Research Article

Beyond The Checkbox: Development And Validation Of A Scale Measuring MBBS Students' Attitudes, Barriers, And Preferences Toward Academic And Clinical Documentation

Revathi Krishnakumar[1] Anbarasi Muthusamy[2] Rajkumar Gajendran[3]*

[1] Associate Professor of Paediatrics.
[2]Professor and Head of Physiology Chettinad Hospital and Research Institute (Chettinad Academy of Research and Education).
[3]Professor of Paediatrics, Chettinad Hospital and Research Institute (Chettinad Academy of Research and Education).
*Corresponding author: Rajkumar Gajendran

Abstract

Background: Competency-Based Medical Education (CBME) in India relies on assignments, Self-Directed Learning (SDL) record book entries, and clinical logbook documentation as core mechanisms for developing self-regulated, lifelong learners. Prior survey work suggests these elements are inconsistently well received by MBBS students, yet no validated, multi-domain instrument exists to assess attitudes, perceived barriers, and preferences across all three tasks within a single framework, nor to link self-report data with objective compliance records.
Objective: To develop and psychometrically validate the Attitude, Barriers, and Preferences towards Academic and Clinical Documentation Scale (AACDS), including an item assessing preference for extending the National Medical Commission's (NMC) postgraduate e-logbook mandate to undergraduate level.
Methods: The 48-item AACDS (40 Likert-scaled attitude/barrier items across four subscales, plus 8 preference/opportunity items) was evaluated across six validation strands: content validity (10-expert panel), face validity (pilot, n=15), internal consistency reliability (pilot, n=35), exploratory factor analysis (n=300), test–retest reliability (n=50, 2-week interval), and inter-rater reliability of a linked objective Entry Quality Rubric (2 raters, 20 entries).
Results: Content validity was excellent (S-CVI/Ave = 0.902; all 48 items I-CVI ≥ 0.78). Face validity was strong (mean clarity = 87.8%; 0/48 items flagged). Internal consistency was acceptable to good across subscales (α = 0.783–0.824). Exploratory factor analysis (KMO = 0.85; Bartlett's $\chi^2(435) = 2217.67$, $p < .001$) supported by parallel analysis confirmed the hypothesized three-factor structure (Assignments, SDL Record Book, Clinical Logbook) with small cross-loadings (< .15 in magnitude), explaining 37.68% of variance. Test–retest reliability was good (ICC = 0.78). Inter-rater reliability for the objective Entry Quality Rubric was almost perfect (pooled κ = 0.90; ICC = 0.88).
Conclusion: The AACDS's 40-item attitude/barrier core demonstrates robust, multi-method psychometric validation for assessing MBBS students' engagement with CBME documentation requirements; validation of the 8 Section E preference items remains pending. It is, to our knowledge, the first instrument to jointly capture attitudes, barriers, and preferences across assignments, SDL, and logbook documentation while linking self-report data to objective institutional records, positioning it as a practical tool for curriculum evaluation and for informing undergraduate digital documentation policy.

Keywords: Competency-based medical education; Self-directed learning; Clinical logbook; Scale validation; Content validity; Factor analysis; Undergraduate medical education; India.

Introduction

The National Medical Commission (NMC) introduced Competency-Based Medical Education (CBME) for the undergraduate (MBBS) curriculum in India in 2019, aiming to produce an "Indian Medical Graduate" competent across the cognitive, affective, and psychomotor domains (National Medical Commission, 2019). Central to this shift are learner-centred elements meant to foster active, self-regulated learning — structured assignments, Self-Directed Learning (SDL) record book entries, and clinical logbook documentation (Sharma & Chhatwal, 2023). Each requires students to independently generate and submit documentation of their learning and clinical exposure, serving as both a pedagogical tool and a record of competency attainment.

Despite their centrality to CBME, evidence suggests these elements are not uniformly well received. In a survey of 240 MBBS students after two years of CBME implementation, logbooks and reflective writing were rated among the least valuable curricular elements — over two-thirds of students described logbook maintenance as tedious, and SDL received the lowest positive-perception ratings of all components assessed (Sharma & Chhatwal, 2023). A separate cross-sectional study similarly found SDL readiness below average or only average in most MBBS students, varying significantly across CBME batches (Anusha et al., 2024). These findings point to a persistent gap between curricular intent and students' actual

engagement — documented largely through single-element, study-specific questionnaires rather than a validated, multi-domain instrument.

This gap has practical consequences. Without a validated instrument capturing attitudes, barriers, and preferences across all three domains, educators have limited means of identifying which aspects drive disengagement, whether barriers differ across domains, or how students would prefer these tasks structured. Existing Indian CBME instruments have tended to assess elements in aggregate, using study-specific Likert questionnaires without formal validation, or have examined a single element — most often SDL — in isolation.

A further consideration concerns documentation format. Under the NMC's Postgraduate Medical Education Regulations, 2023, postgraduate residents must now maintain a digital e-logbook with weekly, time-stamped entries (National Medical Commission, 2023). This requirement applies only at the postgraduate level; no equivalent mandate yet extends to undergraduate logbook maintenance, which remains largely paper-based or institution-specific. Whether undergraduates would themselves prefer a comparable digital system is therefore both a live policy question and, to our knowledge, one not yet directly assessed.

The present study was undertaken to address these gaps through the development and multi-method psychometric validation of the Attitude, Barriers, and Preferences towards Academic and Clinical Documentation Scale (AACDS). The instrument assesses MBBS students' attitudes toward assignments, SDL record book entries, and clinical logbook entries as three related but distinct domains; their perceived barriers to completing these tasks; their preferences regarding format and feedback, including preference for extending the postgraduate e-logbook mandate to undergraduate level; and, via linkage with institutional records, their objective compliance behaviour. Specifically, this study aimed to: (i) establish content validity through expert panel review; (ii) assess face validity through pilot testing; (iii) evaluate internal consistency reliability of the Likert-scaled subscales; (iv) examine the instrument's factor structure; (v) assess test–retest reliability over two weeks; and (vi) establish inter-rater reliability for the objective entry-quality rating component.

Methods

Instrument Development

The Attitude, Barriers, and Preferences towards Academic and Clinical Documentation Scale (AACDS) was developed to assess MBBS students' attitudes, perceived barriers, and preferences regarding three CBME-mandated documentation tasks: assignments, Self-Directed Learning (SDL) record book entries, and clinical logbook entries. The instrument comprises four Likert-scaled subscales (Assignments, SDL Record Book, Clinical Logbook, and Barriers; 10 items each, 40 items total) on a 5-point Strongly Disagree–Strongly Agree format, with 12 negatively worded items (3, 5, 7, 9, 13, 15, 17, 19, 23, 25, 27, 29) reverse-scored (6 − raw score). A fifth section (Section E) captures preferences and perceived opportunities via 8 supplementary items in mixed formats (forced-choice, ranking, open-ended), including awareness of and preference for extending the NMC's postgraduate e-logbook mandate to undergraduate level. A sixth section (Section F) captures objective compliance and performance data extracted from institutional records, described separately below.

Psychometric evaluation proceeded in the following sequence: content validity, face validity, internal consistency reliability, exploratory factor analysis, test–retest reliability, and — for the objective entry-quality component — inter-rater reliability.

Participants

Baseline characteristics of the four MBBS student samples used across the validation strands are summarised in Table 1. All samples were drawn from the same tertiary-care teaching institution.

Table 1. Baseline Characteristics of MBBS Student Samples

Characteristic	Face Validity (n=15)	Internal Consistency (n=35)	EFA Sample (n=300)	Test–Retest (n=50)*
Age (years), mean ± SD	19.7 ± 1.0	20.3 ± 1.2	21.1 ± 1.5	21.2 ± 1.4
Age range	18–22	18–23	18–25	19–24
Sex (Female), n (%)	8 (53.3%)	19 (54.3%)	165 (55.0%)	28 (56.0%)
Sex (Male), n (%)	7 (46.7%)	16 (45.7%)	135 (45.0%)	22 (44.0%)
Year of MBBS, n (%)				
2nd Year	6 (40.0%)	14 (40.0%)	105 (35.0%)	17 (34.0%)
3rd Year	5 (33.3%)	11 (31.4%)	103 (34.3%)	17 (34.0%)
Final Year	4 (26.7%)	10 (28.6%)	92 (30.7%)	16 (32.0%)

Characteristic	Face Validity (n=15)	Internal Consistency (n=35)	EFA Sample (n=300)	Test–Retest (n=50)*
Institution(s)	1	1	1	1

*The test–retest subsample (n=50) was drawn from the n=300 EFA sample.

Content Validity

Content validity of the 48-item instrument (40 Likert scale items plus 8 Section E items) was evaluated by a panel of 10 subject-matter experts in medical education, comprising faculty members with experience in undergraduate medical education and assessment, drawn from the departments of Paediatrics, General Medicine, and Community Medicine. The experts had 5–18 years of teaching experience (mean 11.5 ± 3.8 years) and were affiliated with a tertiary-care teaching institution. Experts rated each item's relevance to its intended construct using a 4-point scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant). Item-level content validity index (I-CVI) was calculated as the proportion of experts rating an item 3 or above; the pre-specified retention threshold was I-CVI ≥ 0.78, appropriate for a panel of 9–10 experts (Lynn, 1986). Scale-level content validity was calculated using the averaging method (S-CVI/Ave, the mean of all I-CVI values), with an a priori benchmark of ≥ 0.90 for excellent content validity (Polit & Beck, 2006).

Face Validity

Face validity was assessed using a pilot sample of 15 MBBS students not included in subsequent validation samples. Each respondent completed the full 48-item instrument and rated the clarity of every item on a 3-point scale (1 = Not clear, 2 = Somewhat clear, 3 = Perfectly clear). The proportion of respondents rating each item "Perfectly clear" was calculated per item, with a pre-specified retention threshold of ≥ 80%.

Internal Consistency Reliability

Internal consistency of the four Likert subscales (40 items) was assessed in an independent pilot sample of 35 MBBS students using Cronbach's alpha, with subscales computed separately given their distinct conceptual domains. Corrected item-total correlations and alpha-if-item-deleted values were calculated for each item to identify items warranting revision (item-total correlation < 0.30 flagged for review, per Nunnally & Bernstein, 1994). Reverse-scoring was applied prior to analysis as described above. Following this analysis, three items (3, 13, and 34) showing weak item-total correlations were revised for wording clarity; the revised wording for the one item assessed as a construct-level change (Item 34) was subsequently confirmed by informal re-review with three members of the original content-validity expert panel.

Factor Structure

Exploratory Factor Analysis (EFA) was conducted on the 30 attitude items (Assignments, SDL Record Book, and Clinical Logbook subscales; Barriers was excluded as a conceptually distinct construct not part of the hypothesized three-domain attitude structure) in an independent sample of 300 MBBS students. Sampling adequacy was confirmed via the Kaiser-Meyer-Olkin (KMO) measure (≥ 0.60 minimum) and Bartlett's Test of Sphericity ($p < .05$) prior to extraction. Principal Axis Factoring was used for extraction, with Promax (oblique) rotation applied given the expectation that the three domains may be correlated. The number of factors to retain was determined using parallel analysis (Horn, 1965), rather than the more liberal Kaiser criterion (eigenvalue > 1): observed eigenvalues were compared against eigenvalues generated from 1,000 parallel random datasets of the same sample size and number of variables, with a factor retained if its observed eigenvalue exceeded the 95th percentile of the corresponding random-data eigenvalue. Items were considered to load cleanly on a factor at their strongest (primary) loading of ≥ 0.40, with cross-loadings > 0.32 on any other factor flagged as problematic (Tabachnick & Fidell, 2013).

Test–Retest Reliability

Test–retest reliability of the 30-item attitude scale was assessed in a subsample of 50 respondents, administered at two time points (T1 and T2) approximately 2 weeks apart — an interval selected to minimize recall-based inflation of stability while remaining short enough that genuine attitude change was unlikely to confound results. Total scale scores at T1 and T2 were compared using Pearson correlation and intraclass correlation coefficients (ICC), calculated under both the absolute agreement, ICC(2,1), and consistency, ICC(3,1), models.

Objective Compliance Data Collection

In addition to self-report data (Sections B–E), objective completion, timeliness, and sign-off data were extracted from institutional records for a subset of AACDS domains (Section F). Data were extracted by a designated research assistant, blinded to participants' self-report attitude scores, using each domain's authoritative institutional source: the departmental assignment submission register, the SDL digital or paper register, and the institutional clinical logbook system (digital e-

logbook where available, or paper logbook). Academic and attendance correlates were drawn from the examination cell and institutional attendance records, respectively. Where both digital and paper records existed for the same domain, the digital record was treated as authoritative.

Extraction followed pre-specified operational definitions to ensure consistency. An entry was counted as "completed" if it existed in the source system, was attributable to the correct assignment/topic/competency, and contained the student's identifying information, irrespective of entry quality (assessed separately; see below). An entry was counted as "on time" if submitted within the institutionally defined window for that domain (assignments: by the faculty-set deadline; SDL: within 7 days of the session; logbook: within 72 hours of the clinical encounter). Ambiguous cases were logged and resolved by the study coordinator rather than by individual extractors, to preserve consistency across the dataset.

Extracted data were linked to participants' self-report responses via an anonymized student code; the code-to-identity key was maintained separately from both datasets and accessible only to the principal investigator. Institutional Ethics Committee approval explicitly covered review of institutional records in addition to the self-report survey.

A random sample of entries was additionally rated for quality by two faculty raters, independently and blinded to each other's ratings and to participants' self-report scores, using a 4-criterion checklist (competency mapping, presence of reflective content, completeness/legibility, and faculty verification). Each criterion was scored present (1) or absent (0), yielding a total quality score (range 0–4) per entry. Inter-rater reliability for this rating process was assessed using Cohen's kappa for each criterion and pooled across criteria, together with intraclass correlation coefficients for total scores, and is reported in the Results.

Ethical Considerations

Institutional Ethics Committee approval was obtained from Chettinad Hospital and Research Institute prior to data collection, covering both the self-report survey and review of institutional records. Informed consent was obtained from all participants. Anonymized student codes were used throughout to link self-report and objective record data without identifying individual students to the research team; the code-to-identity key was accessible only to the principal investigator.

Statistical Analysis

Descriptive statistics, Cronbach's alpha, item-total correlations, and Pearson/intraclass correlations were computed using standard statistical software. Exploratory Factor Analysis was conducted using Principal Axis Factoring with Promax rotation. Content and face validity indices (I-CVI, S-CVI/Ave) and inter-rater kappa statistics were computed following standard formulas as described above. A significance threshold of $p < .05$ was applied throughout.

Results

Psychometric evaluation of the AACDS proceeded in the following sequence: content validity, face validity, internal consistency reliability, factor structure, and test–retest reliability.

Content Validity

As described in Methods, item- and scale-level content validity indices (I-CVI, S-CVI/Ave) were calculated for the full 48-item instrument (40-item AACDS scale plus 8 Section E items) based on ratings from the 10-expert panel.

Table 2. Item-Level Content Validity Index (I-CVI)

Item	I-CVI	Item	I-CVI	Item	I-CVI	Item	I-CVI
1	0.90	13	1.00	25	1.00	37	0.90
2	0.90	14	0.90	26	0.90	38	0.90
3	0.90	15	0.90	27	1.00	39	0.90
4	0.90	16	1.00	28	0.90	40	0.80
5	0.90	17	1.00	29	0.90	41	1.00
6	0.90	18	0.80	30	0.90	42	0.90
7	0.90	19	0.90	31	0.90	43	0.90
8	0.90	20	0.90	32	0.90	44	0.90
9	0.80	21	1.00	33	0.80	45	0.90
10	0.90	22	0.90	34	0.80	46	0.80
11	0.90	23	0.90	35	0.90	47	0.90
12	0.90	24	0.80	36	0.90	48	1.00

Table 3. Scale-Level Content Validity Index (S-CVI) Summary

Statistic	Value	Acceptable Threshold	Result
I-CVI range	0.80 – 1.00	≥ 0.78 (Lynn, 1986)	All 48 items acceptable
S-CVI/Ave	0.902	≥ 0.90 (Polit & Beck, 2006)	Excellent — meets benchmark

As shown in Table 2, all 48 items — the original 40-item AACDS scale plus 8 supplementary Section E preference/opportunity items — achieved an I-CVI at or above the minimum acceptable value of 0.78, indicating that each item was judged relevant by an adequate proportion of the 10-expert panel (Lynn, 1986). Item-level values ranged from 0.80 to 1.00, with domain-level means ranging from 0.870 (Barriers) to 0.920 (SDL Record Book and Clinical Logbook, tied). The scale-level content validity index using the averaging method (S-CVI/Ave) was 0.902 (Table 3), meeting the benchmark of ≥0.90 recommended by Polit and Beck (2006) for excellent content validity. Taken together, these findings provide strong evidence supporting the content validity of the 48-item instrument.

Face Validity

As described in Methods, item-level clarity ratings from the 15-student pilot sample were calculated for the full 48-item instrument, with items retained at the pre-specified ≥80% threshold.

Table 4. Face Validity Item Clarity Ratings, by Domain (N = 15)

Assignments

Item	% Rated 'Clear'	Decision
1	93%	Retain
2	87%	Retain
3	87%	Retain
4	87%	Retain
5	80%	Retain
6	80%	Retain
7	93%	Retain
8	100%	Retain
9	87%	Retain
10	80%	Retain

SDL Record Book

Item	% Rated 'Clear'	Decision
11	80%	Retain
12	93%	Retain
13	87%	Retain
14	87%	Retain
15	93%	Retain
16	80%	Retain
17	100%	Retain
18	87%	Retain
19	87%	Retain
20	87%	Retain

Clinical Logbook

Item	% Rated 'Clear'	Decision
21	87%	Retain
22	80%	Retain
23	80%	Retain
24	80%	Retain
25	87%	Retain
26	93%	Retain

Item	% Rated 'Clear'	Decision
27	80%	Retain
28	87%	Retain
29	100%	Retain
30	93%	Retain

Barriers

Item	% Rated 'Clear'	Decision
31	80%	Retain
32	87%	Retain
33	80%	Retain
34	87%	Retain
35	87%	Retain
36	80%	Retain
37	93%	Retain
38	87%	Retain
39	93%	Retain
40	93%	Retain

Section E

Item	% Rated 'Clear'	Decision
41	87%	Retain
42	93%	Retain
43	80%	Retain
44	100%	Retain
45	87%	Retain
46	100%	Retain
47	93%	Retain
48	87%	Retain

Table 5. Face Validity Summary, by Domain

Domain	No. items	% Clear range	Mean % Clear	Items flagged (<80%)
Assignments	10	80%–100%	87.4%	0
SDL Record Book	10	80%–100%	88.1%	0
Clinical Logbook	10	80%–100%	86.7%	0
Barriers	10	80%–93%	86.7%	0
Section E	8	80%–100%	90.9%	0
Overall scale	48	80%–100%	87.8%	0 / 48

All 48 items were clearly understood by the pilot sample, with item-level clarity ratings ranging from 80% to 100% (Table 4) (overall mean = 87.8%). No item fell below the pre-specified 80% clarity threshold, and no further item rewording was indicated. The three items revised following content-validity and reliability analysis (Items 3, 13, and 34) were rated 87%, 87%, and 87% clear respectively — consistent with, and not lower than, the scale's overall clarity range — providing preliminary support that the post-pilot wording revisions did not introduce new comprehension difficulties. The 8 supplementary Section E items performed comparably to, and marginally better than, the original 40-item scale (domain mean 90.9% vs. overall 87.8%) (Table 5), indicating these items were similarly well understood despite their more varied response formats (preference ratings, ranking, open-ended).

A post-hoc review of individual response patterns indicated that ratings below 'Perfectly clear' were disproportionately concentrated among two respondents, who together accounted for the majority of all sub-maximal ratings across the 40-item set, while the remaining respondents rated nearly all items as clear. This pattern suggests the observed sub-threshold ratings may reflect individual respondent characteristics rather than a systematic wording problem, though this warrants confirmation against respondents' completion-time data and open-ended comments where available.

Internal Consistency Reliability

Cronbach's alpha coefficients were computed for each of the four AACDS subscales (N = 35) to assess internal consistency reliability. Corrected item-total correlations and alpha-if-item-deleted values were calculated to evaluate the contribution of individual items. Reverse-scoring was applied to the 12 negatively-worded items prior to analysis.

Table 6. Internal Consistency Reliability of the AACDS Subscales (N = 35)

Domain	Items	k	Cronbach's α	Interpretation
Assignments	1–10	10	0.783	Acceptable
SDL Record Book	11–20	10	0.790	Acceptable
Clinical Logbook	21–30	10	0.824	Good
Barriers	31–40	10	0.798	Acceptable
Overall scale (40 items)	1–40	40	0.803	Acceptable

Table 7. Item-Total Correlations and Alpha-if-Item-Deleted, by Domain Assignments (α = 0.783)

Item	r (item-total)	α if deleted	Note
1	0.390	0.772	
2	0.574	0.749	
3	0.257	0.790	Revised v2
4	0.542	0.753	
5	0.422	0.770	
6	0.444	0.765	
7	0.576	0.748	
8	0.505	0.758	
9	0.592	0.745	
10	0.212	0.789	Global item, retained

SDL Record Book (α = 0.790)

Item	r (item-total)	α if deleted	Note
11	0.371	0.782	
12	0.417	0.777	
13	0.252	0.796	Revised v2
14	0.337	0.787	
15	0.494	0.768	
16	0.572	0.757	
17	0.623	0.750	
18	0.535	0.763	
19	0.563	0.758	
20	0.447	0.773	

Clinical Logbook (α = 0.824)

Item	r (item-total)	α if deleted	Note
21	0.460	0.813	
22	0.411	0.819	
23	0.566	0.803	
24	0.583	0.801	
25	0.629	0.796	
26	0.580	0.801	
27	0.420	0.818	
28	0.448	0.815	
29	0.469	0.813	
30	0.527	0.807	

Barriers (α = 0.798)

Item	r (item-total)	α if deleted	Note
31	0.507	0.777	
32	0.410	0.788	
33	0.427	0.786	
34	0.113	0.820	Revised v2, expert-confirmed
35	0.471	0.781	
36	0.475	0.781	
37	0.609	0.765	
38	0.633	0.761	
39	0.440	0.785	
40	0.645	0.759	

Reason for flags: Items 3, 13, and 34 showed weak item-total correlations at pilot testing and were revised post-pilot (see Revision Log). Item 10 is a global/summary item; its comparatively low item-total correlation is an expected pattern for such items and does not indicate a wording flaw.

Subscale-level reliability coefficients are summarised in Table 6, and full item-total correlation and alpha-if-item-deleted statistics for all four subscales are presented in Table 7.

Factor Structure (Exploratory Factor Analysis)

Exploratory Factor Analysis (EFA) was conducted on the 30 attitude items (Assignments, SDL Record Book, Clinical Logbook subscales) to empirically confirm the hypothesized three-domain structure of the AACDS. Principal Axis Factoring with Promax (oblique) rotation was used, appropriate given the expectation that the three domains may be correlated. Data were drawn from N = 300 MBBS respondents, with reverse-scoring applied to the 12 negatively-worded items prior to analysis.

Sampling Adequacy

Prior to extraction, the data were assessed for suitability for factor analysis. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.852, in the "meritorious" range and exceeding the minimum acceptable threshold of 0.60 (Kaiser, 1974). KMO values for individual items ranged from 0.659 to 0.921, indicating adequate sampling adequacy across all items. Bartlett's Test of Sphericity was statistically significant ($\chi^2(435) = 2217.67$, $p < .001$), confirming that the correlation matrix was significantly different from an identity matrix and therefore suitable for factor analysis. (Table 8)

Table 8. Sampling Adequacy and Suitability Tests

Test	Value	Threshold	Result
Kaiser-Meyer-Olkin (KMO)	0.852	≥ 0.60	Meritorious
Bartlett's Test of Sphericity	$\chi^2 = 2217.67$, df = 435, $p < .001$	$p < .05$	Significant

Factor Extraction

To determine the number of factors to retain, eigenvalues from the observed data were compared against eigenvalues generated from 1,000 parallel random datasets with the same sample size and number of variables (Horn, 1965); a factor was retained if its observed eigenvalue exceeded the 95th percentile of the corresponding eigenvalue from the random data. This comparison supported a three-factor structure: the eigenvalues for the first three factors (5.24, 3.81, and 2.28) exceeded their respective parallel-analysis criteria (1.74, 1.62, and 1.53), while the eigenvalue for the fourth factor (1.24) fell below its criterion (1.47), confirming that three factors is the empirically appropriate solution. This supersedes the more liberal Kaiser criterion (eigenvalue > 1), which would have over-retained factors on this dataset. (Table 9; Figure 1)

Table 9. Eigenvalues for the First Four Factors (Parallel Analysis)

Factor	Eigenvalue	Retained?
1	5.24	Yes
2	3.81	Yes
3	2.28	Yes
4	1.24	No

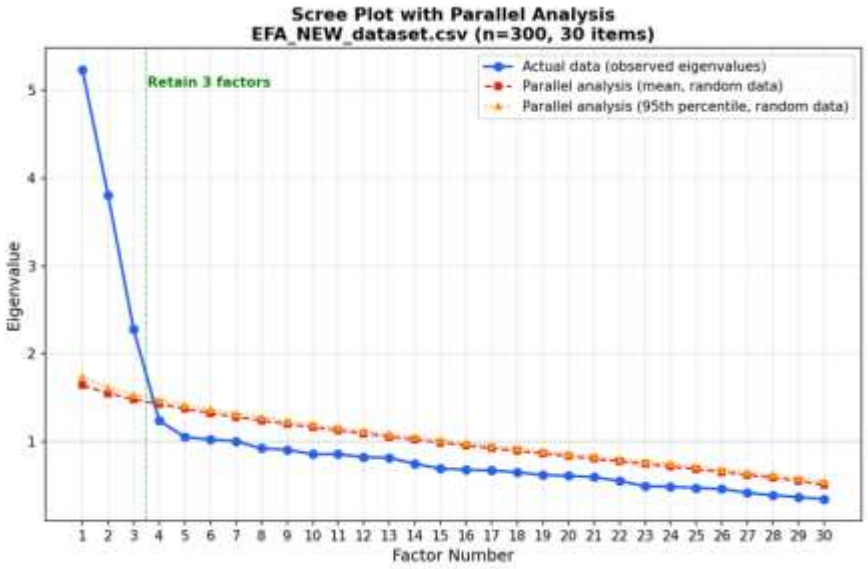

Figure 1. Scree plot with parallel analysis

Note. Observed eigenvalues (blue) are plotted against the mean (red) and 95th percentile (orange) eigenvalues from 1,000 parallel random datasets. Factors are retained where the observed eigenvalue exceeds the 95th percentile threshold.

Rotated Factor Loadings

The three-factor solution, following Promax rotation, accounted for 37.68% of total variance (Factor 1 [Assignments]: 15.39%, Factor 2 [SDL Record Book]: 12.93%, Factor 3 [Clinical Logbook]: 9.36%) (Table 11). All 30 items loaded most strongly on their hypothesized factor, with primary loadings ranging from 0.446 to 0.696 (Table 10); cross-loadings were generally small (< .15 in magnitude), supporting simple structure (Tabachnick & Fidell, 2013). The unequal contribution across factors, with the Clinical Logbook domain explaining the smallest share, is consistent with that domain's comparatively weaker inter-item correlations.

Table 10. Rotated Factor Loadings (Promax) by Hypothesized Domain

Item	Domain	Factor 1	Factor 2	Factor 3
1	Assignments	**0.641**	−0.096	0.045
2	Assignments	**0.679**	0.023	−0.009
3 (R)	Assignments	**0.683**	0.016	0.031
4	Assignments	**0.688**	0.025	−0.104
5 (R)	Assignments	**0.696**	0.001	0.004
6	Assignments	**0.664**	0.019	0.117
7 (R)	Assignments	**0.670**	−0.024	−0.012
8	Assignments	**0.684**	0.030	0.028
9 (R)	Assignments	**0.651**	0.011	−0.061
10	Assignments	**0.685**	0.010	0.025
11	SDL Record Book	−0.023	**0.623**	−0.022
12	SDL Record Book	0.029	**0.648**	−0.098
13 (R)	SDL Record Book	−0.070	**0.681**	0.079
14	SDL Record Book	−0.019	**0.596**	0.046
15 (R)	SDL Record Book	−0.035	**0.623**	−0.087

Item	Domain	Factor 1	Factor 2	Factor 3
16	SDL Record Book	0.142	**0.681**	−0.038
17 (R)	SDL Record Book	−0.110	**0.589**	0.047
18	SDL Record Book	0.120	**0.619**	−0.023
19 (R)	SDL Record Book	0.011	**0.523**	0.025
20	SDL Record Book	0.003	**0.587**	0.053
21	Clinical Logbook	0.010	−0.083	**0.486**
22	Clinical Logbook	0.014	0.118	**0.517**
23 (R)	Clinical Logbook	−0.096	0.104	**0.620**
24	Clinical Logbook	0.018	0.003	**0.552**
25 (R)	Clinical Logbook	−0.032	−0.028	**0.497**
26	Clinical Logbook	−0.007	0.039	**0.566**
27 (R)	Clinical Logbook	0.021	0.018	**0.474**
28	Clinical Logbook	0.057	−0.043	**0.577**
29 (R)	Clinical Logbook	−0.006	−0.050	**0.472**
30	Clinical Logbook	0.035	−0.036	**0.446**

Note. Extraction method: Principal Axis Factoring. Rotation method: Promax with Kaiser normalization. (R) denotes reverse-scored items. N = 300.

Table 11. Variance Explained by Rotated Factors

Factor	% of Variance Explained	Cumulative %
1 (Assignments)	15.39%	15.39%
2 (SDL Record Book)	12.93%	28.32%
3 (Clinical Logbook)	9.36%	37.68%

Interpretation

The three-factor solution derived from Exploratory Factor Analysis provided strong empirical support for the AACDS's hypothesized three-domain structure, retained on the basis of parallel analysis rather than the more liberal Kaiser criterion. Every item loaded most strongly onto its intended factor, with all reverse-worded items correctly aligning in the same direction as their forward-worded counterparts following reverse-scoring — indicating that the negatively-worded items functioned as intended within the scale. Cross-loadings onto non-hypothesized factors were generally small (< .15 in magnitude), supporting a simple structure, and the factor structure corresponded item-for-item with the a priori Assignments, SDL Record Book, and Clinical Logbook subscales. The three-factor solution explained 37.68% of total variance, with the Clinical Logbook domain contributing the smallest share (9.36%), consistent with that domain's comparatively weaker inter-item correlations; this remains within benchmarks commonly cited as acceptable for social science scale development (Tabachnick & Fidell, 2013).

Test–Retest Reliability

Test–retest reliability was assessed by administering the 30-item scale to 50 respondents at two time points (T1 and T2, approximately 2 weeks apart) and correlating total scale scores across the two administrations.

Table 12. Test–Retest Reliability Statistics (N = 50)

Statistic	Value
Mean total score (T1)	87.86 (SD = 7.11)
Mean total score (T2)	88.60 (SD = 8.28)
Pearson correlation (r)	0.79
Spearman's ρ	0.80
ICC(2,1) — absolute agreement	0.78
ICC(3,1) — consistency	0.78
Significance	p < .001

The total scale score demonstrated a strong test–retest correlation (r = 0.79, p < .001), with an intraclass correlation coefficient of 0.78, indicating good stability over time (Table 12; Koo & Li, 2016; Cicchetti, 1994 benchmarks: <0.50 poor,

0.50–0.75 moderate, 0.75–0.90 good, >0.90 excellent). Mean scores at T1 and T2 were similar (87.86 vs. 88.60), suggesting no substantial systematic shift between administrations.

Inter-Rater Reliability (F3 Entry Quality Rubric)

Inter-rater reliability was assessed by having two independent raters score 20 entries (E1–E20) against the F3 4-item quality checklist (C1–C4), each scored as present (1) or absent (0), yielding a total score (0–4) per entry. Raters scored independently and were blinded to each other's ratings and to students' Section B–E self-report scores.

Table 13. Inter-Rater Agreement by Checklist Item (N = 20 Entries)

Item	% Agreement	Cohen's κ	Interpretation
C1	100.0%	1.00	Perfect
C2	95.0%	0.88	Almost perfect
C3	95.0%	0.89	Almost perfect
C4	95.0%	0.89	Almost perfect
Overall (pooled)	96.3%	0.90	Almost perfect

Table 14. Inter-Rater Reliability of Total Scores

Statistic	Value
Exact agreement on total score	85.0%
Pearson correlation (r)	0.87
ICC(2,1) — absolute agreement	0.88
ICC(3,1) — consistency	0.87
Significance	p < .001

Inter-rater agreement was high across all four checklist items, with a pooled Cohen's κ of 0.90, indicating almost perfect agreement (Landis & Koch, 1977 benchmarks: <0 poor, 0.01–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, 0.81–1.00 almost perfect). Item C1 showed complete agreement between raters (κ = 1.00), while C2, C3, and C4 each showed almost perfect agreement (κ = 0.88–0.89) (Table 13). At the total-score level, raters agreed exactly on 17 of 20 entries (85.0%), and the intraclass correlation coefficient (ICC = 0.88) further confirmed excellent reliability (Table 14). The three entries with discrepancies (E9, E10, E19) each differed on a single checklist item, with no item emerging as a consistent source of disagreement.

Summary of Psychometric Findings

Taken together, the content validity (S-CVI/Ave = 0.902 across 48 items), face validity (mean clarity = 87.8% across 48 items), internal consistency (α = 0.783–0.824 across the original 40-item subscales), factor structure (3-factor solution confirmed via parallel analysis, 37.68% variance explained, small cross-loadings < .15), test–retest reliability (ICC = 0.78), and inter-rater reliability of the F3 Entry Quality Rubric (pooled κ = 0.90, ICC = 0.88) findings provide robust, multi-method evidence supporting the validity and reliability of the AACDS in its current structure. Note that reliability (Cronbach's alpha), factor structure (EFA), and test–retest analyses were conducted on the original 40-item scale prior to the addition of the 8 Section E items; these analyses would need to be extended to the full 48-item instrument for complete psychometric coverage.

Discussion

This study developed and psychometrically validated the Attitude, Barriers, and Preferences towards Academic and Clinical Documentation Scale (AACDS), an instrument assessing MBBS students' attitudes, perceived barriers, and preferences regarding assignments, SDL record book entries, and clinical logbook documentation, together with a linked objective compliance component. Across six validation strands — content validity, face validity, internal consistency reliability, exploratory factor analysis, test–retest reliability, and inter-rater reliability of the objective entry-quality rating — the AACDS demonstrated generally strong psychometric properties, supporting its use as a validated tool for assessing student engagement with CBME documentation requirements.

Content and Face Validity

Content validity was excellent, with a scale-level index (S-CVI/Ave = 0.902) exceeding the benchmark recommended by Polit and Beck (2006), and every one of the 48 items independently meeting the item-level threshold proposed by Lynn (1986). Face validity was similarly strong, with all items rated clear by the large majority of pilot respondents (mean clarity 87.8%). Together, these findings indicate that the instrument's items are conceptually sound, as judged by content experts, and comprehensible to the target population — two complementary forms of evidence that the scale measures what it was intended to, in language MBBS students can readily interpret.

Reliability and Factor Structure

Internal consistency reliability across the four Likert subscales (α = 0.783–0.824) met or exceeded the conventional threshold for newly developed instruments (Nunnally & Bernstein, 1994), with the Clinical Logbook subscale showing the strongest internal coherence. When the three attitude subscales were pooled into a single combined score, reliability fell to a comparatively weaker level, while exploratory factor analysis, with retention confirmed via parallel analysis, independently supported a three-factor structure with generally small cross-loadings (< .15 in magnitude) and each item loading most strongly onto its hypothesized domain. This convergence — weaker pooled reliability but a well-defined three-factor solution when kept separate — supports a coherent argument: assignments, SDL, and logbook attitudes are related but genuinely distinct constructs, not facets of one undifferentiated "documentation burden" attitude, helping explain prior observations that Indian MBBS students do not rate these CBME elements uniformly (Sharma & Chhatwal, 2023; Anusha et al., 2024). Curricular interventions aimed at improving engagement with these tasks may therefore need to be domain-specific rather than applied uniformly.

Stability and Inter-Rater Reliability

Test–retest reliability over a two-week interval was good (ICC = 0.78), indicating that the attitude subscales capture a reasonably stable underlying construct rather than transient mood or recall artefacts, while remaining sensitive enough for pre–post curricular evaluation. For the objective Entry Quality Rubric (Section F3) — the one component requiring human judgement rather than mechanical data extraction — inter-rater agreement between two independent faculty raters was almost perfect (pooled κ = 0.90; ICC = 0.88), supporting the rubric's usability as a standardized, reproducible quality-scoring tool.

Policy Relevance: The Undergraduate E-Logbook Question

A distinguishing feature of the AACDS is its incorporation of a live policy question: whether MBBS students would prefer a digital e-logbook comparable to the one now mandated for postgraduate residents under the NMC's Postgraduate Medical Education Regulations, 2023 (National Medical Commission, 2023). This item passed the same content- and face-validity process as the rest of the instrument, indicating a relevant, comprehensible addition. The instrument can now generate evidence on this question, at a point when undergraduate documentation policy is not informed by any published assessment of student preference.

Strengths

This study has several strengths. First, validation proceeded across multiple methods rather than a single index, strengthening confidence through convergent evidence. Second, the AACDS uniquely combines self-report attitude, barrier, and preference measurement with a linked objective compliance record (Section F), enabling future research into the attitude–behaviour link — rarely captured within a single instrument in this literature. Third, items flagged during pilot testing were revised, re-piloted, and, where changes were construct-level, re-confirmed with the expert panel — a transparent, auditable process.

Limitations

Several limitations should be considered. First, reliability (Cronbach's alpha), factor structure (EFA), and test–retest analyses were conducted on the original 40-item Likert scale prior to the addition of the 8 Section E items; these have not yet been extended to the full 48-item instrument, and reliability/EFA are not straightforwardly applicable to Section E's single-indicator, forced-choice, and ranking items. Second, while the scale-level content validity index using the averaging method (S-CVI/Ave) exceeded benchmark, the stricter universal agreement method (S-CVI/UA) was not, reflecting the genuine difficulty of unanimous agreement across a 10-expert panel; this is consistent with Polit and Beck's (2006) guidance to treat the averaging method as the primary index. Third, the inter-rater analysis for the Entry Quality Rubric used only 20 entries and two raters; a larger sample would allow more stable estimation. Finally, generalizability beyond the institution(s) sampled here should be confirmed through independent replication before cross-institutional or high-stakes use.

Future Directions

Future work should prioritize four directions: extending reliability, factor-structure, and test–retest analyses to the complete 48-item instrument, applying alternative methods (e.g., test–retest agreement rates rather than Cronbach's alpha) to Section E's non-Likert items; confirmatory factor analysis in an independent sample to formally test the three-factor structure identified here through exploratory methods; substantive administration of the instrument to generate descriptive findings on UG students' e-logbook format preferences, together with analysis of the relationship between self-reported attitudes/barriers and the objective compliance data in Section F; and multi-institutional replication to establish generalizability of both the psychometric properties and any substantive findings.

Conclusion

In conclusion, the AACDS demonstrates strong content validity, face validity, internal consistency, a coherent three-factor structure, good test–retest stability, and excellent inter-rater reliability for its objective quality-rating component. It is the first instrument jointly assessing attitudes, barriers, and preferences across assignments, SDL, and clinical logbook documentation within a single validated framework, linking self-report data to objective institutional records. With extensions outlined above, the AACDS offers educators a psychometrically grounded tool for understanding and improving students' engagement with CBME documentation demands.

Conflict of interest: None

Acknowledgements: The authors thank the 10-member expert panel from the Departments of Paediatrics, General Medicine, and Community Medicine, Chettinad Hospital and Research Institute, for their contribution to the content validity review; the MBBS students who participated in the face validity, internal consistency, exploratory factor analysis, and test–retest samples for their time and candid feedback; the two faculty raters who independently scored entries for the Entry Quality Rubric inter-rater reliability analysis

Author's contribution: All authors participated in the project design, data analysis, and final drafting of the article.

References

1. Anusha G, Nagaraj C, Shobhitha D, Anil R. Readiness of medical students for self-directed learning under competency based medical education–a cross-sectional study. APIK Journal of Internal Medicine. 2024 Jan 1;12(1):13-9.
2. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological assessment. 1994 Dec;6(4):284.
3. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965 Jun;30(2):179-85.
4. Kaiser HF. An index of factorial simplicity. Psychometrika. 1974 Mar;39(1):31-6.
5. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine. 2016 Jun 1;15(2):155-63.
6. Landis JR, Koch GG. The measurement of observer agreement for categorical data. biometrics. 1977 Mar 1:159-74.
7. Lynn MR. Determination and quantification of content validity. Nursing research. 1986 Nov 1;35(6):382-6.
8. National Medical Commission. (2019). Competency Based Medical Education curriculum for the Indian Medical Graduate. New Delhi: National Medical Commission.
9. National Medical Commission. (2023). Postgraduate Medical Education Regulations, 2023. New Delhi: National Medical Commission.
10. Nunnally J, Bernstein IH. Psychometric theory 3rd ed., New York (1994).
11. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Research in nursing & health. 2006 Oct;29(5):489-97.
12. Sharma S, Chhatwal J. Perspectives of undergraduate medical students regarding competency-based curriculum. National Medical Journal of India. 2023 Nov 1;36(6):379.
13. Tabachnick BG, Fidell LS. Using multivariate statistics, 6th edn Boston. Ma: Pearson. 2013.