

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

Alexandra Ferreira-Valente¹

(mafvalente@gmail.com)

Patrício Costa¹

(pcosta@ecsaude.uminho.pt)

Marta Elorduy²

(melorduy@uic.es)

Montserrat Virumbrales²

(montsev@uic.es)

Manuel J. Costa¹

(mmcosta@ecsaude.uminho.pt)

Jorge Palés³

(jpales@ub.edu)

¹ University of Minho, School of Health Sciences, Life and Health Sciences Research Institute, Campus de Gualtar, 4710-057 Braga, Portugal

² Universitat Internacional de Catalunya, School of Medicine and Health Sciences, Josep Trueta s/n 08165 Sant Cugat del Vallès, Barcelona, Spain

³ University of Barcelona, School of Medicine, Casanova 143, 08036 Barcelona, Spain

Corresponding author: Jorge Palés

Abstract

Background: Empathy is a key aspect of the physician-patient interactions. The Jefferson Scale of Empathy (JSE) is one of the most used empathy measures of medical students. The development of cross-cultural empathy studies depends on valid and reliable translations of the JSE. This study sought to: (1) adapt and assess the psychometric properties in Spanish students of the Spanish JSE validated in Mexican students; (2) test a second order latent factor model. **Methods:** The Spanish JSE was adapted from the Spanish JSE–S, resulting in a final version of the measure. A non-probabilistic sample of 1104 medical students of two Spanish medical schools completed a socio-demographic and the Spanish JSE-S. Descriptive statistics, along with a confirmatory factor analysis, the average variance extracted (AVE), Cronbach’s alphas and composite reliability (CR) coefficients were computed. An independent samples t-test was performed to access sex differences. **Results:** The Spanish JSE-S demonstrated acceptable to good sensitivity (individual items – except for item 2 – and JSE-S total score: $-2.72 < Sk < 0.35$ and $-0.77 < Ku < 7.85$), convergent validity (AVE: between 0.28 and 0.45) and reliability (Cronbach’s alphas: between 0.62 and 0.78; CR: between 0.62 and 0.87). The confirmatory factor analysis supported the three-factor solution and the second order latent factor model. **Conclusions:** The findings provide support for the sensitivity, construct validity and reliability of the adapted Spanish JSE-S with Spanish medical students. Data confirm the hypothesized second order latent factor model. This version may be useful in future research examining empathy in Spanish medical students, as well as in cross-cultural studies.

Keywords: Empathy; Medical students; Validity and reliability; Second order confirmatory factor analysis; Gender.

Background

Empathy has long been considered a key aspect of the therapeutic alliance, and of optimal care [1]. The concept was first introduced by Robert Vicher in 1872[2], and has merited the attention of clinicians and researchers. The crucial role of empathy in the patient-therapist relationship was pointed by Carl Rogers, who considered empathy as the ability to “perceive the internal frame of reference of another with accuracy *as if one were* the other person but without ever losing the ‘as if’ condition” (p. 210)[3]. Rogers underlined the cognitive dimension of empathy, and stressed that empathy is an indispensable condition to the self-actualization and personal growth of the patient.

Ever since Fine and Therrien [4] studied empathy in the context of physician-patient interactions, clinicians and researchers became increasingly interested in empathy in the context of patient care. (see Hojat, 2007 for a comprehensive review [5]). Empirical findings consistently associate empathy with improved accurateness and celerity of diagnosis, patient’s adherence to treatment, better quality of life and well-being [5–8]. The importance of empathy is thus, generally recognized [9] and international recommendations for medical education highlight the need for understanding and developing it in physicians and in medical students [10–12].

There are multiple and often contradictory definitions for the construct of empathy[9]. The inexistence of a consensual definition translates into the co-existence of more than 40 empathy measures [13], that reflect conceptions of the construct as predominantly cognitive [2,14–16], affective [17,18] or both [19,20]. Culture also influences the meanings that people impart to empathy [5,21,22]. The elucidation of cross-cultural differences and similarities related with empathy development during medical training and with the way students and clinicians conceive and manifest empathy would benefit from conceptual clarification of the construct and the application of valid and reliable empathy measures across countries.

One important contribution to the establishment of a widely accepted empathy definition and measure was given by Hojat and colleagues [2]. The authors developed an empathy self-report measure specifically designed to assess physicians' and medical students' attitude towards empathy in patient care [23]. The Jefferson Scale of Physician Empathy – Students version (JSE-S) reflects Hojat's and colleagues [2,5,23] definition of empathy as a predominantly cognitive attribute (as opposed to sympathy) involving the ability to understand the patient's perspective and inner experiences, and the capacity to communicate it. This definition settles in a tripartite view of the construct (attested by the factorial analysis of the JSE-S) comprising the ability to take the patient's perspective (perspective taking), to stand in the patient's shoes (standing in the patient's shoes), and of combining empathy with a sufficient degree of sympathy (compassionate care) [2,23,24]. The JSE-S is currently one of the most commonly used measures in research in medical education worldwide. The measure has proved to have adequate validity and reliability across multiple countries and languages [2,22,23,25–34]. The three factor solution found in the original version [2, 23] has been supported in subsequent studies with the original and the translated versions (*cf.* Table 1). Yet, different factor structures have also emerged. For example, the exploratory factor analysis yielded a five and four factors solution in the Japanese's and German's versions [22, 25], respectively.

[Insert Table 1]

Consistent with Hojat's et al. [2,5,23] definition of the construct, researchers using the JSE-S often report and compare the global score of the JSE-S over the three dimensional scores [22,35–41]. Nonetheless, factorial analysis yielding a more reasonable “correlated multi-factorial model” suggests that empathy is a multidimensional construct [42]. Thus, a total score of the JSE-S relies on the assumption that empathy is a latent second order concept that is manifested through the sub-dimensions of empathy yield by the factorial analysis. Previously reported moderate to strong statistically significant correlations between the three dimensions [28,33] reinforce this possibility [42,43]. Since a “correlated multi-factorial model” is not a measurement model *per se*, as there is

not a common target dimension (i.e. empathy) that directly affects items' variance[42], the test of a second-order model considering and supporting the use of the global JSE-S is needed. Such model, yet to be tested, “places a measurement structure onto the correlations among factors” (p. 4)[42], assuming that the scale's dimensions share a common cause (i.e. empathy) which explains their correlation.

One translation into Spanish of the JSE-S was developed and tested with Mexican students by Alcorta-Garza and colleagues [25]. The authors translated and back-translated into English the JSE-S questionnaire and assessed its psychometric properties in a sample of 1022 undergraduate medical students. Findings supported this version's reliability and construct validity. Alcorta-Garza and colleagues' version has been used to evaluate medical students' attitude towards empathy in other Spanish speaking countries, including a preliminary study conducted in Spain to assess the impact of a communication skills workshop [43]. Nonetheless, the psychometric properties of the JSE-S in Spanish medical students are unknown. Given Mexico and Spain's cultural differences, the adaptation and study of the psychometric properties of the JSE-S with Spanish students is essential to assure the validity and reliability of the measure in this population. Such a study is rather relevant to enable the rigorous development of empathy studies in Spain [30,44], and allow cross-cultural comparisons in medical education research, granting both generalizability of findings and investigation of differences within and between populations. As Portuguese and Italian versions of the JSE-S already exist [26,28,29], the availability of a Spanish version would address country and cultural specificities (e.g. South European *versus* Anglo-Saxon countries) in empathy in patient care interactions, and in empathy evolution during medical training.

The purpose of this study was to: (1) assess the psychometric properties of the Spanish version of the JSE-S in a sample of Spanish undergraduate medical students; (2) test a second order latent factor model for the global JSE-S. Based on previous literature on the validity and reliability of the JSE-S, we predicted that: (1) items would show adequate sensitivity; (2) a factor analysis of JSE-S would yield a three factor solution; (3) a second order confirmatory factor analysis would

confirm the existence of a second order latent factor; (4) the scale would present acceptable convergent validity; (5) the internal consistency and composite reliability for the JSE-S and for each sub-scale would be acceptable to good; and (6) empathy of female students would be higher than their male counterparts.

Methods

Participants

The population of Spanish medical students in the year 2014/2015 comprised 38765 students. Participants were undergraduate medical students enrolled at the University of Barcelona (a public university) and the International University of Catalonia (a private university) in the year 2014/2015. Inclusion criteria included: (1) being at least 18 years old or; (2) attending the first through sixth year of medical school; and (3) being willing to participate. Of a population of 1502 students attending the 1st through 6th year of medical training in the University of Barcelona and in the International University of Catalonia (65% women), 1104 students agreed to participate and were included in our sample. Our sample comprised 1024 participants, 689 of which were enrolled at the University of Barcelona and the remaining 415 attended the International University of Catalonia. Table 2 presents the sample's characteristics. Most students were female (68%) and age ranged from 18 to 44 ($M = 20.7$, $SD = 2.59$). Most participants were in their first through third year of medical school (61%), the pre-clinical period of medical training. Average response rate for the total sample was 74%, with response rate by school and year ranging from 45% (fifth year) to 100% (first year) in the International University of Catalonia (IUC), and 56% (fifth year) to 95% (first year) in the University of Barcelona (UB). The proportion of female and male participants in our sample (68% and 32%, respectively), although approximate to the proportion found in the population of Spanish medical students in the corresponding year (65% and 35%), was statistically significantly different ($\chi^2[1]=5.77$, $p=0.02$).

[Insert Table 2]

Measures

Participants were asked to provide basic demographic and academic information (sex, age, year of medical training, and university entrance score). Students also completed the adapted Spanish version of the JSE-S. The JSE-S is a 20-item self-report questionnaire assessing students' attitude towards empathy in the patient-care context. The original JSE-S comprises three domains: *Perspective Taking* (PT), *Compassionate Care* (CC), and *Standing in the Patient's Shoes* (STS). Participants are asked to report their degree of agreement with each item in a seven-point Likert-type scale, where 1 = "Strongly disagree" and 7 = "Strongly agree". Three partial scores (PT, CC and STS) and one total score may be computed (by the sum of its corresponding items), with higher scores (ranging from 20 to 140 for the total scale) reflecting higher attitude towards empathy.

Previous findings support the validity and reliability of the original and translated JSE-S [2,5,13,23,28,31]. Alcorta-Garza and colleagues tested their Spanish version of the JSE-S in a sample of Mexican medical students, showed adequate internal consistency ($\alpha=0.74$), and the exploratory factor analysis yielded a three factor structure [25].

Procedures

The items of the Alcorta-Garza and colleagues' Spanish version of JSE-S [24] were reviewed by a panel of four European Spanish native speakers, experts in medical education. Minor idiomatic adjustments were carried out in different items in order to correct the idiomatic differences between the idiom in Spain and Mexico. The adjustments were consensual. The Alcorta-Garza and colleagues' Spanish version of JSE-S and the adapted version used in this study are shown in Annex I. A non-probabilistic sample of participants was recruited between September 2014 and May 2015. Students meeting the inclusion criteria were invited to participate by one of the researchers in person at the end of scheduled class time at the beginning of the academic year (first year students), at the beginning of the second semester (second through fifth year students), or at the end of medical training (sixth

year students). Students were specifically informed of the study aims, that participation was voluntary and that responses would be kept anonymous and confidential. Students willing to participate provided oral consent and completed paper-and-pencil versions of the study measures. Students unwilling to participate left the room before the completion of the questionnaire and/or at any point of the questionnaire's completion. These students were excluded from the sample. There was no set time limit to answer the forms.

Research in medical education in our jurisdiction is exempted from formal approval from the university's Ethical Committee on the ground that this type of research does not have the purpose to answer a research question on health or biomedicine, does not imply any procedure or intervention that deserves the need for a formal ethical approval, and that the study followed the ethical guidelines regarding the collection of informed consent and anonymity of data processing, in accordance with the ethical Declaration of Helsinki. This study was confirmed as exempt from formal ethical approval by the Ethics review board of the University of Barcelona – Clinical Hospital Medical School Ethical Research Committee

Data Analysis. Descriptive statistics (means, standard deviations, medians, skewness and kurtosis) were used for the adapted version of the Spanish JSE-S and the individual items. Items sensitivity was assessed through skewness (Sk) and kurtosis (Ku) analysis, with absolute values higher than three and 10, respectively, indicating severe deviance from normal distribution of the items [36].

The hypothesized three-factor model for the JSE-S was tested through a confirmatory factor analysis (CFA). Model quality of fitness was assessed using the Chi Square (χ^2/df), Comparative Fit Index (CFI), Parsimony Comparative Fit Index (PCFI), Goodness of Fit Index (GFI), Parsimony Goodness of Fit Index (PGFI), and Root Mean Square Error of Approximation (RMSEA). The model was considered to have acceptable or good fit, respectively, if χ^2/df was less than 5 or 2 [45],

CFI was higher than 0.8 and 0.9 [46], GFI was higher than 0.9 or 0.95 [47], PCFI and PGFI were higher than 0.6 or 0.8 [48], and RMSEA was lower than 0.08 or 0.05 [47].

Convergent validity was assessed by computing the Average Variance Extracted (AVE)[41]. According to Hair and colleagues' reference values, AVE higher than 0.5 were suggestive of adequate convergent validity [49].

Given the moderate to strong association between factors found in previous research, and since the JSE total score is many times used in medical education research field, we tested a second order latent factor model, considering the global JSE-S [42,50]. The model's adjustment was performed step-by-step, through the analysis of correlation among errors, according to Modification Indices (MI) higher than 11 ($p < 0.001$) [51]. The Chi Square difference test and Expected Cross-Validation Index (MECVI) were computed to compare fit of the initial and final models after adjustments, with statistically significant Chi Square statistic and lower MECVI reflecting better fit [47].

Cronbach's alphas were computed for the total JSE-S and for the three subscales to assess internal consistency of the scale and its domains. Composite reliability (CR) was also determined [49,52]. Cronbach's alpha and CR higher than 0.6 and 0.7 were considered acceptable and good, respectively [53,54].

Finally, in order to detect interaction effects between gender and year of medical training, as well as gender and year of medical school main effects on empathy ratings, we computed a two-way analyses of variance (ANOVA), with JSE-S as the dependent variable, and gender and year of medical school as the independent variables. Prior to these analyses, we evaluated test assumptions, namely normality and homogeneity of variances, by analyzing Sk and Ku, with absolute values of Sk and Ku lower than three and 10 indicating absence of severe violation of normality assumption (55), and Levene's test, respectively. JSE-S total scores presented normal distributions for both men and women and for each year of medical training ($Sk < 1$ and $Ku < 1$), and results for the Levene's test showed no violation of the assumption of homogeneity of variances ($F_{(11,1091)}=1.79, p=0.052$).

In the event that a significant class (year) effect was found, we planned to perform between-temperature comparisons using post hoc Bonferroni tests.

Statistical analyses were computed using software IBM SPSS Statistics (v. 22) and AMOS statistical package (v. 21). Alpha was set at 0.05 for all analyses.

Results

Descriptive Information

Table 3 shows descriptive statistics for JSE-S total score and for JSE-S individual items for the total sample. The seven-point Likert-type scale was entirely used for all items of the questionnaire, with answers ranging from one to seven. With one exception (item 2), items present acceptable skewness (ranging between -2.71 and 0.35; mainly negatively skewed) and kurtosis (ranging between -0.77 and 7.85; mainly leptokurtic) values. The average scores for JSE-S items ranged between 3.67 ($SD=1.75$) for item 18 and 6.65 ($SD=0.72$) for item 2.

[Insert Table 3]

Construct validity: confirmatory factor analysis and convergent validity

All of the six combined fit indices for the CFA supported the three factor structure found for the original JSE-S. Three out of six fit indices indicated acceptable fit ($\chi^2/df = 3.34$; PCFI = 0.79; PGFI = 0.76), and the remaining three suggesting good model fit (CFI = 0.90; GFI = 0.95; RMSEA[HI95%] = 0.05[0.05]) (*cf.* table 4).

The standardized factorial weights and individual items reliability for the model are presented in Figure 1. Nine items showed loadings lower than the reference value of 0.50, indicating that less than 25% of the result of those items were explained by the latent dimension. Yet, 18 out of 20 items exhibited loadings higher than 0.25. Item 18 showed a particularly low saturation level ($\lambda_{ij}^2=0.07$).

[Insert Figure 1]

Convergent validity was assessed through AVE. For all three subscales and for total JSE-S the AVE was lower than 0.50. AVE ranged from 0.23 for *Compassionate Care* subscale to 0.45 for *Standing on the Patients Shoes* subscale (see Table 5).

[Insert Tables 4 and 5]

Second order latent factor model

The second order latent factor model considering the global JSE-S was tested. Since the number of parameters to estimate was the same as the above mentioned modified model, resulting in equal number of degrees of freedom (167), this model presented exactly the same combined fit indexes as the CFA model, suggesting acceptable to good fit.

The inspection of JSE-S items suggests that some items have similar content, as for example item 9 (“I try to imagine myself in my patients’ shoes when providing care to them”) and item 17 (“I try to think like my patients in order to render better care”). Based on the analysis of the modification indexes, specific error terms of these items were correlated, resulting in a new modified model that maintained all the items of the original scale (see Figure 2). Three out of six combined fit indexes suggest good fit of the final model (CFI = 0.93; GFI = 0.96; RMSEA[HI95%]=0.04[0.04]), while the remaining were suggestive of only acceptable fit ($\chi^2/\text{df} = 2.63$; PCFI = 0.82; PGFI = 0.76). The final model presented a goodness of fit higher than the initial one, better than the one found for the initial model ($\Delta\chi^2[1]=120.18$, $p<0.001$; MECVI: 0.59 vs. 0.48).

The standardized factorial weights and individual items reliability for the initial and final models are presented in Figure 2. The SPS first order latent variable presented a loading of 0.17, lower than 0.25, and items’ loadings were similar to those found in the CFA model presented above.

[Insert Figure 2]

Reliability: internal consistency and composite reliability

The internal consistencies (Cronbach's alphas) and composite reliability (CR) for the total JSE-S and its dimensions are listed in Table 4. The overall scale, as the subscales, evidenced acceptable to good reliability (JSE-s: $\alpha=0.78$, $CR=0.87$; TP: $\alpha=0.76$, $CR=0.79$; CC: $\alpha=0.62$, $CR=0.67$; SPS: $\alpha=0.62$, $CR=0.62$) [53,54]. Values if single items are deleted are comparable to the overall and subscales' alphas, suggesting that items did not detract from the reliability of the measure, with, maybe, the exception of item 18.

Gender and year of medical training comparison

Students' global scores in the JSE-S ranged between 50 and 140 ($M=117.56$, $SD=10.43$), with average scores ranging from 113.34 ($SD=12.55$) and 120.06 ($SD=8.46$) for male students attending the third and sixth years respectively, and 115.8 ($SD=11.48$) and 119.88 ($SD=9.41$) for female students attending the fifth and the second years respectively (see Table 6). No significant interaction effects were found between gender and class ($F_{(5, 1091)} = 1.78$, $p = 0.119$, $\eta^2_p = 0.008$, $\pi=0.61$). Significant sex main effects were obtained ($F_{(5, 1091)} = 3.73$, $p = 0.002$, $\eta^2_p = 0.017$, $\pi=0.94$), with females ($M = 118.6$, $SD = 9.64$) tending to score higher than males ($M=115.4$, $SD= 1.69$). There were also significant year of medical training main effects ($F_{(1, 1091)} = 15.88$, $p<0.001$, $\eta^2_p=0.014$, $\pi = 0.98$), sixth year students show higher empathy levels than first and fifth year students ($p's < 0.05$)

[Insert Table 6]

Discussion

The results suggest that the sensitivity, construct validity and reliability of the Spanish JSE-S were acceptable. The convergent validity and individual item sensitivity (item 2) and reliability (item 18) were limited. Even so, findings support the use of the Spanish JSE-S with Spanish

medical students. Considering that previous studies supported the validity and reliability of the JSE-S, this measure may be used in cross-cultural studies on medical students' empathy.

The psychometric sensitivity of the scale and of most items was acceptable. Consistent with previous research in Portugal and Italy [26,28], skewness and kurtosis absolute values for the JSE and for individual items were in the range proposed by Kline [55], except for item 2 (*"Patients feel better when their physicians understand their feelings"*). In fact more than 50% of the participants strongly agreed with the item. The ceiling effect is understandable considering the item's content, as it is reasonable to expect that most people would be more comfortable whenever their feelings are comprehended by others. Item's 2 lack of sensitivity was also reported for the Portuguese JSE [26]. The item's lack of sensitivity, while explicable, redounds in its lower relevance.

The confirmatory factor analysis corroborated that the three-factor structure proposed by the authors of the original version has an adequate fit. The results for item reliability revealed that the factor regression weights for some factors were acceptable and within the range of previous findings [5,26,28,56]. However, these loadings were lower than those of the original JSE. Item 18 showed particularly low and non-significant saturation level, consistent with previous results found for the Portuguese (Portugal and Brazil), Italian, Spanish (Mexico) and Chinese versions [25,26,28–30,40], and also in a recent study assessing the factor structure of the JSE-S in the USA [24]. Differences in data analysis - confirmatory factor analysis (a reflective model) *versus* principal component analysis (a formative model) - might have contributed to these differences [26]. However, other reports on several versions of the measure have identified problematic items (e.g. item 18) suggesting that cross-cultural research would benefit from a modified JSE.

AVEs were lower than the reference values proposed by Hair and colleagues [49], suggesting that the scale has limited convergent validity. The Portuguese version presented AVEs comparable to the ones found in our sample [26]. Such finding supports the use of the JSE-S total score instead of the measure's partial ones.

As hypothesized, the Spanish JSE-S and its dimensions showed acceptable to good internal consistency and composite reliability [53,54], on the range of those found in other translated versions of the measure ($0.74 < \alpha < 0.83$) [22,25–29,31,33,40]. Yet, the Cronbach's alpha of CC if item 18 is deleted is higher than the internal consistency of this dimension, suggesting that this item detracts from the reliability of the subscale. Hence, the eventual elimination of items 2 and 18 could contribute to the improvement of Spanish JSE-S's psychometric properties, suggesting the convenience to continue the study of the Spanish JSE-S. Both items would benefit from some degree of revision in the near future. While item 2 is more comprehensive, item 18 might have different interpretations and its reformulation needs to be considered. In order to enable future cross-country research, we would recommend the preservation of the original structure of the JSE. Nonetheless, as the structure of the scale is, from the beginning of the original JSE-S development, somewhat unbalanced (the number of items *per* dimension is heterogeneous and two dimensions present only inverted items), only modest internal consistency and construct reliability are, in fact, reasonable to expect.

The tested second order latent factor model presented acceptable to good fit. *Perspective Taking* and *Compassionate Care*, with high regression weights on the second order latent variable, contributed equally and largely than *Standing in the Patient's Shoes* to explain the construct of empathy. These results are consistent with the weak inter-scale correlation coefficients found in the confirmatory factor analysis of the correlated multi-factorial model between *Standing in the Patient's Shoes* with the other two factors. Consequently, our results provide limited support to the use of the Spanish JSE-S total score that assures that empathy is a latent (second order) concept that is manifested through *Perspective Taking*, *Compassionate Care* and *Standing in the Patient's Shoes*. Such weak correlations, while in line with the with the results for the Portuguese JSE-S [26], are inconsistent with moderate to strong inter-scale correlations found in the Italian and English versions [28,33]. Hence, our results support the use of the scores for the three dimensions of the Spanish JSE-S over its total score in empathy research in medical education.

As for most of the previous studies using the JSE-S worldwide [28,36,38,41,57], female students reported significantly higher empathy than their male counterparts, suggesting this version's ability to detect differences between individuals. Nonetheless, non-statistically significant results have also emerged [37,58,59].

Taken together, our findings and previous results from other translated versions, suggest that the validity and reliability of the JSE-S generalize across languages and cultures. Nonetheless, our findings are consistent with previous finding of limited convergent validity [26], weak inter-scale correlation coefficients [24,26], item 2 lack of sensitivity [26] and item 18 low saturation level [24–26,28–30]. Such psychometrical limitations reinforce the need to engage in cross-cultural studies comparing: (1) at least, South European countries versions with the Anglo-Saxon countries, and (2) the definition and relevance attributed to the construct of empathy itself across-cultures.

Limitations

There are a number of limitations that should be taken into account when interpreting the results. First, the cross-sectional design did not allow examination of test-retest reliability and sensitivity of the measure to change. Longitudinal studies are needed to clarify the Spanish JSE-S stability over time, and to assess its ability to detect changes in empathy as a result of interventions. The second relevant limitation regards the generalizability of findings. All students were recruited in only two medical schools in Catalonia Community, and sample was non-probabilistic. The authors are not able to determine how representative the sample is of the population of Spanish undergraduate medical students, as the composition of the sample does not take into account the possible differences between students of different regions of Spain. Third, we did not administer other empathy self-report, patient-report and other-report measures, which would help to further establish convergent validity of the measure. Further research addressing this gap would help to determine the extent of overlap between the adapted version of the Spanish JSE-S and other second and third-person empathy measures.

Conclusions

The present study is the first, to our knowledge, to assess the Spanish JSE-S psychometric properties in a sample of Spanish medical students. Our findings provide support for the validity and reliability of the adapted version of the Spanish JSE-S with Spanish medical students, confirm the structural validity of the three-factor model, the scale's satisfactory reliability and ability to discriminate inter-individual differences. Thus, this version may be useful to understand the evolution of empathy in Spanish medical students, as well as in cross-cultural research examining similarities and differences in empathy growth in students from Spain and other countries. Findings provide limited support for the existence of a second order latent factor in the Spanish JSE-S. Based on our findings, it is recommendable that the use of the scores for the three sub-scales of the Spanish JSE-S should prevail over the JSE-S total score.

Abbreviation

AVE: Average Variance Explained

CC: Compassionate Care

CFA: Confirmatory Factor Analysis

CFI: Comparative Fit Index

CR: Composite Reliability

GFI: Goodness of Fit Index

HI: Confidence Interval Higher Limit

I.C.: Confidence Interval

IUC: International University of Catalonia

JSE-S: The Student Version of the Jefferson Scale of Physician Empathy

MECVI: Expected Cross-Validation Index

MI: Modification Indices

PT: Perspective Taking

PCFI: Parsimony Comparative Fit Index

PGFI: Parsimony Goodness of Fit Index

χ^2/df : Chi Square

RMSEA: Root Mean Square Error of Approximation

STS: Standing on the Patients Shoes

UB: University of Barcelona

Declarations

Ethics approval and consent to participate

Research in medical education in our jurisdiction is exempted from formal approval from the university's Ethical Committee on the ground that this type of research does not have the purpose to answer a research question on health or biomedicine, does not imply any procedure or intervention that deserves the need for a formal ethical approval, and that the study followed the ethical guidelines regarding the collection of informed consent and anonymity of data processing, in accordance with the ethical Declaration of Helsinki. This study was confirmed as exempt from formal ethical approval by the Ethics review board of the University of Barcelona – Clinical Hospital Medical School Ethical Research Committee

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

JP, ME and MV designed the study, and administered the surveys. MJC participated in the design of the study. AFV and PC performed the statistical analysis and wrote the first draft of the manuscript. AFV, MJC, JP, ME and MV participated in the discussion of the implications of the results for the Spanish context and future research. All authors critically reviewed the manuscript and gave final approval for the text of the manuscript.

Acknowledgements

The authors thank Jefferson Medical College for permission to adapt the original JSE-S and Adelina Alcorta-Garza for permission to use the original Spanish version. The authors thank all students who participated in this study. **Authors' information**

AFV (PhD) is Post-Doctoral fellow at the pilot line of research in medical education at the Life and Health Sciences Research Institute of the University of Minho, and Assistant Professor at the University Lusíada of Lisbon.

PC (PhD) is Assistant Professor at the Health Sciences School of the University of Minho. Researcher at the pilot line of research in medical education at the Life and Health Sciences Research Institute of the University of Minho.

ME (MD) is Professor and Dean-associated at the Medical School in the Universitat Internacional de Catalunya.

MV (MD, PhD) is Professor of communication skills in clinical practice, Head of the Unit of laboratory and clinical skills simulation.

MJC (PhD) is Associate Professor and Head of the Medical Education Unit at the School Health Sciences of the University of Minho. Researcher at the pilot line of research in medical education at the Life and Health Sciences Research Institute of the University of Minho.

JP (MD, PhD) is Professor and Head of the Medical Education Unit at the Medical School of the University of Barcelona.

References

1. Colliver JA, Conlee MJ, Verhulst SJ, Dorsey JK. Reports of the decline of empathy during medical education are greatly exaggerated: A reexamination of the research. *Acad Med.* 2010;85(4):588–93.
2. Hojat M, Mangione S, Nasca TJ, Cohen MJM, Gonnella JS, Erdmann JB, et al. The Jefferson Scale of Physician Empathy: Development and Preliminary Psychometric Data. *Educ Psychol Meas.* 2001 Apr 1;61(2):349–65.
3. Rogers, C. A theory of therapy: Personality and interpersonal relationships as developed in the client-centered framework. In: S Koch (Ed) *Psychology, a study of science: Foundations of the person and the social context.* New York, NY: McGraw-Hill; 1959. p. 184–256.
4. Fine VK, Therrien ME. Empathy in the doctor-patient relationship: skill training for medical students. *J Med Educ.* 1977 Sep;52(9):752–7.
5. Hojat, M. *Empathy in patient care: antecedents, development, measurement, and outcomes.* Springer; 2007.
6. Del Canale S, Louis DZ, Maio V, Wang X, Rossi G, Hojat M, Gonnella JS. 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;97(9):1243–9.

7. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians' empathy and clinical outcomes for diabetic patients. *Acad Med J Assoc Am Med Coll*. 2011 Mar;86(3):359–64.
8. Pollak, K., Ostbye, T., Alexander, S., Gradison, M., Bastian. L., Brouwer. R., & Lyna. P. Empathy goes a long way in weight loss discussions. Female patients are more likely to step up weight loss efforts when a physician shows empathy and offers support. *J Fam Pract*. 2007;56(12):1031–6.
9. Pedersen R. Empathy: A wolf in sheep's clothing? *Med Health Care Philos*. 2008 Sep;11(3):325–35.
10. Association of American Medical Colleges. Contemporary Issues in Medicine: Communication in Medicine [Internet]. Washington, DC; 1999 [cited 2015 Jun 29]. Report No.: Report III. Available from: https://www.aamc.org/download/130608/data/clinicalskills_oct09.qxd.pdf.pdf
11. Association of American Medical Colleges. Recommendations for preclerkship clinical skills education for undergraduate medical education [Internet]. 2008 [cited 2015 Jun 29]. Available from: https://www.aamc.org/download/130608/data/clinicalskills_oct09.qxd.pdf.pdf
12. General Medical Council. Education and training standards consultation [Internet]. 2014 [cited 2015 Jun 29]. Available from: http://www.gmc-uk.org/16___Education_and_training_standards_consultation.pdf_58923530.pdf
13. Hemmerdinger JM, Stoddart SD, Lilford RJ. A systematic review of tests of empathy in medicine. *BMC Med Educ*. 2007;7(1):24.
14. Hogan, R. Development of an empathy scale. *J Consult Clin Psychol*. 1969;33(3):307–16.
15. Hojat M, Gonnella JS. Eleven Years of Data on the Jefferson Scale of Empathy-Medical Student Version (JSE-S): Proxy Norm Data and Tentative Cutoff Scores. *Med Princ Pract*. 2015 Apr 28;24(4):344–50.

16. Wispé L. The distinction between sympathy and empathy: To call forth a concept, a word is needed. *J Pers Soc Psychol.* 1986;50(2):314–21.
17. Hodges, S.D., Myers, M. Empathy. In: In R F Baumeister and K D Vohs (Eds) *Encyclopedia of social psychology.* Thousand Oaks, CA; 2007.
18. Schafer R. Generative empathy in the treatment situation. *Psychoanal Q.* 1959 Jul;28:342–73.
19. Bennett, M. J. *The empathic healer: an endangered species.* San Diego, CA: Academic Press; 2001.
20. Butters RP. A meta-analysis of empathy training programs for client populations [Internet]. The University of Utah; 2010 [cited 2015 Jun 25]. Available from:
<http://content.lib.utah.edu/utils/getfile/collection/etd2/id/321/filename/755.pdf>
21. Ferreira-Valente, A., Monteiro, J., Barbosa, R., Salgueira, A., Costa, P., Costa, M.J.
Reconsidering declines of empathy during medical education: a scoping review. Submitted;
22. 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):1192–7.
23. 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 Jan;27(7):625–8.
24. Hojat M, LaNoue M. Exploration and confirmation of the latent variable structure of the Jefferson scale of empathy. *Int J Med Educ.* 2014;5:73–81.
25. Alcorta-Garza A, González-Guerrero JF, Tavitas-Herrera SE, Rodríguez-Lara FJ, Hojat M. Validación de la escala de empatía médica de Jefferson en estudiantes de medicina mexicanos. *Salud Ment.* 2005;28(5):57–63.
26. Costa, M.J., Ferreira-Valente, A., Costa, P. Escala de Empatía para Médicos – Estudiantes (JSPE-spv): Jefferson Scale of Physician Empathy student version (JSPE-S). In: Almeida,

Leandro S; Simões, M R; Machado, C; Gonçalves, M M eds Avaliação psicológica: Instrumentos validados para a população portuguesa. in press.

27. Jumroonrojana K, Zartrungpak S. Development of the Jefferson Scale of Physician Empathy-Student Version (Thai Version). *J Psychiatr Assoc Thail.* 57(2):213–24.
28. Leombruni P, Di Lillo M, Miniotti M, Picardi A, Alessandri G, Sica C, et al. Measurement properties and confirmatory factor analysis of the Jefferson Scale of Empathy in Italian medical students. *Perspect Med Educ.* 2014 Dec;3(6):419–30.
29. 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.
30. 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.
31. 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). *Adv Health Sci Educ.* 2013 Oct;18(4):573–87.
32. 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 Nov;32(11):e471–8.
33. Tavakol S, Dennick R, Tavakol M. Psychometric properties and confirmatory factor analysis of the Jefferson Scale of Physician Empathy. *BMC Med Educ.* 2011;11(1):54.
34. 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.
35. Chen DCR, Kirshenbaum DS, Yan J, Kirshenbaum E, Aseltine RH. Characterizing changes in student empathy throughout medical school. *Med Teach.* 2012 Apr;34(4):305–11.

36. Costa P, Magalhães E, Costa MJ. A latent growth model suggests that empathy of medical students does not decline over time. *Adv Health Sci Educ*. 2013 Aug;18(3):509–22.
37. Lim BT, Moriarty H, Huthwaite M, Gray L, Pullon S, Gallagher P. How well do medical students rate and communicate clinical empathy? *Med Teach*. 2013 Feb;35(2):e946–51.
38. Hojat M, Vergare MJ, Maxwell K, Brainard G, Herrine SK, Isenberg GA, et al. The devil is in the third year: a longitudinal study of erosion of empathy in medical school. *Acad Med*. 2009;84(9):1182–91.
39. Hojat M, Michalec B, Veloski JJ, Tykocinski ML. Can Empathy, Other Personality Attributes, and Level of Positive Social Influence in Medical School Identify Potential Leaders in Medicine?: *Acad Med*. 2015 Apr;90(4):505–10.
40. 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.
41. Youssef FF, Nunes P, Sa B, Williams S. An exploration of changes in cognitive and emotional empathy among medical students in the Caribbean. *Int J Med Educ*. 2014 Sep 24;5:185–92.
42. Reise SP, Moore TM, Haviland MG. Bifactor Models and Rotations: Exploring the Extent to which Multidimensional Data Yield Univocal Scale Scores. *J Pers Assess*. 2010 Nov;92(6):544–59.
43. Fernández-Olano, C., Montoya-Fernandez, J., Salinas-Sanchez, A. Impact of a clinical interview training on the empathy level of medical students and medical residentes. *Med Teach*. 2008;30(3):322–4.
44. Esquerda M, Yuguero O, Viñas J, Pifarré J. La empatía médica, ¿nace o se hace? Evolución de la empatía en estudiantes de medicina. *Aten Primaria* [Internet]. 2015 May [cited 2015 Jun 26]; Available from: <http://linkinghub.elsevier.com/retrieve/pii/S021265671500058X>
45. Wheaton B. Assessment of Fit in Overidentified Models with Latent Variables. *Sociol Methods Res*. 1987 Aug 1;16(1):118–54.

46. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990 Mar;107(2):238–46.
47. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods Psychol Res Online.* 2003;8(2):23–74.
48. Blunch N. Introduction to structural equation modelling using SPSS and AMOS. London, UK: SAGE; 2008.
49. Hair, J., Anderson, R., Tatham, R., & Black, W. Multivariate data analysis. 5th ed. New York, NY: Prentice Hall; 1998.
50. Bollen KA. Structural equations with latent variables. New York, NY: Wiley; 1989.
51. Arbuckle, J. L. Amos 17 users' guide. Chicago, IL: SPSS; 2008.
52. Fornell, C., & Larcker, D. F. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J Mark Res JMR.* 1981;18(1):39–50.
53. George, D., Mallery, P. SPSS for Windows step by step: A simple guide and reference. 11.0 update. 4th ed. Boston: Allyn & Bacon; 2003.
54. Kline, P. The handbook of psychological testing. 2nd ed. London: Routledge; 2000.
55. Kline, R. Principles and Practice of Structural Equation Modeling. New York, NY: Guildford; 1998.
56. Shariat SV, Habibi M. Empathy in Iranian medical students: Measurement model of the Jefferson Scale of Empathy. *Med Teach.* 2013 Jan;35(1):e913–8.
57. Quince TA, Parker RA, Wood DF, Benson JA. Stability of empathy among undergraduate medical students: A longitudinal study at one UK medical school. *BMC Med Educ.* 2011;11(1):90.
58. 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):1198–202.

59. Shariat, S.V., Keikhavani, L. Empathy in residents of clinical specialties of Iran University of Medical Sciences. *Iran J Psychiatry Clin Psychol.* 2010;16:248–56.