Reflections on Student Performance in MCQ's and OEQ's in Punjab

By

[1]*Muhammad Bakhtiar, [2]Zahida Habib and [3]Sanaullah Khan*
[1]Senior Subject Specialist. Lahore
[2]Assistant Professor, University of Education, Lahore
[3]University of Education, Lahore

1. Introduction

In December, 2005 The Ministry of Education, Punjab began a project to assess student's performance in all subjects at the end of Grade 5 and Grade 8 in all 36 districts of the Punjab in collaboration with UNICEF. On "SOLO Taxonomy" (Biggs and Collis. 1982).Punjab Examination commission was set up to perform this task.

In the centralised database of Punjab Examination Commission, data of 2.0 million students: 1.2 million of data of Grade 5 and 0.8 million of data of Grade 8 for MCQ's and OEQ's, has been entered at item level spreading over more than four years and on that huge data which is a very authentic rich source for research purpose, Rasch Analysis (Bond and Fox , 2007).   As model of Item Response Theory is being utilised to enhance the educational process and improve mathematics syllabus, the mathematics text books, teacher manuals, and other supplementary teaching resources.

Many researches have been conducted to see the comparative performance of students on different types of exams techniques like multiple choice questions, true/false, completion exercises, short answers/restricted response items and essay types. Different researchers concluded differently. Some declared Multiple choice items easier than Open ended questions while some showed that the facts are other way round, especially in mathematics some showed that students performed equally good on both MCQ's and OEQ's.

In this paper, the focus is therefore on the formulation of the following research questions:
- What is the performance of students of grade 5 and grade 8 in MCQ's and OEQ's for the mathematics curriculum in Punjab, Pakistan
- What is the magnitude of difference in MQ's and OEQ's?
- Does this trend resemble with international trend?

2. Literature Review

Assessment is a pivotal driving force behind, student performance, the measure of student progress and examining the effectiveness in different Grade levels. Therefore, Assessment practices have a strong association/linkage with assessment design, teaching methods for effectiveness of student learning depicted in assessment, devising and also introducing new, innovative and different forms of assessment. The multiple choice question consists of a stem, which presents a precisely defined problem situation free of unnecessary complexity of language and several alternatives, which provide possible solutions to the problem. The stem may be a question or an incomplete statement. The alternatives include the correct answer and several plausible wrong answers, called distracters. The function of the distracters is to distract those students who are uncertain of the answer.

Multiple choice questions (MCQs) also provide faster ways of assessing student learning. MCQs are referred to as objective test items because they can be marked objectively. Multiple choice questions can

be used effectively at the almost all levels to measure a wide range of abilities. Well designed MCQs can measure higher end abilities of SOLO TAXONOMY such as Relational and extended abstract responses. Preparation of a Multiple Choice Question is quite difficult and time consuming that accurately reflects what to measure but its scoring is quite objective. . It also allows for a precise interpretation for content validity, pattern of student learning outcome, difficulty level and reliability of item. It is very easy to conduct. It is also very interesting that student success does not depend on student writing skills but only his analytical skill. On the other hand it also has many limitations. One of them is that it hinders students from expressing creativity or demonstrating original and imaginative thinking. Success of item depends on plausibility of distractors while a few students may guess to answer without understanding it.

Open ended items consist of a stem which represent precisely defined problem situation free of several possible solutions to the problem. Open-ended items require written responses that provide particularly useful insights to students' levels of conceptual understanding. They can also be used to assess students' abilities to communicate in the sciences. In addition, open-ended items, if carefully crafted, can be used to reflect students' abilities to generate rather than recognize information related to scientific concepts and their interconnections. In general, they require complex thinking and yield multiple solutions. Open-ended questions require teachers or evaluators to interpret and use multiple criteria in evaluating responses. Such questions also require more from students than simply memorizing facts.

Open-ended questions let the student verbalize the answer. An open-ended question does not provide the student with a choice of answers. Open-ended questions are also referred to as free-response and they are free to answer the question in any manner they choose. These can lead to repetition, the gathering of irrelevant information, and misunderstandings about the intent of the question. The results obtained from open-ended questions are also more difficult to analyze. Sometimes biasing of evaluator also influences the objectivity of evaluation.

Many researchers have done a lot of studies on the effect of test methods on test performances. Shohamy (1984) found that test methods influenced how readers performed on a test of reading comprehension, and that multiple-choice questions were easier than open questions, and the effect was stronger on low-proficient readers.

Wolf (1993) carried out a similar experiment; he also concluded that multiple-choice questions were easier than open-ended questions.

Samson (1983) used multiple-choice questions, open-ended questions, and summary tests in a reading comprehension test. The results showed there was no significant difference among the three test methods, so she concluded that the three test methods all tested the same ability or trait of the subjects. But she did find that multiple-choice questions were the easiest, and summary test the   most difficult.

In China, much research has been done on testing reading comprehension. Chen & Cao (1999) argue that short answer questions are more effective than multiple–choice questions in testing reading comprehension

Open-ended and multiple-choice scores are strongly correlated, especially for math (Washington State Institute for Public Policy on the 10th-grade WASL 2006.) Students who do well on multiple-choice questions almost always do well on open-ended questions. Similarly, students who do well on open-ended questions also do well on multiple-choice questions. These associations are stronger for math than for reading, which suggests that multiple-choice and open-ended questions assess similar kinds of skills in math, but less so for reading.

(Greg Jamieson. 2005) ESL (English As Second Language) students perform better in multiple-choice assessment formats than they do in open-ended assessment formats whereas EFL (English As First

Language) students perform better in open-ended assessment formats than they do in multiple-choice assessment formats.

It is clear from the above discussion that different researchers concluded differently while developing relationship of scoring between MCQ's and OEQ's. Therefore it will be interesting to see reflection on the performance of students on MCQ's and OEQ's in the Punjab, Pakistan.

3. Method

To explore these issues raised in research questions, Students' data from the database of PEC will be used. This data has been entered at item level i.e. data of 2.0 million students: 1.2 million of data of Grade 5 and 0.8 million of data of Grade 8. Furthermore this data has been manipulated on the Model of Item Response Theory i.e. Rasch Analysis using Quest software(Adams and Khoo 1994) and psychometric properties; item difficulty, item fit, discrimination index at item level, moreover variable map which maps item difficulty to student performance

In PEC examination system for Grade 5 and Grade 8, Paper are developed on three levels of SOLO Taxonomy (Biggs and Collis. 1982). Uni-level requires a single operation, multi-structural demand a series of operations and Relational level that requires an overview and integration of the whole data.
These papers comprises of multiple choice questions MCQ's and Open ended questions OEQ's each OEQ also have three parts representing three levels of SOLO. For Grade 8 the ratio of MCQ's and OEQ's is 30:4 and for Grade 8 40:4. MCQ's are marked as dichotomous data wrong "0" and right "1" while marking structure of OEQ's is more complex and varies from $0 - 5$; each part has an allocation of marks as per the complexity of its SOLO level. SOLO taxonomy model is used to interpret students' responses. Pre structural responses marked as 0.

Keeping in view of item difficulty level(SOLO taxonomy level) whether it is a part of an open ended question or an MCQ has been compared; uni-structural MCQ and uni-structural part of OEQ, and similar way other multi-structural and relational levels of each type has been compared.

4. Results

To see the behaviour of the performance of students in MCQ's and OEQ's for Grade 5 and Grade 8 data from PEC pertaining to paper Mathematics 2009 has been selected.

Grade 8:
Psychometric indices of Grade 8 Paper Mathematics 2009 are given below showing the construct validity and item reliability is up to the standard.

Item estimate

Mean =00   Standard Deviation (SD) = 0.94   Reliability of Estimate = 01 Internal Consistency =.84

Infit Mean Square   Mean=1.00   SD=.10   *Outfit Mean Square*   Mean =1.01   SD =.15

Table 1 of Grade 8 Paper Mathematics 2009 provides the frequency of correct responses to the items exploring students understanding to the associated learning outcome (SLO). For this purpose items either belonging to MCQ's or OEQ's of same SOLO level/item difficulty level are grouped together to compare % correct of each item and means of % correct on MCQ's and OEQ's of that difficulty level/ SOLO level to examine reflection of student's performance on them.

Q.No	1	2	3	4	5	6	7	8	41 a	42 a	43 a	44 a	9	10	11	12	13	14	15
%Correct	59	47	63	62	40	75	89	80	31	10	13	27	71	56	29	65	40	31	60
Mean of %correct	64.83								20.25										
Item Type	MCQ's								OEQ's				MCQ's						
Level	Uni Structural												Multi Structural						
Q.No	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	41b	42b
%Correct	47	77	34	38	59	47	44	30	80	77	48	57	28	11	40	69	26	22	21
Mean of %correct.\	48.25																	18.25	
Item Type	MCQ's																	OEQ's	
Level	Multi Structural																		
Q.No	43 b	44b	33	34	35	36	37	38	39	40	41c	42c	43c	44c					
%Correct	12	28	78	35	50	09	54	65	18	48	15	25	07	07					
Mean of %correct			44.63								13.5								
Item Type	OEQ's		MCQ's								OEQ's								
Level	Relational																		

Exhibit 1 of Grade 8 Paper Mathematics 2009 shows the summary of the table 1 recording the mean of % correct .on MCQ's and OEQ's on each SOLO level/difficulty level; e.g. mean of % correct of Uni structural MCQ is 64.38 while that of OEQ's is 20.25 and mean of % correct of Multi structural MCQ's and OEQ's is 48.25 and 18.25 similarly means of % correct of Relational MCQ's and OEQ's are 44.63 and 13.5 respectively. Accumulative means of MCQ's and OEQ's are 50.9 and 13.5 respectively.

It can easily be seen from this exhibit 1 that performance of students on MCQ's and OEQ's decreases with the increase of complexity of SOLO taxonomy level but on the other hand comparison across MCQ's and OEQ on the same SOLO taxonomy level shows that students performed extremely worse on OEQ's than MCQ's. Therefore, very few students are reflecting the true achievement of the curriculum Student Learning Outcome.
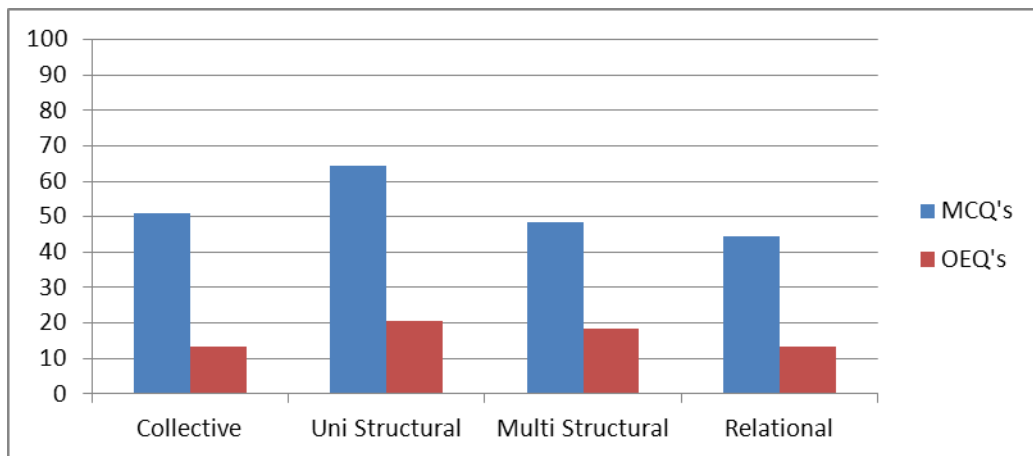

Exhibit 2 of Grade 8 Paper Mathematics 2008 shows the summary of the mean of % correct of each SOLO level.on MCQ's and OEQ's , e.g. means of % correct of Uni structural MCQ are 70.38 while that of OEQ's is 16.25 and means of % correct of Multi structural MCQ's and OEQ's is 47.92 and 22.25 similarly mean of % correct of Relational MCQ's and OEQ's are 36 and 13.4 respectively. Accumulative means of MCQ's and OEQ's are 50 and 13.4 respectively.

Exhibit 3 of Grade 8 Paper Mathematics 2007 shows the summary of the mean of % correct of each SOLO level.on MCQ's and OEQ's , e.g. means of % correct of Uni structural MCQ are 52.8 while that of OEQ's is 18.4 and means of % correct of Multi structural MCQ's and OEQ's is 55.27 and 4.8 similarly mean of % correct of Relational MCQ's and OEQ's are 39.38 and 3.25 respectively. Accumulative means of MCQ's and OEQ's are 51.47 and 3.25 respectively.

Both exhibits show same extremely worse performance on OEQ's as compared to MCQ's.
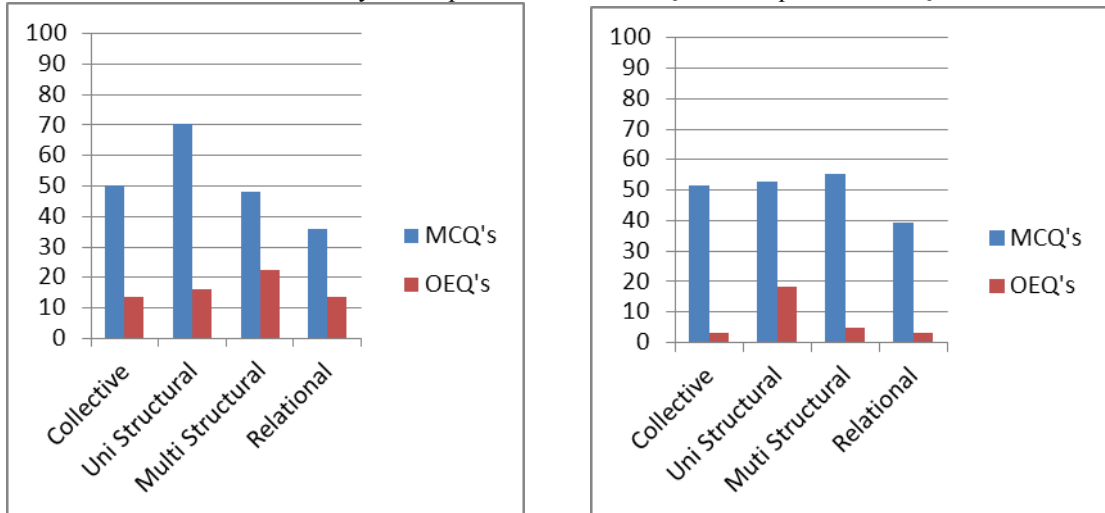
For an SLO of Multi Structural level " Find value of many things if value of many things is given "
An item was asked as an MCQ
    " Kashif eats 20 apples in 5 days. How many apples will he eat in 4 days "

Only 80% provided the correct response.

On the same above SLO when following OEQ was asked
    " if 5 men completes work in 100 days. Then how many men will complete
    the work in 20 days"

Only 22% provided the correct response.

For an SLO of Relational level " Find the value of many things if value of many things is given (Compound Proporation) "

An item was asked as an MCQ
    " if 5 labourers earn Rs. 2000 in 8 days then how much money 3 labourers will earn (
    in Rs.) in 4 days "

Only 46% provided the correct answer.
An item was asked as an OEQ on the above same SLO
    "For a family of 5 members Rs. 2000 are sufficient for 10 days. For how many days
    Rs. 1600 are sufficient for a family of 4 members? "

Only 15% provided the correct answer.

This clearly shows that with the increasing complexity of SOLO level, % of correct responses decrease both in MCQ's and OEQ's but when we compare across MCQ and OEQ of same SOLO level, student's performance is extremely worse in OEQ than MCQ.

Grade 5:
Psychometric indices of Grade 5 Paper Mathematics 2009 are given below showing the construct validity and item reliability is up to the mark.

Item estimate
Mean =00   Standard Deviation (SD) = 0.63   Reliability of Estimate = 01 Internal Consistency =.84
Infit Mean Square   Mean=1.00   SD=.12   Outfit Mean Square   Mean =1.01   SD =.15

Table 2 of Grade 5 Paper Mathematics 2009 also provides frequency of correct responses to the items exploring students understanding to the associated learning outcome (SLO). For this purpose items either belonging to MCQ's or OEQ's of same SOLO level/item difficulty level are grouped together to compare % correct of each item and means of % correct on MCQ's and OEQ's of that difficulty level/ SOLO level to examine reflection of student's performance on them.

Q.No	1	2	3	4	5	6	31a	32a	33a	34a	7	8	9	10	11	12	13	14	15
%Correct	34	73	31	43	26	64	22	13	17	17	58	55	45	65	32	61	36	28	65
Mean of %correct					45.17					17.25									
Item Type			MCQ's					OEQ's							MCQ's				
Level									Uni Structural									Multi Structural	
Q.No	16	17	18	19	20	21	22	23	24	31b	32b	33b	34b	25	26	27	28	29	30
%Correct	52	54	53	35	52	36	35	40	62	15	17	16	09	27	29	54	09	55	35
Mean of %correct.\									48						13.5				34.83
Item Type					MCQ's								OEQ's						
Level																		Multi Structural	
Q.No	31c	32c	33c	34c															
%Correct	05	11	03	03															
Mean of %correct			5.5																
Item Type			OEQ's																
Level																			Relational

Exhibit 4 of Grade 5 Paper Mathematics 2009 shows the summary of the table 2 recording the mean of % correct of each SOLO level on MCQ's and OEQ's , e.g. means of % correct of Uni structural MCQ are 45.17 while that of OEQ's is 17.25 and means of % correct of Multi structural MCQ's and OEQ's is 48.25 and 13.5 similarly mean of % correct of Relational MCQ's and OEQ's are 34.83 and 5.5 respectively. Accumulative means of MCQ's and OEQ's are 44.8 and 5.5 respectively.

It can easily be seen from this exhibit 4 that performance of students on MCQ's and OEQ's decreases with the increase of complexity of SOLO taxonomy level but on the other hand comparison across MCQ's and OEQ on the same SOLO taxonomy level shows that students performed extremely worse on OEQ's than MCQ's. Therefore, very few students are reflecting the true achievement of the curriculum Student Learning Outcome.
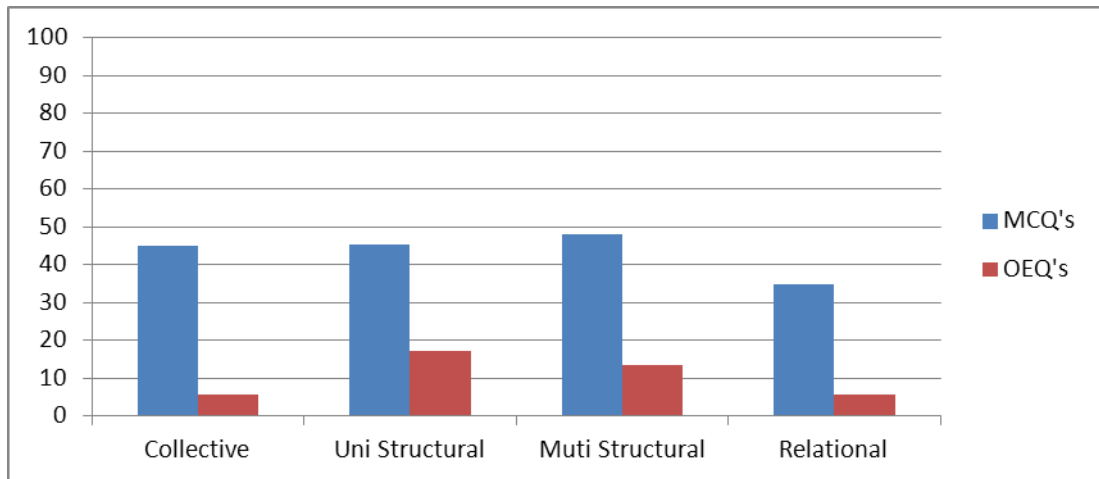
For an SLO of Uni Structural level " Place comma correctly according to given Number   System "   An item was asked as an MCQ
" Correct Placement of comma in million system for number 345678913 is "

Only 76% provided the correct response.
On the same above SLO when following OEQ was asked
" Place comma in 2678172 in million system "

Only 22% provided the correct response.
For an SLO of Multi Structural level " Write the number given in words into numeral "
An item was asked as an MCQ
" Eighty crore seventy lac sixty six thousand three hundred and seventy two in number is "

Only 90% provided the correct response.
On the same above SLO when following OEQ was asked
Write number " seven billion nine hundred eleven million three hundred two thousand " in numeral.

Only 15% provided the correct response.
For an SLO of Relational level " Write the number given in words in one Number system into another Number system in words."

An item was asked as an MCQ
"Nine crore eighty two lac five thousand and twenty one in billion system is"

Only 64% provided the correct answer.
An item was asked as an OEQ on the above same SLO
"Write a number Eighty four areb ninety eight crore seven thousand eight hundred and twelve in million system in words "

Only 5% provided the correct answer.
This also clearly shows that with the increasing complexity of SOLO level, % of correct responses decrease both in MCQ's and OEQ's but when we compare across MCQ and OEQ of same SOLO level, student's performance is worse in OEQ than MCQ.

Exhibit 5 of Grade 5 Paper Mathematics 2008 shows the summary of the table 2 recording the mean of % correct of each SOLO level on MCQ's and OEQ's , e.g. means of % correct of Uni structural MCQ are 68.5 while that of OEQ's is 35.25 and means of % correct of Multi structural MCQ's and OEQ's is 48.67 and 17.25 similarly mean of % correct of Relational MCQ's and OEQ's are 38.67 and 13.5 respectively. Accumulative means of MCQ's and OEQ's are 50.64 and 13.5 respectively.

Exhibit 6 of Grade 5 Paper Mathematics 2007 shows the summary of the table 2 recording the mean of % correct of each SOLO level on MCQ's and OEQ's , e.g. means of % correct of Uni structural MCQ are 66 while that of OEQ's is 16.75 and means of % correct of Multi structural MCQ's and OEQ's is 53.44 and 8 similarly mean of % correct of Relational MCQ's and OEQ's are 62.38 and 14.75 respectively. Accumulative means of MCQ's and OEQ's are 58 and 14.75 respectively.

These both exhibits also show same extremely worse performance on OEQ's as compared to MCQ's.
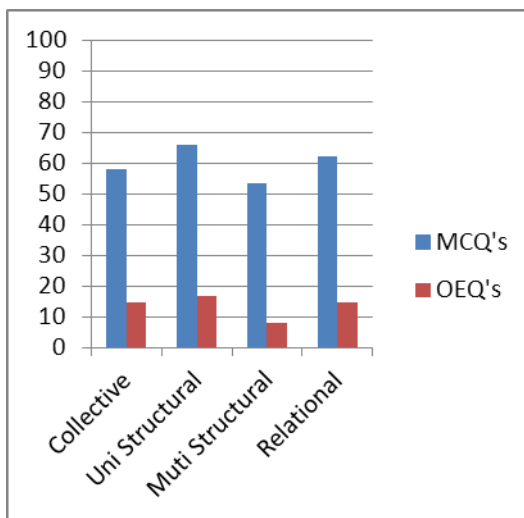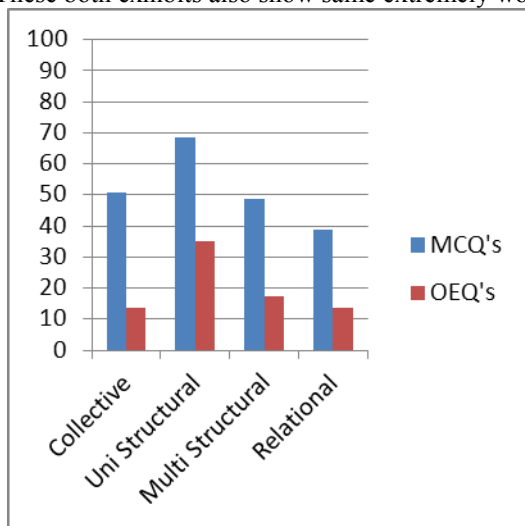
5. Conclusion/Recommendations

Keeping in view the data reported above, there are number of conclusions and recommendations which can be incorporated are discussed below.

The large variation between the performance of students on MCQ's and OEQ's may be an indication of the lack in conceptual understanding or practice of rot learning. Therefore there should be more emphasis on theoretical understanding following with good application or drill.

This may also be an indicator of lack in conceptual heraricy/sequence of concept from lower grade to higher grade in the Curriculum. Therefore it is need of the time for careful investigative look into text books.

One factor may be the non-completion of Syllabus and more focus on Objective than subjective preparation in classroom activities.

It is common observation in Pakistan that academic and professional qualification of Elementary school teachers is low. This may lead to in-effective classroom teaching. Teachers workload may be another factor contributing to poor performance.

There should be an investigation to search the factor of poor performance of students in Mahematics in elementary level.

Teachers refresher courses should be conducted for the teachers and adequate educational resources should be made available for teachers for enhancement of their better understanding.

References

PEC. 2009. http://www.pec.edu.pk/overview.html

Bond. T. and Fox. C., 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* NJ: Lawrence Erlbaum Assoc

Redden. E., Fatima. A., and Bakhtiar M 2009. *Advice for Teachers Mathematics. Grade 5 2009*. Lahore PEC

Redden. E., Fatima. A., and Bakhtiar M 2009. Advice for Teachers Mathematics. Grade 8 2009. Lahore PEC

Adams, R. and Khoo, S. 1996 *Quest* A *Interactive Test Analysis System* Camberwell Victoria. ACER

Biggs, J. and Collis, K. (1982). *Evaluating the Quality of Learning: the SOLO Taxonomy*. New York: Academic Press.

Windham, Douglas M. (1988). *Improving the efficiency of educational systems: Indicators of*

educational effectiveness and efficiency. Tallahassee: Florida State University.

Punjab Text Book Board. (2003). *Mathematics,Grade 5 ,Grade 8*. Lahore.

Pakistan Ministry of Education. (2006). National Curriculum for Mathematics 2006-Grade I to VIII. Islamabad.

Biggs J and Collis K (1982) Evaluating the Quality of Learning: the SOLO taxonomy. New York. Academic Press

Bridge, PD, Musial, J, Frank, R, Roe, T, Sawilowski, S. Measurement practices: methods for developing content-valid student examinations. Med Teach 2003

Crossley J, Humphries G, Jolly B. Assessing health professionals. Med Educ 2002; 36: 800-804.

Haladyna TM. Developing and Validating Multiple-Choice Test Items, 2nd edn. Mahwah, NJ: Lawrence Erlbaum Associates; 1999.

Oosterhoff A. Developing and Using Classroom Assessments, 3rd edn. Upper Saddle River, New Jersey: Pearson Education; 2003.

Newble D I, Baxter A & Elmslie RG(1979) A comparison of multiple Choice and free response test in examination of clinical competence. Medical Education

Ebel, R L. Measuring educational achievement. Prentice Hall, Englewood Cliffs NJ, 1965

Haladyna, T M. Developing and validating multiple choice test items. Lawrence Erlbaum Associates, Hillsdale NJ, 1994

Bennett, R E; Rock, D A & Wang, W C. Equivalence of free-response and multiple choice items. Journal of Educational Measurement, Vo1.28, No.1, ~~77-92,1991

Bridgeman, B & Rock, D A. Relationships among multiple-choice and open-ended analytical questions. Journal of Educational Measurement, Vo1.30, No.4, pp 313-329, 1993

Bennett, R E & Ward, W C (Eds). Construction versus choice in cognitive measurement. Lawrence Erlbaum Associates, Hillsdale NJ, 1993

Fenderson, B A; Damjanov, I; Robeson, M R; Veloski, J; & Rubin, E. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. Human Pathology, Vo1.28, No.5, pp.526- 532, 1997

Schuwerth, L W T; Vandervleuten, C P M; & Donkers H H T, M. A closer look at cueing effects in multiple choice questions. Medical Education, 1996, Vo1.30, No.1, pp.44-49, 1996

Kehoe, J. 1995. 'Basic Item Analysis for Multiple Choice Tests' in *Practical Assessment ,*

Research and Evaluation. Vol 4 No 10.

Chen, Hua & Cao, Jun. (1999). The Feasibility of Subjective Answering Questions in Testing Reading Comprehension. *Shandong Foreign Language Teaching*. 2, 80-84.

Samson, D. M. M. (1983). *Rasch and reading. In J. van Weeren (Ed.), Practice and problems in language testing.* Arnhem: CITO.

Shohamy, E. (1984). Does the testing method make a difference? *The case of reading comprehension. Language Testing*,1 (2), 147-170.

Wolf, D. F. (1993). A comparison of assessment tasks used to measure FL reading comprehension. *The Modern Language Journal*, 77 (4), 473-489.