Sir Syed Journal of Education & Social Research
Vol. 8, Issue 2, 2025 (April – June)
ISSN 2706-6525 (online), ISSN 2706-8285(Print)
ISSN 2706-9362(CD-ROM), ISSN 2706-6525 (ISSN-L)
DOI: https://doi.org/10.36902/sjesr-vol8-iss2-2025(36-42)

SJESR
Sir Syed Journal of Education & Social Research

Development of Computer Science Quality Based Multiple choice Questions at the Secondary Level

* Ghulam Ume Farwa, MPhil Scholar
** Dr. Fareeha Sami, Assistant Professor
*** Muhamad Touseef Khizar, MPhil Scholar

Abstract

*The study examined development of quality based test items of computer science subject at secondary level. The study was descriptive in nature with quantitative research approach. The data were collected by survey method. A self-developed test was used as instrument for study. The population of the Study was all the higher Secondary Schools of Tehsil Sargodha. A hundred (100) samples of students were selected through convenient sampling. Test items were developed by using the Table of Specifications (TOS). The validity was ensured by 5 experts having 3 PhDs and 2 Subject Specialists. The items were also developed by keeping in view the face and content validity. The reliability of the instrument was 0.775 which was drawn through MS-Excel and SPSS.10 items (1,3,5,6,11,15,24, 32,47,55) were rejected and 23 items (*2,4,14,17,18,21,23,26,28,29,30,31,33,35,36,39,40,42,44,45, 51,54,58*) were revised on the basis of difficulty and discrimination. From the conclusion it was found that test was well aligned with TOS as it was valid and according to content. Also it was concluded that test was reliable and it had the caliber to use for further analysis. so it was recommended that QAED academy may arrange the trainings for secondary school teachers to develop quality based test items.*

Keywords: TOS, validity, Reliability, QAED, Difficulty, Discrimination.

Introduction:

In the dynamic realm of education, assessment serves as the cornerstone for measuring students' knowledge, skills, and competencies. Assessment is considered as a methodical approach to collect the relevant information that helps in making decisions about teaching and learning, as well as addressing legal issues when needed (Awuah; AJARR, 2022).

In this era of technology, it is important for students to think logically by breaking down complex problems and finding step by step solutions. Apropos of this there is an increasing demand for fair and effective assessments in computer science to evaluate students' abilities and skills. Many of test formats like selection type test and supply type tests are commonly used in the classroom. Among these, selection type test such as multiple choice, true/false, matching, fill in the blanks are widely favoured. The multiple choice format, in particular, is popular due to its ease of scoring, strong content validity, and suitability for large populations. Exams like SAT, ACT, GRE, GMAT, and TOEFL often use MCQs to assess proficiency and aptitude. Researchers emphasize the benefits of multiple choice formats in disciplines such as computer science, where it is crucial to efficiently assess broad concepts and skills across many students (Maheen et al., 2022). Haladyna, Downing, & Rodriguez (2002) emphasizes that clear question statements, realistic answer options, and simple wording are very important because they reduce confusion and help the test better distinguish between high and low performing students.

Statement of Problem*:*

Teacher made classroom tests often lack content validity, resulting in an inaccurate measure of student learning. This occurs when there is a gap between what is taught and what is tested. Without proper planning, teachers find it difficult to create balanced and representative tests. As a result, many items fail to assess the required learning outcomes effectively. In Computer Science, logical thinking and problem solving are very important skills. Poorly made questions can give a wrong picture of

* Department of Education, University of Lahore Sargodha Campus Email: ghulamfarwa01@gmail.com
** Department of Education, University of Lahore Sargodha Campus Email: fareeha.sami@ed.uol.edu.pk
*** Department of Education, University of Lahore Sargodha Campus Email: touseefkhizar88@gmail.com

students' real abilities. So, it is important to create quality based multiple choice questions that are valid, reliable, and match the curriculum to ensure fair assessment of students' performance.

Significance of Study:

This study is significant because it highlights the importance of developing quality based multiple choice questions in Computer Science at the secondary level. The study provides a model for improving assessment practices by constructing test items through a structured process and by ensuring their validity and reliability. The findings can help teachers design fair and effective tests. It supports students in demonstrating their true abilities. It guides training programs for educators through institutions like QAED. Ultimately, the study contributes to raising the overall quality of assessment in Computer Science education. It may help a Head Teacher assess students' abilities more accurately by considering both item difficulty and student performance. Head teacher may improve test quality by applying it in the School. It provides policymakers with accurate data on student performance by considering item difficulty and discrimination. It may help policymakers to improve educational assessment system.

Research Objectives:

1. To develop quality based multiple choice questions (MCQs) in Computer Science at the secondary level using the Table of Specifications (TOS).
2. To establish the validity of the developed test items through expert review.
3. To determine the reliability of the test through statistical analysis.
4. To analyze the developed MCQs on the basis of difficulty and discrimination indices.

Research Questions:

1. How can quality based multiple choice questions (MCQs) in Computer Science be developed at the secondary level using the Table of Specifications (TOS)?
2. To what extent are the developed MCQs valid as judged by subject experts?
3. What the reliability of the developed test is as measured through statistical analysis?
4. How do the developed MCQs perform in terms of difficulty and discrimination indices?

Methodology:

A sample was selected using multistage sampling .On the basis of 50% selection, 10 Government Higher Secondary Schools of Tehsil Sargodha were selected randomly among 21 schools. A total of 100 students were selected from these 10 schools. 10 students from each school were selected conveniently for sample. Data were collected from students of 9^{th} class.

Research Instrument:

A **self-developed MCQ test** was constructed against a **Table of Specifications (TOS)** aligned to the Punjab Textbook Board Computer Science curriculum for Grade 9, covering core strands such as algorithms, programming basics, hardware, and data representation. Items were written with four options each, keyed, and screened for language clarity and content relevance.

In test construction, the items were distributed across different levels of Bloom's taxonomy. It consists of knowledge, comprehension and application to make quality based test. This approach is consistent with the recommendations of West, J. (2023) who stressed that tests intended to measure conceptual learning should emphasize comprehension and application level items.

Validity of Research Instrument:

The initial pool of MCQs was reviewed by a panel of five experts, including three PhD degree holders in education and two subject specialists of computer science. They examined the items for face and content validity, clarity of language, and alignment with the Grade 9 curriculum of Punjab Textbook Board. Based on their recommendations, minor corrections were made to terminology item phrasing and the appropriateness of distractors to enhance accuracy and readability. After incorporating these expert suggestions, the tool was refined and subjected to a pilot study. It enabled the researcher to conduct item analysis, discrimination analysis, and reliability analysis.

Table 1: Table Of Specification

Topics / Domains	Knowledge	Comprehension	Application and above	Total	Percentage
Problem solving	3	4	5	12	20%
TCP/IP model	2	4	5	11	18%
Topology	2	1	4	7	12%
HTML	1	1	2	4	6%

Binary system	1	2	3	6	10%
Cyber Crime	2	2	3	7	12%
Data communication	1	1	1	3	5%
Simple Encryption	1	1	1	3	5%
Ethical issues related to Data security	1	2	4	7	12%
Total	14	18	28	60	100%
Percentage	23%	30%	47%		100%

The table shows the division of items and scores with respect to selected topics. these topics were categorized as 20% (12 scores) from problem solving , 18%(11scores) from TCP/IP model, 12% (7 scores) from Topology, 6% 94 scores) from Html , 10% (6 scores) from Binary system, 5% (3 scores) from Data communication, 5% (3 scores)from simple Encryption, 12% (7 scores) from Ethical issues related to data security .The topics mentioned in the TOS were taken from the text book of computer science of Punjab textbook board grade 9 (2022-23) .total number of items were 60 which were categorized under the cognitive domain level as knowledge, comprehension and application. These items detail are as follow:

14 items (23%) were developed on knowledge level.
18 items (30%) were developed on comprehension level.
28 items (47%) were developed on application level.

Reliability of Instrument:

Cronbach alpha was calculated Manually by using formulae:

$$\frac{K}{K-1}\left(1-\frac{\sum PQ}{\sigma^2}\right)$$

$$\frac{60}{59}\left(1-\frac{11.5095}{48.43}\right)$$

$$= 0.775$$

Where K=no of items, $\sigma^2 = variance$
PQ= No of correct responses * No of Incorrect Responses
And the reliability was 0.775

SPSS Cronbach Alpha Result

Table shows item wise reliability of computer science subject test of 9th class.

Table 2:Overall Reliability of computer science 9th class test (SPSS)

Instrument	No of Items	Cronbach's Alpha	Decision
Multiple choice test	60	.775	Acceptable

Overall test reliability of the test from SPSS is 0.775 which is considered to be in acceptable range.

MS- Excel and SPSS both results are same with reliability as 0.775

A Cronbach's alpha of **0.775** lies within the commonly accepted range of **$\alpha \geq 0.70$**, which is considered acceptable for internal consistency (Taylor & Francis, 2020; Neurobehavioral Sciences, 2021; Griethuijsen et al., 2016).

Data Analysis:

Item analysis of multiple choice tests of three levels of taxonomy:
Knowledge=K; Comprehension=C; Application=A

Table 3: decision on the basis of difficulty and discrimination indices

Sr#	Level	Difficulty level	Discrimination	Remarks
2	K	0.76	0.22	Revised
4	A	0.72	0.22	Revised
7	C	0.74	0.37	Accepted
8	C	0.75	0.3	Accepted
9	A	0.72	0.48	Accepted
10	C	0.55	0.3	Accepted
12	C	0.72	0.3	Accepted
13	K	0.73	0.3	Accepted
14	K	0.79	0.22	Revised

16	C	0.74	0.37	Accepted
17	C	0.74	0.26	Revised
18	A	0.71	0.22	Revised
19	C	0.71	0.44	Accepted
20	A	0.61	0.48	Accepted
21	A	0.72	0.22	Revised
22	A	0.73	0.33	Accepted
23	A	0.73	0.26	Revised
25	K	0.74	0.33	Accepted
26	A	0.74	0.22	Revised
27	C	0.73	0.3	Accepted
28	A	0.7	0.22	Revised
29	A	0.75	0.26	Revised
30	A	0.72	0.26	Revised
31	K	0.76	0.26	Revised
33	A	0.75	0.26	Revised
34	A	0.7	0.44	Accepted
35	A	0.73	0.22	Revised
36	K	0.73	0.26	Revised
37	C	0.75	0.3	Accepted
38	C	0.75	0.41	Accepted
39	A	0.7	0.22	Revised
40	A	0.69	0.22	Revised
41	A	0.74	0.37	Accepted
42	A	0.71	0.22	Revised
43	C	0.71	0.3	Accepted
44	C	0.72	0.26	Revised
45	K	0.68	0.22	Revised
46	A	0.74	0.3	Accepted
48	K	0.73	0.37	Accepted
49	C	0.69	0.44	Accepted
50	A	0.66	0.37	Accepted
51	K	0.75	0.22	Revised
52	C	0.74	0.37	Accepted
53	C	0.74	0.37	Accepted
54	A	0.71	0.22	Revised
56	A	0.74	0.3	Accepted
57	A	0.75	0.41	Accepted
58	K	0.75	0.22	Revised
59	C	0.69	0.48	Accepted
60	A	0.7	0.3	Accepted

Table presents item-wise analysis of difficulty and discrimination indices across knowledge (K), application (A), and comprehension (C) levels. 10 items (1,3,5,6,11,15,24,32,47,55) were rejected and 23 items (2,4,14,17,18,21,23,26,28,29,30,31,33,35,36,39,40,42,44,45,51,54,58) were revised on the basis of item difficulty and item discrimination.

Item Difficulty Index (P)

The item difficulty index tells us how easy or difficult a question is. It is the proportion of students who answered the item correctly. The value of P ranges from 0 to 1. A higher value means the item is easier, and a lower value means it is more difficult. The formula used for item difficulty is:

$$\mathbf{P} = \frac{R}{\mathrm{N}}$$

Where

P = Difficulty index

R = Number of students who got the item right

N = Total number of students who attempted the item

There are three levels of item difficulty index, as presented here:

Table 4: Item Difficulty Index Range

Range (P)	Decision
0.00 – 0.25	Rejected
0.26 – 0.75	Accepted
0.76 – 1.00	Revised

(Source: Gul, N., Shagufta, S., &Parveen, S. (2022). Item Difficulty in Item Analysis of Intelligence Test Items. Pakistan Social Sciences Review, 6(2), 97–108. https://ojs.pssr.org.pk/journal/article/view/117)

Table 5: Item difficulty of MCQs of Computer Science Subject Test:

P	Decision	Item No	Total
0.00-0.25	Rejected	Nil	-
0.26-0.75	Accepted	1,3,4,5,6,7,8,9,10,11,12,13,15,16,17, 18,19,20,21,22,23,24,25,26,27,28,29,30, 32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52 ,53,54,55,56,57,58.59.60	57
0.76-1	Rewrite	2,14,31	3
		Total	60

The analysis of item difficulty (P-value) revealed that no item fell in the very difficult range (0.00–0.25), while the majority of items (57 out of 60) were within the acceptable range (0.26–0.75). Only 3 items (2, 14, and 31) showed high difficulty values (0.76–1.00) and were recommended for revision. This indicates that most test items were of moderate difficulty and appropriate for assessing student performance.

Item Discrimination Index (D)

The discrimination index tells us how well a question separates high-performing students from low-performing students. A higher D value means the item is good at distinguishing between strong and weak students. It is commonly used in test analysis and Item Response Theory (IRT).the formula for discrimination index $\mathbf{D} = \frac{x_h - x_l}{N}$

$\bar{\mathbf{X}}_\mathbf{h}$ = Average score of the high achievers (e.g., top 27% students)

$\bar{\mathbf{X}}_\mathbf{l}$= Average score of the low achievers (e.g., bottom 27% students)

N = Total number of students

There are four levels of item discrimination index, as presented in Table below.

Table 6: Item Discrimination Index Range

Range (D)	Status
≤ 0.19	Rejected
0.20 – 0.29	Revised
0.30 – 0.39	Marginally Accepted
≥ 0.40	Accepted

(Source: Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., &Rizvi, M. (2017).Difficulty index, discrimination index and distractor efficiency in multiple choice questions. Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University, 13(4), 310–315.)

Table 7: Item Discrimination of MCQs of Computer Science subject Test

D	Decision	Item No	Total
≤ 0.19	Rejected	1,3,5,6,11,15,24,32,47,55	10
0.20-0.29	Rewrite	2,4,14,17,18,21,23,26,28,29,30,31,33,35,36,39,40,42,44,45,51,54 ,58	23
0.30-0.39	Marginally Accepted	7,8,10,12,13,16,22,25,27,37,41,43,46,48,50,52,53,56,60	19
≥ 0.40	Accepted	9,19,20,34,38,49,57,59	8
		Total	60

The item discrimination analysis shows that 10 items (D ≤ 0.19) were rejected due to poor ability to differentiate between high and low achievers. A total of 23 items (D = 0.20–0.29) required revision, while 19 items (D = 0.30–0.39) were marginally accepted. 8 items (D ≥ 0.40) were accepted, demonstrating strong discrimination power. Overall, the test contained a mix of acceptable and revisable items, with a substantial proportion showing potential for improvement.

Conclusion:

The study successfully developed quality based multiple choice questions (MCQs) for Grade 9 Computer Science using a structured Table of Specifications (TOS). It was concluded that the test was found to be both valid and reliable. Most items fall in the acceptable range of difficulty. However, some items required revision or rejection due to weak discrimination indices. Overall, the test reflected the curriculum objectives effectively. It has the potential to serve as a model for developing standardized assessments in Computer Science at the secondary level.

It was concluded that the study highlight the importance of systematic item construction and analysis in educational assessments. The majority of items were of moderate difficulty (P = 0.25–0.75). They were neither too easy nor too hard for students and items supports balanced assessment. Similar results have been reported in previous studies, where items within this range were considered appropriate for measuring student knowledge. (Gul et al, 2022).

In discrimination indices 18 items demonstrated strong discrimination ($D \geq 0.40$), a considerable number of items either needed revision or were rejected. This finding is similar with Mahjabeen et al. (2017), who emphasized that discrimination indices are very important to use in ensuring that test items measure student learning accurately. The overall reliability of 0.775 indicated that the test had acceptable internal consistency. It is aligned with the commonly accepted threshold of $\alpha \geq 0.70$ (Tavakol & Dennick, 2011).

Recommendations:

- Teachers should undergo professional training to ensure quality based item development.
- Items with low discrimination indices should be revised to enhance the effectiveness of the test.
- Institutions like QAED may arrange regular workshops for secondary school teachers on developing valid and reliable MCQs.

Advanced models such as Item Response Theory (IRT) can be applied in future research to provide deeper insights information about the items.

References:

Awuah, L. J. (2022). Assessment for learning: A review of the research evidence. *African Journal of Applied Research, 8*(2), 27–39.

Batool, I., Shah, A. A., &Naseer, S. (2023). Construction, analysis and calibration of multiple-choice questions: IRT versus CTT. *Advanced Research in Education and Society, 6*(2), 56–65.https://doi.org/10.53575/ares.v6.69

Farooq, M., &Mashood, S. (2023).Quality assurance of multiple-choice questions test through item analysis.*Life and Science, 4*(2), 124–130. https://doi.org/10.37185/LnS.1.1.315

Fatima, S., Babar, M. S., Tariq, S., Hassan, S., &Qasim, R. (2024). The relationship of difficulty and discrimination indices with physiology teaching and learning.*Pakistan Journal of Medicine and Dentistry, 13*(1), 25–32. https://doi.org/10.36283/pjmd13-1/005

Gul, N., Shagufta, S., &Parveen, S. (2022). Item difficulty in item analysis of intelligence test items.*Pakistan Social Sciences Review, 6*(2), 97–108. https://doi.org/10.35484/pssr.2022(6-II)09

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002).A review of multiple-choice item-writing guidelines for classroom assessment.*Applied Measurement in Education, 15*(3), 309–334.

Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., &Rizvi, M. (2017).Difficulty index, discrimination index and distractor efficiency in multiple choice questions.*Annals of PIMS-ShaheedZulfiqar Ali Bhutto Medical University, 13*(4), 310–315.

Rahim, M. F., Bham, S. Q., Khan, S., Ansari, T., & Ahmed, M. (2023).Improving the quality of MCQs by enhancing cognitive level and using psychometric analysis.*Pakistan Journal of Health Sciences, 4*(05), 141–146. https://doi.org/10.54393/pjhs.v4i05.700

Ramzan, M., Khan, K. W., Bibi, S., & Imran, S. S. (2021). Difficulty and discrimination analysis of end-of-term multiple-choice questions at Community Medicine Department, Wah Medical College.*Life and Science, 2*(1), 29–33. https://doi.org/10.37185/LnS.1.1.53

Sami, F., Malik, M. A., &Cheema, A. B. (2023). Enhancing Critical Thinking in Physics through Real Engagement in Active Problem Solving (REAPS) Model. *Research Journal of Social Sciences and Economics Review, 4*(2), 47-53.

Tavakol, M., &Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

West, J. (2023). Designing tests for conceptual learning: Balancing knowledge, comprehension, and application. *Journal of Educational Measurement and Evaluation, 40*(1), 45–59.