Received: 24 March 2019 | Revised: 1 July 2019 | Accepted: 6 July 2019

DOI: 10.1111/jep.13293

COMPREHENSIVE REVIEW

WILEY Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

A systematic review of psychometric assessment of the Jefferson Scale of Empathy using the COSMIN Risk of Bias checklist

Brett Williams PhD | **Bronwyn Beovich BN**

Department of Community Emergency Health and Paramedic Practice, Monash University, Melbourne, Victoria, Australia

Correspondence
Brett Williams, Department of Community Emergency Health and Paramedic Practice, Monash University—Peninsula Campus, PO Box 527, McMahons Road, Frankston, Victoria 3199, Australia.
Email: brett.williams@monash.edu

Abstract

Introduction: Empathic communication in health care may enhance positive patient and health care professional relationships, patient satisfaction and can buffer professional burnout. The Jefferson Scale of Empathy (JSE) was developed based on the need to quantitatively measure levels of empathy, particularly in health care settings. Evaluating the utility of empathy is underpinned by the psychometric rigour of the instruments used to measure it. The aim of this study was to critically evaluate the current evidence on the measurement properties of the JSE.

Methods: Two reviewers independently searched six databases for papers describing psychometric assessment of the JSE from January 2000 to July 2018 inclusive. The studies were independently assessed for methodological quality using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist.

Results: The search strategy resulted in the retrieval of 985 papers, of which 59 were included in this study. The majority of papers reported on measures of structural validity and internal consistency, and it was in these areas which the highest quality of reporting was demonstrated. Additionally, there was a generally very good quality in reporting of convergent validity. Reliability, measurement error, cross-cultural validity were reported with less than optimum quality.

Conclusion: The JSE demonstrates robust structural validity, internal consistency, and convergent validity. These measurement properties are generally well reported in the literature in studies of good methodological quality, and thus may be interpreted with relative confidence when used in empathy research. However, current evidence is limited for the properties of reliability, measurement error, and cross-cultural validity. Thus, a degree of caution should be considered in drawing conclusions when using the JSE with regard to these properties. It is recommended that future examinations of the JSE refer to the COSMIN guidelines to ensure complete and well-reported psychometric data are included.

List of abbreviations: COSMIN, COnsensus-based Standards for the selection of health Measurement Instruments; IRI, Interpersonal Reactivity Index; JSE:, Jefferson Scale of Empathy; JSPE:, Jefferson Scale of Physician Empathy; NEO-PI-R:, NEO Personality Inventory-Revised

Trial registration: This study has been registered with PROSPERO: CRD42018111412

KEYWORDS

empathy, psychometrics, systematic review

INTRODUCTION

Empathy is considered important for health care practitioners, and has been empirically shown to improve clinical outcomes [1], and improve communication between caregivers and patients [2]. Therefore, ensuring the tools which measure this characteristic are psychometrically sound, is important. Evaluating the utility of empathy in health care settings is underpinned by the psychometric rigour of the instruments used to measure it.

Empathy is a multidimensional construct and includes both affective and cognitive components. The affective component of empathy involves experiencing the feelings of others [3]. On the other hand, cognitive component is the ability to understand the experiences and feelings of others, but also the capacity to communicate this understanding back to them [3]. Although empathy includes these two components, it is viewed predominately as cognitive attribute in health care settings [4].

Empathy in health care settings is important because it enhances positive patient outcomes and satisfaction [1,5,6]. It has been reported that empathetic communication facilitates positive interaction and relationship between patients and the health care professionals [6,7]. Furthermore, this relationship has a positive effect on physicians and can buffer against professional burnout [8,9].

Hojat and his colleagues developed the Jefferson Scale of Physician Empathy (JSPE) [3] to measure empathy in health care settings. The original scale title was later changed to Jefferson Scale of Empathy (JSE) for use in wider populations of health profession students and practitioners [10]. For the purposes of this study, "JSE" will be used to encompass all subsequent versions of the scale used for various populations.

While there are a number of instruments that measure empathy; arguably the JSE is one of most common measurement tools used within the context of health care. The instrument has been translated into 55 different languages and used worldwide in countries such as Australia, China, America, Brazil, and the Czech Republic, demonstrating its cross-cultural application [11]. While the JSE has been extensively utilized for the study of empathy in health care settings, to the best of our knowledge no systematic review has been undertaken examining the methodological quality of studies that have assessed the JSE psychometric properties.

Development of a scale to measure empathy of health care professionals

Hojat et al. [3] stressed the need for a psychometrically sound instrument to measure empathy among health care professionals and medical students and developed a scale to measure physician empathy. Three groups participated in the development study. Group 1 consisted of 55 physicians, group 2 composed of 41 internal medicine residents, and group 3 comprised of 193 third year medical students. The first version of the JSPE was developed based on an extensive literature review. Subscales from other empathy instruments were also used to test its validity. These included *empathetic concern*, *perspective taking*, and *fantasy scale*, adapted from the Interpersonal Reactivity Index (IRI) [12], *warmth* and *dutifulness*, adapted from the NEO Personality Inventory-Revised (NEO-PI-R) [13], and the Faith in People Scale [14,15].

The IRI was designed to assess empathic tendencies and consists of four separate subscales: perspective taking, fantasy, empathic concern, and personal distress [12]. The NEO PI-R assessed adult personality based on the five-factor model of personality: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness [13]. The Faith in People Scale was designed to assess attitudes to human nature in general [15].

The first version of the 90-item JSPE was tested using the Delphi method to obtain the content of the instrument, its face validity, and its clarity. Fifty-five physicians (group 1) reviewed the first version and crossed out items that they deemed irrelevant in measuring empathy. They edited the items to increase its clarity and also added items that they believed should be included in the measure. Further validity of the scales were obtained from 193 third year medical students (group 3). They included *personal attribute of empathy*, *global sympathy*, *personal attribute of compassion*, *trust*, *tolerance*, *personal growth* (through patient interaction), *communication* (of the understanding), *self-protection*, *humour*, and *clinical neutrality* [3].

The modified JSPE consisted of 45 items, and together with the IRI scales, were completed by 41 resident physicians (group 2). Additionally, the modified 45-items JSPE together with other instruments (IRI scales, personality facets of the NEO PI-R, Faith in People Scale, and personal attributes) were completed by 193 medical students (group 3) for psychometric properties.

To investigate the underlying structure of the JSPE, data from the medical students were subjected to principal component analysis with orthogonal varimax rotation. Four components emerged from this and were retained based on the Kaiser's criteria where a component with an eigenvalue above 1 is retained [16]. The first factor that emerged from this study was labelled as "Physicians view from the patient's perspective"; the second factor was labelled "Understanding patients experiences, feelings, and clues"; factor three which was reverse scored was labelled "Ignoring emotions in patients"; and finally factor four was labelled as "Thinking like the patient". Factors three and four were deemed to be less stable as each factor had less than three items. Among the 20 items that were retained, 17 of these had positive factor structure coefficients with scores ranging from 1 *(strongly disagree)* to 7 *(strongly agree)*. The remaining three items with negative factor structure coefficients were reverse scored on a scale ranging

from 7 *(strongly disagree)* to 1 *(strongly agree)*. Scores can range from 20 to 140 with higher score indicating greater empathy.

The criterion-related validity of JSPE was examined against the IRI. It was found that JSPE correlated with the IRI [12,17] with moderate correlations found between perspective taking and empathetic concern (r = 0.40). Further, empathetic concern was found with compassionate care (r = 0.41) [17]. However, lower correlations were observed between the fantasy and personal distress subscales of the IRI, as these subscales were suggestive of measuring sympathy as opposed to empathy [17].

Kaiser's eigenvalue criterion has been criticized for being prone to over extraction of factors [18], while the practice of determining the names of factors based on item content rather than theory has been questioned [19]. When Hojat et al.'s (2001) [3] two-factor theoretical model of empathy is considered (understanding the experiences of others, and communicating this understanding), it could be argued that there is a lack of congruence between the JSPE's measurement of empathy and the theoretical conceptualisation of empathy.

Subsequent research has proposed various factor structures of the JSE. Many studies have identified a three-factor structure comprising "perspective taking," "compassionate care", and "standing in patient's shoes" [20-22]. However, other authors have reported a four-factor model. For example, a study on Austrian medical students identified three factors which showed similarities to the aforementioned factors, with an additional factor of "negative/no influence of moderating factors and (empathetic) techniques/skills on process/outcome" [23]. A four-factor model was also proposed in a study with dental students in India, encompassing "understanding patient's feelings", "sense of confusion," as well as two factors related to "ignoring the emotional component" [24]. A two-factor structure has also been proposed, consisting of "perspective-taking" and "compassionate care" [25].

Given this lack of congruence, a systematic investigation into the quality of assessment of psychometric validity of the JSE is warranted. [26]

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology was developed to guide the performance of systematic reviews of patient-reported outcome measures (PROMs). A PROM is a measure of "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [27] and they have the potential to improve health care outcomes and increase patient-centred care [28-30]. A checklist was developed by the COSMIN group to enable the evaluation of the methodological quality of studies on measurement properties these self-reported outcome measurement tools [31], where the standards used to assess each of 10 measurement properties are outlined. It should be noted that the COSMIN methodology uses the word "patient" to identify the relevant population, however this may be modified to suit the population under examination [32]. In the present study the terms "health care student" or "health care professional" are the more appropriate terms for outcome measurements examined.

Generally, the COSMIN methodology is employed to make a systematic evaluation of the most suitable tool to measure a certain construct. For example, heart rate recovery in athletes [33], physical function measures used in intensive care units [34], or health-related quality of life in cervical cancer patients[35]. However, it can also be used to review the measurement properties of a single outcome measurement instrument [32,36,37]. In the present review, studies reporting on the measurement properties of the JSE were examined using the guidelines provided by the COSMIN group and their Risk of Bias checklist [31,32].

A review of the literature indicated the majority of published papers using the COSMIN methodology utilized an earlier version of the instrument [38,39]. Since this time the methodology has been updated considerably [31,32]. For detailed explanation of the changes made to the COSMIN instrument, the reader is encouraged to consult Mokkink et al. [32]. To note, the major modifications to the updated COSMIN instrument are:

- Inclusion of poorer quality studies in the summary of published studies.
- Removal of criteria regarding a reasonable gold standard for criterion validity and responsiveness. An exception can be made when a shortened instrument is compared to the original long version; the latter being deemed as the gold standard.
- Removal of standards on formulating hypotheses for hypotheses testing for construct validity and responsiveness
- Removal of standards on adequate sample size for single studies from categories where it is possible to pool the results. Sample sizes can then be considered when formulating conclusions.
- Indicators of the quality of language translation processes for outcome tools are no longer assessed.
- The ratings for each psychometric characteristic have been changed from "excellent", "good", "fair", and "poor", to the terms "very good", "adequate", "doubtful", and "inadequate", with an additional "not applicable" option.
- Removal of standards on missing data and handling missing data as the lack of reporting on number of missing items and on how missing items are handled does not automatically result in biased results of the study.

The present study utilized the updated version of the COSMIN Risk of Bias checklist [31].

Aim

The aim of this study was to conduct a systematic review of studies that have assessed the psychometric properties of the JSE. This type of systematic review has been previously described [40], and primarily involves completing a systematic literature search, assessing the methodological quality of included studies, evaluating the evidence for each

measurement property, and drawing conclusions from best evidence synthesis [40]. The COSMIN Risk of Bias checklist [31] will be utilized for these evaluations. The completion of these processes will enable recommendations to be made regarding the level of confidence with which researchers may use the JSE, and provide benchmarking for use of the scale in future research.

Protocol registration

The study protocol has been registered with PROSPERO: CRD42018111412.

METHODS

Literature search strategy

Six databases (Medline, Embase, PsycINFO, PubMed, Web of Science, and CINAHL) were systematically searched for journal articles published in English, between January 2000 and July 2018 inclusive. The search terms used were "Jefferson" AND "Empathy", "Health Personnel" OR "Student", "Reliability" OR "Validity" OR "Factor Analysis" OR "Classic Test Theory" OR "Item Response Theory".

Eligibility criteria

Studies were included if original data was reported on at least one of the psychometric measurement properties described in the COSMIN Risk of Bias checklist [31], examined the JSE (student, physician, or health profession version), and was published in English. Two reviewers independently assessed the titles and abstracts of the articles to determine whether inclusion criteria were met. Subsequent full-text review resulted in the exclusion of additional articles. Any disagreement that arose was resolved through discussion by the two reviewers until a consensus was reached.

Data collection process

The eligible studies were then examined by the two reviewers and relevant information was captured regarding the general characteristics of the studies. This included authors, language of use, country in which the study took place, population of study participants, mean age, gender, and response rate. Data were then extracted according to guidelines from the COSMIN Risk of Bias checklist [31]. In the present study, the term "Outcome measure instrument development" was used instead of the original "Patient reported outcome measure development" to more accurately reflect that the included studies examined outcomes reported by health care professionals or students, as opposed to patients.

Assessing the risk of bias

Assessment of the methodological quality of the selected studies was carried out using the COSMIN Risk of Bias checklist which contains 10 boxes used to assess a study on standards for good methodological quality (Table 1).

Evaluation of each measurement property

Each measurement property was scored on a four-point scale using the descriptors "very good", "adequate", "doubtful", and "inadequate". A "not applicable" option was also included for each property, and the measurement properties that were relevant to each study were assessed. An overall score for the methodological quality of each measurement property was determined by taking the lowest rating of any of the items in a box, that is "the worst score counts" principle [41]. For example, in the Structural validity box, if a confirmatory factor analysis was undertaken (a "very good" rating), but the sample size was < 5 times the number of items (an "inadequate" rating), the overall quality rating for this box would be judged as "inadequate". The assessment of the methodological quality of the studies was performed independently by two reviewers. Any disagreement that arose was resolved by discussion by the reviewers until a consensus was reached.

As this study examined one outcome instrument, the first two boxes in the Risk of Bias checklist which related to content validity, "outcome measure tool development" and "content validity", were deemed to be applicable to only the original paper which described the development of the JSE [3] and thus excluded from this study as a synthesis of evidence was unable to be drawn. Criterion validity and responsiveness were also excluded as they were not applicable to this evaluation.

Summary and assessing the quality of the evidence

One all included studies were evaluated for their psychometric reporting quality, a synthesis of best existing evidence for each

TABLE 1 COSMIN Risk of Bias checklist summary[31]

Content validity	
Box 1	**Outcome measure tool development**
Box 2	Content validity
Internal structure	
Box 3	Structural validity
Box 4	Internal consistency
Box 5	Cross-cultural validity\measurement invariance
Remaining measurement properties	
Box 6	Reliability
Box 7	Measurement error
Box 8	Criterion validity
Box 9	Hypotheses testing for construct validity
Box 10	Responsiveness

measurement property was then summarized according to the following table (Table 2) adapted from Chiarotto et al., 2016 [42]:

RESULTS

Study selection and characteristics

The search strategy resulted in 985 articles retrieved from six databases; Medline, Embase, PsycINFO, PubMed, Web of Science, and CINAHL. From this, 558 duplicates were found, resulting in a total of 427 articles as shown in Figure 1. After assessing articles for inclusion and exclusion criteria, a total of 59 articles were included in the methodological evaluation using the COSMIN Risk of Bias checklist.

The majority of the cohorts that were previously studied with the JSE consisted of medical students or practitioners, with 44 out of 59 studies involving one or both groups. Other professional cohorts studied included nurses, as well as health care students from nursing, dentistry, pharmacy, paramedicine, and osteopathic medicine disciplines. Thirty-four studies examined a translated version of the JSE. Nine studies involved versions in Spanish, four Iranian, and three in Turkish, Italian, and Chinese. Other language versions examined were Japanese (2 studies), Portuguese (2), French (1), Polish (1), Czech (1), Bengali (1), German (1), and Slovenian (1). The general characteristics of these studies are presented in Table 3.

Risk of bias of JSE measurement properties

Structural validity

Structural validity is a relevant measure only if the outcome instrument to be assessed is based on a reflective model. That is, all items in the instrument are indicators of the one construct undergoing measurement [32]. The JSE was judged to be based on a reflective model, thus it was appropriate to undergo examination for structural validity. Overall, 53 of the 59 studies examined the structural validity of the JSE. Eighteen studies performed a confirmatory factor analysis and thus received a "very good" rating, and 24 studies performed an exploratory factor analysis, so were considered to be "adequate". Eleven studies did not report either type of factor analysis and therefore were scored as "inadequate". A three-factor structure of the JSE was most commonly reported, with 16 studies describing this arrangement [7,10,20,21,43,50,52,58,63,64,66,68,72,79,83,88]. One of these studies reported two different factor structures which was dependent on the medical student cohort involved; a three-factor structure for a preclinical cohort, and four factors for the clinical cohort [79]. One further study supported a four-factor structure, and one study, a two-factor structure [87].

TABLE 2 Levels of evidence for the quality of measurement properties

Level	Rating†	Criteria
Strong	+ or −	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	+ or −	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or −	One study of fair methodological quality
Conflicting	+ or −	Conflicting findings
Unknown	?	Only studies of poor methodological quality or no studies

†+=positive rating, ?=indeterminate rating, −=negative rating

Internal consistency

Fifty-seven studies reported on internal consistency of the tool, of which, 42 studies were assessed as being "inadequate" in this regard. Inadequate ratings for internal consistency were generally due to the internal consistency statistic (most commonly Cronbach's alpha) being calculated for entire scale, not each subscale separately. However, 15 studies received a rating of "very good", which indicated that the internal consistency statistic had been calculated for each subscale within the JSE, as well as adhering to the other requisites for assessing this property. For the studies reporting a three-factor structure, the Cronbach's alpha range for the most commonly utilized factors were as follows: "perspective taking" 0.44 – 0.90, "compassionate care" 0.58 – 0.87, and "standing or walking in the patient's shoes" 0.45 – 0.77.

Cross-cultural validity

As has previously been identified, the COSMIN Risk of Bias checklist no longer includes an assessment on the quality of language translation as in the previous version of the checklist. The cross-cultural validity is now assessed when there is data from at least two different groups, and the scores of these groups are directly compared in one statistical model. Three papers could be evaluated under these provisions. Two of the studies which compared the Spanish version of the JSE in Spain and Latin American countries [43,47] had "doubtful" reporting quality. However cultural comparison involving Brazil, Ireland, New Zealand, Portugal, and the United Kingdom was reported in an "adequate" manner [45].

Reliability and Measurement error

For the majority of studies reported in this paper, reliability and measurement error were not applicable as repeated measures were not utilized. Overall, nine studies reported on repeated measures, and thus were able to be assessed for reliability and measurement error. No study received a "very good" rating for reliability, indicating that the studies lacked evidence regarding the stability of participants between measures, the appropriate time interval between measures, similarity of test conditions, and/or inadequate statistical methods. Three studies were judged as "adequate", five "inadequate," and one "doubtful". Of the nine studies that included repeated measures, no papers calculated the percentage agreement, which was one of the requisites to achieve a "very good" or "adequate" rating on assessment of measurement error. Thus, all nine studies were assessed as "inadequate".

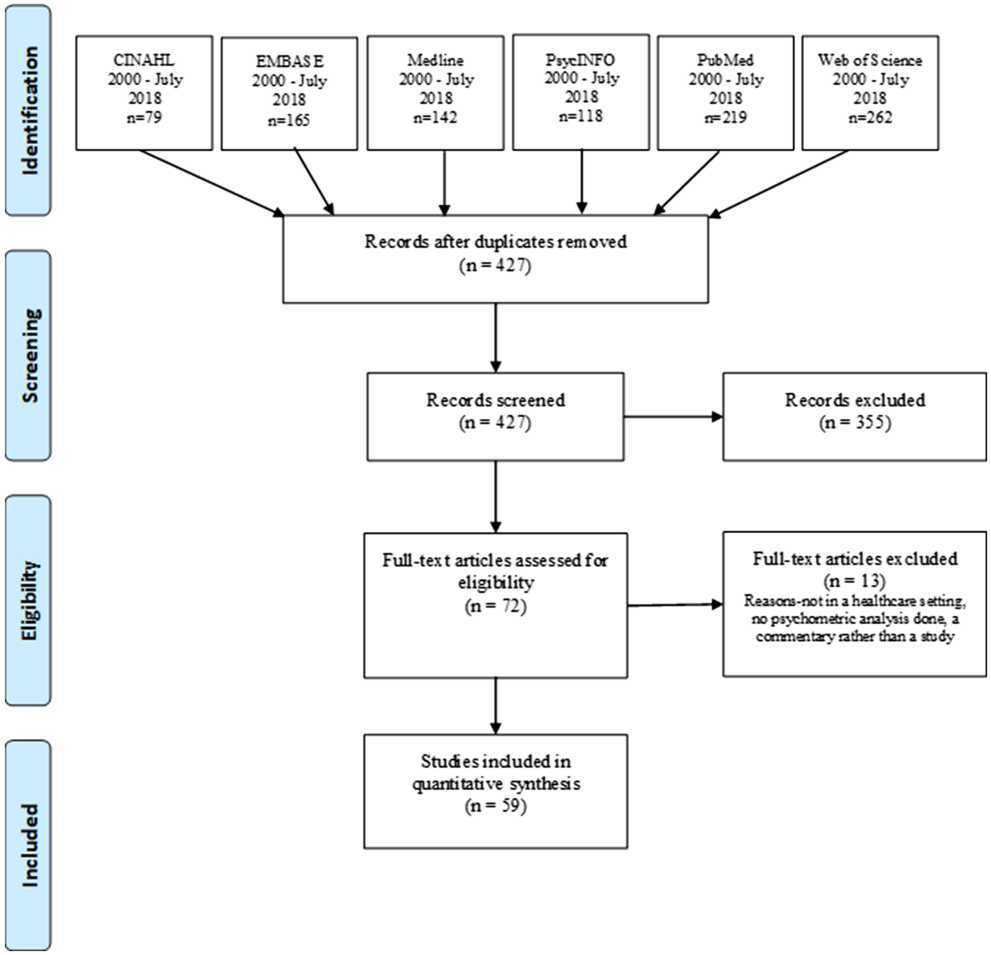

FIGURE 1 PRISMA Flow Diagram

Hypothesis testing for construct validity—Convergent validity

Hypothesis testing for convergent validity assumes that the tool under investigation is a valid measure of the relevant concept and occurs when it is compared to another related outcome measurement tool [32]. Nine studies undertook such a comparison, with six studies investigating comparisons with the IRI [12]. Of these, five were rated "very good" and one "inadequate". Other comparisons were made with the Empathic Tendency Scale [88] and between two different versions of the JSE [7,85]. These latter three scored "very good" for this component. A summary of methodological quality of the studies is presented in Table 4.

DISCUSSION

The JSE has been widely used in many different countries and cultures [91]. However, to the best of our knowledge, this is the first systematic review to examine the risk of bias in the reporting of measurement properties of the JSE (physician version, student version, and other health care versions). Fifty-nine studies were identified for inclusion by a systematic literature search encompassing the dates January 2000 to July 2018 inclusive. The evaluation was undertaken using the COSMIN guidelines and the COSMIN Risk of Bias checklist [31,32].

Despite being a commonly used instrument to measure empathy in health care settings, this study has demonstrated that the quality of reporting of psychometric properties of the JSE is inconsistent and suboptimal.

Structural validity and internal consistency are the most frequently described JSE psychometric characteristics in the literature. Both have been reported well in many studies, demonstrating a strong level of evidence for their use (Table 5). Convergent validity has been assessed in far fewer studies and generally through comparisons with the IRI scale. However, the existing research does demonstrate a strong level of evidence (Table 5) indicating that researchers may have a high level of confidence in results for this property.

In comparison, some other psychometric components of the JSE have not been comprehensively evaluated or reported well as assessed with the COSMIN Risk of Bias checklist. Reliability and measurement error have not generally been well described in existing studies. Nine out of the fifty-nine studies included results of repeated measures [4,22,23,49,54,61,64,68,80], required for assessment of these properties. The results of all nine studies were assessed to be "inadequate" or "doubtful" for both reliability and measurement error, except for three studies where reliability was "adequate" [61,64,80]. Synthesis of these results suggest that reliability has a moderate level of evidence, however for measurement error, it is unknown at present.

TABLE 3 General characteristics of included studies

	Study	Language	Country	Population	Mean (SD) age	Gender (female %)	Number (response rate %)
1.	Alcorta-Garza et al. (2016)[43]	Spanish	Spain and Latin American countries	Physicians and medical students	35 (10.8)	54	715 (80)
2.	Bilgel and Ozcakir (2017)[20]	Turkish	Turkey	Medical students	ns (ns)	ns	ns
3.	Borracci et al. (2015)[44]	Spanish	Argentina	Cardiologists	51.4 (11.4)	24.5	566 (62.9)
4.	Costa et al. (2017)[45]	ns	Portugal, Brazil, United Kingdom, New Zealand, Ireland	Medical students	ns	61.50	3,069 (30.4)
5.	Di Lillo (2009)[46]	Italian	Italy	Physicians	48.7 (8.08)	20.8	289 (37)
6.	Diaz Narvaez et al. (2014)[47]	Spanish	Columbia and Dominican Republic	Medical students	ns	ns	1,838 (55.1)
7.	Diaz-Narvaez et al. (2018)[48]	Spanish	Chile	Dentistry Students	ns	65.2	535 (66.8)
8.	du Vaure et al. (2017)[49]	French	France	Medical students	ns	54.20	299 (74.4)
9.	Ferreira-Valente et al. (2016)[50]	Spanish	Spain	Medical Students	20.7 (2.6)	68	1,104 (73.5)
10.	Fields et al. (2004)[51]	English	United States	Nurses and physicians	ns	100	98 (ns)
11.	Fields et al. (2011)[4]	English	United States	Health profession students	ns	88	265 (93.0)
12.	Fjortoft et al. (2011)[25]	English	United States	Pharmacy students	ns	ns	187 (ns)
13.	Gönüllü, & Öztuna (2012)[52]	Turkish	Turkey	Medical students	ns	48.4	752 (ns)
14.	Gonzalez-Martinez et al. (2018)[53]	Spanish	Colombia	Dentistry students	ns	62.70	322 (79.5)
15.	Hojat & LaNoue (2014)[10]	English	United States	Medical students	ns	51	2,612 (93)
16.	Hojat et al. (2001)[3]	English	United States	Physicians, internal medicine residents, and students	ns	ns	289 (ns)
17.	Hojat et al. (2002)[54]	English	United States	Physicians	46.8 (10.5)	26	704 (70)
18.	Hojat et al. (2002)[55]	English	United States	Physicians	46.8 (10.5)	26	704 (70)
19.	Hojat et al. (2005)[56]	English	United States	Internal medicine residents	ns	ns	93 (ns)
20.	Hojat et al. (2011)[57]	English	United States	Medical students	ns	ns	201 (78)
21.	Hojat et al. (2015)[8]	English	United States	Medical students	ns	49	265 (94)
22.	Hojat, M., et al. (2018)[21]	English	USA	Osteopathic medicine students	24.8 (3.4)	47	6009 (92)
23.	Hsiao, Tsai & Kao (2013)[22]	Chinese	Taiwan	Nursing students	23.26 (5.0)	89.1	613 (83.9)
24.	Jeon & Cho (2015)[58]	Korean	Korea	Pharmacy students	ns	81.9	447 (98.9)
25.	Kataoka et al. (2009)[59]	Japanese	Japan	Medical students	ns	25.8	400 (66.7)
26.	Kataoka et al. (2012)[60]	Japanese	Japan	Physicians	ns	100	285 (21)
27.	Kesbakhi et al. (2017)[61]	Iranian	Iran	Oncology nurses	34.9 (6.8)	88.40	205 (88.3)
28.	Kliszcz, Saver & Sadowska (2006)[62]	Polish	Poland	Physicians, nurses, and students	ns	80	
29.	Kožený & Tišanská (2013)[63]	Czech	Czech Republic	Physicians	46.71 (11.5)	55.4	1,305 (ns)
30.	Leombruni et al. (2014)[64]	Italian	Italy	Medical students	20.6 (ns)	55.6	257 (65.9)
31.	Li et al. (2015)[65]	Chinese	China	Pharmacy students	20.98 (1.55)	65	263 (99.2)

(Continues)

TABLE 3 (Continued)

	Study	Language	Country	Population	Mean (SD) age	Gender (female %)	Number (response rate %)
32.	Magalhães et al. (2011)[66]	Portuguese	Portugal	Medical students	ns	67.4	476 (92)
33.	Mandel & Schweinle (2012)[67]	English	United States	Physician assistant students	24.0 (3.2)	82.3	328 (94.7)
34.	McMillan & Shannon (2011)[7]	English	United States	Nurses	ns	88	598 (83)
35.	Montanari et al. (2015)[68]	Italian	Italy	Nursing Students	22.63 (4.77)	74	797 (92.7)
36.	Mostafa et al. (2014)[69]	Bengali	Bangladesh	Medical students	ns	70.1	348 (81.7)
37.	Paro et al. (2012)[70]	Portuguese	Brazil	Medical students	ns	ns	299 (93.7)
38.	Preusche & Wagher-Menghin (2013)[23]	German	Germany	Medical students	ns	44.9	557 (89.3)
39.	Rahimi-Madiseh (2010)[71]	Iranian	Iran	Medical students	ns	70.2	181 (52.2)
40.	Ren et al. (2016)[72]	English	Singapore	Medical students	ns	53.7	881 (62.9)
41.	Roh et al. (2010)[73]	Korean	Korea	Medical students	ns	32.3	493 (72)
42.	San-Martin et al. (2017)[74]	Spanish	Latin American countries	Physicians and nurses	38 (9.51)	59.5	353 (68)
43.	Shahini et al. (2016)[75]	Iranian	Iran	Student residence assistants	31.49 (4.5)	42	74 (83.5)
44.	Shariat, Eshtad & Ansari (2010)[76]	Iranian	Iran	Physicians	ns	52.7	207 (75)
45.	Shekhawat et al. (2017)[24]	English	India	Postgraduate dental students	ns	65.80	120 (80)
46.	Soler-Gonzalez et al. (2017)[77]	ns	Spain	Physicians and nurses	44 (11)	Not clearly defined	Not clearly defined
47.	Spasenoska, Costello & Williams (2016)[78]	English	Malaysia	Medical students	19.3 (1.2)	57	193 (ns)
48.	Stansfields et al. (2016)[79]	English	United States	Medical students	23.7 (2.7)	49.5	4,749 (77.2)
49.	Ster et al. (2013)[80]	Slovenian	Slovenia	Medical students	19.3 (2.3)	69.2	234 (78.5)
50.	Suh et al. (2012)[81]	Korean	Korea	Physicians	ns	45	229 (100)
51.	Tariq et al. (2018)[82]	English	Pakistan	Medical students	ns	49.60	337 (70.9)
52.	Tavakol, Dennick & Tavakol (2011)[83]	English	United Kingdom	Medical students	ns	55.1	853 (68.2)
53.	Waldrop et al. (2016)[84]	English	United States	Dental students	26.3 (4.2)	43.5	292 (86.9)
54.	Ward et al. (2009)[85]	English	United States	Nursing students	ns	85	333 (71)
55.	Wen et al. (2013)[86]	Chinese	China	Medical students	ns	63.2	753 (83.5)
56.	Williams et al. (2013)[87]	English	Australia	Paramedic students	ns	65.2	330 (ns)
57.	Yanik & Saygili (2014)[88]	Turkish	Turkey	Medical students	20.8 (ns)	74.2	414 (74.2)
58.	Yuguero et al. (2017)[89]	Spanish	Spain	Physicians and nurses	48 (median) (31-65 years)	78.3	267 (52.7)
59.	Yuguero Torres et al. (2015)[90]	Spanish	Spain	Physicians	ns	63.9	108 (49.8)

ns=not specified

TABLE 4 Methodological quality of assessments of the measurement properties of the JSE

	Study (year) (reference no.)	Structural validity	Internal Consistency	Reliability	Measurement error	Convergent validity	Other
1.	Alcorta-Garza et al. (2016) [43]	Very good	Inadequate	NR	NR	NR	Cross cultural validity -Doubtful
2.	Bilgel and Ozcakir (2017) [20]	Very good	Very good	NR	NR	NR	NR
3.	Borracci et al. (2015) [44]	Adequate	Very good	NR	NR	NR	NR
4.	Costa et al. (2017) [45]	Adequate	Very good	NR	NR	Very good	Cross cultural validity -Adequate
5.	Di Lillo (2009) [46]	Adequate	Inadequate	NR	NR	NR	NR
6.	Diaz Narvaez et al. (2014) [47]	N/A	N/A	NR	NR	NR	Cross cultural validity -Doubtful
7.	Diaz-Narvaez et al. (2018) [48]	Inadequate	Inadequate	NR	NR	NR	NR
8.	du Vaure et al. (2017) [49]	Inadequate	Inadequate	Inadequate	Inadequate	NR	NR
9.	Ferreira-Valente et al. (2016) [50]	Very good	Very good	NR	NR	NR	NR
10.	Fields et al. (2004) [51]	NR	Inadequate	NR	NR	NR	NR
11.	Fields et al. (2011) [4]	Inadequate	Inadequate	Inadequate	Inadequate	NR	NR
12.	Fjortoft et al. (2011) [25]	Adequate	Inadequate	NR	NR	NR	NR
13.	Gönüllü, & Öztuna (2012) [52]	Very good	Very good	NR	NR	NR	NR
14.	Gonzalez-Martinez et al. (2018) [53]	Inadequate	Inadequate	NR	NR	NR	NR
15.	Hojat & LaNoue (2014) [10]	Very good	Inadequate	NR	NR	NR	NR
16.	Hojat et al. (2001) [3]	Adequate	Inadequate	NR	NR	Inadequate	Outcome measure development - Inadequate Content validity -Doubtful
17.	Hojat et al. (2002) [54]	Adequate	Inadequate	Inadequate	Inadequate	NR	NR
18.	Hojat et al. (2002) [55]	Inadequate	Inadequate	NR	NR	NR	NR
19.	Hojat et al. (2005) [56]	N/A	N/A	NR	NR	Very good	NR
20.	Hojat et al. (2011) [57]	Adequate	Inadequate	NR	NR	Very good	NR
21.	Hojat et al. (2015) [8]	Adequate	Inadequate	NR	NR	NR	NR
22.	Hojat, M., et al. (2018) [21]	Very good	Very good	NR	NR	NR	NR
23.	Hsiao, Tsai & Kao (2013) [22]	Adequate	Very good	Inadequate	Inadequate	NR	NR
24.	Jeon & Cho (2015) [58]	Very good	Very good	NR	NR	Very good	NR
25.	Kataoka et al. (2009) [59]	Adequate	Inadequate	NR	NR	NR	NR
26.	Kataoka et al. (2012) [60]	Inadequate	Inadequate	NR	NR	NR	NR
27.	Kliszcz, et al. (2006) [62]	Inadequate	Inadequate	NR	NR	NR	NR
28.	Kožený & Tišanská (2013) [63]	Very good	Inadequate	NR	NR	NR	NR
29.	Leombruni et al. (2014) [64]	Very good	Inadequate	Adequate	Inadequate	NR	NR
30.	Li et al. (2015) [65]	Adequate	Inadequate	NR	NR	NR	NR
31.	Magalhães et al. (2011) [66]	Very good	Inadequate	NR	NR	NR	NR
32.	Mandel & Schweinle (2012) [67]	NR	Inadequate	NR	NR	NR	NR
33.	McMillan & Shannon (2011) [7]	Very good	Inadequate	NR	NR	Very good	NR
34.	Montanari et al. (2015) [68]	Very good	Inadequate	Doubtful	Inadequate	NR	NR
35.	Mostafa et al. (2014) [69]	Adequate	Inadequate	NR	NR	NR	NR
36.	Paro et al. (2012) [70]	Adequate	Very good	NR	NR	NR	NR
37.	Preusche & Wagher-Menghin (2013) [23]	Adequate	Inadequate	Inadequate	Inadequate	NR	NR

(Continues)

TABLE 4 (Continued)

	Study (year) (reference no.)	Structural validity	Internal Consistency	Reliability	Measurement error	Convergent validity	Other
38.	Rahimi-Madiseh et al. (2010) [71]	Adequate	Very good	NR	NR	NR	NR
39.	Ren et al. (2016) [72]	Very good	Inadequate	NR	NR	NR	NR
40.	Roh et al. (2010) [73]	Adequate	Inadequate	NR	NR	NR	NR
41.	San-Martin et al. (2017) [74]	NR	Inadequate	NR	NR	NR	NR
42.	Sedaghati Kesbakhi et al. (2017) [61]	Very good	Very good	Adequate	Inadequate	Very good	NR
43.	Shahini et al. (2016) [75]	Inadequate	Inadequate	NR	NR	NR	NR
44.	Shariat, Eshtad & Ansari (2010) [76]	Adequate	Inadequate	NR	NR	NR	NR
45.	Shekhawat et al. (2017) [24]	Adequate	Inadequate	NR	NR	NR	NR
46.	Soler-Gonzalez et al. (2017) [77]	Inadequate	Inadequate	NR	NR	NR	NR
47.	Spasenoska, Costello & Williams (2016) [78]	Adequate	Very good	NR	NR	NR	NR
48.	Stansfield et al. (2016) [79]	Very good	Inadequate	NR	NR	NR	NR
49.	Ster et al. (2014) [80]	Adequate	Inadequate	Adequate	Inadequate	NR	NR
50.	Suh et al. (2012) [81]	Adequate	Inadequate	NR	NR	NR	NR
51.	Tariq et al. (2018) [82]	Inadequate	Inadequate	NR	NR	NR	NR
52.	Tavakol, Dennick & Tavakol (2011) [83]	Very good	Very good	NR	NR	NR	NR
53.	Waldrop et al. (2016) [84]	Adequate	Inadequate	NR	NR	NR	NR
54.	Ward et al. (2009) [85]	Adequate	Very good	NR	NR	Very good	NR
55.	Wen et al. (2013) [86]	Adequate	Inadequate	NR	NR	NR	NR
56.	Williams et al. (2013) [87]	Very good	Inadequate	NR	NR	NR	NR
57.	Yanik & Saygili (2014) [88]	Very good	Very good	NR	NR	Very good	NR
58.	Yuguero et al. (2017) [89]	Inadequate	Inadequate	NR	NR	NR	NR
59.	Yuguero Torres et al. (2015) [90]	NR	Inadequate	NR	NR	NR	NR

NR = not reported

TABLE 5 Best evidence synthesis of measurement properties of the JSE

Measurement property (no. of studies assessing measurement property)	COSMIN Risk of Bias checklist ratings				Level of evidence for the quality of measurement property
	Very good	Adequate	Doubtful	Inadequate	
Structural validity (*n*=53)	18	24	0	11	Strong (+)
Internal consistency (*n*=57)	15	0	0	42	Strong (+)
Reliability (*n*=9)	0	3	1	5	Moderate (+)
Measurement error (*n*=9)	0	0	0	9	Unknown
Convergent validity (*n*=9)	8	0	0	1	Strong (+)
Cross cultural validity (*n*=3)	0	1	2	0	Moderate (+)

Using the updated version of the Risk of Bias checklist [31], only three studies were able to be assessed with regards to cross-cultural validity. Synthesis of the evidence from these studies indicate that there is a moderate level of evidence for this property. The previous version of the Risk of Bias checklist [38] included evaluation of procedures such as language translation, which was not assessed with the

updated tool. Thus, it is possible that some studies may have been previously assessed in the literature as having a higher quality of reporting due to adequate translation procedures used.

Thus, although the JSE is one of the most widely used measures of empathy in a health care environment, a level of caution needs to be employed in the determination of conclusions regarding aspects such as reliability, measurement error, or cross-cultural validity. More consistent reporting of psychometric properties of the JSE would add significantly to the methodological rigour of assessments of the tool.

Most published papers using the COSMIN methodology utilized an earlier the version of the Risk of Bias scale [38,39]. Since this time the methodology has been updated considerably, as previously outlined in the present paper [31,32]. This means that the present study describes different findings that would have been if using the former version of the Risk of Bias checklist.

CONCLUSION

The JSE is a tool to measure empathy in health care, which demonstrates robust structural validity, internal consistency, and convergent validity. These psychometric characteristics are generally well reported in the literature and thus may be interpreted with relative confidence when used in empathy research. However, the quality of evidence for reliability, measurement error, and cross-cultural validity is somewhat poorer and only addressed in a small number of papers. Thus, a degree of caution should be considered in drawing conclusions when using the JSE with regard to these properties until further research is undertaken to strengthen evidence. As the quality of assessment of psychometric properties of the JSE is inconsistent, it is recommended that future examinations of the JSE refer to the COSMIN guidelines to ensure complete and well-reported psychometric data are included.

DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

No ethical approval required for this study.

CONSENT FOR PUBLICATION

Not applicable

AVAILABILITY OF DATA AND MATERIAL

Not applicable

COMPETING INTERESTS

Nil competing interests to declare

FUNDING

No funding was provided for this paper

AUTHORS' CONTRIBUTIONS

BW conceptualized the study, developed the search strategy, evaluated the methodological quality of included studies, wrote sections of the paper, provided feedback on manuscript drafts, and read and approved the final manuscript. BB developed the search strategy, performed the literature search, drafted the manuscript, evaluated the methodological quality of included studies, and read and approved the final manuscript.

ACKNOWLEDGEMENTS

Nil

ORCID

Brett Williams https://orcid.org/0000-0001-6307-1779

REFERENCES

1. Del Canale S, Louis DZ, Maio V, et al. The relationship between physician empathy and disease complications: an empirical study of primary care physicians and their diabetic patients in Parma, Italy. *Acad Med*. 2012;87(9):1243-1249.
2. Beck R, Daughtridge R, Sloane P. Physician-Patient Communication in the Primary Care Office: A Systematic Review. *The Journal of the American Board of Family Practice*. 2002;15(1):25-38.
3. Hojat M, Mangione S, Nasca T, et al. The Jefferson scale of physician empathy: Development and preliminary psychometric data. *Educ Psychol Meas*. 2001;61(2):349-365.
4. Fields S, Mahan P, Tillman P, Harris J, Maxwell K, Hojat M. Measuring empathy in healthcare profession students using the Jefferson Scale of Physician Empathy: health provider-student version. *Journal of Interprofessional Care*. 2011;25(4):287-293.
5. Hojat M, Louis D, Markham F, Wender R, Rabinowitz C, Gonnella J. Physicians' empathy and clinical outcomes for diabetic patients. *Acad Med*. 2011;86(3):359-364.
6. Clark AJ. Empathy: An integral model in the counseling process. *Journal of Counseling and Development*. 2010;88(3):348-356.
7. McMillan LR, Shannon DM. Psychometric analysis of the JSPE nursing student version R: comparison of senior BSN students and medical students attitudes toward empathy in patient care. *ISRN Nurs*. 2011;2011 (1-7):1-7.
8. Hojat M, Vergare M, Isenberg G, Cohen M, Spandorfer J. Underlying construct of empathy, optimism, and burnout in medical students. *International Journal of Medical Education*. 2015;6:12-16.
9. Thirioux B, Birault F, Jaafari N. Empathy is a protective factor of burnout in physicians: New Neuro-Phenomenological hypotheses regarding empathy and sympathy in care relationship. *Front Psychol*. 2016;7:763.
10. Hojat M, LaNoue M. Exploration and confirmation of the latent variable structure of the Jefferson scale of empathy. *International Journal of Medical Education*. 2014;5:73-81.
11. Thomas Jefferson University. Jefferson Scale of Empathy 2017 [http://www.jefferson.edu/university/skmc/research/research-medical-education/jefferson-scale-of-empathy.html. Accessed 11 June 2018.

12. Davis MH. Measuring individual differences in empathy: Evidence for a multidimensional approach. *J Pers Soc Psychol*. 1983;44(1):113-126.

13. Costa P, McCrae R. *Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory (NEO-FFI): Professional Manual*. Psychological Assessment Resources: Odessa, FL; 1992.

14. Rosenberg M. *Occupations and value*. Glencoe, IL: The Free Press; 1957.

15. Rosenberg M. *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press; 1965.

16. Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas*. 1960;20(1):141-151.

17. Hojat M. *Empathy in patient care: Antecedents, development, measurement, and outcomes*. New York: NY Springer; 2007.

18. Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. *Psychol Bull*. 1986;99(3):432-442.

19. Ertel S. *Factor analysis: Healing an ailing model*. Berlin, Germany: University of Göttingen Publisher; 2013.

20. Bilgel N, Ozcakir A. Turkish Version of the Jefferson Scale of Empathy Psychometric Properties. *European Scientific Journal*. 2017;13(20).

21. Hojat M, DeSantis J, Shannon SC, et al. The Jefferson Scale of Empathy: a nationwide study of measurement properties, underlying components, latent variable structure, and national norms in medical students. *Advances in Health Sciences Education*. 2018;3(5):899-920.

22. Hsiao C-Y, Tsai Y-F, Kao Y-C. Psychometric properties of a Chinese version of the Jefferson Scale of Empathy-Health Profession Students. *J Psychiatr Ment Health Nurs*. 2013;20(10):866-873.

23. Preusche I, Wagner-Menghin M. Rising to the challenge: cross-cultural adaptation and psychometric evaluation of the adapted German version of the Jefferson Scale of Physician Empathy for Students (JSPE-S). *Advances in Health Sciences Education*. 2013;18(4):573-587.

24. Shekhawat KS, Chauhan A, Devi SS, Kunjumon S. Construct Validity and Reliability of Jefferson Scale of Empathy-health Care Provider (Student version) among Final Year, Interns and Post-graduate Students of a Dental College in India. *Indian Journal of Public Health Research & Development*. 2017;8(4):184-192.

25. Fjortoft N, Van Winkle LJ, Hojat M. Measuring empathy in pharmacy students. *Am J Pharm Educ*. 2011;75(6):109.

26. COSMIN. Systematic reviews of measurement properties, 2017. [https://www.cosmin.nl/tools/cosmin-taxonomy-measurement-properties/.

27. U.S. Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims 2009 [https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm193282.pdf.

28. Williams K, Sansoni J, Darcy M, Grootemaat P, Thompson C. Patient-reported outcome measures. *Literature review*. Sydney: Australian Commission on Safety and Quality in Health Care; 2016 [https://ahsri.uow.edu.au/content/groups/public/@web/@chsd/documents/doc/uow226913.pdf.

29. Kotronoulas G, Kearney N, Maguire R, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *J Clin Oncol*. 2014;32(14):1480-1510.

30. Black N. Patient reported outcome measures could help transform healthcare. *BMJ* (Clinical Research Ed). 2013;346(jan28 1):f167.

31. Mokkink LB, de Vet HC, Prinsen CA, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27(5):1171-1179.

32. Mokkink LB, Prinsen CAC, Patrick DL, Jordi A, Bouter LM, de Vet HC, et al. *COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs)* 2018 [https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf. .

33. Daanen HA, Lamberts RP, Kallen VL, Jin A, Van Meeteren NL. A systematic review on heart-rate recovery to monitor changes in training status in athletes. *Int J Sports Physiol Perform*. 2012;7(3):251-260.

34. Peterson ML, Lukens K, Fulk G. Psychometric Properties of Physical Function Measures Used in the Intensive Care Unit: A Systematic Review. *Journal of Acute Care Physical Therapy*. 2018;9(2):78-90.

35. Tax C, Steenbergen ME, Zusterzeel PL, Bekkers RL, Rovers MM. Measuring health-related quality of life in cervical cancer patients: a systematic review of the most used questionnaires and their validity. *BMC Med Res Methodol*. 2017;17(1):15.

36. de Araujo FX, Ferreira GE, Schell MS, de Castro MP, Silva MF, Ribeiro DC. Measurement properties of the craniocervical flexion test: a systematic review protocol. *BMJ Open*. 2018;8(2):e019486.

37. Melissant HC, Neijenhuijs KI, Jansen F, et al. A systematic review of the measurement properties of the Body Image Scale (BIS) in cancer patients. *Support Care Cancer*. 2018;26(6):1715-1726.

38. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10(1):22.

39. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-549.

40. de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Systematic reviews of measurement properties*. Measurement in medicine: a practical guide: Cambridge University Press; 2011.

41. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651-657.

42. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. *Phys Ther*. 2016;96(10):1620-1637.

43. Alcorta-Garza A, San-Martin M, Delgado-Bolton R, Soler-González J, Roig H, Vivanco L. Cross-validation of the Spanish HP-version of the Jefferson Scale of Empathy confirmed with some cross-cultural differences. *Front Psychol*. 2016;7:1002.

44. Borracci RA, Doval HC, Nunez C, Samarelli M, Tamini S, Tanus E. Measurement of empathy among Argentine cardiologists: Psychometrics and differences by age, gender, and subspecialty. *Cardiol J*. 2015;22(1):52-56.

45. Costa P, de Carvalho-Filho MA, Schweller M, et al. Measuring medical students' empathy: exploring the underlying constructs of and associations between two widely used self-report instruments in five countries. *Acad Med*. 2017;92(6):860-867.

46. Di Lillo M, Cicchetti A, Scalzo AL, Taroni F, Hojat M. The Jefferson Scale of Physician Empathy: preliminary psychometrics and group comparisons in Italian physicians. *Acad Med*. 2009;84(9):1198-1202.

47. Narváez VPD, Palacio LMA, Caro SE, Silva MG, Castillo JA, Bilbao J. Empathic orientation among medical students from three universities in Barranquilla, Colombia and one university in the Dominican Republic. *Arch Argent Pediatr*. 2014;112(1):41-49.
48. Díaz-Narváez VP, Amezaga-Avitia AC, Sarabia-Alvarez PA, et al. Chilean dentistry students, levels of empathy and empathic erosion: necessary evaluation before a planned intervention: Levels of empathy, evaluation and intervention. *The Saudi Dental Journal*. 2018;30 (2):117-124.
49. Buffel du Vaure C, Lemogne C, Bunge L, et al. Promoting empathy among medical students: A two-site randomized controlled study. *J Psychosom Res*. 2017;103:102-107.
50. Ferreira-Valente A, Costa P, Elorduy M, Virumbrales M, Costa MJ, Pales 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):242.
51. Fields SK, Hojat M, Gonnella JS, Mangione S, Kane G, Magee M. Comparisons of nurses and physicians on an operational measure of empathy. *Eval Health Prof*. 2004;27(1):80-94.
52. Gönüllü İ, Öztuna D. A Turkish Adaptation of the Student Version of the Jefferson Scale of Physician Empathy. *Marmara Medical Journal*. 2012;25(2).
53. González-Martínez F, Tirado-Amador L, Bueno-Hernández J, Chica-Duque B, Díaz-Narváez VP. Changes in Empathy Levels on Dentistry's Students of Public University in Cartagena City, Colombia. *Pesqui Bras Odontopediatria Clin Integr*. 2018;18(1):e3989.
54. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: Definition, components, measurement, and relationship to gender and speciality. *Am J Psychiatry*. 2002;159 (9):1563-1569.
55. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Veloksi JJ, Magee M. The Jefferson Scale of Physician Empathy: further psychometric data and differences by gender and specialty at item level. *Acad Med*. 2002;77 (10):S58-S60.
56. Hojat M, Mangione S, Kane GC, Gonnella JS. Relationships between scores of the Jefferson scale of physician empathy (JSPE) and the interpersonal reactivity index (IRI). *Med Teach*. 2005;27(7):625-628.
57. Hojat M, Spandorfer J, Louis DZ, Gonnella JS. Empathic and sympathetic orientations toward patient care: conceptualization, measurement, and psychometrics. *Acad Med*. 2011;86(8):989-995.
58. Jeon S, Cho E. Assessment of Korean Pharmacy Students' Empathy Using the Jefferson Scale of Empathy. *Am J Pharm Educ*. 2015;79 (5):67.
59. Kataoka HU, Koide N, Ochi K, Hojat M, Gonnella JS. Measurement of empathy among Japanese medical students: psychometrics and score differences by gender and level of medical education. *Acad Med*. 2009;84(9):1192-1197.
60. Kataoka HU, Koide N, Hojat M, Gonnella JS. Measurement and correlates of empathy among female Japanese physicians. *BMC Med Educ*. 2012;12(1):48.
61. Sedaghati Kesbakhi M, Rohani C, Mohtashami J, Nasiri M. Validity and Reliability of the Jefferson Scale of Empathy, Nursing Student Version R in a Sample of Iranian Oncology Nurses. *Nurs Midwifery Stud*. 2017;6 (2):e39505.
62. Kliszcz J, Nowicka-Sauer K, Trzeciak B, Nowak P, Sadowska A. Empathy in health care providers–validation study of the Polish version of the Jefferson Scale of Empathy. *Adv Med Sci*. 2006;51:219-225.
63. Kožený J, Tišanská L. The structure of the Jefferson Scale of Physician Empathy in Czech physicians. *Cesk Psychol*. 2013;57(6):521-532.
64. Leombruni P, Di Lillo M, Miniotti M, et al. Measurement properties and confirmatory factor analysis of the Jefferson Scale of Empathy in Italian medical students. *Perspectives on Medical Education*. 2014;3(6):419-430.
65. Li L, Wang J, Hu XM, Xu C. Empathy in Chinese pharmacy undergraduates: Implication for integrating humanities into professional pharmacy education. *Indian Journal of Pharmaceutical Education and Research*. 2015;49(1):31-39.
66. Magalhães E, Salgueira AP, Costa P, Costa MJ. Empathy in senior year and first year medical students: a cross-sectional study. *BMC Med Educ*. 2011;11(1):52.
67. Mandel ED, Schweinle WE. A study of empathy decline in physician assistant students at completion of first didactic year. *Journal of Physician Assistant Education*. 2012;23(4):16-24.
68. Montanari P, Petrucci C, Russo S, Murray I, Dimonte V, Lancia L. Psychometric properties of the Jefferson Scale of Empathy-Health Professional Student's version: An Italian validation study with nursing students. *Nurs Health Sci*. 2015;17(4):483-491.
69. Mostafa A, Hogue R, Mostafa M, Rana MM, Mostafa F. Empathy in undergraduate medical students of Bangladesh: Psychometric analysis and differences by gender, academic year, and specialty preferences. *ISRN Psychiatry*. 2014;2014:375-439.
70. Paro HB, Daud-Gallotti RM, Tibério IC, Pinto RM, Martins MA. Brazilian version of the Jefferson Scale of Empathy: psychometric properties and factor analysis. *BMC Med Educ*. 2012;12(1):73.
71. 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-e478.
72. Ren GSG, Min JTY, Ping YS, et al. Complex and novel determinants of empathy change in medical students. *Korean Journal of Medical Education*. 2016;28(1):67-78.
73. Roh M-S, Hahm B-J, Lee DH, Suh DH. Evaluation of empathy among Korean medical students: a cross-sectional study using the Korean Version of the Jefferson Scale of Physician Empathy. *Teach Learn Med*. 2010;22(3):167-171.
74. San-Martín M, Delgado-Bolton R, Vivanco L. Professionalism and occupational well-being: similarities and differences among Latin American health professionals. *Front Psychol*. 2017;8:63.
75. Shahini N, Rezayat KA, Behdani F, Shojaei SRH, Rezayat AA, Dadgarmoghaddam M. Empathy Score among Student Residence Assistants in Iran. *Electronic Physician*. 2016;8(12):3357-3362.
76. Shariat SV, Eshtad E, Ansari S. Empathy and its correlates in Iranian physicians: A preliminary psychometric study of the Jefferson Scale of Physician Empathy. *Med Teach*. 2010;32(10):e417-e421.
77. Soler-Gonzalez J, San-Martín M, Delgado-Bolton R, Vivanco L. Human connections and their roles in the occupational well-being of healthcare professionals: a study on loneliness and empathy. *Front Psychol*. 2017;8:1475.
78. Spasenoska M, Costello S, Williams B. Investigating the psychometric properties of the Jefferson Scale of Physician Empathy in a sample of Malaysian medical students. *Advances in Medical Education and Practice*. 2016;7:331-339.
79. Stansfield R, Schwartz A, O'Brien C, 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. *Advances in Health Sciences Education*. 2016;21(1):5-17.
80. Šter M, Šter B, Petek D, Gorup EC. Validation of Slovenian version of Jefferson Scale of Empathy for students. *Slovenian Journal of Public Health*. 2014;53(1):89-100.

81. Suh DH, Hong JS, Lee DH, Gonnella JS, Hojat M. The Jefferson Scale of Physician Empathy: a preliminary psychometric study and group comparisons in Korean physicians. *Med Teach*. 2012;34(6):e464-e468.
82. Tariq N, Tayyab A, Jaffery T. Differences in Empathy Levels of Medical Students Based on Gender, Year of Medical School and Career Choice. *Journal of the College of Physicians and Surgeons Pakistan*. 2018;28 (4):310-313.
83. Tavakol S, Dennick R, Tavakol M. Empathy in UK medical students: differences by gender, medical year and specialty interest. *Education Primary Care*. 2011;22(5):297-303.
84. Waldrop D, Nochajski T, Davis EL, Fabiano J, Goldberg L. Empathy in dentistry: How attitudes and interaction with older adults make a difference. *Gerontol Geriatr Educ*. 2016;37(4):359-380.
85. Ward J, Schaal M, Sullivan J, Bowen ME, Erdmann JB, Hojat M. Reliability and validity of the Jefferson Scale of Empathy in undergraduate nursing students. *J Nurs Meas*. 2009;17(1):73-88.
86. Wen D, Ma X, Li H, Liu Z, Xian B, Liu Y. Empathy in Chinese medical students: Psychometric characteristics and differences by gender and year of medical education. *BMC Med Educ*. 2013;13(1):130.
87. Williams B, Brown T, Boyle M, Dousek S. Psychometric testing of the Jefferson Scale of Empathy health profession students' version with Australian paramedic students. *Nurses Health Science*. 2013;15(1): 45-50.
88. Yanik A, Saygili S. Validity and reliability of the Turkish version of Jefferson Scale of Empathy for nursing students. *Turkiye Klinikleri Journal of Medical Science*. 2014;34(1):111-119.
89. Yuguero O, Ramon Marsal J, Esquerda M, Vivanco L, Soler-González J. Association between low empathy and high burnout among primary care physicians and nurses in Lleida, Spain. *European Journal of General Practice*. 2017;23(1):4-10.
90. Yuguero Torres O, Esquerda Aresté M, Marsal Mora JR, Soler-González J. Association between sick leave prescribing practices and physician burnout and empathy. *PLoS One*. 2015;10(7):e0133379.
91. Hojat M. *Empathy in Health Professions Education and Patient Care*. Switzerland Springer International Publishing; 2016.

How to cite this article: 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*. 2019. https://doi.org/10.1111/jep.13293