https://doi.org/10.29309/TPMJ/2026.33.05.9843 www.theprofesional.com

ORIGINAL ARTICLE

# Item analysis with precise information: An item information function approach in IRT.

Zeeshan Ahmed[1], Anjum Rehman[2], Umme Habiba Rehman[3], Muhammad Maaz Sajid[4], Fasih Ur Rehman[5], Malik Moiz Uddin[6]

**ABSTRACT… Objective:** To identify valid and reliable items providing precise information regarding learner ability to be included in the test pool. **Study Design:** Descriptive Cross-sectional study. **Setting:** SMBB Medical College Lyari Karachi. **Period:** December 2024. **Methods:** Purposive sampling techniques are employed to collect data from 100 MBBS students using physiology SBAs at C1, C2 and C3 level from formative assessment of 2nd Professional MBBS Gastrointestinal tract module held in a public sector medical institute SMBB Medical College, Karachi. The study employed 2 PL dichotomous IRT model for item analysis using discrimination and difficulty index with subsequent application of item information function (IIF) technique to gauge items that rendered precise information employing statistical software STATA 17. **Results:** The item analysis using discrimination (DI) and difficulty index (DIF) found only one acceptable and two desirable items at C1 level while rest are redundant. There is only one desirable item at C2 level while rest are discarded, at C3 level two items are found acceptable, one desirable and rest are discarded. However, subsequent analysis with IIF shows that only one desirable item from C1, C2 and C3 level is appropriate for inclusion in the test pool whilst others need to be revised. **Conclusion:** The study considers item information function (IIF) as a supporting approach to a traditional item analysis for building a quality test pool. It is also established that even acceptable items that are possible candidates for inclusion in test pool by item analysis subsequently disqualified by IIF as are failed to submit precise information regarding learner's ability whereas only one desirable item from C1, C2 and C3 level appeared to provide precise information hence could be included in the test pool.

**Key words:** 2 PL Dichotomous Model, Discrimination and Difficulty Index, IIFs, Kuder Richardson Coefficient.

**Article Citation:** Ahmed Z, Rehman A, Rehman U, Sajid MM, Fasih Ur Rehman, Moiz Uddin M. Item analysis with precise information: An item information function approach in IRT. Professional Med J 2026; 33(05):924-929. https://doi.org/10.29309/TPMJ/2026.33.05.9843

## INTRODUCTION

Contemporary trends in medical education demand innovative, novel and efficient assessment tools for precise evaluation of learner's ability at both undergraduate and postgraduate level. According to Guilbert[1] assessment methodology should be determined and synchronized in accordance with the learning objectives and teaching methodologies as "Assessment is the tail that wags the curriculum dog" which persuades student learning attitude and focus as well.[2] Assessment is a formal action to obtained information about student ability and performance with underlying objectives to develop educational design and content, evaluating and ranking learners ability and performance based on test score.[3] The disposition and attribute of such information largely determine improvement in attributes of the assessment tools.[4] Recent trends in medical education focused on lifelong learning, performance and assessment of learners with enhanced cognitive and learning strategies to gauge knowledge, skills and understanding especially in undergraduate medical education. A valid and reliable assessment tool is sought to properly gauge student continuous improvement by ranking student latent traits hence providing timely feedback to learners and teachers about strengths and weaknesses in attaining learning goals.

The item drafting/designing is based on cognitive levels of learners that need to be assessed, Bloom's taxonomy[5] has underlined 6 levels of cognition with prime focus on first 3 levels i.e. C1 (knowledge), C2 (understanding) & C3 (application) which are used for teaching and assessment in undergraduate medical education. The C1-C3 level is assessed with key tool viz a viz SBAs'/MCQ. In order to promote efficient use of assessment settings the first two levels of Miller's pyramid[6] (Knows, Knows how) confer SBA/MCQ as competent instruments

1. MBBS, MHPE, Director Medical Education, Shaheed Mohtarma Benazir Bhutto Medical College Tannery Road, Lyari, Karachi.
2. MBBS, FCPS, MCPS, Professor Gynaecology and Medical Educationist, Shaheed Mohtarma Benazir Bhutto Medical College, Tannery Road, Lyari, Karachi, Pakistan.
3. MSc, MBA, M.Phil, Director Student Section, Shaheed Mohtarma Benazir Bhutto Medical College Tannery Road, Lyari, Karachi.
4. Mbchb, RMO Oncology, Royal Liverpool University Hospital, United Kingdom.
5. MBA, Director Research, Innovazione Research.
6. MS, Research Associate, Innovazione Research.
**Correspondence Address:**
Dr. Zeeshan Ahmed
Shaheed Mohtarma Benazir Bhutto Medical College Tannery Road, Lyari, Karachi.
dr.zeeshanarehman@gmail.com

**Article received on:**
17/05/2025
**Date of revision:**
16/08/2025
**Accepted for publication:**
01/01/2026

to perk up the assessment process. Item analysis in one of the crucial step in item improvement and corroboration that launch with dimension check and subsequent decision regarding addition/ revision/ deletion of items from the test pool based on the information provided by each item.[7] The item with standard difficulty and discrimination parameters doesn't necessarily denote high-quality items as efficiency of item is also directly related to precision in amount of information imparted by each of the item. Hence, item analyst must have good information about item precision with which it elaborates the information of the latent trait prior to make decision regarding retention, revision or redundancy of item. Previous studies related to item analysis rely mainly on discrimination and difficulty index of items.[8,9,10] Erstwhile studies in Pakistan regarding physiology discipline primarily focused on item analysis and identification of discrimination and difficulty parameters in an endeavor to build valid and reliable test pool with recognition of flaws in item drafting.[11,12] A recent study in Pakistan context incorporated IRT model for assessing the quality of the item to estimate learner competence through discrimination and difficulty measures[13], though there is still a dearth in estimation of accuracy or precision of information gathered by each item in the test.[13] Other studies have also done item analysis at postgraduate level as well to evaluate correlation between measures of item analysis. A statistical model of item response theory (IRT) provides psychometric analysis of assessment tools by corroborating learner performance with the ability appraised with items in the test.[4] Item information function technique (IIFs) in the IRT is a contemporary reliable approach that imparts precise estimations for learner ability based on item difficulty parameter which determines precision in measuring the amount of information granted by each item on the ability scale. This amount of information is reciprocally related to standard error of estimate (SEE).[14,15] Hence at certain ability θ more information will be achieved if the ability level approaches difficulty value due to minimum volatility and standard error near difficulty reaches zero at extremes of the ability scale.[16]

The current study contributes to a great deal in the meticulous development of the test pool with the IIF domain of the IRT which helps in identification of items having precision to provide appropriate information for each item creating a high quality test pool. Furthermore, current study provides insight for policy makers, item drafters and curriculum designers to sketch assessment strategies at the start of planning with a viewpoint of encompassing items that provide precise information in order to ascertain the true ability of the learner. Therefore, in view of the above literature it is pertinent to estimate the precision of information at each ability level by each item to be deemed appropriate for inclusion in the test pool. Hence current study conduct item analysis to estimate item information function for each of the items in the test exhibiting information for C1(precise information for Knowledge), C2(precise information for Understanding) and C3(precise information for Application) based on difficulty parameter with hypothesis:

$H_o$: There is no difference in precise information provided by physiology items at C1, C2 and C3 level.

$H_A$: There is a difference in precise information provided by physiology items at C1, C2 and C3 level.

## METHODS

The current study is quantitative and primary cross-sectional data was collected by purposive sampling employing 19 physiology MCQs from formative assessment of second year MBBS students in SMBB Medical College Lyari, Karachi in Gastrointestinal Tract module at three initial cognitive levels of Bloom's Taxonomy (C1 =7, C2=6 and C3 =6) as physiology award maximum credit hours in GIT module in 2nd professional MBBS. Responses of 100 second professional MBBS students (0=correct and 1=incorrect answer) were collected from a public sector medical college in Karachi, Pakistan. The study employed the 2PL IRT dichotomous model for assessing item analysis and subsequent IIF to identify items providing precise information on ability scale with statistical software STATA 17. The Kuder-Richardson test was used to estimate reliability of items while discrimination and difficulty parameters determine validity of the items at 95% C.I. The items that either found

acceptable or desirable were further analyzed for precision estimation by IIF based on difficulty parameters. Ethical approval was also obtained from Institutional Review Board (IRB/SMBBMC/APPROVAL/2025/625).

## RESULTS

The demographics consist of 53 %females and 47% males. The standard benchmark is high or moderate DI with moderate DIF items. The item analysis for the items at C1 level (Table-I) shows that out of total 7 only 3 items found as appropriate for the test pool such as phc11 is acceptable with both moderate discrimination and difficulty index (S.E 0.5696, p-0.002), phc12 is also found as highly desirable with moderate difficulty (S.E= 0.4154. p=0.000) while phc14 exhibits perfect discrimination and moderate difficulty (S.E=0.1877, p=0.000) with overall test reliability 0.64.

The item analysis for C2 level items (Table-II) shows that out of 6 items only phc25 was found as highly discriminated and moderately difficult (S.E 0.5529, p- 0.006)) with overall test reliability 0.51.

The item analysis for C3 level items (Table-III) shows that out 6 only 3 items found appropriate for inclusion in the test pool i.e. phc31 and phc32 are acceptable with both moderate discrimination and difficulty index (phc31=S.E=0.3776, p=0.015 and phc32= S.E= 0.6604 p=0.017) while phc35 is highly desirable with moderate difficulty (S.E= 0.2148, p=0.003) with overall test reliability 0.45.

**TABLE-I**

**Physiology cognitive level 1(phc1)**

| | Coefficient | | Std.err. | Z | P>z | Alpha | Reliability | |
|---|---|---|---|---|---|---|---|---|
| phc11 | | | | | | | | |
| Discrim | 1.127705 | Moderate | .4505601 | 2.50 | 0.012 | 0.5962 | | Acceptable for test pool |
| Diff | 1.782699 | Moderate | .5696045 | 3.13 | 0.002 | | | |
| phc12 | | | | | | | | |
| Discrim | 1.929745 | Very high | .760927 | 2.54 | 0.011 | 0.5900 | | Higher desirable for test pool |
| Diff | 1.801607 | Moderate | .4154937 | 4.34 | 0.000 | | | |
| phc13 | | | | | | | | |
| Discrim | -2.626306 | Negative | 1.00111 | -2.62 | 0.009 | 0.5628 | | Revise/discard |
| Diff | .6063793 | Moderate | .1670184 | 3.63 | 0.000 | | | |
| phc14 | | | | | | | | |
| Discrim | 5.120488 | Perfect | 4.210913 | 1.22 | 0.224 | 0.5147 | 0.6418 | Higher desirable for test pool |
| Diff | .9431394 | Moderate | .1887658 | 5.00 | 0.000 | | | |
| phc15 | | | | | | | | |
| Discrim | 1.25558 | Moderate | .5122563 | 2.45 | 0.014 | 0.6341 | | Revise |
| Diff | 2.074015 | Very diff | .6307906 | 3.29 | 0.001 | | | |
| phc16 | | | | | | | | |
| Discrim | -1.405978 | Negative | .5350963 | -2.63 | 0.009 | 0.6298 | | Revise/discard |
| Diff | -.5073988 | Moderate | .2299464 | -2.21 | 0.027 | | | |
| phc17 | | | | | | | | |
| Discrim | .3796227 | Low | .4270906 | 0.89 | 0.374 | 0.6823 | | Discard |
| Diff | 5.902237 | Very diff | 6.382597 | 0.92 | 0.355 | | | |

**TABLE-II**

**Physiology Cognitive level 2 (phc2)**

| | Coefficient | | Std.err. | Z | P>z | Alpha | Reliability | |
|---|---|---|---|---|---|---|---|---|
| phc21 | | | | | | | | |
| Discrim | .3122725 | Very low | .2864871 | 1.09 | 0.276 | 0.5545 | | Revise/discard |
| Diff | 1.689778 | Moderate | 1.63 6752 | 1.03 | 0.302 | | | |
| phc22 | | | | | | | | |
| Discrim | -2.181979 | Negative | .7279499 | -3.00 | 0.003 | 0.3472 | | Revise/discard |
| Diff | -1.388154 | Moderate | .2517843 | -5.51 | 0.000 | | | |
| phc23 | | | | | | | | |
| Discrim | .2890314 | Very low | .3351306 | -0.86 | 0.388 | 0.5757 | | Revise/discard |
| Diff | -5.065462 | Very easy | 5.731025 | -0.88 | 0.377 | | | |
| phc24 | | | | | | | 0.5104 | |
| Discrim | -.9206473 | Negative | .385242 | -2.39 | 0.017 | 0.4773 | | Discard |
| Diff | -1.592184 | Moderate | .5717497 | -2.78 | 0.005 | | | |
| phc25 | | | | | | | | |
| Discrim | 1.517775 | High | .55293 | 2.74 | 0.006 | 0.4536 | | Higher desirable for test pool |
| Diff | .1678814 | Moderate | .1855254 | 0.90 | 0.366 | | | |
| phc26 | | | | | | | | |
| Discrim | -.5235263 | Negative | .3262017 | -1.60 | 0.109 | 0.3098 | | Revise/discard |
| Diff | .5205996 | Moderate | .5033383 | 1.03 | 0.301 | | | |

**TABLE-III**

**Physiology Cognitive level 3(phc3)**

| | Coefficient | | Std. errs. | Z | P>Z | Alpha | Reliability | |
|---|---|---|---|---|---|---|---|---|
| phc31 | | | | | | | | |
| Discrim | .9626894 | Moderate | .4046472 | 2.38 | 0.017 | 0.4054 | | Acceptable for test pool |
| Diff | .9175493 | Moderate | .3776165 | 2.43 | 0.015 | | | |
| phc32 | | | | | | | | |
| Discrim | .8281391 | Moderate | .3954518 | 2.09 | 0.036 | 0.3755 | | Acceptable for test pool |
| Diff | 1.569664 | Moderate | .6604524 | 2.38 | 0.017 | | | |
| phc33 | | | | | | | | |
| Discrim | -1.958578 | Negative | 1.035463 | 1.89 | 0.059 | 0.3055 | | Discard |
| Diff | .657408 | Moderate | .2161515 | 3.04 | 0.002 | | | |
| phc34 | | | | | | | | |
| Discrim | .1003121 | Very low | .2821207 | 0.36 | 0.722 | 0.4926 | 0.4504 | Discard |
| Diff | 2.229276 | Very diff | .559636 | 0.34 | 0.734 | | | |
| phc35 | | | | | | | | |
| Discrim | 3.223968 | Very high | 3.357845 | 0.96 | 0.337 | 0.2816 | | Higher desirable for test pool |
| Diff | .6334222 | Moderate | .2148419 | 2.95 | 0.003 | | | |
| phc36 | | | | | | | | |
| Discrim | .1486334 | Very low | .2919805 | 0.51 | 0.611 | 0.4769 | | Discard |
| Diff | 4.383714 | Very diff | 8.637957 | 0.51 | 0.612 | | | |
| phc37 | | | | | | | | |
| Discrim | .8658765 | Moderate | .479219 | 1.81 | 0.071 | 0.4934 | | Revise |
| Diff | 2.367494 | Very diff | 1.087943 | 2.18 | 0.030 | | | |

Discrimination: negative (<0), very low (0.1-0.34), low (0.35-0.64), moderate (0.65-1.34) high (1.35-1.69) very high (>1.70) perfect (+infinity)10 Difficulty: very easy (<-2.00),moderate (-2.00 to 2.00),very difficulty(>2.00) [14]. The widely –accepted cut-off value KR 20 is greater than or equal to 0.7.

**FIGURE-1**

**Physiology IIFs C1 level Item information function curve (Figure-1) shows that only phc14 provides precise information for ability level 1 as compared to other items with lowest standard error while others centered on lower ability levels.**

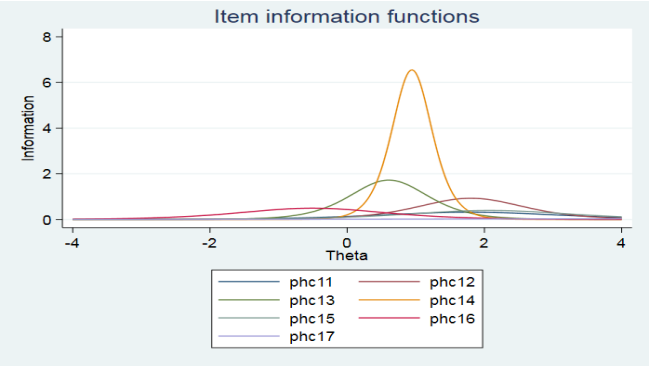



**FIGURE-2**

**Physiology IIFs C2 level shows that phc25 provides precise information for ability level 0.25 as compared to other items with lowest standard error while others centered on lower ability levels.**

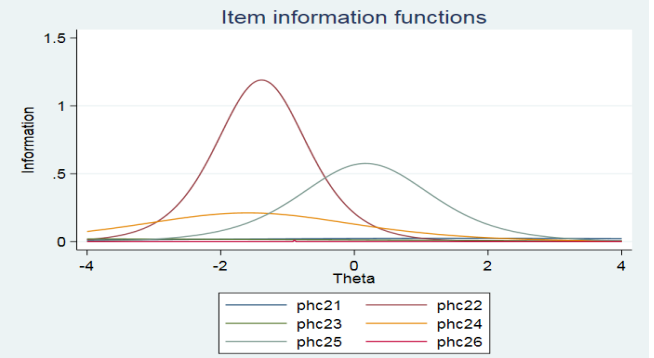
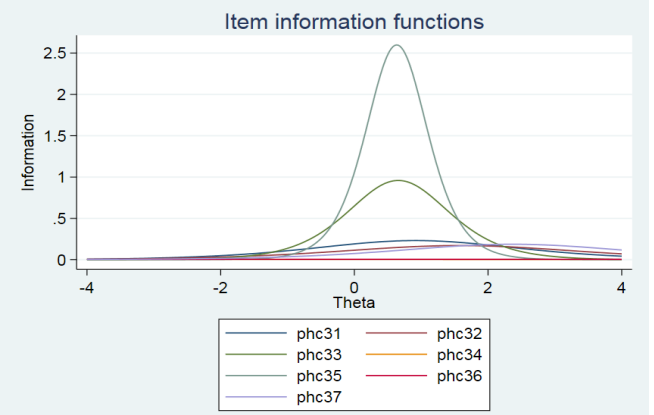



**FIGURE-3**

**Physiology IIFs C3 level shows that only phc35 provides precise information for ability level 0.75 as compared to other items with lowest standard error while others are centered on lower ability levels.**




| Hypothesis Decision | |
|---|---|
| Hyp | Decision |
| Ho | Rejected |

## DISCUSSION

The current study conducted item analysis like previous studies[8,9,10,11,12,13] which depict that C1 level items phc11, 12 and 14 are significant items to be included in test pool but IIFs shows that only phc14 provide precise information with lowest standard error hence should be included in the test pool whilst Phc12 and 11 though highly desirable and acceptable respectively but unable to provide precise information therefore should be reassessed and then included in the pool.

Likewise, phc25 and phc35 from C2 and C3 level physiology items also found as appropriate providing precise information with lowest standard error hence should be included in the test pool. While phc31 and phc32 were acceptable but not giving precise information needs reassessing before inclusion into the pool. This shows that more than half of physiology items at C1, C2 and C3 level for factual knowledge, understanding and application are faulty or weak and thus lack the quality while those acceptable have to be revamped to be used for effective assessment and require major overhauling of the items. Since the precision of information is not equal for physiology items at C1, C2 and C3 cognitive levels, hence Ho is rejected.

## CONCLUSION

The items with very high or moderate DI and moderate DIF are regarded as a benchmark for a high quality, but the items must reflect precise information for the latent trait. The Physiology items for cognition level C1, C2 and C3 are not gauging precise information of knowledge, understanding and application even the ones in the acceptable range. Hence item analysis is not sufficient to gauge quality of item and IIF should be done to provide precise information.

## LIMITATION

The study conducted an item information function test only for formative assessment due to time constraint.

## ACKNOWLEDGEMENT

I would like to acknowledge Shaheed Mohtarma Benazir Bhutto medical college and students for facilitation in conducting the study.

## RECOMMENDATION

The future study should be conducted at a multi-disciplinary level for summative assessment to demonstrate a broader perspective for the overall test.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## SOURCE OF FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.



## REFERENCES

1. Guilbert JJ. **Educational handbook for health personnel.** Revised edition), WHO, Geneva. 1981.
2. Hargreaves A. **The emotional practice of teaching.** Teaching and Teacher Education. 1998 Nov 1; 14(8):835-54.
3. Torre DM, Schuwirth LWT, Van der Vleuten CPM. **Theoretical considerations on programmatic assessment.** Medical Teacher. 2020; 42(2):213-20.
4. Bichi AA, Talib R. **Item response theory: An introduction to latent trait models to test and item development.** International Journal of Evaluation and Research in Education (IJERE). 2018; 7(2):142.
5. Bloom BS. **Taxonomy of educational objectives: The classification of educational goals:** Handbook I Cognitive Do- main, Longmans, Green and Co Ltd. 1956; 250(2025):3.
6. Miller GE. **The assessment of clinical skills/competence/performance.** Acad Med. 1990; 65(9 Suppl):S63-67.
7. Haladyna TM, Rodriguez MC. **Using full-information item analysis to improve item quality.** Educational Assessment. 2021 Jul 3; 26(3):198-211.
8. Zubairi NA, AlAhmadi TS, Ibrahim MH, Hegazi MA, Gadi FU. **Effective use of item analysis to improve the reliability and validity of undergraduate medical examinations: Evaluating the same exam over many years: A different approach.** Pakistan Journal of Medical Sciences. 2025 Mar; 41(3):810.
9. Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V. **Item analysis of multiple choice questions: A quality assurance test for an assessment tool.** Medical Journal Armed Forces India. 2021 Feb 1; 77:S85-9.
10. Rezigalla AA, Eleragi AM, Elhussein AB, Alfaifi J, ALGhamdi MA, Al Ameer AY, et al. **Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items.** BMC Medical Education. 2024 Apr 24; 24(1):445.
11. Rahim AF, Simok AA, Wahab SF. **A guide for writing single best answer questions to assess higher-order thinking skills based on learning outcomes.** Education in Medicine Journal. 2022 Jun 1; 14(2):111-24.
12. Qamar AM, Kanwal W, Perveen R. **Item analysis of tool used for examining the effectiveness of e-modules for academic performance of 7th grade students.** Archives of Educational Studies (ARES). 2022; 2(2):205-22.
13. Rehman ZA, Mehmood SI. **Learner competence across cognitive levels in undergraduate medical education: An item response approach.** JMMC. 2024 Apr 30; 14(2):76-81.
14. **Item Response Theory the Basics of Item Response Theory the.** (n.d.).
15. Wyse AE, RJ DE Ayala (2009) **The Theory and Practice of Item Response Theory. New York: The Guilford Press.** xv+ 448 pages. US $60.00. ISBN: 978-1-59385-869-8. Psychometrika. 2010 Dec; 75(4):778-9.
16. DeMars C. **Item response theory.** Oxford University Press; 2010 Apr 30.

| AUTHORSHIP AND CONTRIBUTION DECLARATION | |
|---|---|
| 1 | **Zeeshan Ahmed:** Study design. |
| 2 | **Anjum Rehman:** Data analysis. |
| 3 | **Umme Habiba Rehman:** Data entry. |
| 4 | **Muhammad Maaz Sajid:** Data collection. |
| 5 | **Fasih Ur Rehman:** Results writing. |
| 6 | **Malik Moiz Uddin:** Draft writing. |