Original Article
http://mjiri.iums.ac.ir
Medical Journal of the Islamic Republic of Iran (MJIRI)

Med J Islam Repub Iran. 2026 (18 May);40.49. https://doi.org/10.47176/mjiri.40.49

Empathy Throughout the Clinical Years in Iranian Generation Z Medical Students: Validity and Reliability Properties of the Persian Version of the Jefferson Scale of Empathy

Somayeh Delavari[1], Farzaneh Barzkar[2,3]*, Seyed Kamran Soltani Arabshahi[1], Zohreh Sohrabi[1], Ahmad Madankan[4], Neda Rahimian[5,6], Mitra Amini[7], Mohammadreza Pourahmadi[8], Seyed Vahid Shariat[9]*

Received: 23 Dec 2025 Accepted: 29 Apr 2026 Published: 18 May 2026

Abstract

Background: The Jefferson Scale of Empathy–Student Version (JSE-S) is among the most widely used instruments for assessing empathy in medical students. This study aimed to evaluate the validity and reliability of the Persian version of the JSE-S among Iranian undergraduate clinical trainees, specifically clerks and interns.

Methods: A cross-sectional study was conducted involving 531 medical students. Content validity was assessed using the CVR, CVI, and kappa coefficient, while face validity was evaluated through student ratings. A PLS-based confirmatory measurement model was implemented in SmartPLS 4 to evaluate construct validity and reliability, which included item loadings, CR, AVE, and discriminant validity.

Results: The Persian JSE-S exhibited excellent content validity, as indicated by the S-CVI/UA of 0.80, I-CVI of 0.97, and $\kappa \geq 0.87$, along with satisfactory face validity (S-FVI/Av = 0.80). Cronbach's alpha coefficients for the subscales of Perspective Taking, Compassionate Care, and Standing in the Patient's Shoes were 0.846, 0.817, and 0.849, respectively, demonstrating strong internal consistency. The CR values ranged from 0.850 to 0.865. Following the removal of item JSE-18, the AVE for Compassionate Care surpassed the 0.50 threshold, whereas Perspective Taking remained slightly below it at 0.42. Discriminant validity was confirmed through cross-loadings, Fornell–Larcker, and HTMT criteria (all <0.90).

Conclusion: The Persian version of the JSE-S demonstrated satisfactory psychometric properties, affirming its reliability and validity for assessing empathy among Iranian Generation Z clinical medical students. Despite minor limitations in convergent validity regarding Perspective Taking, the scale's robust structure, internal consistency, and discriminant validity confirm its utility in educational and research contexts. Further multicenter studies with larger samples are recommended to enhance cross-cultural generalizability.

Keywords: Empathy, Generation Z, Medical students, Medical Education, Measurement Model

Cite this article as: Delavari S, Barzkar F, Soltani Arabshahi SK, Sohrabi Z, Madankan A, Rahimian N, Amini M, Pourahmadi M, Shariat SV. Empathy Throughout the Clinical Years in Iranian Generation Z Medical Students: Validity and Reliability Properties of the Persian Version of the Jefferson Scale of Empathy. *Med J Islam Repub Iran.* 2026 (18 May);40:49. https://doi.org/10.47176/mjiri.40.49

Corresponding author: *Dr Seyed Vahid Shariat, shariat.v@iums.ac.ir*
Dr Farzaneh Barzkar, barzkarfarzane@gmail.com

1. Center for Educational Research in Medical Sciences (CERMS), Medical Education Department, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
2. National Brain Centre, Iran University of Medical Sciences, Tehran, Iran
3. Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran
4. Firoozgar Colorectal Department, GILDRIC, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
5. Department of Internal Medicine, Firoozgar Hospital, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
6. Endocrine Research Center, Institute of Endocrinology and Metabolism, Iran University of Medical Sciences (IUMS), Tehran, Iran
7. Clinical Education Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
8. Rehabilitation Research Center, Department of Physiotherapy, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran
9. Mental Health Research Center, Psychosocial Health Research Institute, and Department of Psychiatry, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

↑What is "already known" in this topic:
The JSE-S is the most widely utilized tool for measuring cognitive empathy in medical education, with previous Persian validations among Iranian medical students confirming its three-factor structure. However, PLS-SEM confirmatory analyses and psychometric properties in Generation Z clinical trainees remain limited, and inconsistent item performance has been noted internationally.

→What this article adds:
This study validates the Persian JSE-S in 531 Generation Z Iranian clinical trainees, confirming excellent reliability, strong loadings, and robust content and discriminant validity (Fornell-Larcker/HTMT), along with a marginal Perspective Taking AVE. These findings affirm the utility of the JSE-S for monitoring empathy during the clinical years among Generation Z trainees, demonstrating superior psychometric properties compared to previous Iranian studies.

[DOI: 10.47176/mjiri.40.49]

Introduction

Physician empathy is widely regarded as a crucial characteristic of effective physicians and a cornerstone of the physician-patient relationship (1). High levels of empathy among physicians are associated with improved patient satisfaction, treatment adherence, and enhanced clinical outcomes (2-4). Beyond benefiting patients, empathy may also contribute to the well-being of physicians. Higher empathy scores in medical practitioners correlate with better mental health, lower psychological distress, and greater professional satisfaction (2, 5). Conversely, decreased empathy has been linked to an increase in self-perceived medical errors, suggesting a potential reciprocal relationship between empathy and clinical performance (2, 6).

Empathy measurement has gained significant importance in medical education for several compelling reasons. First, empathy is recognized as an essential component of clinical competence and medical professionalism (7). Second, the measurement of empathy has become increasingly sought after, particularly for informing and enhancing training methodologies (8). This measurement enables educators to assess the effectiveness of empathy-focused educational interventions. Third, empathy measurement provides concrete metrics for students to evaluate their own progress (9). Fourth, it allows educational institutions to identify concerning trends, such as the documented decline in empathy during medical training. Multiple studies have shown that empathy tends to erode during medical education, particularly in the third year (10, 11). Fifth, empathy measurement is crucial for research linking empathy to clinical outcomes, as studies have demonstrated that physicians with higher empathy scores achieve better clinical outcomes for their patients (12, 13).

Several standardized tools for measuring empathy have been developed, with the Jefferson Scale of Empathy (JSE) being the most widely used in medical education (14). The JSE has been specifically constructed within the context of the doctor-patient relationship and patient care (15, 16). It is a widely studied instrument, featuring a specially designed version to assess empathy in medical students (S-version) (17-20). Given that social constructs are embedded in cultural and historical contexts, the definitions and operationalizations of these constructs in measurement tools (e.g., questionnaires or scales) can vary across generational perspectives (21). While the JSE has been validated in the Iranian medical school context, its psychometric properties among Generation Z—characterized by their egalitarianism, fatalism, and community orientation—remain to be elucidated (15, 22). Furthermore, previous assessments did not specifically focus on clinical trainees. Therefore, we decided to examine the PLS-based confirmatory measurement model of the Persian version of the JSE-S and empathy throughout the Clinical Years among Iranian medical interns and clerks. Thus, this study aimed to assess the validity (convergent and divergent validity) and reliability of the Persian version of the JSE-S questionnaire among Iranian Gen Z medical students.

Methods

Study design

This cross-sectional study was conducted at the Iran University of Medical Sciences (IUMS) and Shiraz University of Medical Sciences (SUMS) during the 2024-2025 academic year.

Study participants, methods for participant selection, and sample size

Undergraduate medical students undergoing clinical training at IUMS and SUMS in both inpatient and outpatient departments were eligible to participate in the study. After explaining the study objectives, a questionnaire link was distributed via a QR code to clerks and interns during morning reports across all major clinical departments. Using the rule of thumb 10x medical students per item and 10% lost to follow-up, minimum sample size estimates about 220 participants as sample size. A total of 540 medical students (IUMS: 435, SUMS:105) completed the questionnaire; however, 9 responses were excluded due to incomplete or poor-quality data (A respondent must answer at least 16 (80%) of the 20 items; otherwise, the form should be regarded as incomplete and excluded from the data analysis) on the JSE-S. Ultimately, 531 clinical trainees were included in the final analysis.

Study measures

The questionnaires included a section on socio-demographic information, encompassing age, sex, marital status, educational status (clerk or intern), clinical department (surgical vs. non-surgical), and the JSE.

Jefferson Scale of Empathy

The JSE is a standardized instrument that has been presented in three versions: the health professionals' students (HPS-version), one for health professionals (HP-version), and another for medical students (S-version). It is a self-reported questionnaire consisting of 20 questions, rated on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). The Persian version of the JSE-S includes three subdomains: Perspective Taking (10 items), Compassionate Care (8 items), and Standing in the Patient's Shoes (2 items) (7, 22, 23). This tool was developed in 2001 by Hojat et al. (23). The psychometric properties of the Persian translation of the S-version of the JSE were assessed by Shariat and Habibi (22) and Rahimi-Madiseh et al. (15). We utilized the Persian version of the JSE provided by the coauthor who standardized this version and obtained formal permission from Thomas Jefferson University for its use in this study.

To ensure the validity of responses on the JSE, participants must answer at least 16 items (80%) out of the total 20. If a participant answers 16 or more items, the missing values will be replaced with the mean score calculated from the items that the participant completed (23).

[DOI: 10.47176/mjiri.40.49]

Statistical analysis

Descriptive analyses and PLS-based confirmatory measurement model were conducted. Content validity was assessed using the Content Validity Ratio (CVR) and Content Validity Index (CVI), which were calculated in Microsoft Excel. Eight expert medical instructors evaluated the relevance and necessity of each item. According to Lawshe's criteria, the minimum acceptable CVR value for eight panelists was 0.75, and the acceptable threshold for CVI was ≥0.83 (24-26). To account for chance agreement among experts, item-level CVIs, the scale-level CVI (S-CVI), and the kappa coefficient were also computed (27).

Face validity was examined by asking 30 medical students to rate the clarity of each item on a 4-point scale ranging from 1 ("not clear") to 4 ("very clear") (28).

PLS-based confirmatory measurement model

To assess the validity, reliability, and measurement model of the JSE-S, Smart-PLS 4 software was utilized in accordance with the recommended procedures for reflective measurement models. The analysis sequentially evaluated indicator loadings (item reliability), internal consistency reliability, convergent validity, and discriminant validity to ensure accurate and reliable construct measurement (29, 30). Loadings greater than 0.70 were deemed desirable, as they indicate that an item explains approximately 50% of the variance, reflecting satisfactory reliability (30). Factor loadings above ±0.30 met the minimum acceptable level, those above ±0.40 were considered more meaningful, and loadings exceeding ±0.50 were regarded as practically significant (31). Given that the sample size exceeded 200, all items with loadings ≤±0.40 were removed from the model (32).

Internal consistency reliability was evaluated using both composite reliability (CR) and Cronbach's alpha (α). CR offers a more precise estimate of reliability as it considers individual indicator loadings, and its interpretation aligns with that of Cronbach's alpha. CR values exceeding 0.70 indicate acceptable reliability, whereas those above 0.80 signify high internal consistency (33).

Convergent validity was assessed using the AVE, defined as the grand mean of the squared loadings of all indicators associated with a construct (33). AVE values greater than 0.50 indicate that the latent variable explains more than half of the variance of its indicators, thereby representing satisfactory convergent validity (30, 34).

Discriminant validity reflects the degree to which a latent variable is distinct from other constructs within the model (35). It was evaluated using cross-loadings, the Fornell–Larcker criterion, and the Heterotrait–Monotrait (HTMT) ratio (35, 36). Cross-loadings indicate that each item should load highest on its respective latent construct compared to others (30). According to the Fornell–Larcker criterion, the square root of a construct's AVE must exceed its correlations with other latent variables (36, 37). The HTMT ratio, a more recent and robust measure of discriminant validity, assesses the average correlations between constructs. High HTMT values suggest potential issues with discriminant validity. Henseler et al. (2015) recommended a threshold of 0.90, with values below this cutoff indicating adequate discriminant validity; a more conservative cutoff of 0.85 may be applied when constructs are conceptually distinct (36, 38).

Results

Participants' Characteristics

A total of 531 undergraduate medical students (288 clerks and 243 interns) participated in this study, with a mean (SD) age of 24.35 (±2.35) years. Females comprised 49% (261/531) of the sample. Approximately 61% of participants were engaged in non-surgical clinical rotations. Table 1 summarizes the demographic characteristics of both clerk and intern groups.

Content Validity

In the present study, the Content Validity Ratio (CVR) exceeded 0.75 for all items of the JSE-S questionnaire. The scale-level Content Validity Index based on Universal Agreement (S-CVI/UA) was 0.80, indicating acceptable content validity, while the average item-level Content Validity Index (I-CVI) was 0.97, as evaluated by eight expert medical educators. The kappa coefficient for all items was ≥0.87, demonstrating excellent agreement beyond chance.

Regarding face validity, the Item-level Face Validity Index (I-FVI) was 0.98, the scale-level Face Validity Index based on Average (S-FVI/Av) was 0.80, and the scale-level Face Validity Index based on Universal Agreement (S-FVI/UA) was 0.98. Furthermore, the average proportion of items deemed clear and comprehensible by 30 medical students was 0.98. Overall, the JSE-S questionnaire exhibited high clarity and comprehensibility, with a proportion of 0.98 across all items.

Descriptive statistics for JSE-S

Table 2 presents the means (SD) and medians [Q25–Q75] for the three components of the JSE-S and the total score. The total mean and median scores for the JSE-S were 100.70 (15.09) and 101 [91–111], respectively.

Reliability

The findings from the primary order measurement mod-

Table 1. Participants' Characteristics

Characteristics	Clerks	Interns	Total	Missing data
	Number (%)	Number (%)	Number (%)	Number (%)
Gender (female)	149 (52%)	112 (46%)	261 (49%)	-
Marital status (single)	265 (92%)	209 (86%)	474 (89.26%)	-
Clinical department (non-surgery)	160 (59%)	143 (63%)	303 (61%)	36 (7%)[1]
Age Mean (SD)	23.47 (2.19)	25.39 (2.11)	24.35 (2.35)	-

[1]. Clinical department information was not reported by fewer than 10% of participants; consequently, these cases were excluded.

[DOI: 10.47176/mjiri.40.49]

Table 2. Descriptive statistics for JSE-S

Subscales	No.	item	Mean (SD)		Median [q_{25} q_{75}]	
perspective taking	JSE16	Physicians' understanding of the emotional status of their patients, as well as that of their families is one important component of the physician-patient relationship	5.36 (1.28)	52.21 (8.78)	5[5_6]	52 [47_58]
	JSE13	Physicians should try to understand what is going on in their patients' minds by paying attention to their non-verbal cues and body language	5.21 (1.31)		5[4_6]	
	JSE20	I believe that empathy is an important therapeutic factor in medical treatment	5.62 (1.19)		6[5_7]	
	JSE15	Empathy is a therapeutic skill without which the physician's success is limited	5.40 (1.33)		6[5_6]	
	JSE10	Patients value a physician's understanding of their feelings which is therapeutic in its own right	5.47(1.29)		6[5_6]	
	JSE2	Patients feel better when their physicians understand their feelings	5.91 (1.12)		6[5_7]	
	JSE4	Understanding body language is as important as verbal communication in physician-patient relationships	5.59 (1.14)		6[5_6]	
	JSE5	A physician's sense of humor contributes to a better clinical Outcome	5.06 (1.44)		5[4_6]	
	JSE9	Physicians should try to stand in their patients' shoes when providing care to them	4.54 (1.73)		5[3_6]	
	JSE17	Physicians should try to think like their patients in order to render better care	4 (1.78)		4[3_5]	
compassionate care	JSE11	Patients' illnesses can be cured only by medical or surgical treatment; therefore, physicians' emotional ties with their patients do not have a significant influence in medical or surgical treatment*	5.01 (1.65)	39.09 (8.52)	5[4_6]	41 [34_45]
	JSE8	Attentiveness to patients' personal experiences does not influence treatment outcomes*	4.8 (1.61)		5[4_6]	
	JSE7	Attention to patients' emotions is not important in history taking*	5.53 (1.51)		6[5_7]	
	JSE14	I believe that emotion has no place in the treatment of medical illness*	5.49(1.59)		6[5_7]	
	JSE18	Physicians should not allow themselves to be influenced by strong personal bonds between their patients and their family members*	2.81(1.52)		3[2_4]	
	JSE1	Physicians' understanding of their patients' feelings and the feelings of their patients' families does not influence medical or surgical treatment*	4.78 (1.62)		5[4_6]	
	JSE19	I do not enjoy reading nonmedical literature or experiencing the arts	5.67 (1.75)		6[5_7]	
	JSE12	Asking patients about what is happening in their personal lives is not helpful in understanding their physical complaints*	4.97 (1.58)		5[4_6]	
standing in the patient's shoes	JSE3	It is difficult for a physician to view things from patients' perspectives*	4.72 (1.54)	9.40 (2.87)	5[4_6]	10 [8_12]
	JSE6	Because people are different, it is difficult to see things from patients' perspectives*	4.67(1.53)		5[4_6]	
JSE_S total			100.70 (15.09)		101 [91_111]	

el indicated that Cronbach's alpha values for perspective-taking, compassionate care, and standing in the patient's shoes were 0.846, 0.817, and 0.849, respectively, demonstrating strong internal consistency among the items (Figure 1, Table 3). Furthermore, the results revealed that the Dijkstra–Henseler's rho (rho_a) scores for the JSE-S subscales collectively ranged from 0.850 to 0.865, underscoring the high reliability of the scales employed in this research. The low loadings (below 0.40) for one of the 20 items suggest that it may be feasible to remove this item, thereby enhancing the conciseness of the scale.

Item loadings greater than 0.70 are recommended, as they adequately explain approximately 50% of the variance of the indicators, thereby providing satisfactory item reliability (36). Ten of the 19 items demonstrated high loadings (≥ 0.70). After the removal of JSE_18, the results of the secondary order confirmatory factor analysis of the measurement model revealed item loadings greater than 0.4, ranging from 0.417 to 0.935 (Figure 2, Table 3). Consequently, it can be concluded that all observed variables in this study, with the exception of JSE_18, exhibit satisfactory reliability. The SRMR in primary and secondary order PLS-based confirmatory measurement models were 0.172 and 0.168, respectively.

Convergent and Discriminant Validity

According to Table 3, the JSE subscales exhibited AVE values ranging from 0.420 to 0.869 in both primary and secondary-order PLS-based confirmatory measurement models. However, some subscales did not meet the recommended AVE threshold of 0.5, indicating that convergent validity was not acceptable for all scales. Specifically, after the removal of item JSE-18, the AVE for the Compassionate Care subscale increased from 0.462 to 0.523, surpassing the 0.5 criterion. In contrast, the Perspective Taking subscale showed no change, maintaining an AVE of 0.420, and therefore failed to meet the suggested threshold for acceptable convergent validity.

The measurement model demonstrated robust internal consistency reliability, with Composite Reliability (CR) values exceeding 0.70 across all constructs, thereby meeting stringent psychometric standards. Although the Average Variance Extracted (AVE) for Perspective Taking was 0.420, which is below the conventional 0.50 threshold,

[DOI: 10.47176/mjiri.40.49]

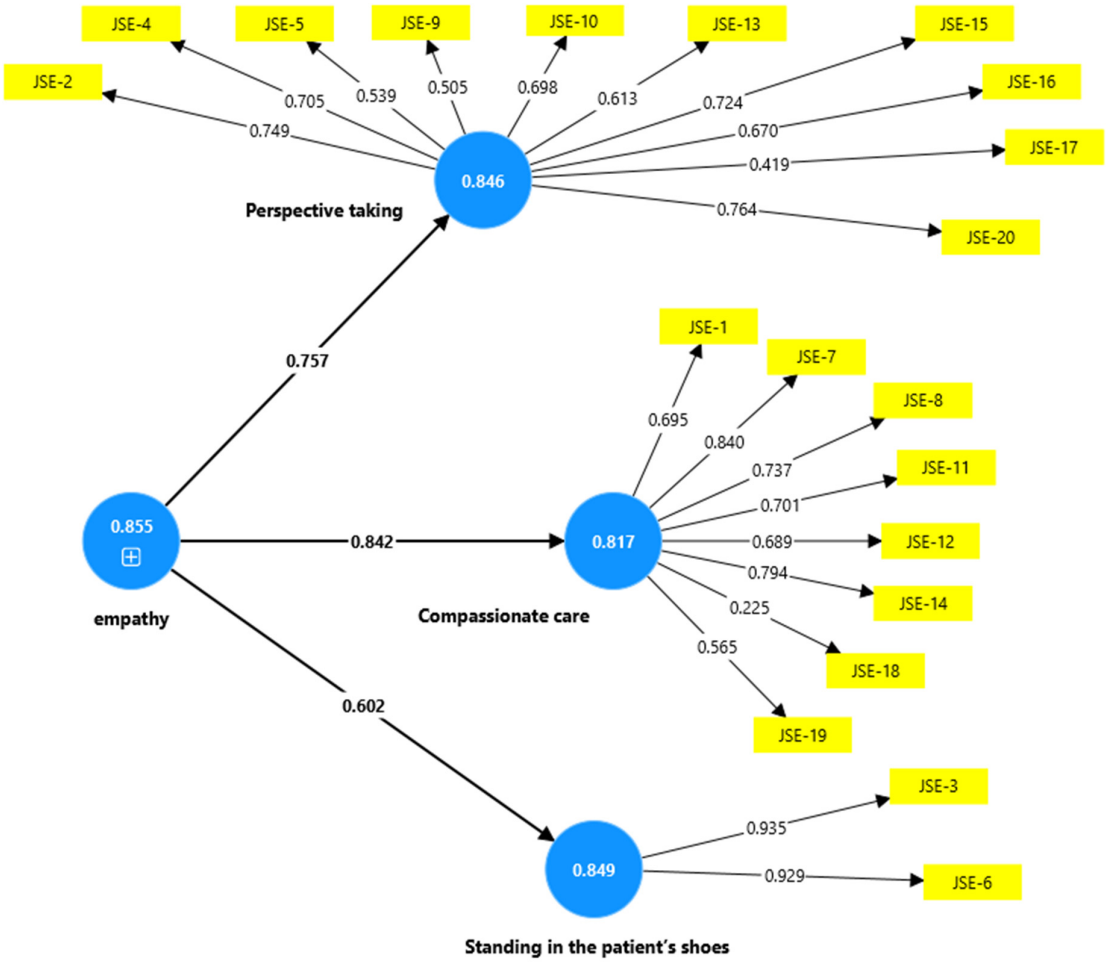

Figure 1. Primary order PLS-based confirmatory measurement model for the Persian version of JSE-S

convergent validity remains acceptable when CR is adequate—a position explicitly supported by foundational PLS-SEM methodology (37). Research indicates that CR serves as a more conservative and reliable indicator of shared variance than AVE, particularly in variance-based structural equation modeling (38, 39). All indicators were systematically retained to preserve content validity and theoretical integrity, as the study utilized the well-validated Jefferson Scale of Empathy, a standardized instrument with established psychometric properties across diverse populations. This decision prioritizes construct comprehensiveness over marginal statistical optimization, consistent with contemporary PLS-SEM guidance that recommends against post-hoc item deletion when reliability metrics (CR ≥ 0.70) and indicator loadings (all P<0.01, loadings ≥ 0.40 for n > 200) are satisfactory (32).

Key thresholds for interpretation: AVE ≥ 0.50 indicates that the construct explains 50 percent or more of the variance of the indicators that comprise the construct (36).

Size of the factor loading (FL): ≥0.71: Excellent, 0.63-0.70: Very Good, 0.55-0.63: Good, 0.45-0.54: Fair, 0.32-0.44: Poor, and ≤ 0.31: Negligible (40).

CR values are interpreted similarly to alpha values, with values exceeding 0.70 considered acceptable and those exceeding 0.80 regarded as excellent (33).

Discriminant validity was assessed as recommended, utilizing cross-loadings, the Fornell–Larcker criterion, and the HTMT. As illustrated in Table 4, all items in both the primary and secondary order PLS-based confirmatory measurement models exhibited the highest loadings with their corresponding latent variables in comparison to other constructs. Consequently, these items satisfied the cross-loading criterion, thereby confirming adequate discriminant validity based on this metric.

As shown in Table 5, the square roots of the AVE exceeded the inter-correlations among the research constructs, thereby indicating acceptable discriminant validity.

As presented in Table 6, the HTMT ratios were below 0.90, indicating no issues with construct discriminant validity. Consequently, the HTMT ratio criterion was also satisfied in this study.

[DOI: 10.47176/mjiri.40.49]

Table 3. Factor Loadings, Cronbach's alpha, CR, and AVE. *PT: perspective taking; **CC: compassionate care; ***SPS: standing in the patient's shoes.

Construct	item	FL**			Primary order PLS-based confirmatory measurement model					Secondary order PLS-based confirmatory measurement model				
		Hojat (41)	Tavakkol, 2010 (42)	Shariat (22)	FL	Cronbach's alpha	CR rho_a	CR rho_c	AVE	FL	Cronbach's alpha	CR rho_a	CR rho_c	AVE
Perspective taking	JSE16	0.70	0.66	0.62	0.670	0.846	0.865	0.875	0.420	0.671	0.846	0.865	0.875	0.420
	JSE13	0.62	0.61	0.51	0.613					0.613				
	JSE20	0.60	0.66	0.64	0.764					0.765				
	JSE15	0.58	0.55	0.64	0.724					0.724				
	JSE10	0.58	0.55	0.58	0.698					0.698				
	JSE2	0.50	0.50	0.49	0.749					0.750				
	JSE4	0.48	·	0.50	0.705					0.705				
	JSE5	0.45	·	0.44	0.539					0.538				
	JSE9	0.46	0.57	0.54	0.505					0.502				
	JSE17	0.46	0.64	0.35	0.419					0.417				
Compassionate care	JSE11	0.60	0.66	0.70	0.701	0.817	0.858	0.865	0.462	0.707	0.845	0.857	0.883	0.523
	JSE8	0.59	0.53	0.28	0.737					0.736				
	JSE7	0.54	0.53	0.59	0.840					0.845				
	JSE14	0.50	0.60	0.60	0.794					0.793				
	JSE18	0.44	0	0.06	0.225					-				
	JSE1	0.43	0.49	0.36	0.695					0.688				
	JSE19	0.37	·	0.40	0.565					0.573				
	JSE12	0.37	0.61	0.53	0.609					0.687				
SPS*	JSE3	0.74	0.76	0.66	0.935	0.849	0.850	0.930	0.869	0.935	0.849	0.850	0.930	0.869
	JSE6	0.66	0.55	0.55	0.929					0.929				

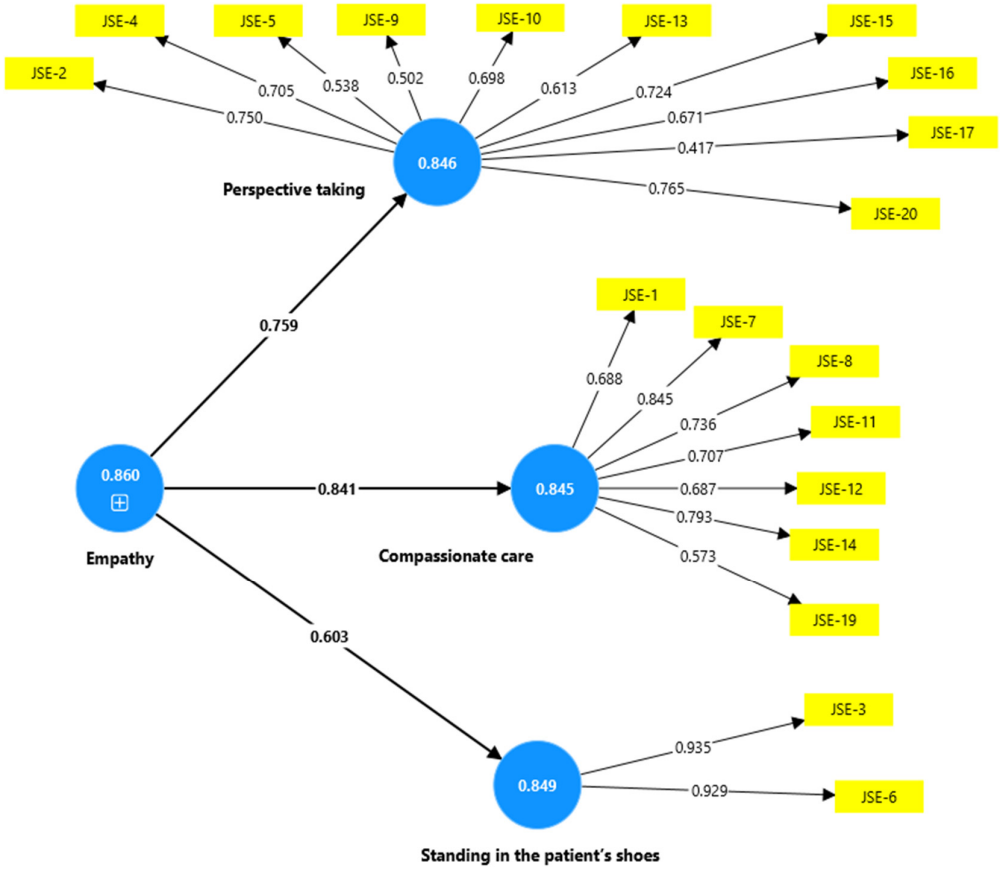

Figure 2. Secondary order PLS-based confirmatory measurement model for the Persian version of the JSE-S.

For conceptually similar constructs, HTMT should be less than 0.90; for conceptually different constructs, HTMT should be less than 0.85 (36).

Discussion

Empathy, a fundamental component of an effective therapeutic relationship, is the emotional ability to comprehend another person's feelings, coupled with the altruistic motivation to address their needs (43). Given the sig-

[DOI: 10.47176/mjiri.40.49]

Table 4. Cross-loading analysis

Construct	Item	Primary order PLS-based confirmatory measurement model			Secondary order PLS-based confirmatory measurement model		
		PT*	CC**	SPS***	PT*	CC**	SPS***
PT	JSE16	0.670	0.206	0.131	0.671	0.210	0.131
	JSE13	0.613	0.156	0.132	0.613	0.157	0.132
	JSE20	0.764	0.292	0.163	0.765	0.292	0.163
	JSE15	0.724	0.245	0.121	0.724	0.244	0.121
	JSE10	0.698	0.274	0.176	0.698	0.274	0.176
	JSE2	0.749	0.342	0.216	0.750	0.346	0.216
	JSE4	0.705	0.271	0.191	0.705	0.275	0.191
	JSE5	0.539	0.152	0.079	0.538	0.149	0.079
	JSE9	0.505	0.031	0.022	0.502	0.021	0.022
	JSE17	0.419	-0.098	-0.044	0.417	-0.106	-0.044
CC	JSE11	0.202	0.701	0.261	0.203	0.707	0.261
	JSE8	0.227	0.737	0.292	0.227	0.736	0.292
	JSE7	0.280	0.840	0.411	0.280	0.845	0.411
	JSE14	0.356	0.794	0.397	0.356	0.793	0.397
	JSE18	0.056	0.225	0.062	-	-	-
	JSE1	0.257	0.695	0.353	0.257	0.688	0.353
	JSE19	0.109	0.565	0.378	0.109	0.573	0.378
	JSE12	0.227	0.689	0.362	0.227	0.687	0.362
SPS	JSE3	0.202	0.466	0.935	0.202	0.467	0.935
	JSE6	0.190	0.434	0.929	0.191	0.436	0.929

Table 5. Fornell-Larcker Criterion Results

	Primary order PLS-based confirmatory measurement model			Secondary order PLS-based confirmatory measurement model		
	CC	PT	SPS	CC	PT	SPS
CC	0.680			0.723		
PT	0.337	0.648		0.338	0.648	
SPS	0.483	0.211	0.932	0.485	0.211	0.932

Table 6. HTMT Discriminant validity

	Primary order PLS-based confirmatory measurement model			Secondary order PLS-based confirmatory measurement model		
	CC	PT	SPS	CC	PT	SPS
CC						
PT	0.393			0.383		
SPS	0.569	0.232		0.575	0.232	

nificant role of empathy in patient-physician relationships, we aimed to assess and determine the psychometric properties and measurement model of the JSE in Iranian students utilizing a PLS-based confirmatory measurement model.

In comparison to earlier Iranian studies conducted among pre-Generation Z medical student cohorts, the psychometric performance of the Persian JSE-S in the current Generation Z sample appears largely comparable in terms of overall reliability and factor structure, while exhibiting subtle differences at the subscale level. Previous Iranian validations, primarily conducted between 2001 and 2015 in mixed or Millennial cohorts, consistently reported acceptable internal consistency and supported a multidimensional structure aligned with the original three-factor model, albeit using earlier psychometric standards that did not routinely report AVE or Composite Reliability (15, 44). Similarly, our findings confirm strong internal consistency and a stable three-factor configuration using a PLS-based confirmatory measurement model, suggesting that the core construct of empathy, as operationalized by the JSE-S, remains psychometrically robust across Iranian cohorts. However, in contrast to earlier studies where construct-level limitations were less explicitly examined, the present analysis identified marginal convergent validity for the Perspective Taking subscale, indicating a potentially greater heterogeneity in how cognitive empathy is expressed or interpreted among Generation Z clinical trainees. Regarding empathy levels, descriptive analysis reveals that the mean empathy score of the present study (100.7 (15.09)) was lower than those reported in previous studies from 2010 (105.1 (12.9)) (45), 2013 (101.4 (14.5)) (22), and 2021 (103.52 (20)) (46); nevertheless, this represents a negligible to small effect (SMD: 0.302 (0.133_0.471), SMD: 0.048 (-0.055_0.15), and SMD: 0.182 (-0.132_0.496), respectively). Subscale patterns also suggest relatively weaker performance in Perspective Taking compared with Compassionate Care, a pattern that was less prominent in prior Iranian cohorts. Collectively, these findings suggest continuity in the overall measurement properties of the Persian JSE-S across generations, accompanied by modest shifts in subscale functioning and empathy expression that may reflect evolving educational, social, and cultural contexts rather than fundamental changes in the construct itself.

The findings from the PLS-based confirmatory measurement model supported the three-factor structure of the JSE within the context of Persian undergraduate clinical education, as reported in the original version (41). The results indicated that the secondary order PLS-based con-

[DOI: 10.47176/mjiri.40.49]

firmatory measurement model for the Persian version of JSE-S possesses acceptable psychometric properties, with the exception of convergent validity as measured by AVE in the perspective-taking subscale. This model is deemed suitable for use among Iranian undergraduate medical students, based on the obtained indicator loadings, internal consistency reliability, and both convergent and discriminant validity. The observed lower discriminant validity in this study may be attributed to the context of Iranian medical students, where cultural norms may further obscure the distinctions between these domains. In many collectivist cultures, empathy is perceived holistically, leading students to potentially not differentiate as distinctly between the cognitive (understanding) and affective (caring) aspects of empathy (15). Although the AVE value for the Perspective Taking construct fell below the conventional threshold of 0.5 (29, 37), the construct demonstrated strong internal consistency reliability, as evidenced by a sufficient Composite Reliability (CR) value exceeding 0.7 (29, 47). In accordance with established methodological guidance in structural equation modeling (SEM), convergent validity can still be supported when the CR is adequate (≥0.7), even if the AVE is marginally below 0.5 (particularly >0.4), especially when utilizing validated instruments (35, 48, 49).

Thus, although the subscales may not be as distinctly reliable as in other contexts, the JSE still consistently measures empathy as a construct.

The PLS-based confirmatory measurement model results for the Persian version of the JSE-S secondary-order measurement model indicated that all observable variables, except for item 18, had factor loadings exceeding the threshold of 0.4, ranging from 0.417 to 0.935, which reflects moderate to excellent loadings. In comparison, factor loadings in the original JSE-S ranged from 0.37 (items 12 and 19) to 0.74 (item 6) (41). Previous studies on the Persian JSE-S reported loadings between 0.06 (JSE 18) and 0.70, with several items (4, 5, 18, and 19) exhibiting loadings ≤0.44, indicating negligible to poor contributions, particularly for JSE 18 (22) (15). Similarly, the Spanish and Brazilian versions of JSE-S demonstrated variable loadings for JSE 18, ranging from 0.07 to 0.68 and 0.30 to 0.85, respectively (20, 50). These inconsistent findings underscore the necessity for further research to elucidate the psychometric properties of the JSE-S across different cultural contexts. Conceptually, JSE-18 appears particularly sensitive to linguistic nuances and cultural interpretation, as its phrasing integrates cognitive appraisal with affective judgment, potentially leading to heterogeneous respondent interpretations across cultures. This semantic ambiguity may compromise its coherence with the reflective measurement assumptions underlying the Compassionate Care construct. In the present study, the removal of JSE-18 resulted in a meaningful improvement in AVE for Compassionate Care, whereas its retention provided no compensatory gain in construct validity. Collectively, the convergence of Persian and cross-cultural evidence supports the treatment of JSE-18 as a psychometrically unstable indicator, warranting cautious interpretation and, where theoretically justified, exclusion to preserve construct validity while maintaining comparability with prior JSE-S research.

Our findings underscore the satisfactory reliability of the secondary-order PLS-based confirmatory measurement model for the JSE-S. The total score exhibited excellent internal consistency, with a Cronbach's alpha surpassing those reported in previous studies involving Iranian medical students (15, 22). The subscales Perspective-Taking, Compassionate Care, and Standing in the Patient's Shoes each demonstrated good internal consistency, with Cronbach's alpha values closely clustered between 0.845 and 0.849. Notably, our Cronbach's alpha values exceed those reported for medical students in the United States (7), Spain (20), and Brazil (50). Composite reliability indices further corroborated the questionnaire's excellent internal consistency, aligning with findings in the existing literature (20).

The results indicated that the JSE-S for undergraduate medical students in clinical settings did not demonstrate acceptable convergent validity for the Perspective Taking and Compassionate Care subscales, as the AVE value did not exceed the 0.50 threshold. For the sensitivity analysis, item JSE-17 was removed due to a low factor loading (<0.44); however, this deletion did not improve the AVE value for the Perspective Taking subscale to reach ≥0.50. Therefore, JSE-17 was retained in the second-order PLS-based confirmatory measurement model. In contrast, within the Compassionate Care subscale, the AVE increased slightly above 0.50 after removing item JSE-18 in the secondary order PLS-based confirmatory measurement model. In comparison, reported AVE values for the Spanish version of the JSE-S ranged from 0.23 to 0.45 (20). These inconsistencies in AVE values highlight the need for further research with larger sample sizes to better establish the construct convergent validity of the Persian version of the JSE-S in clinical undergraduate medical education. Future systematic reviews and meta-analyses are recommended to obtain more robust and conclusive evidence across contexts.

Item-level discriminant validity was assessed through cross-loadings analysis, which revealed that each item loaded more strongly on its intended latent variable than on others, thereby satisfying the cross-loadings criterion (38). This finding indicates that the items are distinct and effectively measure different empathy constructs as intended. Consequently, all items, except JSE 18, were deemed suitable for inclusion in the Persian JSE-S for clinical undergraduate medical students. The cross-loading analysis and the Fornell-Larcker criterion further confirmed discriminant validity by demonstrating clear differentiation among latent variables (36). Additionally, the HTMT ratio was below the threshold of 0.90, supporting the discriminant validity of the Persian JSE-S (38).

The JSE has exhibited strong reliability and validity across diverse cultural and linguistic contexts, including the Persian version validated in the current study (22). This research confirmed high internal consistency, robust content validity, and a stable three-factor structure—Perspective Taking, Compassionate Care, and Standing in the Patient's Shoes—that aligns with prior studies (22).

[DOI: 10.47176/mjiri.40.49]

Additionally, the JSE's sensitivity to changes over time and responsiveness to educational interventions further enhance its utility for tracking empathy development and evaluating curricular innovations in medical training (4, 11).

From a practical standpoint, the JSE offers advantages in terms of brevity and ease of administration. With only 20 items rated on a 7-point Likert scale, it is less time-consuming than instruments such as the Interpersonal Reactivity Index (IRI), which comprises 28 items across multiple subscales (11, 51). The widespread adoption of the JSE in medical schools worldwide has facilitated the establishment of normative data and benchmarks for empathy in clinical training, allowing for meaningful comparisons across institutions and populations (11, 52). This extensive usage supports the integration of the JSE into both formative and summative assessments, as well as in research examining the relationships between physician empathy and patient outcomes, including satisfaction and adherence (52).

Despite these strengths, it is important to acknowledge that the JSE, like other self-report instruments, has certain limitations. Its primary focus on cognitive empathy may underrepresent the affective or emotional components that also contribute to effective patient care (53, 54). Additionally, self-report measures are susceptible to social desirability bias, which can result in inflated empathy scores (4, 55). Some researchers suggest complementing the JSE with behavioral or patient-rated measures to provide a more comprehensive evaluation of clinical empathy (52, 56). Nevertheless, the JSE's robust psychometric foundation, clinical specificity, and practical advantages render it the preferred tool for assessing empathy in medical education, particularly when compared to more general or less validated instruments (12).

Strengths and Limitations

This study possesses several notable strengths. It is among the first to rigorously evaluate the psychometric properties of the Persian version of the Jefferson Scale of Empathy (JSE-S) specifically among Generation Z Iranian clinical medical students, a cohort whose perspectives and educational experiences may differ from those of previous generations. The use of a large and diverse sample of clerks and interns from multiple clinical departments enhances the generalizability of the findings within the context of Iranian undergraduate medical education. Methodologically, the study employed comprehensive assessments of validity and reliability, including content validity indices, confirmatory factor analysis, and advanced reliability metrics such as CR and AVE, thereby ensuring a robust evaluation of the instrument.

However, the study has several limitations. The research was conducted at two institutions with unbalanced sample size, which may restrict the broader applicability of the findings to other Iranian or international contexts. Furthermore, reliance on self-reported data introduces the potential for response bias, including social desirability effects. Lastly, while the JSE-S effectively captures the cognitive aspects of empathy, it may not fully encompass the affective dimensions that are also crucial in clinical care.

Conclusion

This study provides strong evidence that the Persian version of the JSE-S is a reliable and valid instrument for measuring empathy among Generation Z Iranian medical students in clinical training. Our findings confirm robust content validity, high internal consistency, and a stable three-factor structure. The JSE-S effectively captures the cognitive dimension of empathy, which is central to clinical competence and patient care. Overall, the validated JSE-S offers a solid foundation for empathy assessment in Iranian medical education.

The validated Persian JSE-S enables medical educators and researchers to reliably monitor empathy levels in clinical trainees, thereby supporting both formative and summative assessments. Its implementation can help identify students at risk of declining empathy during their clinical years and evaluate the effectiveness of educational interventions designed to foster empathy. Given the established connections between empathy, patient outcomes, and physician well-being, a more comprehensive approach to empathy assessment may prove beneficial. This is particularly pertinent considering the potential relationship between affective empathy, compassion fatigue, and burnout.

Acknowledgment

We are sincerely grateful for the waiver and the comprehensive information that Professor Mohammad Reza Hojat and Dr. Shira Carroll provided on the use of JSE-S. The authors would like to thank medical students at Iran University of Medical Sciences (IUMS) for their participation in the present study.

Conflict of Interests

The authors declare that they have no competing interests.

Authors' Contributions

All authors helped initiate and coordinate the project, gather data, and write the report. MR. P, S. V. Sh., F.B., M.A., and S.D. were involved in the conceptualization of the investigation. Data collection was conducted by S.D., F.B., A.M., M.A. and N.R; S.K.S.A. and Z.S. were involved in the supervision of the investigation. The data was analysed by S.D., and the data interpretation was conducted by S.D., S.V. Sh. and F.B. The manuscript was authored by S.D. and F.B. The final manuscript was perused and endorsed by all authors. Ethics approval and consent to participate.

Ethical Considerations

All participants provided written informed consent, and the study procedures were approved by the Committee on Iran University of Medical Sciences (IR.IUMS.REC.1400.655). The JSE was utilized in this study with permission from Thomas Jefferson University.

[DOI: 10.47176/mjiri.40.49]

Funding Support

This study was approved and supported by Iran University of Medical Sciences, Tehran, Iran.

Data Availability

The datasets applied and/or analyzed in the present study are available from the corresponding authors upon reasonable request.

AI Use Statement

The authors certify that no artificial intelligence (AI) or AI-assisted technologies were used in the design, execution, analysis, interpretation, writing, language editing, or revision of this manuscript. All aspects of the work were performed solely by the authors.

References

1. Hur Y, Cho AR, Kim S. How well do medical students express empathy? Korean J Med Educ. 2014;26 3:217-21.
2. Shi M, Du T. Associations of emotional intelligence and gratitude with empathy in medical students. BMC Med Educ. 2020;20(1):116.
3. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians' empathy and clinical outcomes for diabetic patients. Acad Med. 2011;86(3):359-64.
4. Khajeh Azad MA, Faghankhani M, Asadi A, Golalipour E, Shariat SV, Naserbakht M, et al. The Effects of an Empathic Communication Workshop on Internal Medicine Residents' Self-Perceived Empathy and Their Patients' Perception of Physician's Empathy: A Single-Group Experimental Study. Med J Islam Repub Iran. 2022;36:137.
5. Lamothe M, Boujut E, Zenasni F, Sultan S. To be or not to be empathic: the combined role of empathic concern and perspective taking in understanding burnout in general practice. BMC Fam Pract. 2014;15:15 -
6. West CP, Huschka MM, Novotny PJ, Sloan JA, Kolars JC, Habermann TM, et al. Association of perceived medical errors with resident distress and empathy: a prospective longitudinal study. JAMA. 2006;296(9):1071-8.
7. Hojat M, LaNoue M. Exploration and confirmation of the latent variable structure of the Jefferson scale of empathy. Int J Med Educ. 2014;5:73.
8. Srivastava K, Das R. Empathy: Process of adaptation and change, is it trainable? Ind Psychiatry J. 2016;25:1 - 3.
9. Morris KE, Pappas TN. Creating a Medical School Curriculum to Teach Empathy. Ann Surg Open. 2021;2(3):e085.
10. Mudiyanse RM, editor. Teaching Beyond Knowledge and Skills for the Needs of Society. Sri Lanka J Med 2017.
11. Hojat M, Vergare MJ, Maxwell K, Brainard GC, Herrine SK, Isenberg GA, et al. The Devil is in the Third Year: A Longitudinal Study of Erosion of Empathy in Medical School. Acad Med. 2009;84:1182-91.
12. Hojat M, LaNoue M. Exploration and confirmation of the latent variable structure of the Jefferson scale of empathy. Int J Med Educ. 2014;5:73-81.
13. Derksen F, Bensing J, Lagro-Janssen A. Effectiveness of empathy in general practice: a systematic review. Br J Gen Pract. 2013;63(606):e76-84.
14. Aye SZ, Sein HH, Nyunt MK, Min Y. Comparative Study of Malaysian Medical Students' Empathy Scores Between Preclinical and Clinical Training. Educ Med J. 2022.
15. Rahimi-Madiseh M, Tavakol M, Dennick R, Nasiri J. Empathy in Iranian medical students: a preliminary psychometric analysis and differences by gender and year of medical school. Med Teach. 2010;32(11):e471-e8.
16. Hojat M, Gonnella JS, Mangione S, Nasca TJ, Veloski JJ, Erdmann JB, et al. Empathy in medical students as related to academic performance, clinical competence and gender. Med Educ. 2002;36(6):522-7.
17. Gerbase MW, Gustin MP, Bajwa N, Abbiati M, Baroffio A. Development and Cross-National Validation of a French Version of the Jefferson Scale of Empathy for Students. Eval. Health Prof. 2022;45(3):288-96.
18. Ghardallou M, Zedini C, Sahli J, Ajmi T, Khairi H, Mtiraoui A. Psychometric properties of a French version of the Jefferson Scale of Empathy. Int. J. Med. Educ.2022;13:205.
19. Stansfield RB, Schwartz A, O'Brien CL, Dekhtyar M, Dunham L, Quirk M. Development of a metacognitive effort construct of empathy during clinical training: a longitudinal study of the factor structure of the Jefferson Scale of Empathy. Adv Health Sci Educ. 2016;21(1):5-17.
20. Ferreira-Valente A, Costa P, Elorduy M, Virumbrales M, Costa MJ, Palés J. Psychometric properties of the Spanish version of the Jefferson Scale of Empathy: making sense of the total score through a second order confirmatory factor analysis. BMC Med Educ. 2016;16:1-12.
21. Campbell SM, Twenge JM, Campbell WK. Fuzzy But Useful Constructs: Making Sense of the Differences Between Generations. Work Aging Retire. 2017;3(2):130-9.
22. Shariat SV, Habibi M. Empathy in Iranian medical students: measurement model of the Jefferson scale of empathy. Med Teach. 2013;35(1):e913-8.
23. Hojat M, Mangione S, Nasca TJ, Cohen MJ, Gonnella JS, Erdmann JB, et al. The Jefferson Scale of Physician Empathy: development and preliminary psychometric data. Educ Psychol Meas.2001;61(2):349-65.
24. Lawshe C. A Quantitative Approach to Content Validity. Pers Psychol. 1975.
25. Almanasreh E, Moles R, Chen TF. Evaluation of methods used for estimating content validity. Res Soc Adm Pharm.2019;15(2):214-21.
26. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382-6.
27. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health. 2006;29(5):489-97.
28. Yusoff MSB. ABC of response process validation and face validity index calculation. Educ Med J. 2019;11(10.21315).
29. Sarstedt M, Ringle CM, Hair JF. Partial least squares structural equation modeling. Handbook of market research: Springer; 2021. p. 587-632.
30. Hayat AA, Shateri K, Kamalian Fard S, Sabzi Shahr Babak E, Faraji Dehsorkhi H. Psychometric properties of the persian version of the physician teaching self-efficacy questionnaire. BMC Med Educ. 2023;23(1):163.
31. Peterson RA. A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. Mark Lett. 2000;11:261-75.
32. Meyers LS, Gamst G, Guarino AJ. Applied multivariate research: Design and interpretation: Sage publications; 2016.
33. Purwanto A. Partial least squares structural equation modeling (PLS-SEM) analysis for social and management research: a literature review. J Ind Eng Int. 2021.
34. Ismail K, Nopiah ZM, Mohamad SR, Pang CL. Technical competency among vocational teachers in Malaysian public skills training institutions: Measurement model validation using PLS-SEM. J Tech Educ Train. 2020;12(1).
35. Ab Hamid MR, Sami W, Sidek MM, editors. Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT criterion. Journal of physics: Conference series; 2017: IOP Publishing.
36. Hair Jr JF, Hult GTM, Ringle CM, Sarstedt M, Danks NP, Ray S. Partial least squares structural equation modeling (PLS-SEM) using R: A workbook: Springer Nature; 2021.
37. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res. 1981;18(1):39-50.
38. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci. 2015;43:115-35.
39. Rönkkö M, Evermann J. A critical examination of common beliefs about partial least squares path modeling. Organ Res Methods. 2013;16(3):425-48.
40. Harrington D. Confirmatory factor analysis: Oxford University Press; 2009.
41. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: definition, components, measurement, and relationship to gender and specialty. Am J Psychiatry. 2002;159(9):1563-9.

[DOI: 10.47176/mjiri.40.49]

42. Tavakol M, Wetzel A. Factor Analysis: a means for theory and instrument development in support of construct validity. Int J Med Educ. 2020;11:245-7.
43. Levett-Jones T, Brogan E, Debono D, Goodhew M, Govind N, Pich J, et al. Use and effectiveness of the arts for enhancing healthcare students' empathy skills: A mixed methods systematic review. Nurse Educ Today. 2024:106185.
44. Shariat SV, Habibi M. Empathy in Iranian medical students: measurement model of the Jefferson scale of empathy. Med Teach. 2013;35(1):e913-e8.
45. Rahimi-Madiseh M, Tavakol M, Dennick R, Nasiri J. Empathy in Iranian medical students: A preliminary psychometric analysis and differences by gender and year of medical school. Med Teach. 2010;32(11):e471-8.
46. Hizomi Arani R, Naji Z, Moradi A, Shariat SV, Mirzamohamadi S, Salamati P. Comparison of empathy with patients between first-year and last-year medical students of Tehran University of Medical Sciences. BMC Med Educ. 2021;21(1):460.
47. Tentama F, Anindita WD. Employability scale: Construct validity and reliability. International Journal of Scientific & Technology Research. 2020;9(4):3166-70.
48. Hair Jr JF, Howard MC, Nitzl C. Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. J Bus Res. 2020;109:101-10.
49. Lamm KW, Lamm AJ, Edgar D. Scale. Scale Development and Validation: Methodology and Recommendations. Journal of International Agricultural and Extension Education. 2020;27(2):24-35.
50. Paro HB, Daud-Gallotti RM, Tiberio IC, Pinto RM, Martins MA. Brazilian version of the Jefferson Scale of Empathy: psychometric properties and factor analysis. BMC Med Educ 2012;12:73.
51. Davis MH. Measuring individual differences in empathy: Evidence for a multidimensional approach. J Pers Soc Psychol. 1983;44:113-26.
52. Williams B, Beovich B. A systematic review of psychometric assessment of the Jefferson Scale of Empathy using the COSMIN Risk of Bias checklist. J Eval Clin Pract. 2020;26(4):1302-15.
53. Guidi C, Traversa C. Empathy in patient care: from 'Clinical Empathy' to 'Empathic Concern'. Med Health Care Philos. 2021;24(4):573-85.
54. Larson EB, Yao X. Clinical empathy as emotional labor in the patient-physician relationship. JAMA. 2005;293(9):1100-6.
55. Sassenrath C. “Let me show you how nice I am”: Impression management as bias in empathic responses. Soc Psychol Pers Sci. 2020;11(6):752-60.
56. Wang H, Kline JA, Jackson BE, Laureano-Phillips J, Robinson RD, Cowden CD, et al. Association between emergency physician self-reported empathy and patient satisfaction. PLoS One. 2018;13(9):e0204113.

[DOI: 10.47176/mjiri.40.49]