EVALUATION OF THE VALIDITY AND RELIABILITY OF THE VIETNAMESE VERSION OF THE PROFESSIONALISM ASSESSMENT SCALE FOR MEDICAL STUDENTS

Pham Thi Nga[1*], Nguyen Dinh Phat[1], Bui Thi Bich Ngan[1]

ABSTRACT

Objective: To evaluate the validity and reliability of the Vietnamese version of the Professionalism Assessment Scale for medical students.

Methods: An expert consultation study using the Delphi survey technique was conducted in combination with a cross-sectional descriptive study design.

Results: The I-CVI values for items across all three sections ranged from 0.83 to 0.9. The S-CVI/Ave was 0.87. Regarding reliability, Part 1 showed a Cronbach's α of 0.983, with item–total correlation coefficients ranging from 0.713 to 0.907. Part 2 demonstrated a Cronbach's α of 0.981, with correlation coefficients ranging from 0.648 to 0.910. Part 3 yielded a Cronbach's α of 0.961, with correlation coefficients from 0.713 to 0.878. The intraclass correlation coefficients (ICC) ranged from 0.850 to 0.956 for each section and were 0.869 for the overall scale. All corresponding p-values were > 0.05. Exploratory factor analysis (EFA) showed a KMO coefficient of 0.874, and Bartlett's Test was statistically significant ($p < 0.001$). Three factors had eigenvalues > 1 (6.75, 4.05, 2.70), explaining 61.3% of the cumulative variance. No items were removed; no cross-loadings exceeded 0.4; and all items clearly loaded onto three distinct components.

Conclusion: The instrument demonstrates high reliability, temporal stability, and a clear conceptual structure. It achieves strong content validity (high I-CVI and S-CVI values) and confirms internal consistency as well as a factor structure consistent with the theoretical model of medical professionalism. This tool is appropriate and useful for assessing professionalism in medical students within the context of medical education in Vietnam.

Keywords: Professionalism; content validity; reliability.

I. INTRODUCTION

Competency-based training for general practitioners has become an essential requirement to meet the practical needs of society. The purpose of a competency-based curriculum is to emphasize learner-centered outcomes, ensuring that learners progress from novice to expert within a framework of essential competencies [1][2]. Competencies are commonly grouped into domains including: medical knowledge; critical thinking and inquiry; patient care; professionalism; communication and collaboration skills; and organizational and social determinants of health.

Medical education reform increasingly focuses on teaching medical ethics and professionalism to students, helping them develop appropriate attitudes toward patients and colleagues, as well as a sense of responsibility to the community and society in their future practice [3]. The professionalism module is a relatively new component introduced into medical training programs in Vietnamese medical schools. Therefore, official teaching materials and assessment tools have not yet been formally published, leading to certain challenges for faculty members.

One of the assessment instruments that has demonstrated validity and reliability in evaluating medical students' perceptions of professionalism is the "Professionalism Assessment Scale for Medical Students," developed and validated by Klemenc-Ketiš in Slovenia in 2014 [4]. The professionalism module at Thai Binh University of Medicine and Pharmacy has been implemented since 2022; however, there is currently no standardized instrument to assess students' professionalism.

Therefore, the research team conducted the study entitled: "Evaluation of the Validity and Reliability of the Vietnamese Version of the Professionalism Assessment Scale in Medical Practice," with the research objective: To evaluate the validity and reliability of the Vietnamese version of the Professionalism Assessment Scale for medical students.

II. RESEARCH METHODS

2.1. Study participants

Inclusion criteria:

Third-year medical students who had completed the Professionalism module and at least one clinical course.

1. Trường Đại học Y Dược Thái Bình
*Corresponding author: Pham Thi Nga
Email: phamthingaydtb2025@gmail.com
Received date: 26/02/2026
Revised date: 13/5/2026
Accepted date: 18/5/2026

Students and experts who voluntarily agreed to participate in the study.

Ten experts held at least a master's degree and had experience in medical education, research, or scale validation

Exclusion criteria: Students who were absent from the university during the study period.

Time: From August 2024 to June 2025 at Thai Binh University of Medicine and Pharmacy.

2.2. Study design: A two-round Delphi method combined with a cross-sectional descriptive study design.

Sample size and sampling method: The standard of 10 participants per questionnaire item was applied. The instrument consists of 22 items; therefore, to ensure suitability for test–retest analysis in validity assessment, the sample size was determined as 220 third-year medical students who met the inclusion criteria.

Research instrument

In 2014, Klemenc-Ketiš and Vrecko developed a scale to assess aspects of professionalism in healthcare specifically for medical students. The instrument demonstrated high validity and reliability, with a Cronbach's alpha of 0.88, a Kaiser-Meyer-Olkin (KMO) coefficient of 0.846, and Bartlett's test indicating strong correlations among items ($p < 0.001$).

The instrument consists of 22 items across three domains:

Empathy and humanism (10 items)

Professional relationships and professional development (8 items)

Responsibility and accountability (4 items)

The scale uses a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Data collection procedures

After obtaining permission from Klemenc-Ketiš and Vrecko to use the instrument:

Step 1: The English version of the instrument was translated using forward–backward translation according to the procedure proposed by Richard W. Brislin (1970). The forward translation was performed independently by two bilingual translators fluent in both English and Vietnamese, with experience in medical terminology and health sciences. The back translation was conducted by two other independent translators who had not accessed the original version and had experience in academic or medical translation to ensure semantic equivalence between the original and translated versions.

Step 2: Ten experts were selected using a two-round Delphi method. In Round 1, the experts evaluated the relevance, clarity, and semantic equivalence of the scale items; their feedback was then summarized and used for revision. In Round 2, the experts re-evaluated the revised version to achieve consensus on the content validity of the scale.

Step 3: The final Vietnamese version was pilot-tested among 30 medical students before official data collection. A descriptive study was conducted among 220 third-year medical students to evaluate the construct validity of the Vietnamese version after content agreement by the expert panel. To assess test–retest reliability, 110 students who completed the first survey were randomly selected to complete the second assessment after two weeks. Participants' responses from the two assessments were matched using unique identification codes assigned to each student to ensure anonymity and data consistency.

2.3. Data processing and analysis

Data were entered and analyzed using SPSS version 20.0.

Cronbach's alpha was used to assess internal consistency reliability, with a threshold of $\alpha > 0.7$ (Pallant, 2013) [5].

Intra-class correlation coefficients (ICCs) were used to evaluate stability, with acceptable values ranging from 0.6 (and in some cases 0.5) according to Polit and Beck (2012) [5].

Validityassessment:

Construct validity was evaluated using exploratory factor analysis (EFA). Principal component analysis (PCA) was applied to extract factors based on the following criteria: eigenvalues greater than 1, cumulative percentage of explained variance, and factor distribution on the Scree plot. Orthogonal factor rotation (Varimax) was then used to identify items with factor loadings greater than 0.45, in accordance with established criteria [6].

2.4. Ethical considerations

The study complied with ethical principles, ensuring voluntary participation, confidentiality of personal information, and approval by the Scientific Council of Thai Binh University of Medicine and Pharmacy.

III. RESULTS

Table 1. I-CVI and S-CVI/Ave indices of the Vietnamese version of the instrument

Sub-scale	CVI				
	Relevance	Clarity	Comprehensiveness	Appropriateness of rating scale	I-CVIs
1. Empathy and Humanism	0.83	0.83	1	0.83	0.87
2. Professional Relationships and Professional Development	0.83	0.83	0.83	0.83	0.83
3. Responsibility / Accountability	1	1	0.83	0.83	0.90
S-CVI/Ave: 0.87					

Comment: The I-CVI index for the Empathy and Humanism domain was 0.87; for Professional Relationships and Professional Development, 0.83; and for Responsibility and Accountability, 0.90. The overall average scale-level content validity index (S-CVI/Ave) for the entire instrument was 0.87.

Table 2. Correlation and Internal Consistency Reliability of Items in the Instrument

Sub-scale	Correlation with Items (r)	Cronback's α if Item Deleted	Cronback's α
1. Empathy and Humanism	0.837-0,907	0.982-0,983	0.983
2. Professional Relationships and Professional Development	0.648-0,910	0.982-0,984	0.981
3. Responsibility / Accountability	0,713-0,878	0,983	0.961

Comment: The correlation between items in Part 1 and the item–total correlation index (r) ranged from 0.837 to 0.907, with a Cronbach's α of 0.983. In Part 2, the item–total correlation index (r) ranged from 0.648 to 0.910, with a Cronbach's α of 0.981. In Part 3, the item–total correlation index (r) ranged from 0.713 to 0.878, with a Cronbach's α of 0.961.

Table 3. Test–Retest Reliability Results of the Instrument

Sub-scale	Test Mean (SD)	Retest Mean (SD)	ICC	95% CI	p
Part 1 (10 items)	4.25 (1.45)	4.27 (1.42)	0.851	0.79-0.89	0.13
Part 2 (8 items)	4.35 (1.31)	4.38 (1.27)	0.850	0.78 – 0.89	0.2
Part 3 (4 items)	4.48 (1.05)	4.5 (1.05)	0.956	0.93-0.97	0.31
Total scale (22 items)	4.36 (1.27)	4.38 (1.25)	0.869	0.81-0.9	0.22

Comment: The mean test–retest values ranged from 4.25 to 4.50, ICC ranged from 0.850 to 0.869, and p-values ranged from 0.13 to 0.31.

Table 4. Internal Consistency Coefficients of the Scale

Components of the Scale	Number of Items	Cronback α		Test – retest value Internal Correlation Coefficient			
		Original Version	Vietnamese Version	Internal Correlation Coefficient	95 % CI		F test
					Lower	Upper	p
1. Empathy and Humanism	10	0.84	0.983	0.851	0.825	0.876	0.000
2. Professional Relationships and Professional Development	8	0.78	0.981	0.850	0.823	0.877	0.000
3. Responsibility / Accountability	4	0.75	0.961	0.956	0.944	0.965	0,000

Comment: The internal consistency reliability (Cronbach's α) of the original version ranged from 0.75 to 0.84, while that of the Vietnamese version ranged from 0.961 to 0.983. The internal correlation coefficient (ICC) ranged from 0.850 to 0.965.

Table 5. Results of KMO and Bartlett's Test

Parameter	Value
KMO (Kaiser-Meyer-Olkin)	0.874
Bartlett's Test of Sphericity – Chi-square	2065.4
df	231
Sig. (p-value)	0.000

Comment: KMO > 0.8 indicates that the data are suitable for factor analysis. Bartlett's Test with $p < 0.05$ indicates that the correlation matrix is significantly different from the identity matrix; therefore, the conditions for conducting EFA are satisfied.

Table 6. Total Variance Explained

Instrument Components	Eigenvalue	% of Variance	Cumulative Variance (%)
Part 1	6.75	30.7%	30.7%
Part 2	4.05	18.4%	49.1%
Part 3	2.70	12.2%	61.3%

Comment: A total of three factors explained 61.3% of the overall variance, exceeding the 50% threshold, indicating that the factors adequately explain the data structure.

Table 7. Summary of Unidimensionality Testing and Item Removal

Test	Result
items with cross-loading > 0.4 on ≥ 2 factors	0 items
Items with factor loading< 0.4	0 items
Items removed from the model	No items were removed

Comment: All items were appropriate and consistent with the measurement structure.

IV. DISCUSSION

The study results indicate that the instrument for assessing medical students' professionalism met key standards of content validity, reliability, and structural validity. Based on the evaluation of 10 experts, the item-level content validity index (I-CVI) for most items reached the acceptable threshold (≥ 0.78) as recommended by Lynn (1986) and Denise F. Polit & Cheryl Tatano Beck (2006) [5]. The overall

scale-level content validity index (S-CVI/Ave) was 0.87, reflecting a high level of agreement regarding the relevance and representativeness of the items. A few items with lower I-CVI values (such as those in the "Professional Relationships and Development" domain) suggest the need for further refinement in wording or clarification to improve expert consensus. Nevertheless, overall indices were within acceptable or good ranges, confirming that the content foundation of the instrument is robust.

Across domains, the "Responsibility and Accountability" component achieved an S-CVI/Ave of 0.91—exceeding the 0.90 threshold considered excellent in methodological literature. The domains "Empathy and Humanism" and "Professional Relationships and Development" also demonstrated high levels of appropriateness. Compared with domestic studies such as those by Ho (2023) and Huang (2024) [7],[8], where professionalism scales reported S-CVI/Ave values ranging from 0.88 to 0.89, the present findings are comparable or slightly higher. Internationally, the structure and level of content validity are consistent with the Professionalism Assessment Scale developed by Zalika Klemenc-Ketis and colleagues, as well as the Professionalism Survey developed by Gail F. Blackall in the United States. This suggests that the instrument aligns well with contemporary standards for professionalism assessment in medical education.

Regarding internal consistency reliability, Cronbach's alpha coefficients for all domains exceeded 0.9; notably, the "Empathy and Humanism" domain reached 0.983, indicating very high internal consistency among items. According to the classification of Darren George and Paul Mallery (2003), an alpha > 0.9 is considered "excellent." Item–total correlation coefficients all exceeded the 0.3 threshold recommended by Robert F. DeVellis, demonstrating that each item meaningfully contributed to the overall construct. Compared with studies conducted in South Korea (Kim et al., 2015) and Canada (Cruess et al., 2016) [9], [10], the alpha values in this study are comparable or higher, supporting the instrument's suitability within the Vietnamese medical education context.

Temporal stability was examined using the intraclass correlation coefficient (ICC), which reached 0.869 for the overall scale—classified as "good" according to Tae Kyun Koo and Mae Li. All subscales achieved ICC values above 0.85, with the "Responsibility" domain reaching an excellent level. The similarity between test and retest scores, along with statistical significance, confirms that the instrument demonstrates strong stability and can be reliably used for repeated assessments.

Exploratory factor analysis (EFA) further reinforced the scale's structural validity. The KMO coefficient of 0.874 indicated sampling adequacy according to the criteria of Henry F. Kaiser, and Bartlett's test was statistically significant, confirming sufficient inter-item correlations for factor extraction. Three factors were retained based on the Eigenvalue > 1 criterion, explaining 61.3% of the total variance—exceeding the recommended 60% threshold suggested by Joseph F. Hair et al. (2010) [11]. The three-factor structure included: (1) Empathy and Humanism (10 items), (2) Professional Relationships and Development (8 items), and (3) Responsibility/Accountability (4 items). All factor loadings were above 0.4, with no significant cross-loadings, indicating clear discriminant validity among components and no need to remove any observed variables.

Overall, the findings demonstrate that the 22-item instrument satisfies high standards of content validity, internal consistency reliability, temporal stability, and structural validity. The results are not only consistent with domestic research but also aligned with international trends in professionalism scale development. This provides a strong scientific foundation for implementing the instrument in formal surveys and applying it in teaching and evaluating professionalism competencies among medical students in Vietnam.Top of Form

V. CONCLUSION

The 22-item instrument demonstrated good content validity (S-CVI/Ave = 0.87; I-CVI ≥ 0.78) according to the standards of Lynn and Denise F. Polit & Cheryl Tatano Beck. Internal consistency reliability was very high, with Cronbach's alpha > 0.9 (classified as "excellent" according to Darren George & Paul Mallery), and ICC = 0.869 indicating good stability as defined by Tae Kyun Koo & Mae Li. Exploratory factor analysis (EFA) confirmed a three-factor structure, with a total variance explained of 61.3%, meeting the criteria proposed by Henry F. Kaiser. Further refinement of items with lower I-CVI values is recommended. Confirmatory factor analysis (CFA) should be conducted on a large,

multicenter sample. The instrument should also be applied in periodic assessments in combination with 360-degree evaluation methods to enhance training effectiveness.

REFERENCES

1. **O'Tuathaigh CMP, Idris NA, Duggan E et al. (2019).** Medical students' empathy and attitudes towards professionalism: Relationship with personality, specialty preference and medical programme. PLoS One, 2019. 14(5): p. e0215675.
2. **Joynt GM, Wong W-T, Ling L et al (2017)**. Medical students and professionalism – Do the hidden curriculum and current role models fail our future doctors? Medical Teacher, 2017. 40(4): p. 395-399.
3. **Al-Eraky MM, Donkers J, Wajid G et al. (2015)**. Faculty development for learning and teaching of medical professionalism. Med Teach, 2015. 37 Suppl 1: p. S40-6.
4. **Klemenc-Ketis, Z. and H. Vrecko. (2014).** Development and validation of a professionalism assessment scale for medical students. Int J Med Educ, 2014. 5: p. 205-11.
5. **Polit DF., & Beck CT. (2012).** Nursing Research: Generating and Assessing Evidence for Nursing Practice (9th ed.). Lippincott Williams & Wilkins.
6. **Aliswag EG. (2020).** Applied statistics in Professional Research Practice. 2020, 2268 J. DelaRosa St. Sampaloc Manina. Philippines 1101.
7. **Ho JY, Tuang V, Teo DB, Ponnamperuma G et al (2023).** Development and validation of a new self-assessment tool to measure professionalism among medical students. Annals Academy of Medicine Singapore, 52(9), 457–466. DOI: 10.47102/annals-acadmedsg.2022457.
8. **Huang MC & Pham TPT (2024).** Psychometric validation of the Nursing Professional Values Scale-Revised: Vietnamese version. Jurnal Ners, 19(4). DOI: 10.20473/jn.v19i4.61702.
9. **Kim B, Park J, Ko WS et al. (2024).** Factors influencing nursing professionalism among nursing students in South Korea: a descriptive study. Child Health Nursing Research 2024;30(4):321-330 https://doi.org/10.4094/chnr.2024.036
10. **Cruess, R.L, Cruess, SR, & Steinert, Y. (2016).** Teaching Medical Professionalism: Supporting the Development of a Professional Identity (2nd ed.). Cambridge University Press.
11. **Hair JF, Black WC., Babin, BJ et al. (2010)**. Multivariate data analysis (7th ed.). Pearson.