Bulletin of Education and Research
December 2021, Vol. 43, No.3 pp. 81-94

Evaluation of Question papers by Board of Intermediate and Secondary Education using Item Analysis and Blooms Taxonomy

Ameema Mahroof[*] & Muhammad Saeed[**]

Abstract

This small scale study aims to analyze the question papers of Board of Intermediate and Secondary Education in the subject of computer science with reference to item analysis and Bloom's taxonomy. Data were collected from 100 students of Grade 9th and 10th from the schools of Lahore city using convenient sampling technique. Data collected on the papers developed by Board of Intermediate and Secondary Education for the year of 2015 and 2016. Item analyses were performed using Conquest software. Findings of the study shows that in the question papers conducted by Board of Intermediate and Secondary Education the majority questions were measuring the student abilities of knowledge and comprehension and only few questions were given to measure the student abilities to analyze, synthesize and evaluate, and this can be very helpful for the policy makers. Result of item analysis shows that many questions were not in the acceptable range of item difficulty and item discrimination. Items in the question papers were either too easy or too difficult. Findings revealed that the papers conducted and administered by Board of Intermediate and Secondary Education were not up to the mark, with reference to Bloom's taxonomy. The researcher recommended to train the assessment committee/panel developing the items.

Keywords: Bloom's Taxonomy, Question Papers, Item Analysis, Cognitive Domain, Computer Science, assessment of school students

[*]PhD Scholar, Institute of Education and Research, University of the Punjab, Lahore
[**] Professor, University of Lahore, Lahore

Introduction

In Pakistan, Educators may face the challenges in analyzing that whether the question items fulfilling the requirements of Blooms taxonomy at different cognitive levels. This article analyzed the papers of computer science of secondary level according to the Bloom's taxonomy of cognitive domain followed by item analysis to measure psychometric qualities of the test items. The study has two parts; first the researcher analyzed the papers with reference to Blooms taxonomy and second the measurement of psychometrics properties (item difficulty and discrimination index) of each test item.

Taxonomy is an attempt to classify the levels and forms of learning. Bloom's taxonomy is developed by Bloom, he was an educational Psychologist, and he was totally against the rote learning and memorization, so he formed a taxonomy which is knows as Blooms taxonomy. (Mehmood, Iqbal, Abdullah & Farooq, 2016). Bloom's taxonomy has three levels. The cognitive domain, affective domain and psychomotor domain, and each of these domains has levels. It is recommended that one cannot achieve the higher levels until below them is covered. Bloom's taxonomy of cognitive domain has six levels from easy to complex. These are Knowledge, Understanding, application, analysis, synthesis and evaluation. Each one of them is described below.

Knowledge is the lowest level of cognitive domain. Memorization comes in it. For example, if a teacher teaches a topic to the student and the other day raised some questions related to that topic, this recall of the lesson will be 'knowledge'. Understanding is not simply based on rote learning, rather the learner is expected to interpret the information in his/her own words. For example, if a student is able to interpret information in his/her own words and it understandable to the listener that we can say the students developed understanding ability. Application means that students should be able to apply the knowledge in different situations. Analysis means to break down the information into parts and then find out its relationship, select the important points, and exclude all irrelevant information. Synthesis means to build a new thing and give shape to knowledge. It includes creativity. Creation of new things by a student with their own understanding is synthesis. Evaluation is the highest level and it include judgment of something.

It is very important to analyze the exam papers according to Bloom's taxonomy to check that how far they are measuring student abilities of knowledge, understanding, application, synthesis and evaluation. Furthermore, the balanced paper covers the difficulty and discrimination level of items to identify the hidden capacities of students.

One of the most powerful techniques to the teachers to check the quality of items is item analysis. According to Shakil (2008), Item analysis is a procedure to check the quality of an item and a test as a whole by examining student responses towards individual item. Bichi, A. A. (2013) states (As cited in Suruchi & Rana, 2014) that item analysis has two purposes, first to identify the bad items and second to analyzed the areas where the students have mastered or not. He further states that it measures the performance of an individual test item in terms of its difficulty and discrimination power (means to distinguish between high and low achievers). So item analysis helps us to select the best test items by excluding the poor test items. Item analysis is usually associated with three qualities: item difficulty, item discrimination, and power of distractors (only for MCQs).

Item difficulty, as defined by Kohoe (1995), is the proportion of examinees respond to the item correctly is item difficulty. It is also known as P-value. The formula for item difficulty is: P = number of test takers who pass the item / total numbers of test takers. Its value ranges from 0 to 1. The higher the value of P, the easier the item. A zero means no one got the item right while 1 means that everyone got the item right. The closer an item gets to 0 or 1, the less information it contributes about test takers. The most acceptable value is of item difficulty is between 0.27 and 0.84.

Item discrimination means that how the item is discriminating between high and low achievers. It is a relationship between how well a student performed in an item and on a whole test scores. The range of item discrimination is -1 to 1. The higher the value, the more the item is discriminating. It works like if an item has a high discriminating value then it means the student performed well on the test got the item correct, and who had low score got the item incorrect. The items with zero or negative discriminating value should be removed because it shows that student who did poor on the overall exam, got the item right whereas students who overall did better on the exam got the item incorrect. Its acceptable value is from 0.20 or higher. The formula for item discrimination is divide the examines in two halves (upper and lower achievers), then count number in the high group who got the item right and number in the low group who got the item right and divide it by the number of examinees in one group.

Distractor analysis is very important for the quality of MCQs because the quality of distractors effect the student performance in exam It addresses the performance of these incorrect responses option. Just as the key or the correct answer must be definitely correct, the distractors must be definitely incorrect.

Computer science was introduced at secondary level as an elective subject in order to develop the skills of software, networking, hardware, graphics and programming. With the advancement of technology, the Computer has gained so much popularity that everyone now supports computer science education. Computer Science is a discipline with a set of rules and principles that can be used to solve problems in real world. (Report of Curriculum Improvement Task Force, 2005). Computer Science was introduced as an elective subject at Secondary level in 90s.

Much work has attempted to analyze exam questions against Bloom's taxonomy. Lahari and Mukherjee (2015) conducted a research on Analysis of Multiple Choice Questions and they conclude that the items analyzed in their study had optimum difficult level but distractor efficiency is poor .Shahzad, Qadoos, Naeem, Badshah, Muhammad, and Ramzan (2011), proposed to analyze the Biology paper of class intermediate with reference to Bloom's taxonomy and on the basis of the finding the researchers strongly recommended that BISE Bannu set the papers by those paper setters that have full command on Bloom's Taxonomy. Bichi, A. A (2013) also studied on an item analysis and his study revealed that 12 items out of 40 failed to meet the set criteria. Veeravagu, Muthusany, Marimuthu, and Subrayand (2010) found that students performed better in questions with low level thinking process compared to high order questions. Iqbal, Ullah and Nisar (2019), conducted a research on Physics paper and suggested that raining must be provided to the paper setters in such a way that lead them to include such items which can measure different abilities of the students to achieve the required objectives.

Statement of the Problem

The purpose of this study is to analyze the papers of computer science at secondary level according to the Bloom's taxonomy of cognitive domain with the focus on analyzing the Multiple Choice Questions through item analysis using the software Conquest to identify the weaknesses in the item.

Objectives of the study

Objectives of the study were to:

- Evaluate the question paper of computer science at secondary level with reference to Bloom's cognitive domain.
- Determine the strengths and weaknesses of items in the question papers in terms of psychomotor properties of test.
- Judge the overall quality of question papers in the subject of computer science in terms of basic rules/principles of test construction.

Research questions

Following are the research questions of the study

- What are the levels of Question papers of Computer Science with reference to Bloom's cognitive domain learning?
- What are the strengths and weaknesses in the questions in regard to different psychometric properties of tests?
- How much paper developers follows the basic rules of test construction?

Limitation of the Study

Although the researcher has reached its aims, but researchers confronted some unavoidable limitations. Primarily researcher decided to perform item analysis on students' responses of computer science papers conducted by (BISE) Board of Intermediate and Secondary Education, but due to some reasons, the researchers couldn't succeed to get the data of students from Board of Intermediate and Secondary Education so researchers selected a sample of 100 students to collect data on MCQs.

Delimitation

The study was delimited to the:

1. Question papers of BISE Lahore in the years of 20l5 and 2016.
2. Objective part of the BISE Computer science papers.

Methodology

This study is quantitative and qualitative in nature. The question papers of BISE for year 2015 and 2016 are analyzed.

Population and Sampling

The population of study was all the students in grade 9^{th} and 10^{th} studying the subject of computer science. This was a small scale study and sample size was 100 students, fifty for grade 9 and fifty for grade 10. Data were collected from 100 students of seven public schools using convenient sampling technique.

Data Collection and Ethical Considerations:

The purposes of this study were clearly communicated to the Principals and subject teachers of the selected schools. Consent was taken from the Principal by explaining that collected data will be solely used for research purpose and no information will be used against their schools. Data were collected on the question paper of Board of Intermediate and Secondary Education.

Grade 9th Question paper of computer science were consisting of 15 MCQs so the data were collected from the students on 2015 and 2016 question papers. Both question papers were given to the students, 20 minutes given for each paper. Objective type paper of computer science for grade 10th were also consist of 15 MCQs to be solved in 20 minutes.

Data Analysis and Findings

The data were analyzed using a checklist develop by the researcher. Checklist were validated from two expert opinions. The two experts were doing their Ph.D. in the area of assessment and have been teaching for more than 12 years. In this tool each item was measured against the levels of Bloom's taxonomy. For the item analysis, the data collected from 100 students were analyzed using the software Conquest. For the analysis of data using conquest, data were entered in note sheet using the correct option and then code is written in conquest using the answer keys.

Table 1

Levels of Bloom's taxonomy in papers

Year	Knowledge	Understanding	Application	Analysis	Synthesis	Evaluation
2015 Grade 9th	32%	47%	4%	11%	4%	-
Grade 10th	51%	43%	4%	2%	-	-
2016 Grade 9th	53%	34%	13%	-	-	-
Grade 10th	40%	47%	13%	-	-	-

Table 1 shows the percentage of the questions being developed against the levels of Blooms taxonomy. It shows that in 2015 paper of Computer Science for grade 9th and 10thmaximum items measured just students' knowledge and understanding about the content whereas very small number of items assessed students' application, analysis and synthesis ability and there is no single item found that assess the student evaluation ability. For the year of 2016, the result shows that no single item was measured the student abilities of analysis, synthesis and evaluation. Graphical representation is shown below:

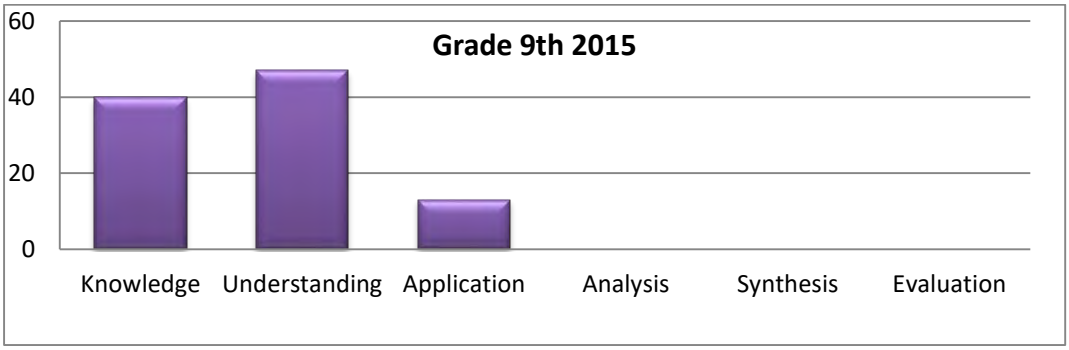

Fig 1. *Grade 9th Paper as per levels of Blooms Taxonomy*

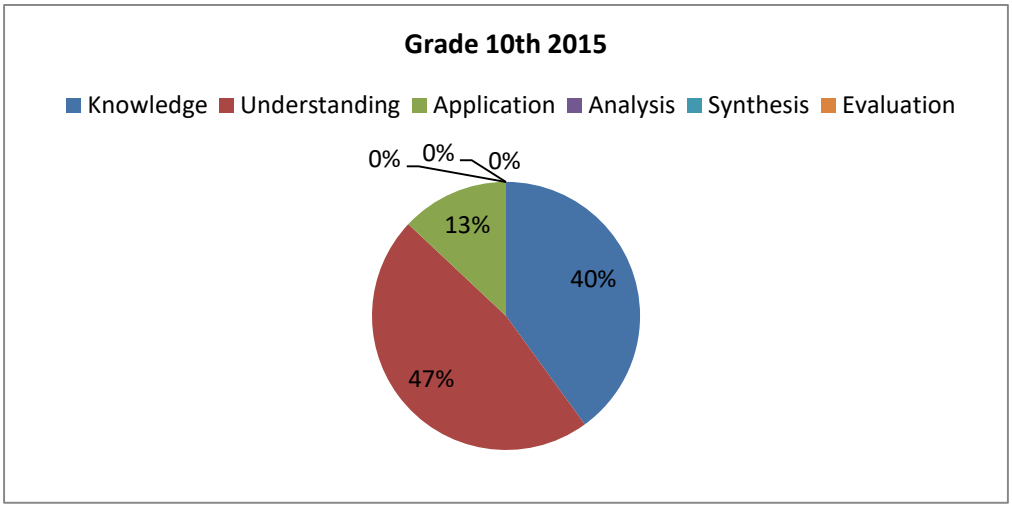

Fig 2. *Grade 10th Paper as per levels of Blooms Taxonomy*

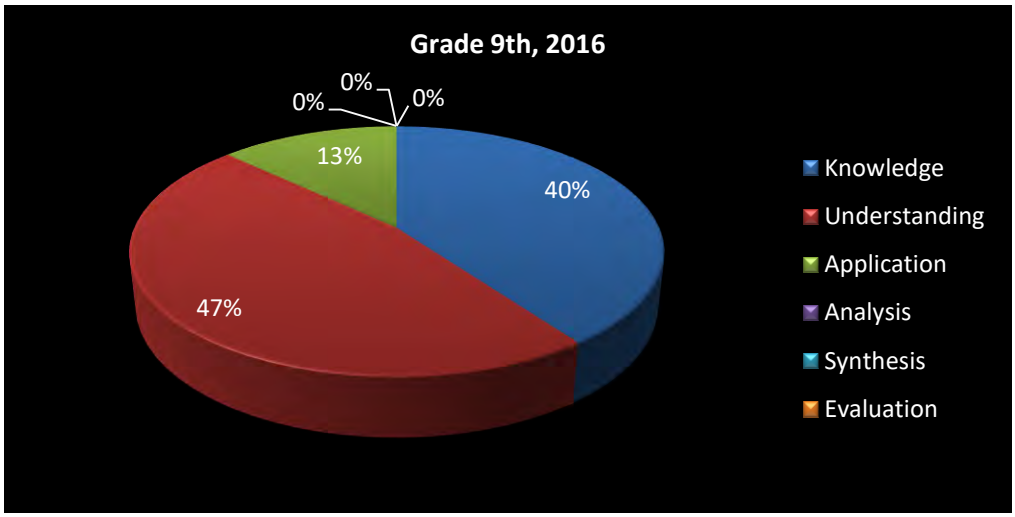

Fig 3. *Grade 9th Paper as per levels of Blooms Taxonomy for year 2016*

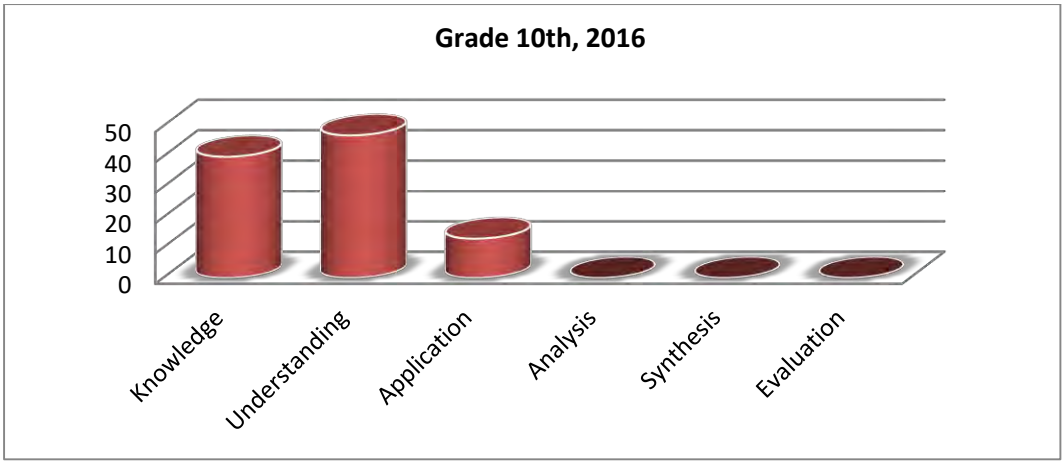

Fig 4. *Grade 10th Paper as per levels of Blooms Taxonomy for year 2016*

Table 2

Item Difficulty and Item Discrimination values for Grade 9th 2015

Question Number	Item Difficulty	Item Discrimination
Q1	0.63333	0.3333
Q2	0.16666	0.06666
Q3	0.63333	0.333
Q4	0.53333	0.6666
Q5	0.5	0.73333
Q6	0.26666	0
Q7	0.86666	0.26666
Q8	0.83333	0.3333
Q9	0.73333	0.26666
Q10	0.66666	0.6666
Q11	0.76666	-0.0666
Q12	0.83333	0.2
Q13	0.9	0.2
Q14	0.93333	0.13333
Q15	0.9	0.2

Table 2 shows the item difficulty and item discrimination values for grade 9thfor the year of 2015. As per literature the acceptable range for item difficulty is from 0.3 to 0.7. if Items are below 0.3, then these items are difficult and need to be modified and if difficulty index ranges above 0.7 then those items are easy and need to be changed.

Acceptable value for item discrimination is above 3. If any item value is below 3 then this item is not discriminating between lower abilities student and higher abilities students means either students with low ability can correct the answer of high difficulty or vice versa.

In this paper item 4 and 5 are desirable difficulty and also discriminating between upper and lower achievers. Difficulty index and discriminating values shows that this many items in this paper needs revision i.e. 12, 13, 14, 15. And many items need to be omitted such as item no 2 and 6.

Below Table 3 shows the conquest analysis of item no 11 which is a bad fit item and fig 5 demonstrates its graphical representation.

Table 3
Bad Fit Item

Label	Score	Count	% of tot	Pt Bis t(p)	PV1Avg:1	PV1	SD:1
a	0.00	1	3.33	0.09	0.49(.628)	1.40	0.00
b	1.00	25	83.33	-0.07	-0.39(.698)	1.24	1.12
c	0.00	3	10.00	-0.04	-0.24(.815)	1.32	0.47
d	0.00	1	3.33	0.14	0.73(.473)	3.16	0.00

The table 3 shows the bad fit item. Total number of respondents who answer this question is 30, discrimination of this question is -0.07 with item threshold at -0.65 and MNSQ at 1.34

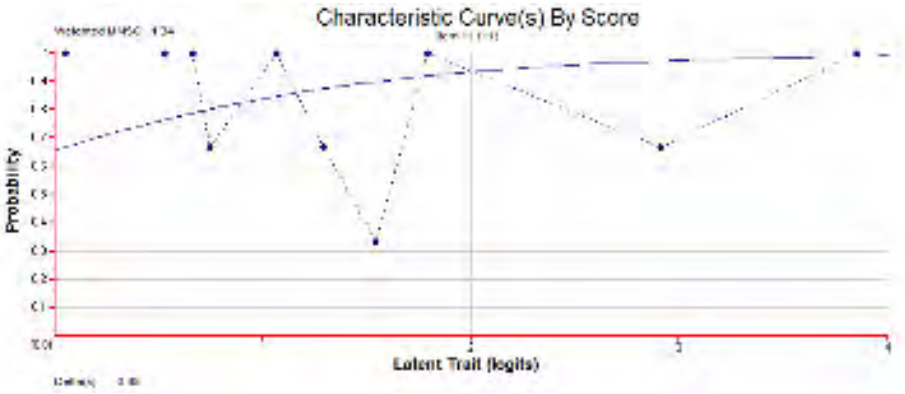

Fig 5. *Item Characteristic Curve*

The figure 5 of this item shows the bad fit item. This item is not good discriminator in high achievers and low achievers with discrimination index -0.07 and MNSQ=1. 34.. The correct responses for this item is 25 which shows that the item is easy, every distracter should select 5 or above the 5 percent of each but in this item 83.33 students select the key which shows that this item should be removed from the paper

Table 4

Good Fit Item

Item: 5 (5)

Label	Score	Count	% of tot	Pt Bis t(p)	PV1Avg:1	PV1	SD:1
a	0.00	10	3.33	-.40	-2.32(.028)	0.79	0.33
b	1.00	17	56.67	0.69	5.11(.000)	1.81	1.18
c	0.00	3	10.00	-0.52	-3.20(.003)	0.25	0.37

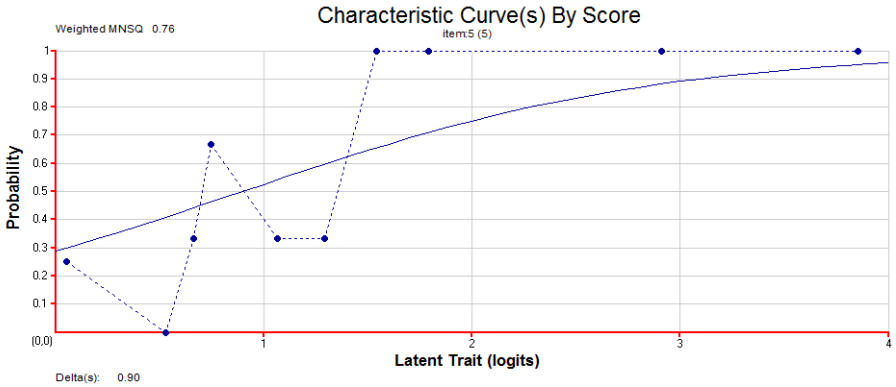

Fig 6. *Item Characteristic Curve*

Table 5

Grade 10th 2015

Question Number	Item Difficulty	Item Discrimination
Q1	0.93333	0.13333
Q2	0.23333	-0.06666
Q3	0.26666	0.13333
Q4	0.73333	0.53333
Q5	0.8333	0.3333
Q6	0.93333	0.13333
Q7	1	0
Q8	0.93333	0.13333
Q9	0.96666	0.06666
Q10	0.86666	0
Q11	0.7	0.6
Q12	0.83333	0.3333
Q13	0.43333	-0.2
Q14	0.26666	-0.13333
Q15	0.7	0.6

The table 5 of this item shows the Good fit item. This item is good discriminator in high achievers and low achievers with discrimination index 0.6 and MNSQ=0.76. The correct responses for this item is 17 which shows that the item is moderate difficulty.

Table 6
Grade 9th 2016

Question Number	Item Difficulty	Item Discrimination
Q1	0.73333	0.4
Q2	0.53333	0.8
Q3	0.6	0.2666
Q4	0.4666	0.53333
Q5	0.56666	0.46666
Q6	0.43333	0.46666
Q7	0.8	0.4
Q8	0.3333	0.26666
Q9	0.43333	0.2
Q10	0.36666	0.46666
Q11	0.53333	0.4
Q12	0.56666	0.6
Q13	0.6666	0.6666
Q14	0.3333	0
Q15	0.73333	0.26666

Table 6 shows the Item Analysis of grade 9th paper for the year of 2016. Difficuly index and dicrmination value shows that this paper is better then the 2015 paper of grade 9th. But still some items in this paper needs to be revised that is 3, 8, 9, 14.

Bad fit item in this is item no 14 because its difficult and not discriminating between the abilities level of students.

Table 7
Grade 10th 2016

Question Number	Item Difficulty	Item Discrimination
Q1	0.53333	0.53333
Q2	0.6666	0.6666
Q3	0.5666	0.3333
Q4	0.7	0.3333
Q5	0.4	0.26666
Q6	0.63333	0.3333
Q7	0.5	0.6
Q8	0.5666	0.2
Q9	0.6666	0.6666
Q10	0.56666	0.46666
Q11	0.43333	0.6
Q12	0.53333	0.5333
Q13	0.36666	0.6
Q14	0.7	0.6
Q15	0.43333	0.2

Table 7 shows the Item Analysis of grade 10th paper for the year of 2016. Difficuly index and dicrmination value shows that this paper is better then the 2015 paper of grade 10th as there is no negative discrimination value, But still some items in this paper needs to be revised that is 5, 8, 15. Bad fit item in this is item no 4 because its difficult and not discriminating is low as compared to others. These values shows that the paper is well constructed, all questions have an acceptable value of difficulty and discrimination except one or two, this maybe because all the MCQS of this paper is of knowledge and understanding category and previous researchers proves that student mark the correct answer to lower cognitive ability.

Table 8

Itemization of Correct Responses with cognitive level of domain for 2015 paper

Question Number	Level of cognitive Domain	Number of students with correct responses
Grade 9th		
1	Knowledge	7
2	Knowledge	8
3	Understanding	7
4	Knowledge	8
5	Application	3
6	Understanding	6
7	Understanding	6
8	Knowledge	6
9	Knowledge	9
10	Knowledge	7
11	Understanding	7
12	Application	5
13	Understanding	7
14	Knowledge	8
15	Knowledge	7
Grade 10th		
1	Understanding	5
2	Analysis	4
3	Knowledge	6
4	Knowledge	5
5	Knowledge	8
6	Synthesis	1
7	Analysis	2
8	Analysis	2
9	Analysis	1
10	Understanding	6
11	Understanding	6
12	Knowledge	8
13	Knowledge	8
14	Understanding	7
15	Understanding	6

This table shows the correct responses of students according to the levels of thinking process in the paper of computer science 2015 for grade 9th and 10th.The researcher didn't add the data of 2016 papers because in 2016 there is no higher order items the student performance in this table is measured according to the levels of Bloom's Taxonomy. And the results show that student performance is better with low level thinking items like knowledge, understanding. Students face difficulties in answering the correct responses for higher level questions (application, analysis, and synthesis). The findings conclude that there's a relationship between the level of thinking and the students' ability to answer them correctly.

Conclusion

It is obvious from the findings that maximum focus in all the papers were on lower order thinking skills, very less attention is given to assess the student abilities of analysis and synthesis. Evaluation level of cognitive domain of Bloom's taxonomy was totally neglected. As a result of item analysis of Computer Science, it is concluded that paper for the year of 2015 is not well constructed, there are many negative discrimination items that's need to be omitted, and as compare to 2015, papers of 2016 for grade 9th and 10th are comparatively better, as there were no negative discrimination and many items have acceptable values for difficulty index and discrimination. But it can also be due to the lower level categories questions. In 2016 paper, not a single item was measuring student abilities of analysis, synthesis and evaluation. The findings of the study further revealed that performance of the students get affected by the increase of level of thinking process advocated by Bloom. It indicates that there is a relationship between the student responses and level of thinking. In Higher level thinking items, mostly learners got the item wrong, and in lower order thinking, most of the students got the item correct. It can also be concluded that paper setter didn't develop the paper carefully, there are many unintended clues in the distractor, spelling and grammatical mistakes as well, options are not homogeneous, and questions with "not" statement also present in the paper.

Recommendations

Following recommendation were made in the light of findings:

- Items of higher order thinking skills should be included in the paper to assess the student's abilities of analysis, synthesis, and evaluation
- Paper setters should be trained in developing questions that measure different abilities of learners.
- Item having high difficulty value and low difficulty value should be excluded from the paper.
- Items should be piloted before administration

- It is suggested that while writing the MCQS avoid providing clues, keep options independent of one another, always word the stem positively and keep all options homogeneous in content.
- It is also recommended to conduct further study on it for distracter analysis.

References

Bichi, A. A. (2013). Item analysis using a derived science achievement test data. *International Journal of Science and Research*

Bloom B. S. (1956). Taxonomy of Educational Objectives. Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

Iqbal, J., Ullah, U., and Nisar. M. (2019). Analysis of Question Papers of Physics on Bloom's Taxonomy at Secondary Level, *Global Regional Review, 4*(4), 29-36.

Kehoe, Jerard (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research & Evaluation, 4*(10). http://PAREonline.net/getvn.asp?v=4&n=10.

Mukherjee, P. and Lahari, S. K (2015). Analysis of multiple choice questions (MCQs): Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. *Journal of Dental and Medical Sciences, 14*(12), 47-52.

Report of Curriculum Improvement Task Force, (2005). Ministry of Education, Government of Pakistan.

Shakil, M. (2008). Assessing Student Performance Using Test Item Analysis and its Relevance to the State Exit Final Exams of MAT0024 Classes - An Action Research Project. A Paper presented on MDC Conference Day, at MDC, Kendall Campus.

Shahzad, S., Qadoos, A., Badshah, S. N., Muhammad, H., and Ramzan, S. M. (2011). Analytical study of question papers on Bloom's taxonomy. *Interdisciplinary Journal of Contemporary Research in Business, 3*(8).

Tariq. M., Iqbal, M., Abdullah., & Farooq. M (2016). Analysis of question papers of physics at secondary level in Pakistan in the light of revised bloom's taxonomy. *Journal of Applied Environmental and Biological Sciences, 6(*4S), 124-131