Iran J Public Health, Vol. 54, No.6, Jun 2025, pp.1179-1192

Review Article

Psychometric Validation and Reliability of the 9-Item Shared Decision-Making Questionnaire: A Systematic Review

*Pavan Kumar Narapaka [1], Manisha Singh [2], Sarasa Meenakshi [1], Jaseena CV [1], Manasa Goudicherla [1], Krishna Murti [1], *Sameer Dhingra [1]*

1. *Department of Pharmacy Practice, National Institute of Pharmaceutical Education and Research (NIPER) Hajipur, Dist. Vaishali, Bihar, India*
2. *Department of Medical Oncology, Mahavir Cancer Sansthan and Research Centre, Patna, Bihar, India*

***Corresponding Author:** Email: sameerdhingra78@gmail.com

(Received 20 Aug 2024; accepted 19 Nov 2024)

Abstract

Background: This study aimed to provide comprehensive information on translated versions of the 9-item shared decision-making questionnaire, widely used to measure patient involvement in shared decision-making, by combining psychometric validation information.

Methods: We searched various databases such as PubMed, Scopus, Google Scholar, along with developer website to gather pertinent literature published until Feb, 2024. This psychometric validation carried out based on item characteristics, content validity, and factor analysis results of individual studies. Our evaluation was based on predetermined cut-off values for item difficulty, discrimination index, Cronbach's alpha, Kaiser Meyer Olkin factor (KMO), Bartlett's test of sphericity, and factor extraction and rotation, confirmatory factor analysis range. The European Social Research Council guidance on the conduct of narrative synthesis in systematic reviews was employed for synthesis of validation results.

Results: The final analysis included nine studies with 3090 participants from various countries, and most participants had adequate literacy, and age range was 30-60 yr. Most model versions had a good fit, and all studies reported satisfactory results, except for one study's discrimination index values. The tool's validity was satisfactory. Most of the studies reported a convenient sample was the main limitation, along with recall bias in the final responses.

Conclusion: The 9-item shared decision-making tool can be used to measure patient involvement in shared decision-making in validated language-respected countries, as proper evaluation procedures reported satisfactory results in the confirmatory analysis models and reliability testing.

Keywords: Shared decision-making; Patient involvement; Validity and reliability; Psychometric validation; 9-Item shared decision-making

Introduction

Shared decision-making (SDM) has become a crucial component of patient-centered care, enabling patients to actively participate in the decision-making process and articulate their preferences regarding treatment options (1). By providing patients with evidence-based information,

DOI: https://doi.org/10.18502/ijph.v54i6.18896

healthcare providers can empower them to comprehend the advantages and disadvantages of various treatment procedures (2, 3). The number of research studies conducted on shared decision-making and its impact on healthcare has increased considerably (4). Some nations' healthcare systems have made it mandatory to involve patients in their healthcare plan decisions, leading to better outcomes. However, the level of patients' perceived experiences varies from one nation to another, depending on their healthcare system's development and people's knowledge (5-8). Therefore, there is a need to develop tools that can measure patients' actual involvement in clinical encounters. This information can help nations plan towards improvisation and making mandatory policies.

A systematic review identified 16 questionnaires that can help measure patients' experiences of SDM (9). Among these, the nine-item shared decision-making questionnaire (SDM-Q-9) is one of the most frequently used instruments for assessing the extent to which clinicians involve patients in decision-making. The questionnaire has two versions: a patient SDM-Q-9 and a physician SDMQ-Doc, allowing for the assessment of patients' involvement in decision-making from two perspectives (10, 11). The questionnaire has been widely used in various clinical settings, including primary and specialized care, clinical trials, and national surveys (12, 13). SDMQ-9 is a useful measure in various areas of medicine, such as anesthesiology, cardiovascular diseases, dermatology, mental illnesses, oncology, otolaryngology, and traumatology (14).

There is a clear need for quality improvement in validation studies, such as sample sizes, methodological quality, and finding ways to quantify validity and compare its measurement properties across different healthcare system levels (9, 15). Therefore, this study aimed to provide comprehensive information on all translated versions of the SDM-Q-9, including item characteristics and factor analysis, to aid the research community in comprehending the translated versions of SDM-Q-9. Psychometric validation ensure that the instrument is both reliable and valid for its intended purpose, it accurately measures the constructs it is designed to assess, providing credible and reproducible results across different populations and settings (16).

Furthermore, the study will provide a detailed analysis of the questionnaire's psychometric properties, including reliability, validity, and responsiveness, to evaluate the questionnaire's suitability for use in various settings. These findings are instrumental as evidence-based practices that facilitate cross-cultural comparisons of SDM-Q-9 validation results and assess the suitability of the tool among different populations.

Methods

The study adhered to PRISMA guidelines for systematic reviews and meta-analyses to ensure transparency and accuracy, providing valuable insights for future research and clinical decisions (17).

Literature retrieval

To ensure that we found all pertinent studies for our topic, we implemented an extensive search strategy that involved using a combination of keywords and MeSH terms related to our research question. We conducted the search across multiple databases, including PubMed, Scopus, and Google Scholar. To prevent missing any pertinent articles, we used a variety of search terms such as "shared decision-making" OR "clinical decision-making" OR "medical decision-making" OR "decision-making" AND "Validity and Reliability" OR "Psychometric Validation" OR "Psychometric Evaluation" OR "Psychometric Practice" AND "SDM-Q-9" OR "9-item shared decision-making questionnaire" as these are pertinent to our research question. Furthermore, we searched the official questionnaire website (www.patient-als-partner.de) and reviewed relevant reference lists to gather as much information as possible. We limited our search to studies that published until Feb, 2024, which were in English or translated into other languages from the inception of the databases. To ensure

accuracy and reliability, two independent reviewers conducted the search.

Selection Criteria

In order to be considered for inclusion in this systematic literature review, a study must meet certain criteria. These criteria include relevance to the research question, publication in a peer-reviewed journal, use of a suitable study design (factor analysis attempted, item characteristics should be reported), appropriate methodology, translation of tools with proper procedures, and reporting of outcomes that are pertinent to the research question. Any study that fails to meet these criteria is excluded from the review.

Examination of literature

The study's inclusion and exclusion criteria were meticulously reviewed. During the screening process, the titles and abstracts were initially examined, and any discrepancies were resolved through collaboration between the reviewers. The same researchers then conducted a full-text screening of the eligible studies. Any further disagreements were settled through discussions, and if required, consultation with another reviewer was sought. Ultimately, a consensus was reached based on the established eligibility criteria, ensuring a meticulous and unbiased evaluation.

Data Extraction

The authors collaborated to develop a data collection grid that took into account all essential factors, ensuring accurate and reliable results. The literature was also reviewed for general information such as the first author, year of research, the population included, research location, the language of questionnaire translation, and sample size, along with demographic details that were extracted (gender, age, education level, marital status), and some information related to health literacy, the self-reported status of health, and employment status, living area. Additionally, validation evidence was collected, including item characteristic parameters (item difficulty, discrimination index, corrected item-total correlation, and internal consistency), factor extraction details including the exploratory factor analysis (EFA), and principal component analysis (PCA) parameters along with factors loading, and confirmatory factor analysis factors various indices information were collected, including Degree of Freedom (D.F), Chi-square value (X^2), comparative adjustment index (CFI), Roots Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square residual (SRMR), Goodness of fit index (GFI), Tucker Lewis Index (TLI). The data collection grid used was comprehensive and robust, and the results obtained can be confidently relied upon.

Considerations for evaluation of validated studies

In this systematic literature review, we have planned to evaluate the validation of each study through three stages of work. Firstly, we evaluate the item characteristics of the translated tools into the various languages. Secondly, validate the factor extraction and rotation. Finally, we will assess the goodness of fit index for the final validation.

Item characteristics measurement

The evaluation of the characteristics of translated versions involves the measurement of item characteristics, such as item difficulty, discrimination index (corrected item-total correlation), and internal consistency (Cronbach's Alpha). The discrimination index aids in assessing the contribution of each item to the scale's efficiency, with suggested values ranging from 0.30 to 0.70 (18). On the other hand, Cronbach's Alpha determines if a group of items consistently assesses the same characteristic, with a value of 0.7 to 0.9 considered good (>0.7), better (>0.8), and excellent (>0.9) (19). Item difficulty, meanwhile, pertains to the mean score of each item, considered difficult if it is less than 50% of the total score (15).

Factor extraction

Factor extraction is an important step in factor analysis, which involves principal component analysis and exploratory factor analysis. In order to accurately extract factors, the Eagan value

should be considered (ranging from 0.5 to 1.0) (20). Additionally, the Kaiser-Meyer-Olkin (KMO) test, which determines sample adequacy, should be applied, with a range of 0.8-1.0 (21). Bartlett's test of sphericity is also applied, with a *P*-value<0.005, to determine the suitability of the data for factor analysis. These methods provide a reliable approach to factor extraction and ensure accurate results (22).

Final validation

For the purpose of final validation, a confirmatory factor analysis findings considered to assess the goodness of fit index by considering various parameters, including the degree of freedom (DF), chi-square value (x2), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), goodness of fit index (GFI), comparative fit index (CFI), and Tucker Lewis index (TLI). The limited indices considered in individual studies along with the results of the final validation.

Methodological quality assessment

As the present study primarily focused on questionnaire validation, a rigorous evaluation of the methodological quality of the studies included was not feasible. Despite our best efforts, we could not locate any sources that could aid in the said assessment. As a result, any conclusions drawn from the study must be interpreted with caution, given the limitations of the available data.

Representation of outcome

Our team carried out a comprehensive literature review in accordance with the guidelines for narrative synthesis in systematic reviews established by the European Social Research Council.

Results

Results of literature search

Overall, 981 studies were initially retrieved from various databases. After eliminating duplicates (n=421), primary screening was conducted by two independent authors, followed by a full-text screening based on the established inclusion and exclusion criteria. Ultimately, 20 studies were selected for the final analysis. Out of these, nine studies were included in the systematic literature review for final analysis (Fig. 1).

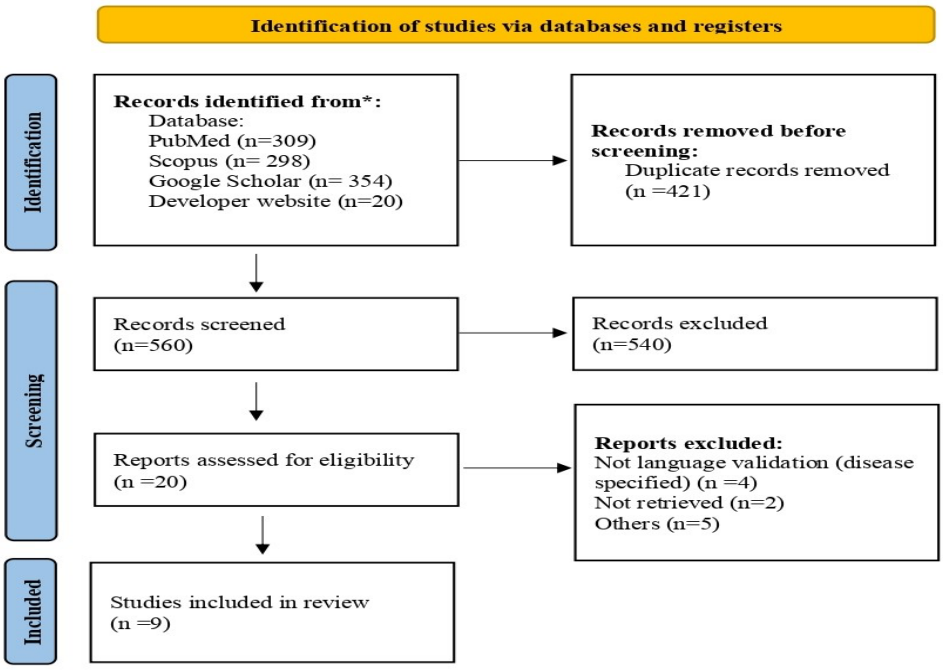

Fig. 1: PRISMA Flow Chart

Demographic information of studies included

The final manuscript included nine studies that validated the SDM-Q-9 questionnaire. These studies translated the SDM-Q-9 into different languages, such as Dutch, Spanish, Arabic, Italian, Japanese, Hungarian, Romanian, and Hindi, and were published between 2015-2024 (14, 15, 23-29). Of these nine studies, only three provided information about the included disease population, while the others did not disclose it (15, 26, 29). The total study population was 3090, with the Romanian study having the highest population (n=665) and the Spanish version study having the lowest sample (n=76) (23, 28). Most of the studies reported male participants distribution around 30%-40%, and females 60-70%. All study populations were aged between 30 and 60 years old, except for one study which included 58.7% of participants aged between 50 and 70 (25). Six out of eight studies reported the education levels of the study population, with most having adequate literacy ranging from high school to university degrees. Five studies revealed the marital status of the population, with most being married, except for the Italian population study, which involved 50.9% of singles (n=147) (26). Other relevant demographic details were also recorded, such as employment status, health literacy, type of disease, self-reported health status, and chronic comorbidities status, all listed in Table 1.

Table 1: Demographic details of studies analyzing the 9-Item shared decision-making questionnaire

First author	Year of Research	Sample Size	Gender (Male, %)	Education level (Highest % participants with qualification)	Marital status (Married, %)	Other demographic information
Rodenburg-Vandenbussche, et al. [15]	2015	180	34%	Not available	Not available	Type 2 diabetes 41%, Psychiatric 20%, Ophthalmic 20%, Gynaecologic 13%, Breast Cancer 7%
Alvarado et al. [23]	2019	76	42.1%	42.1% had secondary education	Not available	Not available
De las Cuevas et al. [24]	2015	540	84.4%	40.9% had completed primary education, 33.7% had completed secondary education	Not available	Not available
Alzubaidi et al [25].	2019	516	43.6%	Most of the participants (45%) with university degree	68.2%	Full-time employment 41.9%, Inadequate health literacy 60.7%.
de Filippis et al. [26]	2022	289	41.2%	Not available	43.9%.	Schizophrenia Spectrum Disorder 20.8%, Major Depressive Disorder 21.5%, anxiety disorder 14.9%, and bipolar disorder 11.1%.
Goto et al. [27]	2020	131	44.3%	>70% of the patients had completed either high school or university-level education	60.3%	49.6% were employed among the included study population.
Rencz et al. [14]	2019	537	46%	Most of the study population were studied above high school	Not available	62.4% of population have chronic morbidities, and 82.3% of the population have fair to good self-reported health.
Baicus et al. [28]	2019	665	34.2%	Most of the study population were studied above high school	60.2%	67% included the Urban area, 64.3% autoimmune diseases followed by atrial fibrillation.
PK Narapaka et al. [29]	2024	156	64%	NA	96.7%	13% of the population have family history, 25.3% have comorbidities, and highest percentage of participants involved in between age 56-65

Results of Item Characteristics

Several studies have analyzed the item characteristics of SDM-Q-9. However, only three characteristics were defined in all manuscripts, which are Item Difficulty, Discrimination Index (Corrected Item-Total Correlation), and Internal Consistency (Cronbach's Alpha). Unfortunately, the ceiling effect and floor effect data were not included in the studies due to unavailability of data.

Out of all the studies, only four calculated the Item Difficulty (14, 15, 28, 29). The Dutch studies showed no item difficulty with a mean score range above 2.5 or 50%. The Dutch study reported the highest mean score of 4.3 for Item 9, while the lowest mean score for Items 2, 3, and 7 was 3.5 (15). The Hungarian version reported a high score for Item 9, and the lowest score for Item 2 was 2.9 (14). Lastly, the Romanian version also reported the highest score of 3.8 for Item 1 and the lowest score of 2.7 for Item 6, and the Hindi version of SDM-Q-9 also reported 3.35 for item-1, and 4.12 for item-4 (28, 29). De Filippis et al did not report the Discrimination Index. The discrimination index values mentioned in Table 2 ranged from 0.36-0.84 among all the studies, except for Item 1 in De las Cuevas et al (24).

All studies reported Cronbach alpha between the range of 0.80-0.92, except for the Arabic version, which reported values greater than 0.90 (0.915 and 0.929) (25). The studies represented the Cronbach alpha values in the mean of all items and individual items as well. A single study mentioned Cronbach alpha in the form of mean scores, while the rest of the studies reported it individually for each item. In summary, all versions of SDM-Q-9 showed good internal consistency, with good corrected item-total correlation, and no item difficulty Table 2.

Table 2: Psychometric validation of studies analyzing the 9-Item shared decision-making questionnaire

Variables					Results of fitness of the tool (CFA)			
Study reference	Internal consistency (Cronbach's Alpha)	Method for factor extraction	Factor extracted	Kaiser–meyer–olkin (KMO)	Tested models	Considered Cut-off Values for CFA parameters (>5)	Results of CFA	Conclusion and limitations
(15)	0.85-0.88	PCA with Eigen value >1	PCA resulted with 2 components component 1: 1-3 component 2 :4-9	> 0.85, except Item 1-(0.67)	**Model-1** (the one-factor model) **Model-2** (one-factor model excluding Item 1) **Model-3** (a one-factor model excluding Item 9) **Model-4** (a one-factor model excluding both Items 1 and 9)	The CFA models were regarded as acceptable to good when the fit indices met the following cut-off criteria: RMSEA <***0.06***; CFI >***0.95***and SRMR <***0.08***	For the patients, none of the fit indices cut-off criteria were met by Model 1, while Model 2 produced acceptable fit indices by meeting the cut-off criteria for two of them (RMSEA and SRMR) and improving the others ($\chi 2$ and CFI). Model 4 provided the best fit indices.	This study demonstrated the good acceptability and reliability of the Dutch version of SDM, used in primary and specialized care. The authors noted that there was no comparable validation study in the same population to confirm the correlation results. Additionally, they did not consider a large enough sample from primary and secondary regions for separate validation, and the use of a convenient sample was

Table 2: Continued …

								another limitation.
(23)	0.80-0.84	PCA with Varimax rotation	Resulted in 2 components. Items 3–6 and 9 in the first component; on the other hand, items 1, 2, 7 and 8 in the second component.	0.833	**Model-1** (excluding the first item) **Model-2** (excluding the nine item) **Model-3** (in the first factor, items 3–6, 9 and in the second factor, items 1, 2, 7, 8) **Model-4** (corresponds to the first factor with items 3–6 and the second factor with Items 1, 2, 7–9) **Model-5** (consists of the first factor's items 3-6 and the second factor's items 2, 7, 8, and 9, except for item 1**)**	The standards used to gauge how well the CFA-derived models' match: CFI >0.9 and RMSEA < 0.08	Model 2 resulted with best indices in CFI, RMSEA, AIC, BIC, and CAIC.	This study found good validity. However, the major limitation was the sample size which may not represent the population. Another limitation was that this tool was validated in the diabetes population only, making it unsuitable for a variety of populations in the nation.
(24)	0.885	PCA with oblimin rotation	Resulted in one component	0.821,	**Model-1** (including all the nine items) **Model-2** (the structure obtained with PCA in the current sample, composed by two factors) **Model-3** (a mono factorial structure but excluding item 1)	The following criteria were used to indicate the fit of the CFA models to the data. CFI and GFI > 0.90, and RMSEA and SRMR < 0.08.	The best solution was achieved with the one-factor model, excluding item 1 (Model 3), which yielded the best fit indices (CFI and GFI).	This study has shown adequate reliability and validity, using a convenient sample.
(25)	0.915 and 0.929	PCA, EFA with obliman rotation along with Eigen value >1	Resulted with one component	0.907	**Model-1** (Including all the items) **Model-2** (excludes one item 1)	GFI and CFI values above ***0.90*** and RMR and RMSEA values below ***0.05*** indicated that the CFA model was a good fit	Both models were adequate since they produced similar indices, although model 2 demonstrated slightly better indices.	The validity and reliability parameters displayed were excellent, suggesting suitability for use in the 22-member states of the Arabic League. However, there were some limitations, such as the use of a convenient sample, which may have led to biased respons-

Table 2: Continued …

								es. Additionally, conducting EFA and CFA with the same sample may have also impacted the findings.
(26)	0.86	Not available	Not available	Not available	NA	For TLI and CFI, values of 0.90 and above are considered adequate, while values of .95 or above are considered excellent; for RMSEA values of .08 and below are considered adequate and values of .05 or below are considered excellent; for SRMR a cut-off values close to .08 are considered adequate. Values of $\chi2$ /df<3.0 are considered good, while values <2.0 are excellent.	The CFA exhibited an outstanding fit with a relative chi-square ($\chi2$ /df) of 1.69, CFI of .98, TLI of .97, RMSEA of .05, and SRMR of .08. These results indicate that the Italian version of SDM-Q-9 model is appropriate.	The Italian version of SDM-Q-9 yielded better indices compared to other language-validated studies. However, there was a recall bias, and the protocol did not include a test-retest reliability. Additionally, the sample had a broad clinical variability in terms of disease characteristics, making it more heterogeneous.
(27)	0.89-0.91	PCA	Resulted in one component with all items > 0.4	Not available	NA	NA		A chi-squared value (discrepancy chi-square value divided by the degrees of freedom) of 11.84 (p = 0.619), GFI of 0.981, AGFI of 0.938, RMSEA of 0.0, and CFI of 1.0, suggesting good fit to the data.	This study confirmed that the tools can be used in the Japanese population. However, they mentioned that the study was conducted in a single area, so the results may not be applicable to a larger population. The sample size considered for CFA was also relatively small.
(14)	0.90-0.93	EFA and CFA with >1 Eigen value	one factor	0.910	**Model-1** (One-factor model including all items)	Df=27, X^2=387.39, CFI=0.89, RMSEA=0.15, SRMR=0.05, GFI=NA, TLI=NA	The desired threshold values were***>0.90*** for CFI and***≤0.8*** for both RMSEA and SRMR	The overall performance of the four models was quite similar. Almost all the models fulfilled the cut-off criteria of $\chi2$, CFI, and SRMR; however, none of them attained an acceptable RMSEA value. Model 1, which included all nine items, exhibited a relatively satisfactory	Overall results indicate that the tools are suitable for measuring SDM among primary and secondary care. However, the shorter duration from the decision taken time and recalling may cause bias in responses
					Model-2 (One-factor model excluding item 1)	Df=20, X^2=260.28, CFI=0.92, RMSEA=0.15, SRMR=0.03, GFI=NA, TLI=NA			
					Model-3 (One-factor model excluding item 9)	Df=20, X^2=324.0, CFI=0.90, RMSEA=0.16, SRMR=0.05, GFI=NA, TLI=NA			
					Model-4 (One-factor	Df=14, X^2=195.81,			

Table 2: Continued …

					model excluding items 1 and 9)	CFI=0.93, RMSEA=0.15, SRMR=0.03, GFI=NA, TLI=NA	performance with CFI=0.899, RMSEA =0.158, and SRMR = 0.052.	
(28)	0.95-0.96	PCA with Varimax	One component	0.94	NA	NA	NA	This study reported good validity among the Romanian population. Furthermore, it can be used for reporting the SDM, although the sample may not be representative of the entire Romanian population. The hospital survey was non-anonymous, so there were chances for social desirability bias.
(29)	0.80-0.83	EFA with varimax	One component	0.825	One factorial model includes all One factorial model without Item-1 One factorial model without Item-9 One factorial model without Item-1 and 9 One factorial model without items below 0.5 on factor loading	Chisquare statistic ($\chi 2$), comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The desired threshold values were greater than 0.90 for CFI and less than or equal to 0.8 for both RMSEA and SRMR, Goodness of fit index (GFI), Adjusted goodness of fit index (AGFI), and Tucker Lewis index (TLI)	All models almost reached the desired value cutoff, but the model- 5 showed all desired indices cut off values	This study reported accepted model fit indices values for the translated version that indicates the tool suitability for measurement of patients involvement in shared decision-making among Indian oncology patients, the major limitation of the study is same sample was used for the EFA and CFA and its not a representative sample of the population.

NA- Not available, D.F-Degree of Freedom, X2-Chi-square value, RMSEA- Roots Mean Square Error of Approximation, SRMR- Standardized Root Mean Square residual, GFI-Goodness of fit index, CFI- comparative adjustment index, TLI-Tucker Lewis Index

Results of factor extraction

The results obtained from factor extraction are significant in determining the number of constructs to include in a questionnaire. In most studies, Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA) were the preferred methods, with an Eigen value of 1.0 serving as the cut-off point. While one study deviated from this norm by not conducting factor extraction (24), most of the studies used either orthogonal (Varimax) or oblique (Obliman) rotation to determine factor rotation. Among the four studies that provided details about factor rotation, three employed orthogonal rotation, while the remaining two used oblique rotation (24, 25).

Single-factor extraction was the outcome of six studies, while two studies produced two-factor

extractions. A study produced two-factor extraction (Factor 1: Item 1-3, Factor 2: Item 4-9) (15) and Alvarado et al produced two components. Items 3–6 and 9 had readings above 0.5 in the first component, while items 1, 2, 7, and 8 had their highest readings in the second component (23).

Overall, all the studies adhered to standard procedures, either PCA or EFA, with an Eigen value of 0.5-1. Rodenburg-Vandenbussche et al and Alvarado et al which produced two components. However, the rest of the studies produced a single factorial model similar to the developer model. Additionally, all studies had a KMO value above 0.80, which indicates an adequate sample in the validation study, and that the data was suitable for factor analysis Table 2.

Results of Confirmatory factor analysis

This study analyzed several Confirmatory Factor Analysis (CFA) models used in various studies, except for the Romain version (27). Four parameters were commonly used to measure the scale fit: X^2, CFI, RMSEA, and SRMR. However, some studies also added GFI and TLI to this list. The Arabic version of the validation evaluated two models, and both models reported good fit indices (25). Another version created four models to evaluate the Dutch version of SDM-Q-9 for CFA. One of the models yielded acceptable fit indices for RMSEA and SRMR (15). Rencz et al created four models, and almost all of them met the cut-off values of X^2, CFI, and SRMR, but none of them had a good RMSEA value (14). De Filippis et al tested a single model, including TLI along with SRMR, CFI, and RMSEA, and all the values showed a good fit for the Italian version (26). The Japanese study also showed a good fit for the Japanese version (27). Alvarado et al prepared five models, and the best adjustment indices were found with Model 2 (23). De Las Cuevas conducted CFA with three models, and they reported that Model-3 showed a better fit (24). PK Narapaka et al conducted CFA with five various models, all of them shown an acceptable level of fit indices on the final analysis (29).

To conclude, Alzubaidi et al. (model-2), De las CC et al. (model-3) (24, 25) indicated that these versions produced better indices on CFA without Item-1. Rodenburg-Vandenbussche et al. produced acceptable results. There was a problem in the fit of Item-1 in the respected versions. The Spanish version reported the best indices without Item-9, the Dutch version resulted in a problem for fit with Item-1, 9 (15). The Hungarian version created four different models, but Model-1 with the inclusion of all items reported good results on CFA (26). The Italian and Japanese versions validated only one model and showed good fit results (24, 25).

Discussion

The SDM-Q-9 was extensively utilized in healthcare studies over the past decade and translated into various languages, with validation conducted for use in different healthcare systems. This study provides a comprehensive overview of all the validated versions of the SDM-Q-9, facilitating researchers' utilization of this tool in other languages while validating it.

All studies methodology involved consolidating the results with item characteristics, factor analysis (including factor extraction, rotation, and good of fit of tool). All studies mentioned about three item characteristic parameters that including item difficulty, discrimination index (corrected item-total correlation), and internal consistency (Cronbach's Alpha). Item difficulty in the validation of a questionnaire refers to the measure of how challenging a particular item (or question) is for the respondents to answer correctly or as intended (30), discrimination index is a psychometric statistic used to assess how well an individual questionnaire item differentiates between respondents who perform well overall and those who perform poorly on the overall test or construct being measured (31), and cronbach's alpha is a measure of internal consistency, which indicates how well a set of items in a questionnaire or scale measures a single unidimensional construct (32). While three studies reported acceptable lev-

els of item difficulty, one reported a score below the expected range for item 1. The internal consistency of all studies was deemed well, with nearly all studies reporting a Cronbach alpha value close to 0.9. For factor extraction and rotation, all studies used PCA, EFA, and varimax and oblimin rotation, with a cutoff for factor extraction of 1. Most studies resulted in a single component or factor extraction, but two studies resulted in more than one factor.

All studies had KMO values above the cutoff, indicating adequate sample adequacy. Sample adequacy is an essential validation criterion as it necessitates that a defined sample be included in the final analysis. However, this systematic review did not observe significant differences in study results with lower or higher sample sizes, and Bartlett's test significance was also below 0.05. For confirmatory analysis, different models were created, with most studies including two models that excluded items 1 and 9. Almost all studies showed a good fit for CFA, but the factors varied across the studies. Most models in CFA showed a good fit after excluding items 1 and 9, indicating that these items were problematic for measuring patient involvement. A study conducted among the Spanish population using this validated questionnaire reported good SDM experience results (28).

Higher education levels among the study populations may act as an influencing factor for good response in the clinical setting. No information was mentioned in the manuscripts regarding the reason for choosing factor extraction methods between PCA and EFA, factor rotation (varimax and oblimin). After validation of the translated questionnaire, limited studies were conducted using the translated questionnaire in each translated language, and most of the included studies reported that there may be a chance of bias in the questionnaire's output as it measures outcomes based on patient recall of the last clinical visit. Therefore, conducting studies among diverse populations for validation will provide a better outcome on the validation.

All of the studies provided good validity and accepted reliability, as well as good model fit through confirmatory factor analysis (CFA). However, individual studies mentioned limitations related to the study. One of the most commonly mentioned limitations was recall bias, which occurred due to the questionnaire being designed as a closed-ended tool to report perceived experiences by recalling past experiences. To avoid this limitation, validation study responses can be collected for a fixed duration of time after the clinical encounter. A short duration of data collection may also cause bias. Another limitation was the sample size. Most of the studies used a convenient sample for validation and mentioned that the results may not be representative of the entire country's population. There were different criteria for the consideration of sample size. One criterion is that researchers should consider a 1:10 ratio of responses to each item for validation. According to epidemiological studies, various factors related to the disease population need to be considered for sample calculation, and then the required number of responses should be collected. One study mentioned that test-retest reliability considerations may also improve the results. The selected disease population is also a limitation, whereby the tool cannot be applied to other disease populations. Thus, considering these limitations in the next validation studies will improve the accuracy of the psychometric evaluation.

A literature review study was conducted to compare the SDM-Q-9 with CollaboRATE, including five studies with 442 patients. Among this population, the CollaboRATE scale reported a median score of 88.9%, IQR 81.5%–100%, with a 12.5-point higher score when compared to SDM-Q-9. The questionnaire also reported 37.5% of high scores, revealing a substantial ceiling effect. Both tools were subjective-based analysis tools, and there may be a chance that they are influenced by other factors and the risk of missing values (14). SDM-Q-9 has better internal consistency when compared with the OPTION scale and found a weak correlation between the total scores of each scale (33). Even though a review of the literature mentioned that SDM-Q-9 and its associated tool

SDM-DOC were used in a wide range of studies (34).

Limitations and future recommendations

This study was the first to compile validation-related information for the SDM-Q-9 questionnaire, but there may be some bias in the results because we excluded studies that validated questionnaires for various medical conditions. To gain a better understanding of this tool, a study with more flexible inclusion criteria covering all studies could be conducted. In the future, conducting a systematic review with updated or multilingual studies would provide more reliable results for authors and researchers on the SDM-Q-9. We also observed variations in the validation procedures adopted across studies. Following a standard protocol by contacting the developer team will provide more accurate results in the future. Additionally, limited studies have been conducted in countries that have validated the SDM-Q-9 questionnaire. Therefore, conducting clinical-based studies using the validated questionnaire in various diseased populations will help to better understand the tool's characteristics.

Conclusion

All versions of the included studies demonstrated good validation results. However, the models with items 1 and 2 had a problem with CFA fit results. All item characteristic parameters, factor extraction, and rotation results showed positive outcomes. Furthermore, internal consistency results were better among all the validated versions.

Journalism Ethics considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Conflict of interest

The authors declare that there is no conflict of interests.

References

1. Jacobs S, Davies N, Butterick KL, et al (2023). Shared decision-making for children with medical complexity in community health services: a scoping review. *BMJ Paediatr Open,* 7 (1): e001866.
2. Egan C, Naughton C, Caples M, Mulcahy H (2023). Shared decision-making with adults transitioning to long-term care: A scoping review. *Int J Older People Nurs,* 18(1):e12518.
3. Hoffmann TC, Legare F, Simmons MB, et al (2014). Shared decision making: what do clinicians need to know and why should they bother? *Med J Aust,* 201 (1):35-39.
4. Lu C, Li X, Yang K (2019). Trends in shared decision-making studies from 2009 to 2018: a bibliometric analysis. *Front Public Health,* 7:384.
5. Alrawiai S, Aljaffary A, Al-Rayes S, et al (2020). The OPTION scale: measuring patients' perceptions of shared decision-making in the Kingdom of Saudi Arabia. *J Multidiscip Healthc*, 13:1337-1346.
6. Ismail MA, Midin M (2021). Shared decision-making and role preference among patients with schizophrenia in malaysia: a cross-sectional study. *Front Psychiatry,* 12:680800.
7. Driever EM, Stiggelbout AM, Brand PL (2022). Patients' preferred and perceived decision-making roles, and observed patient involvement in videotaped encounters with medical specialists. *Patient Educ Couns,* 105 (8):2702-2707.
8. Schoenfeld EM, Goff SL, Downs G, et al (2018). A Qualitative analysis of patients' perceptions of shared decision making in the emergency department:"let me know i have a choice". *Acad Emerg Med,* 25 (7):716-727.
9. Gärtner FR, Bomhof-Roordink H, Smith IP, et al (2018). The quality of instruments to assess the process of shared decision making: a systematic review. *PLoS One,* 13 (2):e0191747.

10. Kriston L, Scholl I, Hölzel L, et al (2010). The 9-item Shared Decision Making Questionnaire (SDM-Q-9). Development and psychometric properties in a primary care sample. *Patient Educ Couns,* 80 (1):94-99.
11. Scholl I, Kriston L, Dirmaier J, et al (2012). Development and psychometric properties of the Shared Decision Making Questionnaire–physician version (SDM-Q-Doc). *Patient Educ Couns,* 88 (2):284-290.
12. Doherr H, Christalle E, Kriston L, et al (2017). Use of the 9-item Shared Decision Making Questionnaire (SDM-Q-9 and SDM-Q-Doc) in intervention studies—a systematic review. *PLoS One,* 12 (3):e0173904.
13. Glass KE, Wills CE, Holloman C, et al (2012). Shared decision making and other variables as correlates of satisfaction with health care decisions in a United States national survey. *Patient Educ Couns,* 88 (1):100-105.
14. Rencz F, Tamási B, Brodszky V, et al (2019). Validity and reliability of the 9-item Shared Decision Making Questionnaire (SDM-Q-9) in a national survey in Hungary. *Eur J Health Econ,* 20(Suppl 1):43-55.
15. Rodenburg-Vandenbussche S, Pieterse AH, Kroonenberg PM, et al (2015). Dutch translation and psychometric testing of the 9-item shared decision making questionnaire (SDM-Q-9) and shared decision making questionnaire-physician version (SDM-Q-doc) in primary and secondary care. *PLoS One,* 10 (7):e0132158.
16. Souza ACd, Alexandre NMC, Guirardello EdB (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol Serv Saude,* 26:649-659.
17. Page MJ, McKenzie JE, Bossuyt PM, et al (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71.
18. Ferketich S (1990). Internal consistency estimates of reliability. *Res Nurs Health,* 13 (6):437-440.
19. Taber KS (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education,* 48:1273-1296.
20. Karami H (2015). Exploratory factor analysis as a construct validation tool:(Mis) applications in applied linguistics research. *TESOL Journal,* 6 (3):476-498.
21. Kaiser HF (1970). A second generation little jiffy. *Psychometrika,* 35:401–415.
22. Hair J, Black W, Babin B, Anderson R (2010). Multivariate data analysis, Prentice Hall, New Jersey. *Technometrics,* 15 (3).
23. Alvarado-Villa GE, Moncayo-Rizzo JD, Gallardo-Rumbea JA (2019). Spanish validation endorsement of SDM-Q-9, a new approach. *BMC Public Health,* 19:106.
24. De las Cuevas C, Perestelo-Perez L, Rivero-Santana A, et al (2015). Validation of the S panish version of the 9-item S hared D ecision-M aking Q uestionnaire. *Health Expect,* 18 (6):2143-2153.
25. Alzubaidi H, Hussein A, Mc Namara K, Scholl I (2019). Psychometric properties of the Arabic version of the 9-item Shared Decision-Making Questionnaire: the entire process from translation to validation. *BMJ Open,* 9 (4):e026672.
26. de Filippis R, Aloi M, Pilieci AM, et al (2022). Psychometric Properties of the 9-Item Shared Decision-Making Questionnaire (SDM-Q-9): validation of the Italian version in a large psychiatric clinical sample. *Clin Neuropsychiatry,* 19 (4):264-271.
27. Goto Y, Miura H, Son D, et al (2020). Psychometric evaluation of the Japanese 9-item shared decision-making questionnaire and its association with decision conflict and patient factors in Japanese primary care. *JMA J,* 3 (3):208-215.
28. Baicus C, Balanescu P, Gurghean A, et al (2019). Romanian version of SDM-Q-9 validation in Internal Medicine and Cardiology setting: a multicentric cross-sectional study. *Rom J Intern Med,* 57 (2):195-200.
29. Narapaka PK, Singh M, Murti K, Dhingra S (2024). Validity and reliability of the 9-item shared decision-making questionnaire (SDM-Q-9) among Indian oncology patients in a tertiary care hospital. *Clin Epidemiol Glob Health,* 27:101626.
30. DeVellis RF, Thorpe CT (2021). *Scale development: Theory and applications*. ed. Sage publications.
31. Cohen RJ, Swerdlik ME, Phillips SM (1996). *Psychological testing and assessment: An introduction to tests and measurement*. ed. Mayfield Publishing Co.

32. Cronbach LJ (1951). Coefficient alpha and the internal structure of tests. *Psychometrika,* 16 (3):297-334.
33. Ubbink DT, van Asbeck EV, Aarts JW, et al (2022). Comparison of the CollaboRATE and SDM-Q-9 questionnaires to appreciate the patient-reported level of shared decision-making. *Patient Educ Couns,* 105 (7):2475-2479.
34. Scholl I, Kriston L, Dirmaier J, Haerter M (2015). Comparing the nine-item Shared Decision-Making Questionnaire to the OPTION Scale–an attempt to establish convergent validity. *Health Expect,* 18 (1):137-150.