

Review began 05/04/2026
Review ended 05/10/2026
Published 05/19/2026

© Copyright 2026

Mondal et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.109190

ItemSight: Development of a Shiny Application for Simplified Item Analysis of Type A Multiple-Choice Questions

Himel Mondal ¹, Pradosh Kumar Sarangi ², Shaikat Mondal ³, Anupkumar D. Dhanvijay ¹, Rintu K. Gayen ⁴, Aritra Mukhopadhyay ⁴, Raunak Mishra ⁴

1. Physiology, All India Institute of Medical Sciences, Deoghar, Deoghar, IND 2. Radiodiagnosis, All India Institute of Medical Sciences, Deoghar, Deoghar, IND 3. Physiology, Raiganj Government Medical College and Hospital, Raiganj, IND 4. Electronics and Communication Engineering, Institute of Engineering and Management, Kolkata, IND

Corresponding author: Himel Mondal, himelmkg@gmail.com

Abstract

Item analysis is a vital step in educational assessment to check the quality of multiple-choice questions (items). Traditional item analysis methods using spreadsheets or statistical software are often time-consuming, error-prone, and challenging for educators with limited programming expertise. To address these issues, ItemSight, an intuitive software for item analysis, was developed to automate and simplify the psychometric evaluation of test items. The software was developed using the R-Shiny framework. It computes item difficulty, discrimination index, distractor efficiency, internal consistency (if an item is deleted), and point-biserial correlation. To run the software, users need R and RStudio with installed packages and the code we provide. After analysis, the report can be easily exported in comma-separated values (CSV) format for further analysis. All data remains stored locally on the user's computer, ensuring privacy. Currently, the software supports only type A multiple-choice questions (four options with a single best answer). It can be accessed and used through RStudio without any cost. The tool serves as a valuable resource for educators and researchers in medical and allied health sciences, promoting evidence-based improvement of assessment quality.

Categories: Medical Education

Keywords: educational assessment, item analysis, multiple-choice questions, reliability, shiny application, test quality

Introduction

Item analysis is a systematic approach used to evaluate the quality and performance of examination questions, particularly multiple-choice questions (MCQs), which are widely used in medical education because of their objectivity, reliability, and ability to assess a broad range of knowledge domains [1,2]. In item analysis, several psychometric indicators are examined to determine whether a question functions appropriately. The difficulty index reflects the proportion of students answering an item correctly, indicating whether the question is too easy or too difficult. The discrimination index indicates how effectively an item differentiates between high- and low-performing students, while reliability coefficients assess the internal consistency of the overall examination [3,4]. For example, a question answered correctly by nearly all students may contribute little to distinguishing competency levels, whereas a question predominantly answered correctly by high-performing students is generally considered a high-quality item.

High-quality MCQs are essential because assessment outcomes directly influence student progression, competency evaluation, curriculum review, and accreditation standards in medical education [5]. Poorly constructed questions may produce misleading estimates of student ability, compromise examination validity, and result in unfair pass-fail decisions. Item analysis helps educators identify flawed or ambiguous questions, revise ineffective distractors, and improve the overall quality of assessment programs [6,7].

Despite its recognized importance, the routine implementation of item analysis remains challenging in many educational settings. Previous studies have identified barriers such as limited faculty training in psychometrics, insufficient time, lack of institutional support, and difficulties in using statistical software for routine assessment evaluation [8]. In practice, educators often conduct item analysis using spreadsheet software such as Microsoft Excel or dedicated psychometric software. However, spreadsheet-based workflows typically require manual formula entry and repetitive data handling, increasing the risk of calculation errors and making the process time-consuming. Advanced statistical software may offer comprehensive analytical capabilities, but it often requires licensing costs and familiarity with statistical procedures, limiting accessibility for educators with limited technical expertise [9].

To address these challenges, we developed ItemSight, an interactive R Shiny-based application designed to automate item analysis within a user-friendly interface. The application enables educators to calculate and

How to cite this article

Mondal H, Sarangi P, Mondal S, et al. (May 19, 2026) ItemSight: Development of a Shiny Application for Simplified Item Analysis of Type A Multiple-Choice Questions. Cureus 18(5): e109190. DOI 10.7759/cureus.109190

interpret commonly used psychometric indices without requiring advanced statistical knowledge, thereby facilitating routine quality assessment of MCQs in medical education.

This article was presented at the 12th State Association of Physiologists and Pharmacologists of India Conference (APPICON) 2026, conducted on February 6-7, 2026, at Calcutta National Medical College and Hospital, Kolkata, India.

Technical Report

Parts of MCQ

A well-constructed MCQ consists of several key parts that together ensure clarity and effectiveness in assessment. The parts are described briefly in Table 1 [10].

Part	Description	Example
Stem	The question or statement that presents the problem or task to be answered	A 25-year-old male presents with a sudden onset of chest pain and shortness of breath after lifting heavy weights. On examination, he has decreased breath sounds on the right side. A chest X-ray reveals a collapsed right lung with a visible pleural line.
Lead-in	The specific question or instruction directing the student what to do (sometimes part of the stem)	What is the most likely diagnosis?
Options/choices	The set of possible answers provided, including one correct answer and several distractors	A. Pulmonary embolism B. Acute myocardial infarction C. Spontaneous pneumothorax D. Bronchial asthma
Key/correct answer	The single best or correct option among the choices	C. Spontaneous pneumothorax
Distractors	The incorrect but plausible options are designed to distract those who do not know the correct answer	A. Pulmonary embolism B. Acute myocardial infarction D. Bronchial asthma

TABLE 1: Parts of a multiple-choice question

This is an example of a multiple-choice question with a single best answer and three distractors.

Item analysis metrics

The item analysis metrics that form the basis of the Shiny application's calculation methods are shown in Table 2 and described below.

Metric	Formula/method	Explanation
Difficulty index (p-value)	Number of correct responses ÷ total number of students × 100	Percentage of students who answered the item correctly
Difficulty index (27% method)	$(H + L) \div N \times 100$	H = correct answers in top 27% scorers; L = correct answers in bottom 27% scorers; N = total students in upper and lower 27%
Discrimination index	$2 \times (H - L) \div N$	
Distractor efficiency	Functional distractors ÷ 3 × 100	A distractor is "functional" if selected by at least 5% of students
Point-biserial correlation	Correlation between item score (0 or 1) and total test score	Pearson's correlation with one dichotomous variable
Cronbach's alpha	$(k \div (k - 1)) \times (1 - (\text{sum of item variances} \div \text{variance of total scores}))$	k = total number of items

TABLE 2: Calculation method of item analysis metrics

This is only for multiple-choice questions with four response options (three distractors and one answer key).

Difficulty Index (Proportion Correct/P-Value)

The difficulty index is calculated as the proportion of students who answered an item correctly. It is obtained by dividing the number of correct responses by the total number of students and expressing it as a percentage [11]. This measure is calculated to know how easy or difficult an item is. A higher value means the item is easier, while a lower value indicates more difficulty. It gives us an idea of whether the item is appropriate for the level of students being assessed.

Difficulty Index (27% Method)

This method compares item performance between the top 27% of scorers and the bottom 27% [12]. The number of correct responses in both groups is added and divided by the total number of students in these groups. This calculation is used because extreme groups provide a clearer picture of how well an item differentiates ability levels. It gives us an idea of the item's effectiveness at distinguishing stronger from weaker students. This method is more commonly used.

Discrimination Index

The discrimination index is calculated as the difference in correct responses between the top 27% and bottom 27% groups, standardized by the total number of students in those groups [13]. It is calculated to see how well an item separates high-performing from low-performing students. A high positive value means the item is good at discrimination, while a negative value suggests a faulty question. This gives us an idea about the quality of the item in terms of its ability to differentiate student performance.

Distractor Efficiency

Distractor efficiency is based on how many of the wrong options (distractors) are chosen by at least 5% of students [14]. The percentage of functional distractors is calculated for each item. It is calculated to check if the incorrect options are plausible enough to attract students who do not know the answer. This gives us an idea of whether distractors are effective in testing knowledge or if they are too obvious to be avoided.

Point-Biserial Correlation

The point-biserial correlation is the Pearson correlation between the item score (0 for wrong; 1 for correct) and the total test score [15]. It is calculated to measure the consistency of an item with the overall performance of students. A high positive correlation means that students who did well overall also got this item correct, indicating a good item. This gives us an idea about the alignment of the item with the overall test quality.

Cronbach's Alpha (and Alpha if Item Deleted)

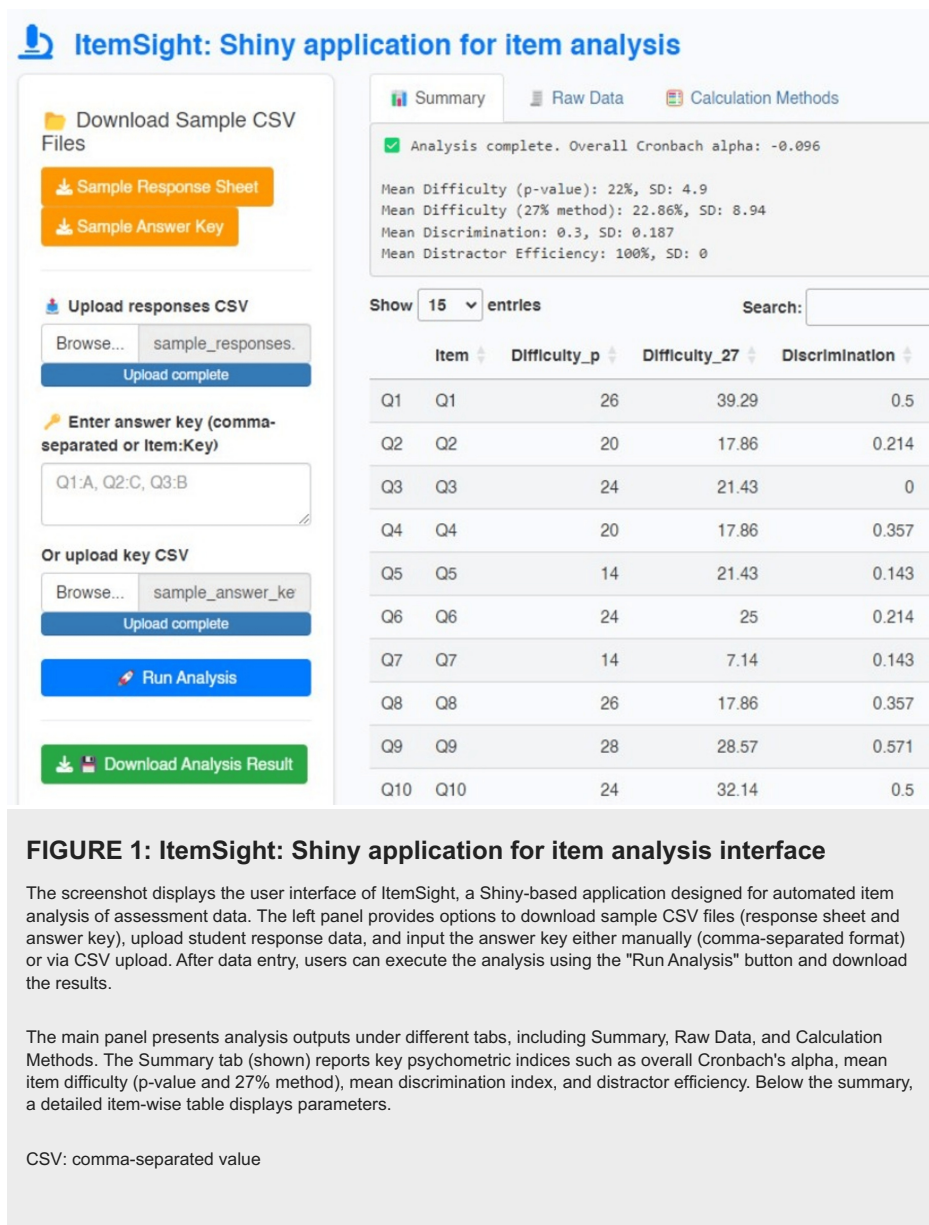
Cronbach's alpha is calculated using item variances and the variance of the total test score. It measures the internal consistency or reliability of the test as a whole. It is calculated to assess whether the items collectively measure the same construct. A higher alpha indicates more reliable items [16]. The "alpha if item deleted" shows how the overall reliability would change if one item were removed, giving us an idea of which items weaken the test's consistency. In item analysis, the Kuder-Richardson Formula 20 (KR-20) is used for measuring the reliability [2]. If the data is dichotomous (1 and 0 in the MCQ score), then both yield the same result [17]. In this application, Cronbach's alpha is being calculated.

Development of the application

The ItemSight application was developed using the R programming language and the Shiny framework to provide an interactive and accessible platform, and it is being utilized in biomedical application development [18]. The app integrates several R packages, including Shiny for building the user interface and server logic, DT for rendering interactive tables, dplyr, and tidyr for efficient data manipulation. The code of the application is available from Mendeley Data [19].

Ten sets of MCQ response sheets were created to evaluate the software's performance. Each set contained varying numbers of student responses (100-250) and MCQ items (20-100). For every dataset, both the students' MCQ answers and the corresponding answer keys were entered into the software. To assess accuracy, five of the 10 response sets were also scored manually using the same answer keys. Manual calculations were independently performed and cross-verified by the authors, who have prior experience (>5 years) in medical education research, psychometric assessment, and item analysis methodology. Since the application is based on established mathematical formulas with deterministic outputs, the possibility of computational error is virtually negligible when calculations are performed correctly. The manually derived item-wise scores were then compared with the software-generated outputs. Concordance between manual calculations and the software results was examined to ensure the accuracy of the statistical analysis. As there was no difference between manually calculated values and software-generated values, no statistical test was applied.

The workflow was performed on a standard ASUS Vivobook (X1502ZA) laptop. The system is powered by a 12th Generation Intel Core i3-1215U processor with 16 GB RAM. It includes integrated 128 MB Intel UHD Graphics and 477 GB storage. The device runs on a 64-bit Windows 11 operating system. ItemSight successfully processed (analyzed data without software glitch) all 10 datasets without errors, indicating that the software is capable of handling both moderate and large volumes of data as designed by its underlying algorithm. The analysis was done without any visible time lag. One such output is shown in Figure 1.



Across all tested response sets, the item statistics by the software were fully consistent with those obtained through standard manual calculations. The generated item analysis report can be exported in comma-separated values (CSV) format, allowing users to conduct additional statistical analysis or archival storage as needed. All files, namely, student responses, answer keys, and generated reports, remain stored locally on the user's computer, thereby ensuring data confidentiality and eliminating reliance on external servers.

How to use the application

Before using the ItemSight Shiny application, users need to have R (version 4.0 or above recommended) and RStudio installed on their computer. The app requires several R packages to be installed before running. The application code contains a prompt to install all dependencies. Hence, the first time, it will take some time to install the packages.

To use ItemSight, users need to open RStudio (Figure 2) and copy the code from the downloaded Microsoft Word document and paste it into the RStudio console and press Enter on a Windows PC or Return on a Macintosh computer.

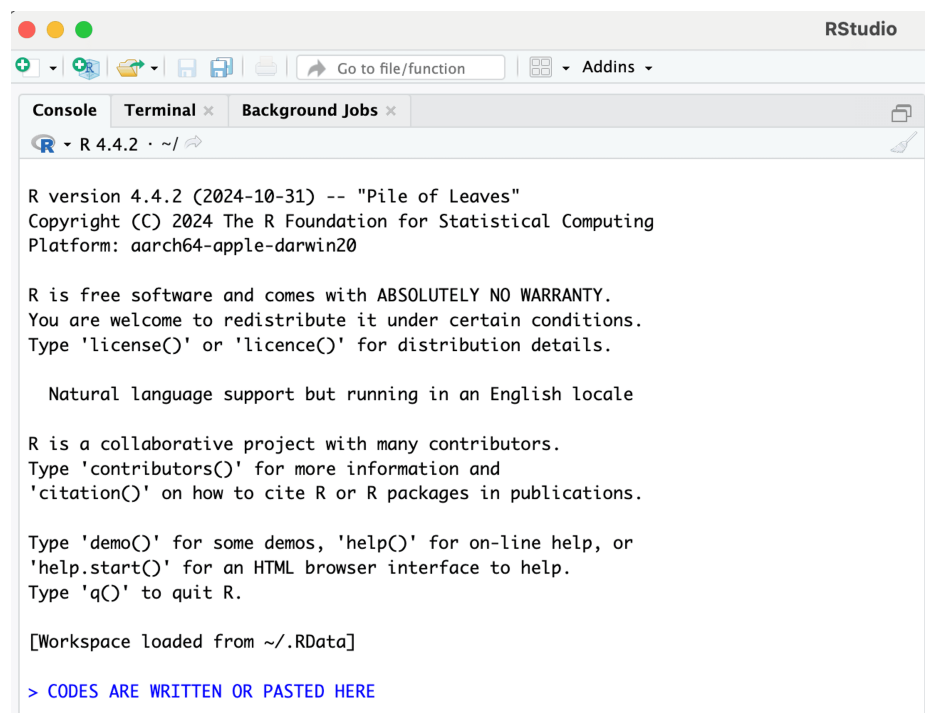


FIGURE 2: RStudio console showing the place where code is written or pasted from elsewhere

The screenshot depicts the console panel of RStudio with R version 4.4.2 ("Pile of Leaves") successfully launched. The interface displays startup information, including platform details, licensing notice, and guidance commands (e.g., `help()`, `citation()`, `demo()`). The console (bottom-left pane) is the primary workspace where users write or paste R code, execute commands, and view outputs. Tabs for Console, Terminal, and Background Jobs are visible, indicating support for interactive coding, system-level commands, and parallel task execution. This environment serves as the foundational workspace for developing scripts, running analyses, and building Shiny applications.

To use the ItemSight app, the user needs to have two files in CSV format: a spreadsheet containing the students' responses (each row representing a student and each column representing an item) and a corresponding answer key file. The user needs to first download the provided "Sample Response Sheet" and "Sample Answer Key" from within the app (Figure 1), which can be opened in Microsoft Excel or similar spreadsheet applications. The user needs to enter their own student responses in the response sheet and fill in the correct answers in the answer key.

After preparing the files, the user needs to upload the response sheet by clicking the Browse button under "Upload responses CSV", followed by uploading the answer key through the Browse button under "Or upload key CSV". Next, the user needs to click on Run Analysis, after which the results will automatically appear under the Summary tab, showing difficulty indices, discrimination index, point-biserial correlation, distractor efficiency, and Cronbach's alpha (Figure 1). Finally, to save the output, the user needs to click on Download Analysis Result, which will generate a CSV file containing the complete item analysis.

Discussion

The application offers several advantages for teachers and researchers conducting item analysis. It automates the entire process, allowing users to upload student responses and answer keys and instantly obtain key metrics such as difficulty, discrimination, distractor efficiency, point-biserial correlation, and reliability. This saves time and minimizes calculation errors compared to manual methods. The interface is user-friendly, and teachers only require moderate computer skills and no programming knowledge. The app also allows downloading of results for further review or reporting, making it practical for routine assessment evaluation. These data can also be used for research purposes. In addition, RStudio and this application run on a personal computer, and no data is shared with anyone. Hence, the data security is there.

The current version of the application has some practical limitations. It is designed to handle type A MCQs with fixed options (A-D) [20] and does not support other response formats like three-option or five-option MCQ or partial marking. The app requires properly formatted CSV files for input, and any mismatch in file structure or key entry can lead to errors. It does not include advanced error-checking or automated handling of incomplete or irregular datasets, so the user must ensure data quality before analysis. The app also runs

locally in R Shiny, which means performance may slow down when analyzing very large datasets.

We did not include a built-in interpretation facility, as different institutions may adopt varying criteria to categorize MCQs [21,22]. The application provides the raw scores, allowing educators or administrators to make further decisions regarding item quality or classification according to the type of examination and the respective institutional policy.

How to use the application for educational and research purposes

The Shiny application is an open-source application, and anyone can use it for any purpose. No permission is required for using the application for educational purposes. Researchers, while using the application for analysis and publication of the results, can cite this research paper for attribution. For any help, they can contact the corresponding author of the article.

Conclusions

ItemSight provides a freely accessible platform for conducting item analysis of MCQs. By integrating key metrics such as difficulty index, discrimination index, distractor efficiency, point-biserial correlation, and Cronbach's alpha into a single user-friendly Shiny interface, the application reduces the burden of manual calculation and ensures accuracy in assessment evaluation. Its ability to generate structured outputs and downloadable reports makes it a practical tool for teachers and researchers seeking evidence-based improvement of test items. While it has certain technical limitations, ItemSight offers a valuable first-line solution for routine item analysis and supports the broader goal of enhancing the quality of student assessments through data-driven insights.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Pradosh Kumar Sarangi, Himel Mondal, Anupkumar D. Dhanvijay, Rintu K. Gayen, Aritra Mukhopadhyay, Raunak Mishra, Shaikat Mondal

Acquisition, analysis, or interpretation of data: Pradosh Kumar Sarangi, Himel Mondal, Anupkumar D. Dhanvijay, Rintu K. Gayen, Aritra Mukhopadhyay, Raunak Mishra, Shaikat Mondal

Drafting of the manuscript: Pradosh Kumar Sarangi, Himel Mondal, Anupkumar D. Dhanvijay, Rintu K. Gayen, Aritra Mukhopadhyay, Raunak Mishra, Shaikat Mondal

Critical review of the manuscript for important intellectual content: Pradosh Kumar Sarangi, Himel Mondal, Anupkumar D. Dhanvijay, Rintu K. Gayen, Aritra Mukhopadhyay, Raunak Mishra, Shaikat Mondal

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.

Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Bhattacharjee S, Mukherjee A, Bhandari K, Rout AJ: Evaluation of multiple-choice questions by item analysis, from an online internal assessment of 6th semester medical students in a rural medical college, West Bengal. *Indian J Community Med.* 2022, 47:92-5. [10.4103/ijcm.ijcm_1156_21](https://doi.org/10.4103/ijcm.ijcm_1156_21)
2. Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V: Item analysis of multiple choice questions: a quality assurance test for an assessment tool. *Med J Armed Forces India.* 2021, 77:S85-9. [10.1016/j.mjafi.2020.11.007](https://doi.org/10.1016/j.mjafi.2020.11.007)
3. AlKhatib HS, Brazeau G, Akour A, Almuhaissen SA: Evaluation of the effect of items' format and type on psychometric properties of sixth year pharmacy students clinical clerkship assessment items. *BMC Med Educ.* 2020, 20:190. [10.1186/s12909-020-02107-3](https://doi.org/10.1186/s12909-020-02107-3)
4. Foster RC: KR20 and KR21 for some nondichotomous data (it's not just Cronbach's alpha) . *Educ Psychol Meas.* 2021, 81:1172-202. [10.1177/0013164421992535](https://doi.org/10.1177/0013164421992535)
5. Parekh P, Bahadoor V: The utility of multiple-choice assessment in current medical education: a critical review. *Cureus.* 2024, 16:e59778. [10.7759/cureus.59778](https://doi.org/10.7759/cureus.59778)

6. Bhat SK, Prasad KH: Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: a cross-sectional study. *Indian J Ophthalmol*. 2021, 69:343-6. [10.4103/ijo.IJO_1610_20](https://doi.org/10.4103/ijo.IJO_1610_20)
7. Mondal H, Kumari A, Priya A, Dhanvijay AK, Gayen RK, Mondal S, Sarangi PK: Psychometric properties of multiple-choice questions used in undergraduate medical education in India: a systematic review and meta-analysis. *Cureus*. 2025, 17:e99299. [10.7759/cureus.99299](https://doi.org/10.7759/cureus.99299)
8. Karthikeyan S, O'Connor E, Hu W: Barriers and facilitators to writing quality items for medical school assessments - a scoping review. *BMC Med Educ*. 2019, 19:123. [10.1186/s12909-019-1544-8](https://doi.org/10.1186/s12909-019-1544-8)
9. Ames A: Item analysis. *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc., Thousand Oaks (CA); 2018. [10.4135/9781506326139.n357](https://doi.org/10.4135/9781506326139.n357)
10. Gottlieb M, Bailitz J, Fix M, Shappell E, Wagner MJ: Educator's blueprint: a how-to guide for developing high-quality multiple-choice questions. *AEM Educ Train*. 2023, 7:e10836. [10.1002/aet2.10836](https://doi.org/10.1002/aet2.10836)
11. Chauhan GR, Chauhan BR, Vaza JV, Chauhan PR: Relations of the number of functioning distractors with the item difficulty index and the item discrimination power in the multiple choice questions. *Cureus*. 2023, 15:e42492. [10.7759/cureus.42492](https://doi.org/10.7759/cureus.42492)
12. Kim YH, Kim BH, Kim J, Jung B, Bae S: Item difficulty index, discrimination index, and reliability of the 26 health professions licensing examinations in 2022, Korea: a psychometric study. *J Educ Eval Health Prof*. 2023, 20:31. [10.3352/jeehp.2023.20.31](https://doi.org/10.3352/jeehp.2023.20.31)
13. Dhanvijay AK, Dhokane N, Balgote S, Kumari A, Juhi A, Mondal H, Gupta P: The effect of a one-day workshop on the quality of framing multiple choice questions in physiology in a medical college in India. *Cureus*. 2023, 15:e44049. [10.7759/cureus.44049](https://doi.org/10.7759/cureus.44049)
14. Rezigalla AA, Eleragi AM, Elhussein AB, et al.: Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Med Educ*. 2024, 24:445. [10.1186/s12909-024-05433-y](https://doi.org/10.1186/s12909-024-05433-y)
15. Ali SH, Ruit KG: The impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. *Perspect Med Educ*. 2015, 4:244-51. [10.1007/s40037-015-0212-x](https://doi.org/10.1007/s40037-015-0212-x)
16. Tavakol M, Dennick R: Making sense of Cronbach's alpha. *Int J Med Educ*. 2011, 2:53-5. [10.5116/ijme.4dfb.8dfd](https://doi.org/10.5116/ijme.4dfb.8dfd)
17. Anselmi P, Colledani D, Robusto E: A comparison of classical and modern measures of internal consistency. *Front Psychol*. 2019, 10:2714. [10.3389/fpsyg.2019.02714](https://doi.org/10.3389/fpsyg.2019.02714)
18. Meyer EL, Kumaus C, Majka M, Koenig F: An interactive R-Shiny app for quickly visualizing a tidy, long dataset with multiple dimensions with an application in clinical trial simulations for platform trials. *SoftwareX*. 2023, 22:101347. [10.1016/j.softx.2023.101347](https://doi.org/10.1016/j.softx.2023.101347)
19. ItemSight: A Shiny Application for Item Analysis. (2026). Accessed: 2026: <https://data.mendeley.com/datasets/3pdn4c3gph/2>.
20. Al-Rukban MO: Guidelines for the construction of multiple choice questions tests. *J Family Community Med*. 2006, 13:125-33.
21. Understanding item analyses. (2026). Accessed: 2026: <https://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/>.
22. Rudolph MJ, Daugherty KK, Ray ME, Shuford VP, Lebovitz L, DiVall MV: Best practices related to examination item construction and post-hoc review. *Am J Pharm Educ*. 2019, 83:7204. [10.5688/ajpe7204](https://doi.org/10.5688/ajpe7204)