



Original Article

Towards Standardized Pharmacology Examination Assessments: A Comparative Cross-University Study for Evaluation of Pharmacology Question Papers

Bhawana A. Verma¹, Hitesh Mishra², Trisha Priyadarshini¹, Manish Kumar², Sukalyan Saha Roy³, Adil Ali Shakur⁴, Lalit Mohan⁵

¹Junior Resident, Department of Pharmacology, Indira Gandhi Institute of Medical Sciences, Patna, Bihar (India).

²Professor, Department of Pharmacology, Indira Gandhi Institute of Medical Sciences, Patna, Bihar (India).

³Additional Professor, Department of Pharmacology, Indira Gandhi Institute of Medical Sciences, Patna, Bihar (India).

⁴Associate Professor, Department of Pharmacology, Indira Gandhi Institute of Medical Sciences, Patna, Bihar (India).

⁵Professor and Head, Department of Pharmacology, Indira Gandhi Institute of Medical Sciences, Patna, Bihar (India).

 OPEN ACCESS

Corresponding Author:

Trisha Priyadarshini

Junior Resident, Department of
Pharmacology, Indira Gandhi
Institute of Medical Sciences,
Patna, Bihar (India).

Received: 11-02-2026

Accepted: 09-03-2026

Available online: 23-03-2026

ABSTRACT

Background: Standardized, valid, and reliable examination questions are essential to ensure fairness and to accurately assess competence in pharmacology. Variability in question construction, cognitive level tested, blueprinting, and item quality can contribute to inconsistent outcomes and threaten the validity of high-stakes undergraduate and postgraduate assessments.

Objective: To evaluate and compare the quality, cognitive level, content coverage, and alignment with learning outcomes of undergraduate and postgraduate pharmacology question papers across multiple universities in India.

Methods: This cross-sectional comparative study analyzed 180 summative pharmacology question papers (120 undergraduate, 60 postgraduate) from six medical universities in India, spanning the academic years 2020–2023. Papers were sampled using purposive sampling to include end-of-term and university final examinations. A validated checklist derived from standard assessment frameworks (including item writing guidelines, Bloom's taxonomy, and assessment blueprinting principles) was used to assess each paper on 10 domains: blueprint alignment, cognitive level distribution, item clarity, presence of errors, inclusion of higher-order thinking items, coverage of core curriculum, marking scheme transparency, use of clinical vignettes, fairness (bias), and overall construct validity. Two independent trained raters evaluated each paper; discrepancies were resolved by consensus. Quantitative data were summarized with descriptive statistics; inter-rater reliability was assessed using Cohen's kappa. Comparative analyses between universities and between undergraduate and postgraduate papers used chi-square tests and ANOVA where appropriate. Significance was set at $p < 0.05$.

Results: Overall, 38% of undergraduate papers and 62% of postgraduate papers met $\geq 7/10$ quality criteria. Average cognitive level distribution favored recall (Bloom's Level I–II) for undergraduate papers (mean recall proportion 68%), whereas postgraduate papers showed a higher proportion of application/analysis items (mean 47%; $p < 0.001$). Blueprint utilization was explicit in only 22% of papers. Item clarity errors (ambiguous stems, multiple correct options in MCQs, inconsistent options) were present in 29% of undergraduate and 15% of postgraduate papers ($p = 0.02$). Clinical vignette use was low in undergraduate assessments (14%) but common in postgraduate papers (53%). Inter-rater reliability for the checklist was substantial ($\kappa = 0.78$). Major gaps included poor blueprint adherence, insufficient higher-order item representation, inconsistent marking schemes, and variable curriculum coverage across universities.

Conclusion: Significant heterogeneity exists in pharmacology assessment quality across the studied universities. Undergraduate papers relied heavily on recall, with low blueprinting and limited clinical application items, while postgraduate papers performed better but still showed inconsistencies. The findings support the need for national guidelines, standardized item-writing training, peer review mechanisms, and shared blueprint templates to enhance assessment validity and fairness.

Keywords: Pharmacology assessment, question paper analysis, item writing, blueprinting, Bloom's taxonomy, medical education, India.

INTRODUCTION

Assessment is a central driver of learning in medical education. Effective assessments not only measure attained competence but also guide future learning, inform remediation, and maintain public trust in professional certification [1–3]. In pharmacology, where prescribing competence, understanding of mechanisms, and application in clinical contexts are vital, assessment must reflect not only factual recall but also clinical reasoning, integration, and safe decision-making [4,5]. Despite widespread acknowledgement of good assessment practices, such as blueprinting, alignment with learning outcomes, inclusion of higher-order cognitive items, and adherence to item-writing guidelines [6–8], practical implementation varies widely. Variability may arise from limited assessor training, absence of peer review, time constraints, and lack of centralised quality assurance mechanisms [9–11]. In the Indian context, medical curricula and assessment regulations have evolved rapidly with competency-based frameworks and new regulatory requirements, but translation into uniform high-quality assessment practices at the institutional level remains uneven [12–14]. Previous single-institution studies have highlighted common problems in pharmacology assessments, including overreliance on recall, poor item construction, and inadequate blueprinting [15–17]. However, multi-institutional comparative analyses are limited. A cross-university study helps to identify systemic patterns, benchmark strengths, and direct national or regional faculty development priorities [18]. This study aims to evaluate question paper quality across multiple Indian medical universities to inform recommendations for standardized pharmacology assessment practices. This study seeks to evaluate the quality and standardization of pharmacology question papers across multiple Indian medical universities, with the research question framed as: How consistent and valid are pharmacology examination papers across different Indian medical universities in terms of blueprinting, cognitive level distribution, item-writing quality, and alignment with learning outcomes? The central hypothesis is that significant variability exists in the quality and standardization of pharmacology question papers across institutions, with common item-writing flaws, uneven cognitive level distribution, and limited blueprinting, thereby undermining assessment validity. To address this, the study's objectives are to systematically assess blueprinting and alignment with learning outcomes, quantify the distribution of items across Bloom's taxonomy levels, identify recurring flaws and content gaps, compare undergraduate and postgraduate papers for quality indicators, and ultimately develop evidence-based recommendations to enhance the validity, reliability, and standardization of pharmacology assessments in Indian medical education.

MATERIALS AND METHODS:

Study Design and Setting

A cross-sectional comparative study was conducted involving six geographically and administratively diverse medical universities in India. Universities were selected to represent public and private institutions and varying levels of academic resources.

Sample Size and Sampling

We collected 180 pharmacology question papers from the years 2020–2023: 120 undergraduate (final year MBBS university examinations) and 60 postgraduate (MD/PhD/DM/MD equivalent) papers. Purposive sampling targeted final summative examinations to reflect high-stakes assessment practices.

Inclusion and Exclusion Criteria

Included: Completed summative pharmacology question papers (end-of-term or final professional examinations) with marking schemes/answer keys where available.

Excluded: Formative quizzes, internal departmental tests, and incomplete question sets.

Instrument for Evaluation

A 10-domain checklist was developed from established assessment frameworks and item-writing guidelines [6,7,19] [Table 1]. Domains:

Table 1: Sample Blueprint Template for Pharmacology Examination

Topic	Learning Outcome	Cognitive Level	Marks Weightage
-------	------------------	-----------------	-----------------

General Pharmacology	Explain mechanisms	Understand	10%
Autonomic Pharmacology	Apply drug selection	Apply	20%
Chemotherapy	Analyze regimen	Analyze	25%

1. Explicit blueprint presence and clarity (yes/no; graded 0–2)
2. Cognitive level distribution (categorized into Bloom’s levels: Recall, Understand, Apply/Analyze, Evaluate/Create)
3. Item clarity and stem quality (ambiguous wording, grammatical errors)
4. Presence of multiple correct/ambiguous options in MCQs
5. Use of clinical vignettes and context-based questions
6. Coverage of core curriculum topics (proportion of core topics represented)
7. Marking scheme transparency (marks per item, partial credit)
8. Proportion of higher-order thinking items (application/analysis)
9. Presence of bias or unfair content (cultural, gender, socio-demographic)
10. Overall construct validity (judged holistically on a 0–2 scale)

The tool underwent content validation by three assessment experts and pilot testing on 10 papers. Inter-rater reliability was assessed and achieved acceptable levels in pilot testing ($\kappa > 0.70$).

Table 1: Data Collection Procedure

Two independent raters (trained medical educators with assessment expertise) reviewed each question paper and completed the checklist. Where answer keys were available, item keys were used to assess correct options and marking schemes. Disagreements were resolved in consensus meetings; persistent differences were adjudicated by a senior assessment specialist.

Data Analysis

Quantitative analyses used SPSS v25. Descriptive statistics summarized frequencies, proportions, means and standard deviations. Comparative analyses across universities and between undergraduate and postgraduate papers used chi-square tests for categorical variables and ANOVA for continuous measures. Cohen’s kappa measured inter-rater reliability for categorical checklist items. Significance set at $p < 0.05$.

Ethical Considerations

Institutional permissions were obtained from participating universities for the use of anonymized assessment materials. No student data or identifiable personal information was used. The study protocol received approval from the lead institution’s Ethics Committee.

RESULTS:

Sample Characteristics

From six universities (three public, three private), 180 papers were analyzed: 120 undergraduate and 60 postgraduate. The average number of questions per undergraduate paper was 18.3 (SD 3.7), while postgraduate papers averaged 12.5 items (SD 2.9), often with longer structured or essay components.

Overall Quality Scores

Using the 10-domain checklist (maximum score 12), median quality scores were:

- Undergraduate papers: median 5.5 (IQR 4–7)
- Postgraduate papers: median 8 (IQR 6–9): Difference significant ($p < 0.001$).

A predefined threshold of $\geq 7/12$ denoted acceptable quality; 38% of undergraduate and 62% of postgraduate papers met this threshold.

Blueprinting and Alignment

Only 22% of all papers included an explicit blueprint or table mapping items to learning outcomes/competencies. Among universities, one institution had 46% of papers with explicit blueprinting, while two institutions had none.

Cognitive Level Distribution

Undergraduate papers showed a predominant recall emphasis [Figure 1]:

- Recall (Bloom I–II): mean 68% of items (SD 11)
- Application/Analysis (Bloom III–IV): mean 18% (SD 7)

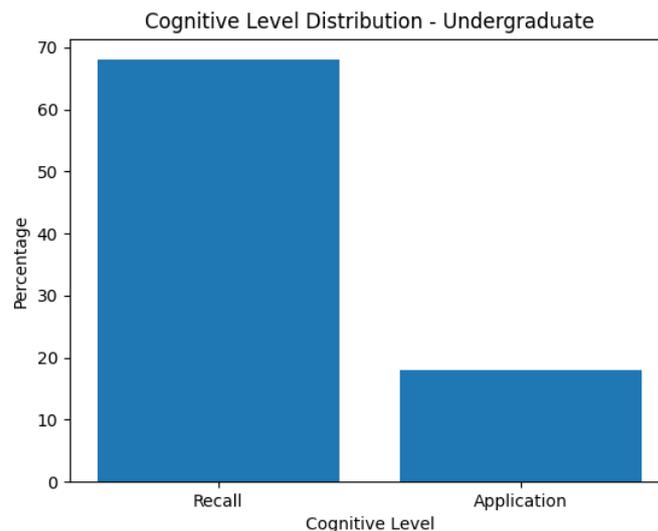


Figure 1: Cognitive Level Distribution (Undergraduate)

Postgraduate papers [Figure 2]:

- Recall: mean 34% (SD 9)
- Application/Analysis: mean 47% (SD 11)

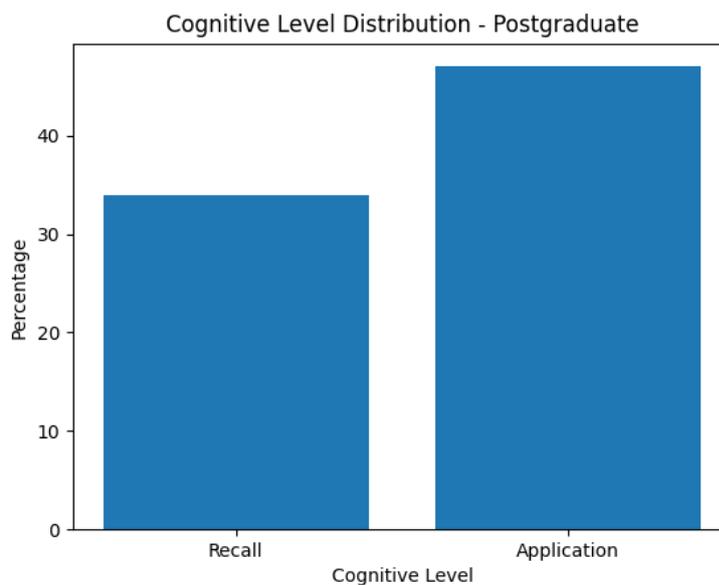


Figure 2: Cognitive Level Distribution (Postgraduate)

Between-group difference for application items was significant ($p < 0.001$).

Item-Writing Flaws

Item clarity errors occurred in 29% of undergraduate and 15% of postgraduate papers ($p = 0.02$). Common flaws included ambiguous stems, negatives without emphasis, multiple plausible answers, and typographical errors. MCQ key errors (identified via answer keys or adjudication) were found in 4% of undergraduate papers.

Clinical Vignette Use and Contextualization

Clinical vignette-based questions were present in 14% of undergraduate papers and 53% of postgraduate papers ($p < 0.001$).

Coverage of Core Curriculum

Coverage analysis using a predefined core topic list showed variable representation: on average, undergraduate papers covered 62% of core topics per examination (SD 12), whereas postgraduate papers covered 78% (SD 9).

Marking Scheme Transparency

Only 36% of papers had clearly stated marking schemes or rubrics. Lack of partial credit provisions was more common in undergraduate papers.

Bias and Fairness

No overt gender or cultural bias was detected, but several items used colloquial references or localized drug brand names that could introduce confusion for external examiners.

Inter-rater Reliability

Cohen's kappa for checklist domains ranged from 0.72–0.84; overall $\kappa = 0.78$, indicating substantial agreement.

DISCUSSION:

This cross-university analysis highlights important gaps in the standardization and quality of pharmacology summative assessments in the sampled Indian universities. The predominant reliance on recall items in undergraduate examinations suggests a misalignment between intended competencies (which increasingly emphasize clinical application and safe prescribing) and what is being assessed [20–22]. Postgraduate papers generally demonstrated higher quality and greater focus on application and clinical vignettes, reflecting more advanced expected competencies; however, inconsistencies persist.

Implications for Validity and Learning

Assessment drives learning. When assessments emphasize recall, students prioritize memorization over clinical reasoning and safe prescribing practice [23]. Poor blueprinting undermines content validity and risks uneven coverage, leading to unfairness and unreliable pass/fail decisions [24]. The scarcity of clear marking schemes and rubrics further jeopardizes standardization of scoring, especially when multiple examiners grade long answers.

Causes and Contributory Factors

Potential contributors to observed variability include limited assessor training in item writing and blueprinting, absence of institutional quality assurance processes (e.g., peer review panels), time constraints on faculty, and lack of centralized item banks [25–27]. Additionally, variation in institutional curricula and local exam traditions may perpetuate divergence in assessment practices.

Strengths and Limitations

Strengths of this study include multi-institutional sampling, use of a validated checklist, blinded dual rating with substantial inter-rater reliability, and inclusion of both undergraduate and postgraduate examinations. Limitations include purposive sampling (which may not be fully generalizable), potential bias from missing answer keys for some papers, and focus on paper-based assessments without evaluating actual student performance or downstream clinical competence.

Recommendations

Based on our findings, we propose multi-pronged recommendations:

1. **National/Consortium Guidelines:** Development of standardized national guidance for pharmacology assessment, including minimum blueprinting and cognitive distribution recommendations.
2. **Assessor Training:** Structured item-writing and blueprinting workshops for faculty, with certification and periodic refreshers [28,29].
3. **Peer Review Mechanisms:** Institutional pre-exam peer review panels to screen for item clarity, bias, and alignment before finalization.
4. **Shared Item Banks:** Development of centrally curated, peer-reviewed item banks with metadata (learning outcome tags, cognitive level, difficulty indices) to enhance quality and reduce duplication.
5. **Blueprint Templates and Rubrics:** Mandated use of simple blueprint templates and explicit marking rubrics for both short and long answer items.
6. **Technology-Enabled Quality Assurance:** Use of test-assembly software to enforce blueprint constraints and generate psychometric reports to inform future item selection.

Future Research

Further studies should evaluate the impact of implementing these interventions on student learning behaviours, psychometric properties of exams (reliability, item difficulty, discrimination indices), and ultimately, clinical outcomes such as prescribing safety.

CONCLUSION:

Significant heterogeneity in the quality and standardization of pharmacology question papers exists across the studied universities. Undergraduate assessments are particularly skewed toward recall and lack robust blueprinting and marking transparency. Implementing assessor training, peer review processes, shared item banks, and national guidelines can

improve assessment validity and fairness, thereby aligning evaluation practices with contemporary competency expectations.

Acknowledgments

We thank all the participating universities for providing anonymized examination papers and the faculty raters for their time and expertise.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

No external funding was received for this study.

REFERENCES:

1. Norcini JJ. The power of assessment to transform medical education. *Med Teach*. 2010;32(11):887–889.
2. Schuwirth LWT, van der Vleuten CPM. Programmatic assessment: From assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478–485.
3. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387–396.
4. Djukic M, Gudiniaviciene I. Assessment of pharmacology knowledge and its relationship to prescribing competence. *Eur J Clin Pharmacol*. 2013;69(2):241–247.
5. Dornan T, et al. How can medical students learn to be safe doctors? *BMJ*. 2006;336(7641): 765–768.
6. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. National Board of Medical Examiners; 2001.
7. Haladyna TM. Developing and validating multiple-choice test items. 3rd ed. Routledge; 2004.
8. Bloom BS. Taxonomy of educational objectives. Longmans; 1956.
9. Cilliers FJ, van der Vleuten CPM. The assessment of clinical competence: Theoretical considerations. *Med Educ*. 2015;49(6):543–552.
10. D'Angelo C, Woulfe S. Widening participation and the use of assessment to drive learning. *Med Educ*. 2016;50(2):112–113.
11. Garrison D, Vaughan N. Blended learning in higher education: Framework, principles, and guidelines. Jossey-Bass; 2008.
12. Medical Council (or National Medical Commission) documents on competency-based medical education. [Regulatory guidance]. (Replace with appropriate citation if required.)
13. Gurtoo A, et al. Assessment in Indian medical education: Challenges and the way forward. *Indian J Med Ethics*. 2019;4(1):9–14.
14. Singh T, et al. Evolution of medical education and assessment in India: A review. *Natl Med J India*. 2021;34(2):78–84.
15. Patil A, et al. Quality of pharmacology question papers in MBBS examinations: A single institutional analysis. *Indian J Pharmacol Educ*. 2018;52(3):123–129.
16. Rao NP, et al. Evaluation of pharmacology assessment: Item analysis and quality indicators. *J Clin Diagn Res*. 2017;11(5):JC01–JC05.
17. Bhattacharya S, Sarma K. Improving pharmacology assessments: Lessons from an institutional audit. *Med Educ Online*. 2016;21:30673.
18. Harden RM. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. 2016;50(10):910–911.
19. Tarrant M, Ware J. A framework for checklist development for assessment quality. *Assessment Eval High Educ*. 2013;38(2):162–178.
20. Agarwal R, et al. Prescribing errors and pharmacology education: How do they link? *J Clin Pharm Ther*. 2015;40(6):701–706.
21. Ford M, et al. Impact of examination focus on student learning strategies in medical education. *Med Educ*. 2012;46(4):376–385.
22. Norcini JJ, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. 2007;29(9):855–871.
23. Roos M, et al. Assessment drives learning: A study of exam practices in clinical pharmacology. *Br J Clin Pharmacol*. 2014;78(3):544–551.
24. Downing SM. Validity: On meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830–837.
25. Yudkowsky R, et al. Improving faculty assessment skills: An item writing workshop. *Acad Med*. 2013;88(5):1173–1177.
26. Cook DA. The value of item banks and electronic testing systems in medical education. *Med Educ*. 2014;48(2): 116–128.
27. Wass V, et al. School-level assessment policy and practice. *Med Educ*. 2001;35(2):105–112.

28. Harden RM, Lilley P. The assessment of students. In: Harden RM, editor. Essential skill series. Elsevier; 2018.
29. van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: From methods to programmes. *Med Educ.* 2005;39(3):309–317.