Society of Medical Education & Research

2024, Volume 4, Page No: 165-171
ISSN: 3108-4850

Annals of Pharmacy Education, Safety, and Public Health Advocacy

Assessing the Validity of a Situational Judgment Test Targeting the Pharmacy Affective Domain

Erik J. Svensson[1]*, Lina K. Holm[1]

[1]Department of Social Pharmacy, Faculty of Pharmacy, University of Oslo, Oslo, Norway.

***E-mail** ✉ erik.svensson@outlook.com

Abstract

Following the formal inclusion of the affective domain within pharmacy accreditation requirements, pharmacy programs must adopt effective approaches to support formative evaluation of student competence in these competencies. This investigation aimed to generate validity evidence for a newly developed situational judgment test designed to measure affective domain outcomes specified in the Accreditation Council for Pharmacy Education (ACPE) Standards 2016.After preliminary item piloting across several pharmacy institutions, a refined situational judgment assessment consisting of 15 items was constructed to capture affective attributes relevant to professional pharmacy practice. The finalized instrument, termed the Pharmacy Affective Domain Situational Judgment Test (PAD-S), was administered to students enrolled at three universities. Evidence related to internal structure validity was evaluated using descriptive item statistics, Cronbach's alpha, and Rasch measurement modeling.A total of 559 complete and analyzable student responses were included. Item difficulty varied across the scale, and Rasch modeling supported an ordered hierarchy of items while indicating meaningful item contribution to the construct being measured. Internal consistency reliability, as estimated by Cronbach's alpha, reached 0.70, reflecting acceptable reliability. In contrast, Rasch-based reliability indices were slightly lower (0.65 and 0.66), suggesting limited precision in distinguishing levels of student ability, potentially attributable to a shortage of highly challenging items.Overall, the PAD-S demonstrated satisfactory performance as an instrument for assessing the affective domain in pharmacy education. The findings suggest that PAD-S can serve as a practical formative assessment tool within a broader assessment framework and may require fewer institutional resources compared with alternative evaluation methods.

Keywords: Affective domain, Situational judgment test, Instrument validity, Rasch analysis

Introduction

Within pharmacy education, the affective domain encompasses competencies related not only to direct patient care and professional practice skills, but also to personal growth and professional identity formation. Consequently, the cultivation and evaluation of this domain are critical from both patient-centered and health system perspectives [1]. Many colleges and schools of pharmacy have embedded affective domain development into cocurricular activities, defined as structured learning experiences occurring outside traditional classroom instruction and required experiential placements [2, 3]. Doctor of Pharmacy (PharmD) programs have implemented diverse strategies to deliver, document, and assess affective learning outcomes. A national survey of pharmacy programs identified reflective writing and self-reported surveys as the most frequently employed assessment approaches [3]. Although these strategies offer meaningful insights, they present notable challenges, including difficulties in objectively quantifying outcomes and the substantial faculty workload required for consistent and timely evaluation. Faculty across pharmacy education have identified cocurricular assessment as a significant concern [3]. The absence of standardized, objective, and comprehensive assessment instruments limits the ability to systematically evaluate affective competencies and monitor student development longitudinally. In response,

Access this article online https://smerpub.com/

Received: 18 January 2024; **Accepted:** 27 April 2024

How to cite this article: Svensson EJ, Holm LK. Assessing the Validity of a Situational Judgment Test Targeting the Pharmacy Affective Domain. Ann Pharm Educ Saf Public Health Advocacy. 2024;4:165-71. https://doi.org/10.51847/PheOaIEJ9G

interest has grown in the application of situational judgment tests as tools for measuring affective constructs such as empathy, professionalism, ethical behavior, and problem-solving [4–6].

Situational judgment tests are designed to simulate realistic professional scenarios in which respondents select or rank potential responses based on perceived effectiveness. These assessments typically employ multiple-choice or ranking formats that require examinees to judge appropriate actions in practice-relevant contexts [7]. Accumulating evidence supports the validity of situational judgment tests as standardized and formative assessment tools, and their use has expanded across health professions education, including pharmacy programs [4–8].

An earlier version of a pharmacy-focused situational judgment test was developed through cognitive interviewing with pharmacy students to refine item clarity and relevance [9]. That instrument required approximately 20–30 minutes to complete and consisted of nine scenarios, each accompanied by four response options to be ranked by participants. It was intended to comprehensively address the affective competencies outlined in ACPE Standards 3 and 4, including problem-solving, patient advocacy, interprofessional collaboration, cultural sensitivity, communication, self-awareness, leadership, innovation, entrepreneurship, and professionalism. However, that study did not evaluate internal structure validity evidence for the instrument [9].

A concise and psychometrically sound tool capable of assessing affective competencies while minimizing institutional burden would be valuable for formative curriculum evaluation and for tracking student progression. Accordingly, the present study sought to examine the psychometric characteristics of a revised situational judgment test designed to assess the affective domain.

Materials and Methods

The evaluation of measurement validity involves the synthesis of multiple forms of evidence [10]. Common sources include test content validity, which addresses the relevance and representativeness of items; response process validity, which considers how examinees interpret and respond to items; and internal structure validity, which relies on psychometric analyses to assess item functioning and reliability. **Figure 1** summarizes the sequential procedures used to develop the situational judgment test and to examine its validity evidence.

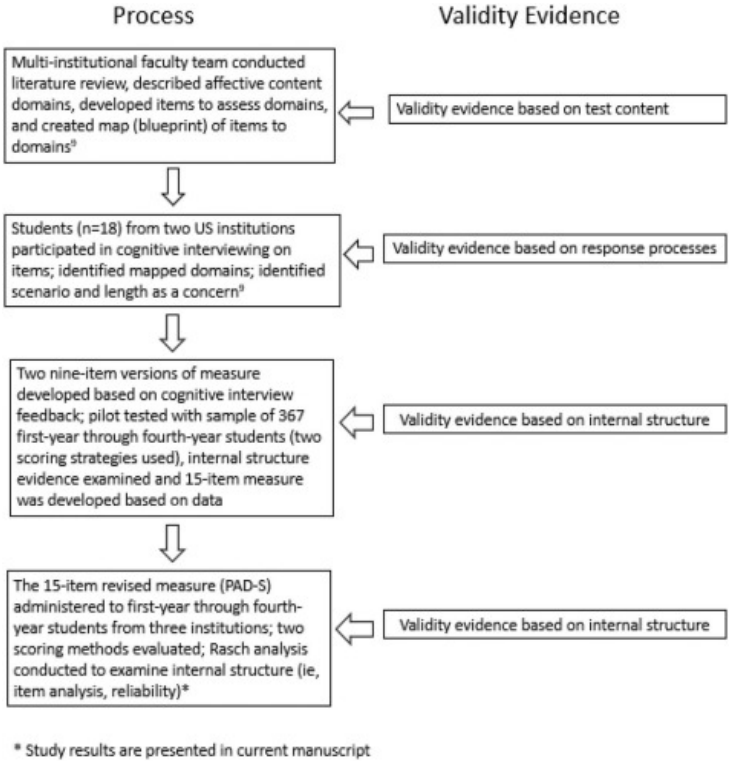

Figure 1. Overview of the stages undertaken to develop and evaluate the validity of the Pharmacy Affective Domain Situational Judgment Test (PAD-S).

In the present investigation, an existing situational judgment test that had undergone prior development and evaluation was modified and expanded to a 15-item format to assess affective domain competencies. This revised instrument, designated the Pharmacy Affective Domain Situational Judgment Test (PAD-S), is available at no cost to interested institutions upon request. Alignment of PAD-S items with the individual components of ACPE Standards 3 and 4 is summarized in **Table 1**. The PAD-S was administered electronically via the ExamSoft platform (ExamSoft Worldwide LLC) to pharmacy students spanning the first professional year (P1) through the fourth professional year (P4) at three academic institutions: Manchester University, Cedarville University, and the University at Buffalo. These institutions reflected diverse characteristics within the Academy. Manchester University is a private Midwestern institution where approximately 40% of enrolled students identify as non-White; Cedarville University is also a private Midwestern institution with about 33% non-White student representation; and the University at Buffalo is a public Northeastern institution with roughly 57% of students identifying as non-White. Participation in the PAD-S was encouraged for all enrolled students, with the option to decline inclusion of individual results in the research analysis. Of the 630 students who completed the assessment, 559 (88.7%) consented to the use of their data and were included in the final analyses. Evidence related to internal structure validity was evaluated for the full 15-item PAD-S. Given the absence of standardized scoring conventions for situational judgment tests, two scoring approaches were applied: the first awarded one point for each correctly sequenced response option (yielding a possible range of zero to three points), while the second assigned one point each for correctly identifying the most effective and least effective responses (yielding a range of zero to two points) [11].

Table 1. Alignment of Pharmacy Affective Domain Situational judgment test (PAD-S) items with ACPE standards 3 and 4 affective domains

ACPE Affective Domain	PAD-S Item Coverage
3.1 Problem Solving	Q1–Q15
3.2 Education	Q1, Q4, Q5, Q10, Q12, Q15
3.3 Patient Advocacy	Q2
3.4 Interprofessional Collaboration	Q3, Q9, Q12
3.5 Cultural Sensitivity	Q4
3.6 Communication	Q1, Q3, Q4, Q5, Q6, Q7, Q9, Q10, Q11, Q12, Q14
4.1 Self-Awareness	Q8, Q10, Q12, Q15
4.2 Leadership	Q8, Q11
4.3 Innovation and Entrepreneurship	Q9, Q10, Q14
4.4 Professionalism	Q11, Q13, Q14

Abbreviations: PAD-S = Pharmacy Affective Domain Situational Judgment Test; ACPE = Accreditation Council for Pharmacy Education.

As an initial analytic step, response distributions for each PAD-S item were reviewed alongside overall test scores and internal consistency estimates calculated using Cronbach's alpha in SPSS (version 25.0; IBM Corp). Cronbach's alpha is commonly applied to evaluate internal consistency reliability, with values of 0.70 or higher typically regarded as satisfactory [12]. To complement this approach, Rasch measurement modeling was employed to account simultaneously for individual student ability (reflecting overall affective domain competence) and item difficulty, while also examining patterns of student–item interaction across ability levels [12–14]. Rasch modeling yields multiple indicators of reliability and internal structure validity, including fit indices and person–item maps, and assesses whether items operate cohesively as a single underlying construct through evaluation of item fit [14–16]. Item fit was assessed using infit and outfit mean-square (MNSQ) statistics, which indicate the extent to which an item meaningfully contributes to the measurement construct; MNSQ values between 0.6 and 1.4 were considered acceptable [16]. Item difficulty parameters were expressed in logits (mean = 0, SD = 1), with negative values representing less challenging items and positive values indicating greater difficulty [16].

In addition, Rasch analyses generated estimates of both item and person reliability, as well as separation indices for items and respondents. Person reliability is conceptually analogous to traditional reliability coefficients such as Cronbach's alpha and is interpreted similarly, with values of 0.70 or higher viewed as acceptable [16]. Separation indices provide further insight into reliability by indicating the number of distinguishable strata along the instrument's measurement continuum, with values below 2.0 suggesting limited discrimination [16, 17]. The Rasch output also included an item–person map, which visually represents the correspondence between item difficulty and respondent ability on a shared logit scale [18]. This

map is critical for evaluating whether item difficulty levels appropriately span the range of respondent abilities; ideally, items should be distributed across the full ability continuum, with a concentration aligned near the central range of respondent abilities. Rasch analyses were conducted using Winsteps software (version 3.72.0; Winsteps.com) [19]. Ethical approval for the study was obtained from the institutional review board of Manchester University, Indiana.

Results and Discussion

The final analytic sample consisted of 559 valid student responses, including 191 first-year (P1), 154 second-year (P2), 172 third-year (P3), and 42 fourth-year (P4) pharmacy students drawn from the three participating institutions. Under scoring method 1, which awarded one point for each correctly sequenced response option (maximum score = 45), the mean (SD) score was 30.10 (5.65). Using scoring method 2, which assigned one point each for correctly identifying the most effective and least effective options (maximum score = 30), the mean (SD) score was 20.40 (4.26). Cronbach's alpha was 0.70 for both scoring approaches, supporting acceptable internal consistency regardless of scoring strategy. Item-level score distributions for each question are summarized in **Table 2**. Across all items and both scoring methods, students typically earned at least one point, although noticeable variability in scores was observed, reflecting differences in item difficulty. This variability indicates that the items were not uniformly easy or difficult, thereby supporting the suitability of Rasch modeling to further investigate the relationship between student ability and item difficulty.

Table 2. Item difficulty estimates and fit indices for Pharmacy Affective Domain Situational judgment test items across scoring approaches

	Scoring option 1							Scoring option 2					
PAD-S	0 points	1 point	2 points	3 points	Infit	Outfit	Item	0 points	1 point	2 points	Infit	Outfit	Item
Item	%	%	%	%	MNSQ[a]	MNSQ[a]	difficulty	%	%	%	MNSQ[a]	MNSQ[a]	difficulty
Q1	15.9	31.1	35.2	17.7	1.15	1.51	0.72	21.6	58.0	20.4	1.01	1.37	1.03
Q2	23.6	17.7	42.4	16.3	1.36	1.44	0.76	32.6	44.7	22.7	1.26	1.43	1.23
Q3	4.3	9.1	31.1	55.5	1.12	1.07	-0.63	5.5	22.5	71.9	1.01	0.89	-0.91
Q4	10.6	46.3	36.0	7.2	0.78	0.83	0.91	18.1	71.9	10.0	0.66	0.71	1.19
Q5	4.7	5.4	23.8	66.2	1.27	1.14	-0.97	5.2	24.3	70.5	0.98	0.91	-0.86
Q6	3.2	4.7	32.0	60.1	1.00	0.95	-0.90	4.1	20.0	75.8	0.94	0.86	-1.15
Q7	5.4	5.7	31.8	57.1	1.05	0.94	-0.69	7.7	22.0	70.3	1.03	0.90	-0.75
Q8	15.4	9.8	33.8	41.0	1.35	1.29	0.06	19.3	33.8	46.9	1.11	1.07	0.32
Q9	15.7	23.6	36.9	23.8	1.06	1.05	0.53	29.3	33.3	37.4	1.21	1.18	0.81
Q10	8.9	21.1	51.2	18.8	0.83	0.84	0.37	14.5	30.2	55.3	1.13	1.12	-0.04
Q11	2.9	11.3	42.4	43.5	0.95	0.93	-0.39	4.1	25.4	70.5	0.94	0.90	-0.91
Q12	4.1	10.0	63.5	22.4	0.65	0.63	0.00	5.2	65.1	29.7	0.74	0.76	0.40
Q13	5.0	14.3	38.8	41.9	0.94	0.92	-0.23	9.5	19.3	71.2	1.12	0.97	-0.72
Q14	8.8	24.3	40.3	26.7	1.01	1.02	0.30	14.7	53.5	31.8	0.94	0.94	0.58
Q15	6.8	14.3	56.4	22.5	0.80	0.78	0.15	8.2	36.5	55.3	1.09	1.05	-0.22

Abbreviations: PAD-S = Pharmacy Affective Domain Situational Judgment Test; MNSQ = mean-square.

[a] Item fit was evaluated using infit and outfit mean-square (MNSQ) statistics, which indicate the extent to which each item meaningfully contributes to the overall measurement; values ranging from 0.6 to 1.4 were considered acceptable. Item difficulty was expressed in logits (mean = 0, SD = 1) and reflects the relative challenge posed by each item, with negative logit values representing easier items and positive values indicating greater difficulty.

Rasch analysis results for both scoring approaches are summarized in **Tables 2 and 3**, encompassing item fit statistics, difficulty parameters, and reliability indices.

Across the majority of items, infit and outfit mean square (MNSQ) values fell within the commonly accepted bounds of 0.6 to 1.4, indicating satisfactory model fit and appropriate item functioning, with only two items under scoring option 1 (Q1 and Q2) marginally exceeding this interval. Item difficulty estimates demonstrated that scoring option 1 spanned from −0.97 to 0.91 logits, whereas scoring option 2 extended from −1.15 to 1.23 logits, reflecting the presence of relatively easy, average, and more challenging items in both methods, although the second option exhibited greater dispersion along the difficulty continuum. As shown in **Table 3**, item reliability was exceptionally high (0.99) for both scoring strategies, with item separation values exceeding 10, supporting a stable and well-defined hierarchy of item difficulties within the situational judgment test. In contrast, person-level indices were more modest: reliability coefficients were 0.65 for scoring option 1 and 0.66 for scoring option 2, slightly below the conventional benchmark of 0.70, while person separation indices (1.37 and 1.38, respectively) did not reach the desired threshold of 2.0, suggesting that the instrument has limited precision in consistently differentiating and ranking individuals by ability.

Table 3. Reliability and internal structure evidence for the Pharmacy Affective Domain Situational Judgment test by scoring method

Scoring Method	Sample Size (N)	Cronbach's α	Item Reliability	Item Separation	Person Reliability	Person Separation
Option 1	559	0.70	0.99	10.05	0.65	1.37
Option 2	559	0.70	0.99	10.48	0.66	1.38

Notes:

- **Cronbach's α:** Indicates internal consistency of the scale; values above 0.70 are generally considered acceptable for reliability.
- **Item reliability & separation:** Low item separation (<2) or item reliability (<0.70) suggests that the sample size may be insufficient to firmly establish the item difficulty hierarchy.
- **Person reliability & separation:** Low person separation (<2) or reliability (<0.70) indicates that the test may have limited ability to distinguish between higher- and lower-performing individuals, potentially requiring more items to improve discrimination.

For both scoring approaches, very few students had ability estimates below −1, indicating a limited number of low-performing students. Item difficulties ranged from −1 to 1 logit, while student abilities spanned −1 to 2 logits. The alignment between most items and student abilities was adequate; however, the highest-performing students exceeded the difficulty of the most challenging items, suggesting that the instrument does not fully capture high-level abilities. This limitation in accurately ranking top performers is reflected in the lower person reliability values. Person-item maps illustrating the relationship between student ability and item difficulty were examined and are available upon request.

Pharmacists who demonstrate strong affective domain competencies are better equipped to respond effectively to diverse human experiences, identities, and practices [20]. With growing emphasis on cultivating the affective domain, assessing students' competence in this area is critical, as it informs both curricular and co-curricular development. The present study adds validity evidence for a situational judgment test designed to evaluate the affective domain in pharmacy students. The PAD-S can serve as a practical tool for educators within a multi-faceted assessment strategy, offering advantages such as efficiency, objectivity, and lower resource demands compared with traditional measures like reflective essays.

Internal consistency (reliability) estimates for the PAD-S were generally acceptable. Although other situational judgment tests report higher internal consistency, these instruments often focus on a single domain (e.g., empathy or professionalism), whereas the PAD-S assesses multiple elements [4, 5]. Additionally, the PAD-S consists of 15 items, compared with longer measures that use 25 scenarios or up to 290 items across 40 scenarios [5, 6]. While adding more items could enhance reliability, the PAD-S was intentionally limited to a 20–30 minute completion time to reduce participant burden, an issue noted during cognitive interviews of the longer version [9]. Based on person reliability and separation findings, one potential improvement could be the inclusion of more challenging items to better assess higher-performing students. Nonetheless, the PAD-S aims to comprehensively cover key affective domain

elements without being overly time-consuming, providing actionable information on student skill gaps that can inform curricular or co-curricular interventions.

Several limitations should be considered. Data were collected from only three schools, which may limit the generalizability of findings to all pharmacy students. Variations in administration across institutions could have influenced responses; for instance, at one school, the PAD-S was administered alongside other annual assessments, whereas at another, it was delivered concurrently with an OSCE, with students informed that performance would not affect their OSCE grade. This may have resulted in responses that did not fully reflect students' maximum performance, though arguably offered a more authentic reflection of real-world behaviors. The limited number of items may also restrict discrimination among student abilities; however, prior feedback from longer versions indicated that increasing item numbers added cognitive burden [9]. Finally, responses may reflect what students perceive as the expected behavior rather than their own development, an inherent challenge in measuring internal growth.

Despite these constraints, the PAD-S retains utility for its intended purpose as a formative tool rather than a high-stakes assessment. Its development included multiple pilot tests, cognitive interviews, and iterative revisions [9]. Rasch modeling was employed to estimate reliability and examine internal structure validity using fit statistics and person-item maps, providing more detailed evidence than traditional metrics such as Cronbach's alpha [16].

No specific scoring method is recommended, as both approaches yielded comparable reliability and internal validity. The two scoring methods were examined due to debates on scoring complexity and whether distinguishing between two middle-ranked items adds meaningful information. Institutions may select the approach aligned with their assessment goals: for simpler scoring focusing on recognizing best and worst responses, option 2 is suitable; for greater discrimination, option 1, which scores all four items, may be preferred.

Currently, the PAD-S is integrated into assessment plans at one institution to evaluate affective domain skills at the P1 (baseline) and P3 levels, while another school uses it in annual data collection for triangulation with other measures. Future research should examine students' growth in the affective domain over time and determine whether the PAD-S can consistently measure progression. Additionally, investigations could explore the frequency and consistency of demonstrated skills and whether competency thresholds can be established.

Conclusion

The 15-item PAD-S was developed to formatively assess affective domain competencies aligned with Standards 3 and 4 for pharmacy students. Overall, the instrument performed well under both scoring methods. Validity evidence supports the PAD-S as a practical, resource-efficient formative tool within a comprehensive assessment framework to identify student skill gaps. Future studies should evaluate the instrument with diverse student populations and explore its ability to measure growth or change in the affective domain over time.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Pagatpatan, Jr, C.P. Valdezco, J.A.T. · Lauron, J.D.C. Teaching the affective domain in community-based medical education: A scoping review.Med Teach. 2020; 42:507-514
2. Accreditation Council for Pharmacy Education.Accreditation standards and key elements for the professional program in pharmacy leading to the Doctor of Pharmacy degree: "Standards. 2016;https://www.acpe-accredit.org/pdf/Standards2016FINAL.pdf
3. Maerten-Rivera, J.L. Chen, A. · Augustine, J.A survey of co-curriculum implementation and assessment in accredited pharmacy programs. Am J Pharm Educ. 2021; 85:ajpe7569
4. Wolcott, M.D. · Lupton-Smith, C. Cox, W.C. . A five-minute situational judgment test to assess empathy in first-year student pharmacists. Am J Pharm Educ. 2019; 83:6960
5. Smith, K.J. Flaxman, C. · Farland, M.Z.Development and validation of a situational judgement test to assess professionalism. Am J Pharm Educ. 2020; 84:7771

6. Patterson, F. · Galbraith, K. · Flaxman, C. .Evaluation of a situational judgement test to develop non-academic skills in pharmacy students.Am J Pharm Educ. 2019; 83:7074.
7. Patterson, F. ·Zibarras, L. · Ashworth, V. Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100.Med Teach. 2016; 38:3-17.
8. Webster, E.S. · Paton, L.W. · Crampton, P.E.S. ...Situational judgement test validity for selection: A systematic review and meta-analysis. Med Educ. 2020; 54:888-902.
9. DeLellis, T. · Noureldin, M. · Park, S.K. ...Development of a situational judgment test to assess ACPE Standards 3 and 4. Am J Pharm Educ. 2021; 85:8511.
10. American Educational Research Association (AERA). American Psychological Association (APA), National Council on Measurement in Education (NCME). AERA Publications, 2014.
11. De Leng, W.E. · Stegers-Jager, K.M. · Husbands, A. ...Scoring method of a Situational Judgment Test: Influence on internal consistency reliability, adverse impact and correlation with personality? Adv Health Sci Educ Theory Pract. 2017; 22:243-265.
12. Santos, J.R.A. Cronbach's alpha: A tool for assessing the reliability of scales. J Ext. 1999; 37:1-5.
13. Peeters, M.J. · Beltyukova, S.A. · Martin, B.A. Educational testing and validity of conclusions in the scholarship of teaching and learning.Am J Pharm Educ. 2013; 77:186.
14. Tennant, A. · Conaghan, P.G. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheumatol. 2007; 57:1358-62.
15. Smith, Jr, E.V. Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective.J Appl Meas. 2001; 2:281-311.
16. Bond, T.G. · Fox, C.M.Applying the Rasch model: Fundamental measurement in the human sciences.Routledge/Taylor & Francis Group, New York, NY, US, 2015.
17. Gallegos, P. · Peeters, J.M.A measure of teamwork perceptions for team-based learning.Curr Pharm Teach Learn. 2011; 3:30-35.
18. Liu, X.Using and developing measurement instruments in science education: A Rasch modeling approach.IAP Information Age Publishing, Charlotte, NC, US, 2010.
19. Linacre, J. Reliability and separation of measures. 2023.
20. Droege, M. · Assa-Eley, M.T.Pharmacists as care providers: Personal attributes of recent pharmacy graduates. Am J Pharm Educ. 2005; 69:290-295.