INTRODUCTION
Assessment is an important part of curriculum which helps make several decisions for curriculum planning, promotion to the next level and learning needs. During past few decades a lot of research has gone into study psychometric properties of various assessment tools and thus the concepts of reliability and validity of assessments. Multiple Choice Questions (MCQs) have become increasingly popular as a tool of assessment in both under and postgraduate medical education because of being more reliable and valid tool for written assessment. This format of assessment allows to efficiently assess large number of candidates with a wide range of content. In the meantime, various theories were developed to analyze MCQ responses and several psychometric indicators were developed to gauge the performance of an MCQ and the candidates’ cohort. Thus, MCQ analysis based on these psychometric theories has evolved into an important step in feedback, quality assurance and identifying learning needs.

MCQs analysis holds an important role in constructing assessments and basic knowledge of its elements is essentially required by most of the faculty members. Academic faculty members actively involved in developing assessments require know how of various parameters used in analyzing and interpreting MCQ analysis. Classical Test Theory (CTT) and the item response theory have served the measurement community for almost a century. It is easy to apply in many testing situations. Item statistics (item difficulty and item discrimination) are important components of the CTT model. This model collectively considers a pool of examinees and examines their success rate on an item, well known as the p-value which is an indicator of item difficulty. The ability of an item to discriminate between higher ability and lower ability examinees is the item discrimination. Because the faculty often finds it difficult to interpret statistic-heavy reports, it was decided to develop and test a simpler and more user-friendly format for better understanding and meaningful interpretation of assessment results.

METHODOLOGY
A survey about the terms required for interpretation of the report was conducted through a focus group discussion with the course directors to explore about their perceptions about utility of computer-generated MCQs analysis data (based on CTT), understanding of basic psychometric parameters used in these reports, and use of assessment analysis through a questionnaire for feedback to the students and faculty members as a means for closing teaching-learning cycles. It turned out that the majority of directors had limited understanding of this report. Based on this information, the authors devised a strategy where a pre-interpreted report was generated for easy understanding and use by course directors.

This computer generated analysis, with all the parameters mentioned in survey instrument, was sent to
all course directors as a routine. It was found that the knowledge about MCQ related statistics among the faculty was lacking. The computer generated analysis was too complicated for review. There was lack of awareness of MCQ analysis by the faculty.

An interpretive guide based on analysis of MCQ assessments in a simplified format was designed. The simplified MCQ guide comprised of a visual format, descriptive summary of parameters and narrative summary report at the end which was prepared by the examination department.

The visual format consisted of spread of scores, mean score and standard deviation (Figure 1). An analysis of distractors of MCQs with the key is shown in Figure 2. Each bar would represent an MCQ and distractors used were color-coded. The width of each colored bar represented the percentage that distractor was used. The key was also provided in front of each MCQ bar. Descriptive summary of parameters was based on Classical test theory. Test-retest Reliability indicated by Cronbach’s alpha was provided. The computer generated report based on OMR scoring was processed through the software Remark-classic OMR version 2.5. The software used built-in standard formula to calculate Cronbach’s alpha. The data was generated by the examination department and provided to the course directors for review. Discrimination Indices (DI) of MCQs with subcategory of items with negative discrimination index, and Facility Indices (FI) of MCQs with subcategories as easy (FI > 0.70) and hard (FI < 0.30) were reported. Those MCQs with DI < 0.20 were also reported with subcategories of easy and hard. Based on the above mentioned data, a narrative summary was provided addressing explanations and possible remedial measures for students. Poorly performing MCQs were highlighted for review by course coordinators during their modular team meetings for reviewing MCQs for content validity, structure, and language issues. The information was also utilized for finalizing results prior to informing students.

Course directors were asked to present course feedback from students and assessment result details in weekly faculty forum meetings. These meetings were attended by both junior and senior faculty members. The discussions during these meetings were utilized to educate other faculty members involved in assessment MCQs writing. Faculty members discussed possible reasons for low and negative DIs. Issues addressed in these meetings were related to blue printing (matching objectives with assessment), content of MCQs, vetting process in the module meetings and problems with functioning of module teams. Instructional design came under discussion.

A faculty survey was done 6 months after introducing the new format looking at their perceptions about understanding and utility of MCQ analysis. Data was analyzed using SPSS version 19.0. Descriptive statistics were calculated in terms of percentage and frequency for categorical variables.

**RESULTS**

Simplified MCQs interpretive guide was provided to course directors after each end-of modular MCQ assessment. More than 30 reports have been generated so far. A year after implementation, 16 faculty members were surveyed regarding the utility and level of understanding of various parameters used in the guide (Table I).

Eighty five percent of course coordinators (n=14) agreed that they could understand the MCQ analysis. All of the course coordinators agreed that they found the revised format easy to understand. Eighteen percent (n=3) faculty strongly, while 82% (n=13) agreed that analysis is helping them to refine the process of MCQ making. Forty five percent (n=7) faculty agreed that analysis of MCQ was helpful in identifying knowledge gaps in students. Fifty five percent (n=9) of the coordinators were uncertain about its use in identifying knowledge gaps. All agreed that this analysis can be utilized in refining future learning strategies for intended objectives.

Second part of the study explored the level of understanding of statistical measures and psychometric principles used in the analysis. Results are shown in Table I, and expressed as percentages. Some (10%) of the faculty members found it hard to understand content validity and use of distractor analysis. A minority (9%) had problem with understanding distribution curve of scores, reliability co-efficient and discrimination index.

<table>
<thead>
<tr>
<th>Table I: Understanding of the statistical parameters by faculty (n=16).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert (%)</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Distribution</td>
</tr>
<tr>
<td>Reliability</td>
</tr>
<tr>
<td>Discrimination index</td>
</tr>
<tr>
<td>Difficulty index</td>
</tr>
<tr>
<td>Distractors</td>
</tr>
<tr>
<td>Content validity</td>
</tr>
</tbody>
</table>
Assessment of medical students by MCQs is well recognized in medical education. If properly constructed they are able to test higher levels of cognitive domains and are thus able to accurately discriminate between high and low achieving. An important component of quality assurance in any assessment method is the post-hoc analysis of these MCQs. Traditionally Classical Test Theory (CTT) and the item response theory have been used for psychometric analysis of the MCQs. The concepts of reliability and validity of assessments were thus introduced.

Tarrant et al. highlighted the importance of reviewing item performance after test administration and using these results to improve them for future administration. Not only is this necessary for quality assurance, a further threat may be that in future universities would need to justify the quality of their assessment to avoid legal actions.

When Shifa College of Medicine moved towards an integrated system-based modular curriculum in 2008, assessment was also intended to be designed in the same line, but it was realized that MCQ related statistics among the faculty was lacking and computer generated analysis was too complicated for review. Knight indicated that faculty of any institute is generally not specialists in research discipline of assessment and they do not analyze their assessments based on standards associated with validity and reliability. They relied on their clinical acumen rather than quantitative evidence.

Realizing this fact, the authors introduced this simplified interpretation guide. Presentation of the end of module results to the Faculty in Faculty forums in the form of these interpretation guides helped the module coordinators and team members in understanding of the statistical parameters in a simple manner. Further half way through the term, a Faculty survey showed that the simplified interpretation guide was helping the faculty in the process of refining the MCQs.

Crisp mentions the use of simple analytical tools and easily understood frameworks which facilitates engagement of the faculty with the underlying theoretical issues related to assessment and student learning.

During the module meetings in the post-hoc phase, it was also realized that this interpretation guide was also helping the junior faculty members who were involved in the planning phase of the module in understanding what was basically wrong with the MCQs. This was encouraging since the junior faculty, though involved with the preparation of the assessment results, was not at all familiar with simple statistical terms.

In literature descriptions of item analysis in a visual format for academics has been reported. However, it does not allow an adequate overall appraisal of the whole assessment process. Visual format of the distractor analysis was facilitating our faculty in finding functioning and non-functioning distractors in an uncomplicated matter. It was further realized that reliability of the assessment method by using Cronbach's alpha in conjunction with other parameters was creating an environment of healthy competition among the course coordinators which was giving them enough confidence to further improve upon their assessment techniques.

Development of this interpretation guide appears to be a unique experiment since the authors could find very little literature on using such a methodology in health professional domains.

In future, the impact of this effort could be the development of customized MCQ bank, helping in creation of an assessment cell with faculty members taking ownership of their own assessment strategies, a process already in pipeline at our institute, and quality assurance would be further strengthened which is our intended ultimate goal.
CONCLUSION

The interpretation report for MCQs was considered useful and was thought to be helpful in identifying knowledge gaps and refining MCQ structure. Thus, this kind of model is very practical for a relatively novice faculty.

REFERENCES