper reliability are displayed.

B. *Paper correlation coefficient*: Figures of correlation coefficients across different papers within the same examination are compiled.

C. *Mark distribution*: Paper marks' distribution is illustrated in both boxplot and histogram

D. *Paper marks equating*: The graph for equating marks of various papers is shown, which is derived from calibration of item difficulties by Item Response Theory (IRT). The graph shows expected marks of different papers for any specific values of student ability.

E. *Summary statistics*: Summary report will be generated for users, which provides an overall picture of profile of various item statistics for papers concerned.

From the information provided, it is supposed that users could capture the following:

- Overall student performance in various papers
- Display of outliers (extremely poor and extremely good students)
- Paper reliability
- Relative difficulties of various papers, and
- Overall matching between student ability and difficulty of each paper

Paper analysis results can be exported in MS Office Excel format by using the export function.

Figure 1a: Paper level analysis: Basic statistics and boxplot of paper marks

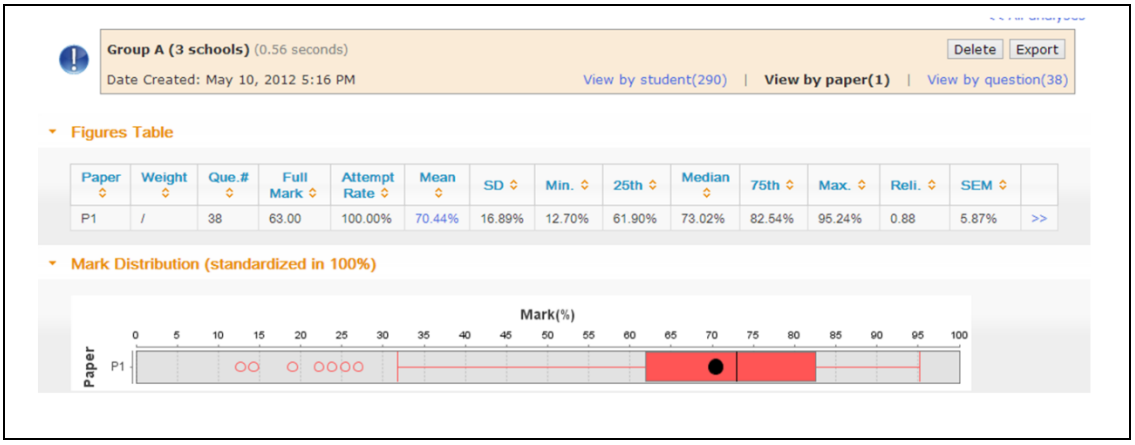


Figure 1b: Paper level analysis: Histogram of paper marks and graphs from Item Response Theory (IRT)

Item Level Analysis

In the item level analysis, the statistics are mainly divided into four sections, namely:

A. *Figures of basic statistics*: Basic item statistics such as mean, standard deviation of item marks and the corresponding 25%, 50%, 75% percentiles and statistics on item difficulty, item discrimination, and item fit are displayed.

B. *Mark distributions for all items*: The mark distributions of all items are graphically displayed; e.g., boxplots and pie charts. Besides, for each MC item counts for each option are provided by student performance level, in term of marks of the whole paper.

C. *Item person map*: The map shows the calibrated item difficulties and student abilities on the same scale based on IRT.

D. *Boxplots over student dimensions*: Boxplots over student dimensions (e.g. student class and gender) are generated so that comparison across the student dimension concerned could be achieved. Summary statistics of each boxplot, such as mean, min, 25% percentile, median, 75% percentile, max are also provided.

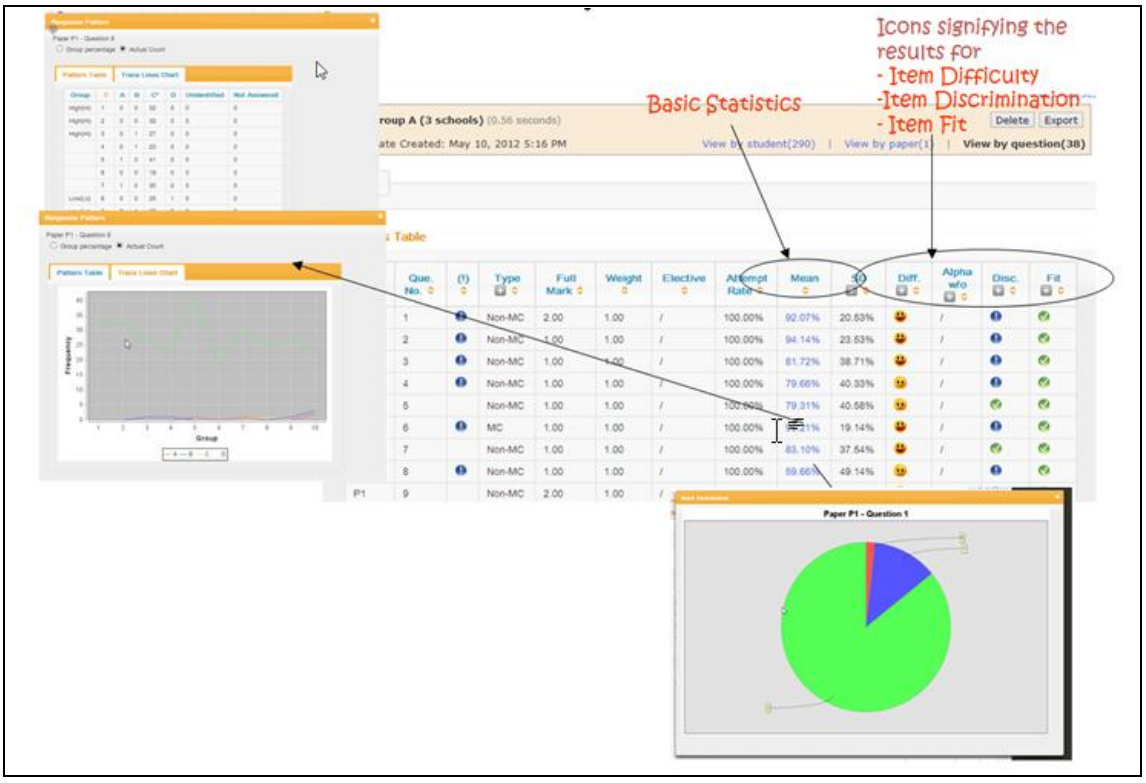
It should be noted that in item level analysis, users may change the filter preference with respect to paper, item category and item level (*Advanced, Intermediate, Fundamental*).

From the information provided, it is supposed that users could capture the following:

- Quality of each individual item, in terms of item difficulty, item discrimination and item fit
- The effectiveness of each option count of a MC item
- Student performance of a non-MC item, in term of mark distribution

Item level analysis results can be exported in MS Office Excel format by using the export function.

Figure 2: Item level analysis: Basic statistics on item quality, MC option counts and distribution of item marks



Student Level Analysis

In the student level analysis, the statistics are divided into six sections, namely:

A. *Figures of basic statistics*: Basic student statistics, such as student subject marks, ranking and student ability based on IRT. Moreover, for each student, his/her actual mark of an item, the corresponding expected mark (derived using IRT), and the mark difference between them are provided.

B. *Item person map*: On the left of the map are students (persons) taking part in the assessment and on the right are items. Persons on the left are arranged from low to high ability level, and items on the right are arranged from easy to difficult level. The ability and difficulty levels are calibrated using Item-Response Theory (IRT).

C. *Student-problem chat (SP chart)*: The S-P chart is a student-by-item (row x column) data matrix in which the students have been arranged in a descending order of their total numbers of questions answered correctly from top to bottom, and the items have been arranged in an ascending order of difficulties, in terms of the total numbers of students who answered the questions concerned correctly; i.e., the student with highest ability is on the top and the easiest item is on the left (in term of numbers of correctness). On the SP-chart, a student curve is shown. The curve is drawn in accordance with students' total numbers of questions answered correctly.

D. *MCI chart*: The MCI chart is useful in the diagnosis of abnormal pattern in student performance. An associated index related to the SP-chart is the Modified Caution Index (MCI). The MCI has a value between 0 and 1. The greater is the value of the MCI, the larger the departure from a normal pattern is. For example, if a student can answer almost all the difficult items correctly but answers most of the easy ones incorrectly, the MCI of the student will be close to 1. When the Modified Caution Index of a student is greater than 0.3, it suggests some abnormality in the response pattern of the student. When the subject marks of a student are above 50% of the corresponding full marks, it suggests good performance of the student. By applying these two rules, the MCI-Chart can be constructed

E. *Boxplots over item dimensions*: Boxplots over item dimensions (i.e. item category and item level) are generated so that comparison across the item dimension concerned could be achieved. Summary statistics such as mean, min, 25% percentile, median, 75% percentile, max are also provided.

F. *Summary statistics*: Summary report is generated for users, which provides an overall picture of the profile of various summary statistics on student

performance by class, with standardized remarks provided.

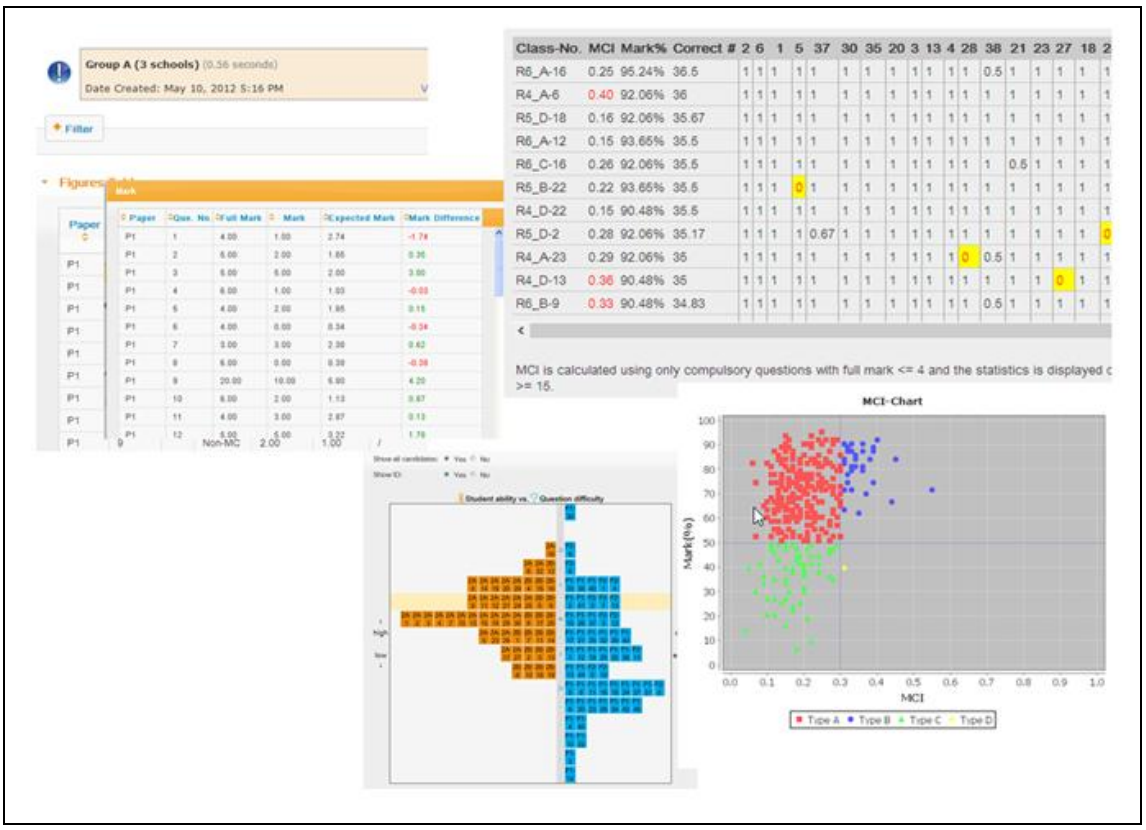
It should be noted that users may change the filter preference with respect to class (or other user-defined grouping) in the student level analysis.

From the information provided, it is supposed that users could capture the following:

- Student overall performance in terms of subject marks, ranking, student ability (from IRT)
- Student performance of each item by comparing the expected mark derived from IRT with the actual mark
- For each student, the relative difficulties of each item, as compared with the student ability concerned
- Identification of a student with odd behavior; i.e., do wrongly in easy questions and do correctly in difficult questions

The student level analysis results can be exported in MS Office Excel format by using the export function.

Figure 3: Student level analysis: Basic statistics on student performance, expected mark vs. actual mark, SP chart, and MCI chart



Item Banking

Users may make use of an item bank to set up and maintain a repository of their own test items for further test development. Users are able to store items in item banks, and various item statistics of these items, each of which is identified by a unique item ID for proper data management. After creating item banks and granting access rights to relevant users, the appropriate users can add, update or delete items in the item banks.

After storing a sufficient number of items in an item bank, users can generate a test by using the utility, Automated Test Assembly, which draws items from a chosen item bank in accordance with a number of user-defined criteria.

Automated Test Assembly

Automatic test assembly is introduced for automatically constructing a test based on items from item banks in accordance with a number of criteria. Such a facility can help reduce the workload of test developers and ensure the quality of tests.

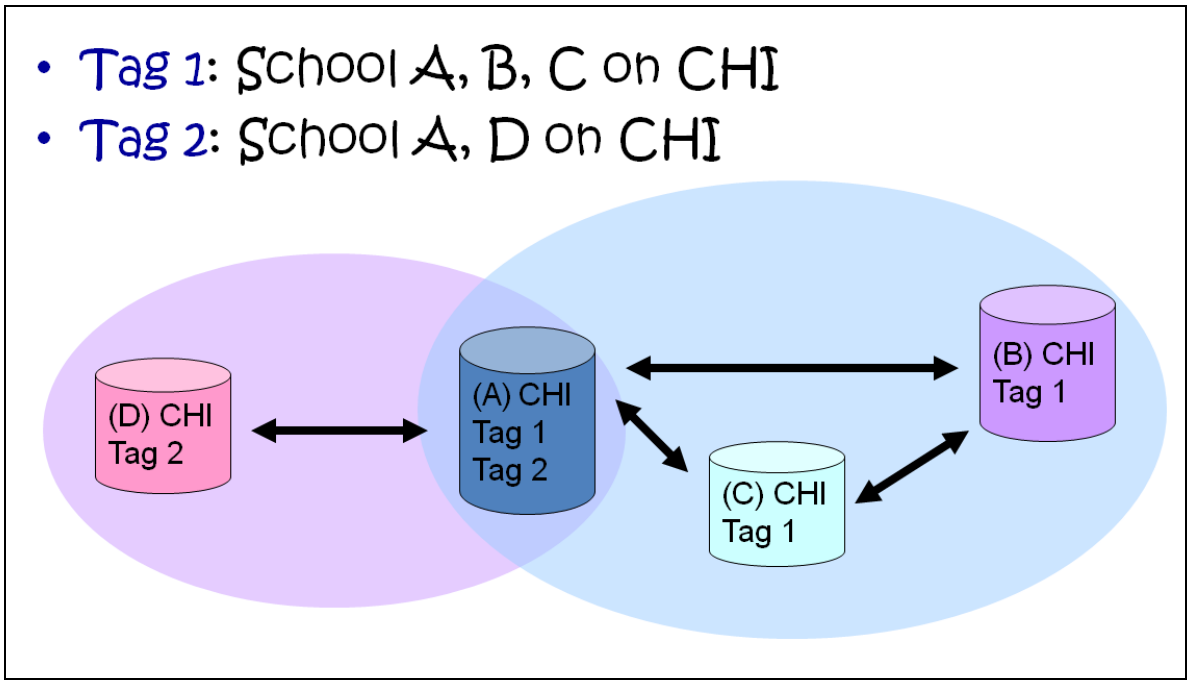
Users are required to provide the following criteria for the selection of items. Only items fulfilling all these criteria will be selected.

- Number of questions:
    - Total number of questions (a compulsory field with the only the option ”=” provided)
    - Number of questions in each category (user-defined categories)
    - Number of questions in each level (fundamental, intermediate, advanced)
    - Number of questions in each question type (MC, Non-MC)
- Specific questions: Users can specify items to be included or excluded by entering the item ID
- Item statistics: Users can select items with item statistics falling within the ranges (Lower bound and/or upper bound) specified. The statistics related with item difficulty, item discrimination, item fit
- Full Mark: Users can specify the range of paper mark (Lower bound and/or upper bound) of the test.
- Optimization Criterion: This is the mandatory field in Automated Test Assembly. Users have to define the mean mark expected in the test. The items will be selected such that the mean mark will be as close to the optimization criterion as possible using integer linear programming algorithms.

Item bank Sharing

Item bank sharing is to create virtual pools of items from different organizations. Under such an arrangement, each organization manages its own items; but could generate a test from items pooled from different organizations. Item bank sharing is achieved by using the notion of share tag. A share tag, in fact, is a unique ID of a virtual pool. Items in item banks from different organizations, which have the same share tag, would be shared with each other. It should be noted that an item bank could have more than one share tag. It implies that items could be consolidated from more than one virtual pool of items when generating a written test.

Figure 4: Item bank sharing: An illustration of an item bank that could have more than one share tag



Project Progress

For the sake of efficiency, school groups are the key audience of AQP promotion for the time being. Due to such an effort, the HKEAA and *The Catholic Education Office* have reached an agreement on launching the AQP to primary and secondary schools of the group. At the beginning, 52 primary schools and 20 secondary schools are involved. The first phase of the project will be launched starting from Early 2014 to the end of August 2014. It is supposed that the renewal of subscription will be on an annual basis, starting from September 2014 onwards. In order to further expand the use of AQP, the HKEAA has met with another school group early this year to discuss the possibility of deployment of AQP at their schools.