

1 **Title Page**

2 **Measuring Medical Students' Empathy: Exploring the Underlying Constructs of**  
3 **and Associations Between, Two Widely Used Self-Report Instruments in Five**  
4 **Countries.**

5 Patrício Costa, PhD, Marco Carvalho-Filho, MD, PhD, Marcelo Schweller, MD, PhD,  
6 Pia Thiemann, PsyD, Ana Salgueira, PsyD, John Benson, MD, MRCP, FRCGP,  
7 Manuel João Costa, PhD, Thelma Quince, PhD.

8 Patrício Costa is Assistant Professor at the Health Sciences School of the University of  
9 Minho and researcher in medical education at the Life and Health Sciences Research  
10 Institute of the University of Minho, Portugal.

11 Marco Antonio de Carvalho-Filho is Professor of Medicine, Department of Emergency  
12 Medicine, Faculty of Medical Sciences, State University of Campinas (Unicamp),  
13 Campinas, Brazil.

14 Marcelo Schweller is an Emergency Medicine Physician and Preceptor, in Faculty of  
15 Medical Sciences, University of Campinas (Unicamp), Campinas, Brazil.

16 Pia Thiemann is Research Assistant in the Palliative Care Group, Primary Care Unit,  
17 Department of Public Health and Primary Care, University of Cambridge, UK.

18 Ana Salgueira is a PhD student at the Health Sciences School of the University of  
19 Minho and researcher in medical education at the Life and Health Sciences Research  
20 Institute of the University of Minho, Portugal.

21 John Benson is a Senior Lecturer in General Practice and Director of the General  
22 Practice Education Group, Primary Care Unit, Department of Public Health and Primary  
23 Care, University of Cambridge, UK.

24 Manuel Costa is Associate Professor at the Health Sciences School of the University of  
25 Minho and researcher in medical education at the Life and Health Sciences Research  
26 Institute of the University of Minho, Portugal.

27 Thelma Quince is Medical Educational Research Associate in the Primary Care Unit,  
28 Department of Public Health and Primary Care and the School of Clinical Medicine,  
29 University of Cambridge, UK.

30 Corresponding author: Dr. Thelma Quince,

31 The Primary Care Unit, Department of Public Health and Primary Care, University of  
32 Cambridge, IPH, Forvie Site, Robinson Way, Cambridge CB2 0SR. UK

33 [taq1000@medschl.cam.ac.uk](mailto:taq1000@medschl.cam.ac.uk)

34 Tele: 0044(0)1223 330364 Fax: 0044(0) 1223 762512

35 Word Count 3047

36

## 37 **Abstract**

38 **Purpose:** The definition of clinical empathy is unclear and evidence about its  
39 development among undergraduate medical students conflicting. These problems may  
40 stem from the instruments used to measure medical students' empathy. The authors  
41 sought to enhance understanding of the underlying constructs of two of the most widely  
42 used self-report instruments: Davis's Interpersonal Reactivity Index (IRI) and Jefferson  
43 Scale of Empathy (Student version) JSE (S), and of the distinctions and associations  
44 between these instruments.

45 **Methods:** IRI and JSE-S were administered in three separate studies between 2007 and  
46 2014, in 5 countries (Brazil, Ireland, New Zealand, Portugal, United Kingdom). Data  
47 from 3069 undergraduate medical students were collected. Exploratory factor analyses,  
48 correlation analyses and multiple linear regression analyses were performed.

49 **Results:** Exploratory Factor Analysis yielded identical results in each country,  
50 confirming the subscale structures of each instrument. Results of correlation analyses  
51 indicated significant but weak correlations ( $r=.313$ ) between the total IRI and JSE-S  
52 scores. All inter-correlations of IRI and JSE-S subscale scores were statistically  
53 significant but also weak (range  $r=-.040$  -  $r=.306$ ). Multiple linear regression models  
54 revealed that the IRI subscales were weak predictors of all JSE-S subscale and total  
55 scores. The IRI subscales explained between 8.9% and 15.3% of variance for JSE-S  
56 subscales and 19.4% for JSE total score.

57 **Conclusions:** The IRI and JSE-S are only weakly related, suggesting that they measure  
58 different constructs. Research into, and interventions addressing medical student  
59 empathy need clearer understanding and definition of the construct under consideration  
60 as results from the two scales are not comparable.

## 61 **Introduction**

62 Empathy is a core element in patient care. It may enhance patients' satisfaction and  
63 trust, so facilitating compliance and adherence to therapy.<sup>1-4</sup> Receipt of empathy may  
64 be therapeutic in its own right.<sup>5-7</sup> Greater trust by the patient may encourage better  
65 exchange of information in consultations, enabling better diagnosis and shared decision  
66 making.<sup>1,4,8</sup> From the doctor's perspective, empathy may lead to better clinical  
67 decisions,<sup>9-11</sup> greater job satisfaction, and enhanced psychological well-being.<sup>10,12-15</sup>  
68 The development of empathy among medical students would seem crucial to future  
69 patient care.

70 However "clinical empathy" is poorly defined and measured.<sup>9,16-18</sup> It has been seen as  
71 the ability to:

- 72 1. understand the patient's situation, perspective and feelings (and their attached  
73 meanings)
- 74 2. communicate that understanding and check its accuracy
- 75 3. act on that understanding with the patient in a helpful (therapeutic) way.<sup>19</sup>

76 This definition implies a multi-dimensional construct incorporating affective, cognitive,  
77 behavioural and moral components.

78 For patients it is the empathetic behaviour they receive which is important. However  
79 asking patients to assess medical students' empathy is problematic and studies using  
80 simulated or standardised patients have produced mixed results.<sup>20-22</sup> Most studies of  
81 medical student empathy rely on self-report measures, rather than direct observations.<sup>18</sup>  
82 The most widely used instruments are Davis's Interpersonal Reactivity Index (IRI) and  
83 the Jefferson Scale of Empathy (Student version) JSE- S.<sup>23,24</sup>

84 Davis considered empathy to be a set of related constructs, concerning responsivity to  
85 others, but each discriminable from each other.<sup>23</sup> The IRI comprises 28 items (9  
86 negative) forming four, 7 item, subscales: Perspective Taking (IRI-PT) assessing  
87 consideration for the psychological point of view of the other person; Empathetic  
88 Concern (IRI-EC) assesses consideration for their feelings and concerns, Personal  
89 Distress (IRI-PD) assessing personal anxiety in tense interpersonal settings and Fantasy  
90 Scale (IRI-FS) assessing tendencies to transpose oneself imaginatively into the feelings  
91 and actions of fictional characters.<sup>23</sup> IRI-EC and IRI-PT have be seen as “other-  
92 oriented” and IRI-PD and IRI-FS as “self-oriented”.<sup>25,26</sup> IRI-EC and IRI-PD relate to  
93 affective aspects while IRI-PT and IRI-FS to cognitive aspects. Respondents rate the  
94 extent to which statements apply, from “Does not describe me very well” to “Describes  
95 me very well” on a 5-point Likert scale.<sup>23</sup>

96 The IRI has been used in a wide variety of contexts including neurological studies,<sup>27</sup>  
97 clinical conditions<sup>28,29</sup> and criminology.<sup>30</sup> It has been found to have good psychometric  
98 properties and is regarded as a valid, and reliable instrument for measuring empathy.<sup>31</sup>  
99 Although less well used with medical students the factorial structure proposed by Davis  
100 has been supported in studies among college students.<sup>32</sup>

101 The JSE was developed as measure of empathy applicable to patient care.<sup>24</sup> This 20  
102 item scale comprises three underlying factors: Perspective Taking (10 positively  
103 worded), Compassionate Care (8 negatively worded) and Standing in the Patient’s  
104 Shoes (2 negatively worded). Most studies of medical student empathy report only the  
105 total JSE-S score.<sup>33</sup> Respondents rate their level of agreement with each statement on  
106 an ascending 7-point Likert scale (1 to 7). Used in a variety of cultural settings for  
107 assessing the empathy of medical students, nurses and other healthcare students its  
108 validity and reliability have been well supported<sup>18,34-38</sup>

109 Among healthcare students and practitioners the IRI subscales commonly used are the  
110 “other oriented” scales of IRI-EC and IRI-PT.<sup>39,40</sup> The distinction between cognitive  
111 and affective components is less clear in the JSE with both “Standing in the patient’s  
112 shoes” and “Perspective Taking” appearing to reflect the cognitive component of  
113 empathy.<sup>41</sup>

114 The IRI and JSE-S were conceived with different populations in mind. The IRI is  
115 applicable to the general population and seen to reflect generic or dispositional  
116 empathy.<sup>23</sup> The JSE is applicable to those engaged in healthcare and hence seen to  
117 measure empathy specific to that context.<sup>24</sup> Studies of undergraduate medical students  
118 in different countries using the IRI have shown that they fall within the norms for IRI-  
119 EC and IRI-PT.<sup>42</sup> It would seem reasonable to expect at least a moderate associations  
120 between some of the IRI and JSE-S subscales. Further, a study of medical students  
121 found a moderate correlation between the total scores of the JSE-S and IRI ( $r = .45$ ,  
122  $p < .01$ ).<sup>44</sup> However, unlike the JSE-S, the IRI subscales are not normally summed to a  
123 total score.<sup>18</sup>

124 To consider the underlying structural and conceptual differences of the IRI and JSE this  
125 study asked:

- 126 1] Whether the underlying factorial structures of the IRI and JSE-S reflected the  
127 dimensional constructs of empathy indicated by their respective subscales:
- 128 2] How the scales related to each other in terms of their total and subscale  
129 scores:
- 130 3] Whether scores on the IRI as a generic measure of empathy predicted scores  
131 on the JSE-S as a measure of empathy specific for healthcare.

## 132 **Method**

133 Data were obtained from three separate studies.

134 Study one included data from 16 UK medical schools, one in Ireland and one in New  
135 Zealand. All students beginning and all students approaching the end of, their  
136 undergraduate medical education were invited by email to participate in an international  
137 comparison. An online questionnaire survey took place between September 2013 and  
138 July 2014, and examined empathy, (IRI and JSE-S), psychological wellbeing, death  
139 anxiety and attitudes towards end of life care. Overall ethical approval was granted by  
140 the Psychology Research Ethics Committee of the University of Cambridge and by the  
141 relevant bodies in each participating school.

142 Study two was based in one Portuguese University. For each year between 2007 and  
143 2014 students beginning their undergraduate medical education were invited to  
144 complete a paper questionnaire covering the JSE-S. In January and February 2013  
145 students in all years were invited to complete a paper questionnaire covering the IRI.  
146 Data collection and storage were authorized by the Portuguese Commission for Data  
147 Protection (CNDP: 10432/2011). Retrospective approval was obtained: - Subcomissão  
148 de ética para as Ciências da Vida, process SECVS - 071/2013.

149 Study three was undertaken in one university in Brazil. In 2011 and 2012, all  
150 undergraduate medical students in years 4 and 6 were invited to complete paper  
151 questionnaires covering both the IRI and JSE-S. Ethical approval was granted by the  
152 Research Ethics Committee in Human Beings at the Faculty of Medical Sciences of  
153 Unicamp.

154 In all studies participants gave prior consent either in writing or online, and  
155 participation was voluntary and anonymous with no incentives offered.

## 156 **Participants:**

157 **Table 1.** Sample characterization (Sex, Country and Entry Scheme)

158

159 Medical schools in the studies offered “standard” courses lasting 5/6 years, with  
160 students typically aged 18 or 19 on entry. Some schools also offered 4 year accelerated  
161 “graduate entry” courses for students typically aged 21 or over on entry who had  
162 obtained a first degree.

163 The timing and balance of biomedical science and clinical course components in the  
164 participating schools varied. Some schools devoted the early years largely to biomedical  
165 sciences, others adopted a more integrated approach. This study did not set out examine  
166 in detail the nature of the courses offered and simple labels such as “integrated” may not  
167 fully represent course content and structure.

168 The sample comprised 3,069 medical students (Table 1) of whom 2059 (67.1%) were  
169 from the UK and 1887 (61.5%) were female. The majority of students (2619, or 85.3%),  
170 had entered standard courses. A statistically significant, but small in terms of effect size,  
171 difference in gender composition of samples in each country was found, with  
172 proportionately fewer males among the Portuguese sample and proportionately more  
173 males among the Brazilian sample. ( $\chi^2_{(4, n=3069)} = 9.6, p = .047, \text{Cramer's } V = .056$ ).

## 174 **Instruments:**

175 We used the JSE-S (student version) in all countries. The IRI Portuguese version is 24  
176 items as opposed to 28 and was the result of a validation study which demonstrated

177 factor loadings  $<.35$  for items numbered 1, 15, 18 and high standardized residual for  
178 item 10. <sup>44</sup> We adjusted the item numbers of the 28 item IRI used in other countries to  
179 those of the Portuguese version to allow IRI data to be merged.

## 180 **Data analysis and modelling strategy:**

181 We merged the JSE and IRI items and converted them into in the same scale using z  
182 scores. We used exploratory factor analysis (EFA) to explore the factorial structure of  
183 all IRI and JSE items, using the scree plot, the Kaiser's eigenvalue  $> 1$  method and  
184 Parallel Analysis (PA) to explore the optimal number of factors and principal axis  
185 factoring (PAF) with oblimin rotation. We used Cronbach's Alpha to measure internal  
186 consistency and Pearson correlation coefficient to examine associations between  
187 subscale and total scores of each scale. We examined the effects of country and sex on  
188 subscales scores using MANOVA and on the total JSE-S scores using ANOVA. (The  
189 results of these are presented in the appendix.) We used multiple linear regression to  
190 examine the extent to which IRI scores predicted JSE scores (total and subscale) with  
191 IRI subscales, country, sex and entry scheme being independent variables. Data  
192 analyses were performed using IBM SPSS Statistics v22 and the R.Commander <sup>45</sup> and  
193 the *psych* package. <sup>46</sup> We considered P values of 5% as significant and interpreted  
194 effect sizes according to values given by Cohen(1988). <sup>47</sup>

## 195 **Results**

### 196 **Latent dimensions of the IRI and JSE: (Table 2)**

197 **Table 2.** Exploratory Factor Analysis (communalities and loadings) for IRI and JSE  
198 items and Cronbach's Alpha scores

199 An EFA performed on the combined IRI and JSE-S datasets z-scores, resulted a **nine**  
200 factor solution according to the Kaiser's eigenvalue  $>1$  method and an **eight** factor

201 solution according to the PA and scree plot analysis (the line straightens after the eighth  
202 factor). Both solutions produced some dimensions with critical internal consistency  
203 values (Cronbach's alphas lower than 0.60). Therefore, the theoretically anticipated  
204 solution of **seven** factors accounting for 44.6% of variance was tested and led to higher  
205 and more acceptable internal consistency values. For the final EFA **seven** factors  
206 solution, measures of appropriateness of factor analysis were checked including KMO =  
207 .873 and Bartlett's test ( $\chi^2_{(946)} = 33016, p < .001$ ).

208 Considering the theoretical structure, a practical significance of 5% and an acceptable  
209 factor loading of  $\geq 0.224$  were found for all item. All items clustered as expected and  
210 recorded the highest loading on their original dimension with the exception of JSE-S  
211 item 14, which loaded higher onto JSE-PT than onto its original JSE-CC dimension.  
212 Nine items showed significant double loadings, but none crossed the two scales. In each  
213 of the five countries the seven factor structure revealed a satisfactory fit, (Table 2) with  
214 the exception for JSE-SPS dimension in Ireland (Cronbach's alpha=0.472).

### 215 **Pearson correlations for all IRI and JSE subscales: (Table 3)**

216 **Table 3.** Pearson correlations for IRI and JSE subscales and total scores.

217 **Within scale associations:** For both the IRI and JSE-S correlations between each  
218 subscale score and the total score were statistically significant: for the IRI generally  
219 strong ( $r = .431$  to  $r = .712$ ), for the JSE-S, moderate ( $r = .377$ ) to very strong ( $r = .854$ ).  
220 Correlations between the subscales within each scale were significant but less strong.  
221 For the IRI these ranged from  $r = .061$  between IRI-EC and IRI-PD to  $r = .403$  between  
222 IRI-EC and IRI-PT. A negative association was found between IRI-PD and IRI-PT. For  
223 the JSE the range was  $r = .114$  between JSE-SPS and JSE-PT and  $r = .467$  between JSE-  
224 PT and JSE-CC.

225 **Between scale associations:** The correlation between total scores of JSE-S and IRI was  
226 positive and significant, but weak  $r=.313$ . All inter-correlations of JSE-S and IRI  
227 subscale scores were statistically significant but weak, ranging from  $r=-.040$  (JSE-PT  
228 with IRI-PD) to  $r=.306$  (JSE-PT with IRI-EC). The only exception was the non-  
229 significant, negative correlation between IRI-PD and JSE-CC ( $r=-.016$ ). The correlation  
230 between the subscales scores of one scale and the total score of the other scale were also  
231 all statistically significant but weak. IRI-PD was negatively associated with all JSE  
232 subscales scores.

### 233 **Multiple linear regression models:**

234 **Table 4.** Multiple linear regression models for JSE dimensions.

235 The multiple linear regression analyses tested whether the IRI subscales, gender,  
236 country, and entry scheme significantly predicted JSE subscale and total scores. The  
237 reference categories were female, UK and standard entry (Table 4.) All regression  
238 models were significant, with a relatively low adjusted R squared, varying between  
239 8.9% and 15.3% of explained variance for JSE-S subscales and 19.4% for JSE total  
240 score.

241 With the exception of IRI-PD, all IRI subscales were significant, positive, predictors of  
242 each JSE subscale. Sex, was significant in all regression models except for JSE-PT. The  
243 extent to which students in countries differed from those in the reference country (UK)  
244 varied between instruments and between subscales of each instrument. Overall students  
245 in Brazil differed most from those in the UK whereas students in Ireland differed least.  
246 Entry scheme was not significant in any of the four tested models. The most pronounced  
247 predictor of total JSE-S score was IRI-EC.

248

## 249 **Discussion**

250 This study found that the dimensional structure of each instrument reflected its  
251 composite subscales with strong internal consistencies. The EFA results supported the  
252 cross-cultural construct validity and stability of both scales. For the IRI, our study  
253 confirmed Davis's 4 factor structure in 5 countries. To the authors' knowledge this  
254 factorial structure has been confirmed in studies of college students albeit with minor  
255 variations <sup>32</sup> but never before among medical students .

256 For the JSE, our results broadly accord with Hojat's original 3 factor structure and  
257 within that, the prominence of Perspective Taking (JSE-PT).<sup>24</sup> The only exception to  
258 this was the result for JSE-SPS in Ireland, possibly resulting from a combination of  
259 small sample size and small number of contributing items (n=2).

260 Our findings accord well with international JSE-S studies of medical students which,  
261 broadly support the 3 factor structure and their respective relative importance<sup>37, 38</sup> but  
262 with minor variations. For example studies of German and Japanese medical students  
263 support the JSE-PT construct but report variations in JSE-CC, possibly attributable to  
264 cultural differences. <sup>37,49</sup> A recent US study found the factorial structure of the JSE-S  
265 varied between preclinical and clinical medical students. Such analysis was beyond the  
266 scope of our study. <sup>41</sup>

267 The shared variance between the scales and subscales found in this study support the  
268 view that the scales measure different but related constructs. This view is further  
269 supported the correlation results which revealed only weak correlations despite an  
270 expectation of moderate correlations particularly in respect of subscale scores of IRI-  
271 PT and JSE-S-PT and IRI-EC and JSE-S-CC. Multiple linear regression models  
272 similarly suggested that all IRI subscales were weak predictors of the JSE-S subscale

273 scores and total score, with the strongest predictor of the JSE-S total score being the  
274 IRI-EC.

275 The study supports the view of gender differences in respect of empathy with women  
276 recording higher scores on self-report measures.

277 The suggestion that the two scales measure different but related constructs has  
278 implications for medical education, and medical education research. Care is needed in  
279 comparing studies using different scales. Conflicting results of studies of the trajectory  
280 of empathy during undergraduate medical education may, in part, be attributable to the  
281 use of instruments which are not comparable.<sup>33,39</sup> Similar implications may apply to  
282 intervention studies.

283 The suggested difference between to the two scales points to the need to clarify the  
284 constructs being measured. Whereas the IRI measures generic empathy the JSE-S may  
285 measure some idealized view of an empathic doctor-patient relationship. This  
286 distinction is reflected in differences in the wording of the scales. The IRI asks  
287 respondents the extent to which each statement “describes” his or herself, with all items  
288 containing the words “I” or “me”. The JSE-S asks respondents for their level of  
289 agreement with statements about either how “doctors” should behave or the doctor-  
290 patient relationship, with only 4 items relating to the individual.

291 The IRI and JSE-S were conceived with different populations in mind. Generic  
292 empathy may be shaped by personality, certain life experiences and possibly culture.  
293 Studies in various cultures suggest that psychological conditions exert the largest  
294 influence.<sup>42</sup> As an idealized view of an empathetic doctor-patient relationship JSE-S  
295 scores may be shaped by cultural influences affecting both medical education and  
296 patient expectations.<sup>49</sup> These may be more amenable to training and education than IRI  
297 scores.<sup>50</sup> Studies examining the impact of educational interventions aimed at enhancing

298 empathy have found a larger increase in JSE-S scores than in IRI scores.<sup>22</sup> However  
299 idealized views may also be more vulnerable to the hidden curriculum.<sup>16</sup>  
300 To characterize and clarify how the IRI and JSE-S constructs relate to each other, and  
301 how they change during medical education there is a need for more studies using both  
302 instruments, for more qualitative and mixed methods work and for more longitudinal  
303 work. If, as suggested the JSE-S measures context specific empathy then greater  
304 attention needs to be paid to that context including perhaps critical incidents and  
305 medical course content and structure. Our study only included undergraduate students.  
306 Comparable studies of post graduate medical students and/or physicians are needed.  
307  
308 This is one of the few studies of medical students using both the IRI and JSE-S and to  
309 the authors' knowledge the only study to include European, Brazilian and New Zealand  
310 data. One of its strengths is the large number of participants drawn from 5 countries.  
311 Whilst sample size in each country differed this was not a major limitation since one a  
312 main goal of the study was to explore the latent structure of IRI and JSE-S. Another  
313 limitation is that the analyses were run on the 24-item version of the IRI and did not  
314 include age per se. Our study drew data from countries with essentially "European"  
315 values which may explain the absence of marked cultural differences. Studies  
316 comparing the IRI and JSE-S among medical students in countries with very different  
317 cultural backgrounds, particularly those in which extreme scores have been recorded  
318 would be valuable in identifying differences between generic empathy and what is  
319 perceived to be an appropriate empathetic doctor/patient relationship.<sup>42</sup>

## 320 **Conclusions**

321 The factor analysis undertaken in this study supports the accepted factorial structure of  
322 the IRI and JSE-S and reaffirms the relationship of their respective subscales to the  
323 underlying dimensions of empathy: affective and cognitive, and for the IRI self-oriented  
324 versus other-oriented. These results are enhanced by being confirmed in 5 countries.  
325 However, this study suggests that the IRI and JSE are structurally different, weakly  
326 related concepts: the former generic or dispositional empathy, the latter context specific  
327 empathy. Consideration of this distinction may give rise to implications for medical  
328 education and may have implications for patient care. There is a need for more studies  
329 using both instruments, involving those at different stages in medical training, and for  
330 more longitudinal and qualitative studies in order to understand the practical  
331 implications of this distinction.

### 332 **Conflict of interest statement**

333 The authors declare that the research was conducted in the absence of any commercial  
334 or financial relationships that could be construed as a potential conflict of interest.

335

### 336 **Acknowledgements**

337 Patrício Costa had full access to of the data in the study and takes responsibility for the  
338 integrity of the data and the accuracy of the data analysis.

339 The authors are indebted to the following for their help and support in data collection:  
340 James Brimicombe, BSc., University of Cambridge, David Bristow, PhD., University of  
341 Plymouth, Ana Da Silva, PhD., University of Swansea, Jonathan Hales, MBChB  
342 MRCGP., University of Leicester, Joanne Hart, PhD. C Psychol., University of  
343 Manchester, Peter Hayes, MBChB BAO MICGP, University of Galway, Republic of  
344 Ireland, Kathy Jackson, MBChB MRCGP MD FHEA., University of St Andrews,  
345 Susan Jamieson, PhD., University of Glasgow, Margaret Jones, MB ChB MRCGP.,  
346 University of Nottingham, Paul Kinnersley, MBChB., University of Cardiff, Paula  
347 McDonald, MB ChB MRCGP MPH MFPHM Brighton & Sussex Medical School, Jean

348 McKendree, PhD., Hull/York Medical School, Helen Moriarty, MBChB MGP DPH.,  
349 University of Otago, New Zealand, Anne Stephenson, MBChB MRCGP PhD FHEA.,  
350 King's College University of London, Anna Stickland, MB ChB MRCGP., University  
351 of Liverpool, Kate Thomas, MBChB FRCGP., University of Birmingham, Trevor  
352 Thompson, MBBS MRCGP, PhD., University of Bristol, Helen Watson, PhD.,  
353 University of Exeter.

354

355

356

357

## 358 **References**

- 359 1. Kim SS, Kaplowitz S, Johnston MV. The effects of physician empathy on  
360 patient satisfaction and compliance. *Eval Health Prof.* 2004;27:237-51.
- 361 2. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to  
362 treatment: Three decades of research. A comprehensive review. *J Clin Pharm*  
363 *Ther.* 2001;26:331-42.
- 364 3. Derksen F, Bensing J, Lagro-Janssen A. Effectiveness of empathy in general  
365 practice: A systematic review. *Br J Gen Pract.* 2013;63:e76-e84.
- 366 4. Parkin T, de Looy A, Farrand P. Greater professional empathy leads to higher  
367 agreement about decisions made in the consultation. *Pat Educ Couns.*  
368 2014;96:144-150.
- 369 5. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS.  
370 Physicians' empathy and clinical outcomes for diabetic patients. *Acad Med.*  
371 2011;86:359-64.
- 372 6. Rakel D, Barrett B, Zhang Z. et al. Perception of empathy in the therapeutic  
373 encounter: Effects on the common cold. *Pat Educ Couns.* 2011;85:390-7.
- 374 7. Canale SD, Louis DZ, Maio V. et al. The relationship between physician  
375 empathy and disease complications: An empirical study of primary care  
376 physicians and their diabetic patients in Parma, Italy. *Acad Med.* 2012;87:1243-  
377 9.
- 378 8. Steinhausen S, Ommen O, Thüm S. et al. Physician empathy and subjective  
379 evaluation of medical treatment outcome in trauma surgery patients. *Pat Educ*  
380 *Couns.* 2014; 95:53-60.
- 381 9. Larson EB, Yao X. Clinical empathy as emotional labor in the patient-physician  
382 relationship. *JAMA.* 2005;293:1100-6.

- 383 10. Coulehan JL, Platt FW, Egner B, et al. "Let me see if I have this right.. ":  
384 Words that help build empathy. *Ann Intern Med.* 2001;135:221-7.
- 385 11. Beckman HB, Frankel RM. Training practitioners to communicate effectively in  
386 cancer care: It is the relationship that counts. *Pat Educ Couns.* 2003;50:85-9.
- 387 12. Thomas MR, Dyrbye LN, Huntington JL, et al. How do distress and well-being  
388 relate to medical student empathy? A multicenter study. *J Gen Intern Med.*  
389 2007;22:177-83.
- 390 13. DiLalla LF, Hull SK, Dorsey JK. Effect of gender, age, and relevant course  
391 work on attitudes toward empathy, patient spirituality, and physician wellness.  
392 *Teach Learn Med.* 2004;16:165-70.
- 393 14. Bellini LM, Baime M, Shea JA. Variation of mood and empathy during  
394 internship. *JAMA.* 2002;287:3143-6.
- 395 15. West CP, Huschka MM, Novotny PJ, et al. Association of perceived medical  
396 errors with resident distress and empathy: A prospective longitudinal study.  
397 *JAMA.* 2006;296:1071-8.
- 398 16. Neumann M, Bensing J, Mercer S, Ernstmann N, Ommen O, Pfaff H. Analysing  
399 the "nature" and "specific effectiveness" of clinical empathy: A theoretical  
400 overview and contribution towards a theory based research agenda. *Pat Educ*  
401 *Couns.* 2009; 74:339-346.
- 402 17. Pedersen R. Empirical research on empathy in medicine-A critical review. *Pat*  
403 *Educ Couns.* 2009;76:307-22.
- 404 18. Hemmerdinger J, Stoddart SD, Lilford RJ. A systematic review of tests of  
405 empathy in medicine. *BMC Med Educ.* 2007;7, doi:10.1186/1472-6920-7-24.
- 406 19. Mercer SW, Reynolds WJ. Empathy and the quality of care. *Br J Gen Pract.*  
407 2002;52 (suppl):S9-13.

- 408 20. O'Connor K, King R, Malone KM, Guerandel A. Clinical examiners, simulated  
409 patients and student self-assessed empathy in medical students during a  
410 psychiatry Objective Structured Clinical Examination. *Acad Psychiatry*. 2014;  
411 38:451-457.
- 412 21. Berg K, Majden JF, Berg D, Veloski J, Hojat M. Medical students self-reported  
413 empathy and simulated patients' assessment of student empathy: An analysis by  
414 gender and ethnicity. *Acad Med*. 2011;86:984-988.
- 415 22. Schweller M, Costa FO, Antonio MARGM, Amaral EM, De Carvalho-Filho  
416 MA. The impact of simulated medical consultations on empathy levels of  
417 students at one medical school. *Acad Med*. 2014 89:632-637.
- 418 23. Davis MH. Measuring Individual differences in Empathy: Evidence for a  
419 Multidimensional Approach. *J Pers Soc Psychol*. 1983 44:113-126.
- 420 24. Hojat M. *Empathy in Patient Care: Antecedents, Development, Measurement,  
421 and Outcomes*. Philadelphia, PA: Springer; 2007.
- 422 25. Banissy MJ, Kanai R, Walsh V, Rees G. Inter-individual differences in empathy  
423 are reflected in human brain structure. *Neuroimage*. 2012;62:2034-2039.
- 424 26. Juujärvi S, Myyry L, Pessa K. Does care reasoning make a difference? Relations  
425 between care, justice and dispositional empathy. *J Moral Educ* 2010;39:469-489.
- 426 27. Lamm C, Batson CD, Decety J. The neural substrate of Human Empathy:  
427 Effects of Perspective-taking and Cognitive Appraisal. *J Cogn Neurosci*.  
428 2007;19:42-58.
- 429 28. Lehmann AQ, Bahcesular K, Brockmann TM, et al. Subjective experience of  
430 emotions and emotional empathy in paranoid schizophrenia. *Psychiatry Res*.  
431 2014;220:825-33.

- 432 29. Narme P, Mouras H, Duru C, Krystowiak KP, Godfrey O. Emotional and  
433 cognitive social processes are impaired in Parkinson's disease and are related to  
434 behavioural disorders. *Neuropsychology*. 2013;27:182-192.
- 435 30. Schaffer M, Clark S, Jeglic EL. The role of empathy and parenting style in the  
436 development of antisocial behaviors. *Crime Delinq*. 2009;55:586-599.
- 437 31. Baron-Cohen, S, Wheelwright, S. The empathy quotient: an investigation of  
438 adults with Asperger syndrome or high functioning autism and normal sex  
439 differences. *J Autism Dev Disord*. 2004;34:163-175.
- 440 32. Fernández AM, Dufey M, Kramp U. Testing the psychometric properties of the  
441 Interpersonal Reactivity Index (IRI) in Chile. *Eur J Psychol Assess*.  
442 2011;27:179-185.
- 443 33. Roff S. Reconsidering the "decline" of medical student empathy as reported in  
444 studies using the Jefferson Scale of Physician Empathy-Student version (JSPE-  
445 S). *Med Teach*. 2015;37:783-786.
- 446 34. Loureiro J, Goncalves-Pereira M, Trancas B, Caldas-de-Almeida JM, Castro-  
447 Caldas A. Empathy in the doctor patient relationships as viewed by first year  
448 medical students: data on validity and sensibility to change of the Jefferson  
449 Measure in Portugal. *Acta Med Port*. 2011, 24(Suppl 2);431-42.
- 450 35. Fields SK, Mahan P, Tillman P, Harris J, Maxwell K, Hojat M. Measuring  
451 empathy in healthcare profession students using the Jefferson Scale of  
452 Physician Empathy: healthcare provider- student version. *J Interprof Care*.  
453 2011;25:287-93.
- 454 36. Ward J, Schaal M, Sullivan J, et al. Reliability and validity of the Jefferson  
455 Scales of Physician Empathy in undergraduate nursing students. *J Nurs Meas*.  
456 2009;17:73-88.

- 457 37. Preusche I, Wagner-Menghin M. Rising to the challenge: cross-cultural adaption  
458 and psychometric evaluation of the adapted German version of the Jefferson  
459 Scale of Physician Empathy for Students –JSPE-S. *Adv Health Sci Educ Theory  
460 Pract.* 2013;18:573-587.
- 461 38. Myoung-Sun R, Bong-Jin H, Dong Hun L, et al. Evaluation of empathy among  
462 Korean medical students: a cross-sectional study using the Korean version of the  
463 Jefferson Scale of Physician Empathy. *Teach Learn Med.* 2010;22:167-171.
- 464 39. Quince T A, Parker RA, Wood DF, Benson JA. Stability of empathy among  
465 undergraduate medical students: a longitudinal study at one UK medical school.  
466 *BMC Med Educ.* 2011;11: 90. doi 10.1186/1472-6920-11-90.
- 467 40. Paro HBMS, Silveira PSP, Perotta B. et al. Empathy among medical students: Is  
468 there a relation with quality of life and burnout? *PLoS ONE.* 2014;9:e94133.
- 469 41. Stansfield RB, Schwartz A, O'Brien CL, et al. Development of a metacognitive  
470 effort construct of empathy during clinical training: a longitudinal study of the  
471 factor structure of the Jefferson Scale of Empathy. *Adv Health Sci Educ Theory  
472 Pract.* 2015; doi 10.1007/s10459-015-9605-1.
- 473 42. Quince T, Thiemann P, Benson J, Hyde S. Undergraduate Medical Student  
474 Empathy: Current Perspectives. *Advances in Medical Education and Practice.*  
475 2016 Forthcoming.
- 476 43. Hojat M, Mangione S, Kane GC, Gonnella JS. Relationship between scores on  
477 the Jefferson Scale of Physician Empathy JSPE and the Interpersonal Reactivity  
478 Index (IRI). *Med Teach.* 2005;27:625-8.
- 479 44. Limpo T, Alves RA, Castro SL. Medir a empatia: Adaptação portuguesa do  
480 Índice de Reactividade Interpessoal. *Laboratório de Psicologia.* 2013;8:171-184.

- 481 45. Fox J. The R Commander: A Basic Statistics Graphical User Interface to R. Journal of  
482 Statistical Software 2005;14:1-42.
- 483 46. Revelle W. Psych: Procedures for Personality and Psychological Research.  
484 Northwestern University, Evanston, Illinois, USE, [http://CRAN.R-](http://CRAN.R-project.org/package=psych)  
485 [project.org/package=psych](http://CRAN.R-project.org/package=psych) Version=1.5.8.
- 486 47. Cohen, J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ:  
487 Lawrence Earlbaum Associates. 1988
- 488 48. Roh MS, Hahm BJ, Lee DH, Suh DH. Evaluation of empathy among medical  
489 students: A cross-sectional study using the Korean version of the Jefferson Scale  
490 of Physician Empathy. Teach Learn Med. 2010; 22:167-171.
- 491 49. Kataoka HU, Koide N, Ochi K, Hojat M, Gonnella JS. Measurement of empathy  
492 among Japanese medical students: psychometrics and score differences by  
493 gender and level of medical education. Acad Med. 2009;84:1192-7.
- 494 50. Hegazi I, Wilson I. Maintaining empathy in medical schools: it is possible. Med  
495 Teach. 2013;35:1002-8.
- 496