




REVIEW

Clarifying changes in student empathy throughout medical school: a scoping review

Alexandra Ferreira-Valente^{1,2,3} · Joana S. Monteiro^{4,5} ·
Rita M. Barbosa⁶ · Ana Salgueira^{1,2} · Patrício Costa^{1,2,7} ·
Manuel J. Costa^{1,2} 

Received: 8 June 2016 / Accepted: 21 July 2016 / Published online: 27 July 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Despite the increasing awareness of the relevance of empathy in patient care, some findings suggest that medical schools may be contributing to the deterioration of students' empathy. Therefore, it is important to clarify the magnitude and direction of changes in empathy during medical school. We employed a scoping review to elucidate trends in students' empathy changes/differences throughout medical school and examine potential bias associated with research design. The literature published in English, Spanish, Portuguese and French from 2009 to 2016 was searched. Two-hundred and nine potentially relevant citations were identified. Twenty articles met the inclusion criteria. Effect sizes of empathy scores variations were calculated to assess the practical significance of results. Our results demonstrate that scoped studies differed considerably in their design, measures used, sample sizes and results. Most studies (12 out of 20 studies) reported either positive or non-statistically significant changes/differences in empathy regardless of the measure used. The predominant trend in cross-sectional studies (ten out of 13 studies) was of significantly higher empathy scores in later years or of similar empathy scores across years, while most longitudinal studies presented either mixed-results or empathy declines. There was not a generalized international trend in changes in students' empathy throughout medical school. Although statistically significant changes/differences were detected in 13 out of 20 studies, the calculated effect sizes were small in all but two studies, suggesting

✉ Manuel J. Costa
mmcosta@ecsau.de.uminho.pt

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

² 3B PT Government Associate Laboratory, Braga, Guimarães, Portugal

³ University Lusíada of Lisbon, Lisbon, Portugal

⁴ USF Odisseia, Maia, Portugal

⁵ Faculty of Medicine, Oporto University, Porto, Portugal

⁶ USF Rainha D. Amélia, Porto, Portugal

⁷ Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal

little practical significance. At the present moment, the literature does not offer clear conclusions relative to changes in student empathy throughout medical school.

Keywords Empathy · Medical education · Undergraduate students · Scoping review

Introduction

Medical schools are increasingly aware of their crucial role in the development of empathy of undergraduate medical students. However, there are concerns that student empathy may deteriorate during undergraduate medical education (Austin et al. 2007; Hojat et al. 2004; Newton et al. 2008; Roff 2015). The worrying signs emerged initially from a longitudinal study in one North American medical school, which reported score declines in the Jefferson Scale of Physician Empathy (JSE-S) that were concurrent with a pre-clinical to clinical transition in the curriculum (Hojat et al. 2004). The declines were statistically significant with a small magnitude and were independent of “demographic and performance variables”. The conclusions were confirmed within the same population with a larger data series (Hojat et al. 2009). Other studies developed in the USA, the UK, Poland, in Iran and in India reached similar conclusions (Chen et al. 2007; Kliszcz et al. 1998; Shariat and Habibi 2013; Shashikumar et al. 2014; Todres et al. 2010). An important limitation of these studies is that the reported declines have been grounded on self-reported empathy measures, rather than on observational measures of student empathic behavior (Colliver et al. 2010). Also, the reported effect sizes of changes in empathy are small, which reinforces the doubt of whether self-report measures scores are reflected in actual behaviors at the bedside (*ibidem*). These findings hurt the accountability of medical education and therefore deserve attention. Even though findings originated from a limited sampling of international contexts and were obtained in only few countries, the potential detrimental effect of medical education on student empathy echoed widely in the international literature.

A recent surge of studies in multiple countries has originated findings that do not seem to confirm the universality of empathy declines. In fact, studies with the two most widely used instruments—the JSE-S or the Interpersonal Reactivity Index (IRI; Davis 1983)—have revealed either a statistically significant growth in empathy or non-statistically significant (n.s.) variations (Austin et al. 2007; Kataoka et al. 2009; Roh et al. 2010). At the present moment, the findings concerning the changes in students’ empathy through medical are unclear. Not only are the results divergent but there are further important confounding variables which compromise inter-study comparisons. For example, there are ample inter-study variations in lengths of undergraduate programs, time points of data collection and methodologies also vary, some being cross-sectional and others longitudinal. Even though further research is necessary to clarify whether findings about empathy development are generalizable across countries, it is timely to conduct a more rigorous appraisal of the existing evidence.

The primary goal of this study was to appraise available international evidence and to assess the existence of international generalizable trends concerning changes in students’ empathy during medical training as revealed by quantitative studies focusing on formal undergraduate programs. As the intention was to synthesize the scattered evidence available, the option was to conduct a scoping review. By definition, the scoping review

methodology is a fit for purpose reviewing technique to summarize primary literature, whenever the aim is to conduct a relatively rapid review and the research designs and methods of the literature are heterogeneous (Arksey and O'Malley 2005; Mays et al. 2001, Levac et al. 2010). Scoping reviews may be developed as a stand-alone studies- as in the present case—or as one part of an ongoing review process aiming at determining the worth and feasibility of conducting more thorough reviews, such as full systematic reviews (Arksey and O'Malley 2005; Levac et al. 2010; Mays et al. 2001). Even though the scoping review methodology is not widely used in medical education, it is well established for synthesizing heterogeneous research evidence (Pham et al. 2014). We took advantage of the language fluency of the authors—English, Spanish, Portuguese and French—and surveyed literature in multiple languages. The specific aims were: (1) to characterize the research designs of empathy studies of undergraduate medical education; (2) to evaluate the associations between research designs (type of study, measures, inclusion of preclinical/clinical transition) and empathy outcomes; and (3) to examine eventual trends of empathy development during medical education across countries.

Methods

The scoping review followed six-stages, as proposed by Levac et al. (2010): (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; (5) collating, summarizing and reporting results; (6) consultation.

Research question and relevant studies

The research question and a search strategy were defined in the first meeting and considered the language fluencies of the authors. Searches were conducted in: 1. PubMed with the MeSH terms 'empathy' and 'educational, medical' for papers published in English, Portuguese, Spanish and French, for articles published between 2009 and 2016 (previous papers were reviewed by Pedersen (2009)); 2. online engines within official journals of medical education societies in France and Portuguese and Spanish speaking countries (see Table 1). Reference lists were hand searched.

Table 1 List of official journals of national education societies consulted

Revista Mexicana de Educación Médica
Revista de Investigación en Educación Médica
Revista Brasileira de Educação Médica
Revista Fundación Educación Médica
Revista Argentina de Educación Médica
Educación Médica y Salud
Pedagogie Medicale
Revista de Educación en Ciencias de la Salud
South East Asian Journal of Medical Education
Best Evidence Medical Education
Perspectives on Medical Education
International Journal of Medical Education
Educación Médica Superior

JSM, RMB and AFV independently reviewed all abstracts for inclusion in the review. The inclusion criteria were: (1) participants were undergraduate medical students (UMS); the studies were either (2) longitudinal, or (3) cross-sectional, with participants attending two or more years of medical training. Exclusion criteria were: (1) studies aimed exclusively at assessing psychometrical properties of instruments or at assessing interventions to enhance empathy; (2) comments, editorials or related manuscripts. Full texts were recovered and exclusions/inclusion criteria were applied once again. Disagreements between reviewers were settled through discussion with a third party.

Charting, collating, and summarizing

The data extraction sheet was iteratively developed by JSM, RMB and AFV, summarizing the country of origin, design and methods, and institutional context (see Table 1). MJC was “third party” throughout the process.

To identify moments of pre-clinical to clinical transition, the study plans were recovered from institutional websites and analyzed by MJC and AFV. Moments corresponding to a considerable increase in the number of hours allocated to teaching in clinical workplaces were considered the transition moment.

Data analysis

To assess the practical significance of statistical differences, PC and AFV computed Hedges' g , Cohen's d_z and r^2 as measures of effect size. Large and small effect sizes

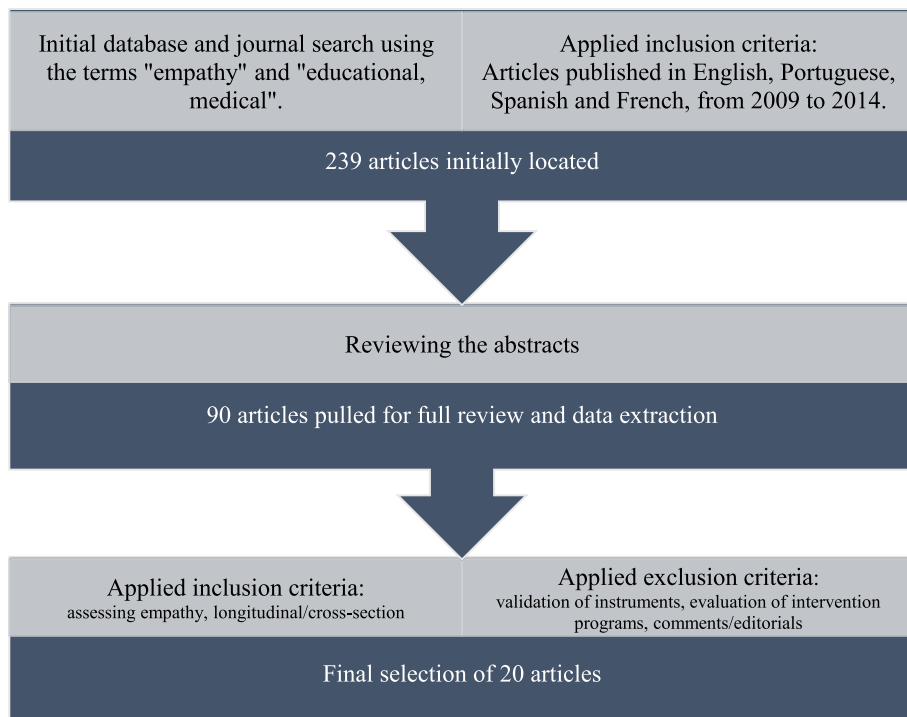


Fig. 1 Flowchart of search and results

corresponded, respectively, to g/d_z higher than 0.5 and r^2 higher than 0.25, g/d_z lower than 0.2 and r^2 lower than 0.1 (Cohen 1988).

Results

Figure 1 outlines the search results. The original search yielded 239 potentially relevant citations. After reviewing the abstracts of all identified articles, 90 were pulled for full review and data extraction. Twenty articles met the inclusion criteria and were included in the analysis.

Research designs

The final selection comprised institutional programs with diverse characteristics. For example, degree lengths varied between four years in the USA (Chen et al. 2012) and eight years in the Dominican Republic (Silva et al. 2014). There were large variations across studies of participants' year of study: some studies had participants from all school years (e.g. Kataoka et al. 2009; Toto et al. 2015), others from only one (Loureiro et al. 2011). Sample sizes ranged from 72 (Lim et al. 2013) to 1162 participants (Loureiro et al. 2011). In 14 studies samples were large (over 150 participants) (Bratek et al. 2015; Chen et al. 2012; Duarte et al. 2015; Esquerda et al. 2016; Hegazi and Wilson 2013; Kataoka et al. 2009; Magalhães et al. 2011; Montilva et al. 2015; Nunes et al. 2011; Quince et al. 2011; Toto et al. 2015; Wen et al. 2013; Youssef et al. 2014), while the remaining six had fewer than 150 participants. There were 13 countries represented: Australia (2 studies), Brazil (1), China (1), Dominican Republic (1), Japan (1), New Zealand (1), Portugal (4), Poland (1), Spain (1), Trinidad and Tobago (2), the UK (1), the USA (3) and Venezuela (1) (cf. Table 2). Of the 20 articles included in this scoping review, 14 were written in English, 4 were written in Spanish, 2 were written in Portuguese, and none in French. All studies were questionnaire-based, using the IRI, the JSE-S, with two exceptions (Esquerda et al. 2016; Thomazi et al. 2014). Some studies reported additional empathy measures (Handford et al. 2013; Youssef et al. 2014) such as the Empathy Quotient Questionnaire (EQ-60), the Reading the Mind in the Eyes Test (RMET), and the Toronto Empathy Questionnaire (TEQ).

In what concerns the methodological designs, there were 13 cross-sectional (Bratek et al. 2015; Duarte et al. 2015; Esquerda et al. 2016; Handford et al. 2013; Hegazi and Wilson 2013; Kataoka et al. 2009; Magalhães et al. 2011; Montilva et al. 2015; Nunes et al. 2011; Silva et al. 2014; Toto et al. 2015) and seven longitudinal studies (Chen et al. 2012; Costa et al. 2013; Hojat et al. 2009; Lim et al. 2013; Loureiro et al. 2011; Quince et al. 2011; Thomazi et al. 2014). The samples of longitudinal studies varied in terms of number of cohorts and years of study, from one cohort sampled in two moments in year 1 (Loureiro et al. 2011) to four cohorts sampled annually as students progressed in the curriculum (Quince et al. 2011). Fourteen studies comprised the pre-clinical to clinical transition (Chen et al. 2012; Costa et al. 2013; Duarte et al. 2015; Esquerda et al. 2016; Handford et al. 2013; Hegazi and Wilson 2013; Hojat et al. 2009; Kataoka et al. 2009; Magalhães et al. 2011; Montilva et al. 2015; Quince et al. 2011; Thomazi et al. 2014; Toto et al. 2015; Youssef et al. 2014). Nine studies analyzed gender specificities in empathy growth (Bratek et al. 2015; Chen et al. 2012; Costa et al. 2013; Esquerda et al. 2016; Hegazi and Wilson 2013; Hojat et al. 2009; Lim et al. 2013; Quince et al. 2011; Silva et al. 2014).

Table 2 Articles included in the scope review

Authors	Country	Sample (N)	Measure	Design	Duration of medical school (years)	Sampled school years
Bratek et al. (2015)	Poland	509 UMS (n = 331), candidates, trainees, residents, specialists	IRI	CS	6	1st–6th
Chen et al. (2012)	USA	1162	JSE-S	L	4	Admission through 4th
Costa et al. (2013)	Portugal	77	JSE-S	L	6	Admission 3rd 4th
Duarte et al. (2015)	Portugal	208	JSE-S	CS	6	1st 3rd 4th
Esquerda et al. (2016)	Spain	191	TECA	CS	6	1st 3rd
Handford et al. (2013)	Australia	100 5 groups: UMS phase 1 UMS phase 3 Medical practitioners Other students Other adults	IRI EQ-60 RMET	CS	6	2nd cycle (4th and 5th) Phase 1: 1st–2nd Phase 3: 5th–6th
Hegazi and Wilson (2013)	Australia	404	JSE-S	CS	5	1st–5th
Hojat et al. (2009)	USA	456 Matched cohort (n = 121)	JSE-S	L	4	Admission 1st 2nd 3rd 4th
Kataoka et al. (2009)	Japan	400	JSE-S	CS	6	1st–6th

Table 2 continued

Authors	Country	Sample (N)	Measure	Design	Duration of medical school (years)	Sampled school years
Lim et al. (2013)	New Zealand	72	JSE	L	6	5th (beginning) 5th (after MI/BI training session) 6th
Loureiro et al. (2011)	Portugal	81	JSE-S	L	6	Admission 1st (after attending the course of medical psychology)
Magalhães et al. (2011)	Portugal	476	JSE-S	CS	6	Admission 6th
Montilva et al. (2015)	Venezuela	246 UMS (n = 175), and nursing.	JSE-S	CS	6	3rd 4th 6th
Nunes et al. (2011)	Trinidad and Tobago	Sample 1: 355 UMS (n = 162), dentistry, pharmacy, veterinary medicine and nursing. Sample 2: 366 (UMS: n = 162)	JSE-S	CS	5	Sample 1: Admission Sample 2: 1st (end)
Quince et al. (2011)	UK	Total Core science entrants: 1111 students Total clinical entrants: 542 students	IRI-EC ("Emphatic concern") IRI-PT ("Perspective taking")	L	6	Annually (1st–6th)
Silva et al. (2014)	Dominican Republic	515	JSE-S	CS	8	1st–5th
Thomazi et al. (2014)	Brazil	79	Empathy Inventory	L	6	1st 4th
Toto et al. (2015)	USA	460	IRI	CS	4	1st through 4th
Wen et al. (2013)	China	820	JSE-S	CS	5	1st–4th
Youssef et al. (2014)	Trinidad and Tobago	669	JSE-S TEQ RMET	CS	5	1st–5th

Table 2 continued

Authors	Transition moment	Assessment moments	Statistical analysis	Average empathy M (SD)	Effect size	Conclusions
Bratek et al. (2015)	Not reported	–	Mann–Whitney test	Not reported	–	n.s.
Chen et al. (2012)	2nd	5	Linear mixed models	T0: 113.0 (12.5) T1: 114.6 (11.7) T2: 116.0 (12.4) T3: 114.2 (13.7) T4: 113.3 (12.6)	$d_z = 0.04^a$ $r^2 = 0.0003$	Admission and 1st < 2nd 2nd > 3rd and 4th
Costa et al. (2013)	4th	3	Latent growth model	T1: 109.3 (11.6) T2: 111.2 (10.6) T3: 110.8 (10.8) $r^2_{T1T2} = 0.27$	$d_z = 0.11$ $r^2 = 0.003$	n.s. differences between assessment moments Smooth linear growth in empathy between T1 and T3
Duarte et al. (2015)	4th	–	ANOVA	1st (n = 75): 112.9 (10.5) 3rd (n = 66): 113.7 (11.6) 6th (n = 67): 116.8 (10.4)	$g = -0.37$ $r^2 = 0.03$	n.s.
Esquerda et al. (2016)	4th	–	Mann–Whitney test	Women: 1st (n = 44): 58.5 (26.1) 3rd (n = 61): 60.0 (27.6) 4th/5th (n = 37): 71.7 (23.8)	Women: $g = -0.52$ $r^2 = 0.07$	Total sample and Men: n.s. Women: TECA empathy global score increased from the 1st through 4th/5th

Table 2 continued

Authors	Transition moment	Assessment moments	Statistical analysis	Average empathy M (SD)	Effect size	Conclusions
Handford et al. (2013)	3 rd	–	ANOVA	IRI-EC Phase1 (n = 20): 20.5 (3.9) Phase3 (n = 20): 20.8 (3.5) IRI-PT Phase1: 17.9 (6.0) Phase3: 18.8 (3.3) EQ-60 Phase1: 41.2 (11.2) Phase3: 45.1 (13.9) RMET Phase1: 53.6 (5.0) Phase3: 50.5 (6.1)	IRI-EC $g = -0.08$ $r^2 = 0.002$ IRI-PT $g = -0.19$ $r^2 = 0.01$ EQ-60 $g = -0.30$ $r^2 = 0.02$ RMET $g = 0.53$ $r^2 = 0.07$	EQ-60, IRI-EC, IRI-PT: Phase1 < Phase3 RMET: n.s.
Hegazi and Wilson (2013)	3 rd	–	Kruskal–Wallis test	Not reported	$\eta^2 = 0.02$	n.s.
Hojat et al. (2009)	2 nd	5	ANOVA Dependent samples <i>t</i> test	T0: 114.3 (9.0) T1: 115.8 (10.3) T2: 115.7 (8.9) T3: 108.5 (10.7) T4: 110.5 (11.5)	$d_s = 0.36^a$ $r^2 = 0.03$	Admission = 1st = 2nd > 3rd = 4th

Table 2 continued

Authors	Transition moment	Assessment moments	Statistical analysis	Average empathy M (SD)	Effect size	Conclusions
Kataoka et al. (2009)	5th	-	ANOVA	1st (n = 50): 98.5 (15.4) 2nd (n = 56): 103.8 (15.4) 3rd (n = 96): 105.0 (10.8) 4th (n = 65): 102.8 (14.4) 5th (n = 47): 105.4 (13.6) 6th (n = 86): 107.8(12.1)	$g = -0.69$ $r^2 = 0.01$	1st < other (except for 4th)
Lim et al. (2013)	-	3	ANOVA	T1: 114.2 (9.5) T2: 113.6 (10.5) T3: 82.5 (17.9) $r_{T1T3} = -0.05$	$d_z = 1.15$ $r^2 = 0.37$	T ₁ > T ₃ T ₂ > T ₃
Loureiro et al. (2011)	-	2	Wilcoxon	T1: 108.8 (9.5) T2: 112.4 (8.7)	$d_z = 0.39^a$ $r^2 = 0.04$	T1 < T2 (except for "Standing in the patient's shoes") T1 > T2 ("Standing in the patient's shoes").
Magalhães et al. (2011)	4th	-	ANOVA	1 st (n = 536): 110.3 (10.4) 6th (n = 120): 118.2 (9.1)	$g = 0.78$ $r^2 = 0.14$	1th < 6st
Montilva et al. (2015)	4th	-	ANOVA	3st (n = 40): 120.0 (2.0) 4th (n = 56): 119.6 (2.1) 6th (n = 79): 120.0 (1.2)	$g < 0.001$ $r^2 < 0.001$	n.s.

Table 2 continued

Authors	Transition moment	Assessment moments	Statistical analysis	Average empathy M (SD)	Effect size	Conclusions
Nunes et al. (2011)	-	-	Independent sample <i>t</i> -test	Sample 1 (n = 355): 110.8(13.2) Sample 2 (n = 366): 107.3(13.7)	$g = -0.69$ $r^2 = 0.01$	Admission > 1st (medical students)
Quince et al. (2011)	4th	1st cohort: 6 2nd cohort: 6 3rd cohort: 5 4th cohort: 4	Regression analysis	Men: IRI-EC 1st (n = 309): 19.4 (4.0) 2nd (n = 145): 18.1(4.9) 3rd (n = 86): 18.8(4.2) 4th (n = 154): 19.5(4.1) 5th (n = 81): 18.9(4.5) 6th (n = 45): 19.0(3.5) IRI-PT 1st: 18.0 (4.2) 2nd: 17.6 (4.9) 3rd: 18.4 (4.0) 4th: 17.9 (4.2) 5th: 17.8 (4.5) 6th: 18.2 (4.7)	Men: IRI-EC $g = -0.06$ $r^2 = 0.001$ IRI-PT $g = 0.03$ $r^2 = 0.002$	Men: IRI-EC declined slightly during 1st through 6th Women: n.s.
Silva et al. (2014)	-	-	ANOVA	1st (n = 108): 101.6(12.3) 2nd (n = 151): 99.8(12.8) 3rd (n = 109): 102.7(13.6) 4th (n = 94): 103.6(11.4) 5th (n = 53): 101.8(11.4)	$g = -0.02$ $r^2 < 0.001$	n.s.

Table 2 continued

Authors	Transition moment	Assessment moments	Statistical analysis	Average empathy M (SD)	Effect size	Conclusions
Thomazi et al. (2014)	3rd	2	-	T1: 142.1 T2: 142.0	-	n.s.
Toto et al. (2015)	2nd	-	ANCOVA	IRI-EC 1st (n = 155): 3.17 2nd (n = 128): 3.08 3rd (n = 95): 3.11 4th (n = 72): 3.39 IRI-PT 1st: 2.75 2nd: 2.85 3rd: 2.80 4th: 3.30	-	1st = 2nd = 3rd < 4th
Wen et al. (2013)	Not reported	-	ANOVA	1st (n = 211): 107.4(13.4) 2nd (n = 267): 109.2(11.4) 3rd (n = 275): 110.5(10.8) 4th (n = 149): 112.1(13.6)	$g = -0.35$ $r^2 = 0.03$	1st < 4th

Table 2 continued

Authors	Transition moment	Assessment moments	Statistical analysis	Average empathy M (SD)	M	Effect size	Conclusions
Youssef et al. (2014)	3rd	-	Kruskal–Wallis test	JSE-S 1st (n = 224): 108.1(11.3) 2nd (n = 136):106.7(11.0) 3rd (n = 102): 102.4(12.4) 4th (n = 94): 105.3(12.4) 5th (n = 113): 104.6(11.7) TEQ ^b 1st:48.6(11.3) 2nd:47.6(11.0) 3rd: 44.8(12.4) 4th: 46.8(12.4) 5th: 45.9(11.7) RMET 1st: 25.9(4.3) 2nd: 26.8(3.5) 3rd: 25.3(4.8) 4th:24.9(3.9) 5th: 25.7(4.2)	JSE-S $g = 0.31$ $r^2 = 0.02$ TEQ $g = 0.24$ $r^2 = 0.01$ RMET $g = 0.05$ $r^2 = 0.001$	JSE and TEQ: 1st > 3rd 2nd > 3rd RMET: n.s.	

^a Whenever correlation coefficient between measures was not reported we assumed $r = 0.50$; ^b We have detected equal SD values for JSE and TEQ and which we think might be an editing error—however we based our calculations on the reported scores; T0: Admission

In terms of reported results, thirteen studies presented full descriptive statistics. Mean empathy differences between scores were assessed with parametric or non-parametric tests of hypothesis in all but three studies (Chen et al. 2012; Costa et al. 2013; Thomazi et al. 2014). Three studies reported multivariate analysis (Chen et al. 2012; Costa et al. 2013; Quince et al. 2011).

Are empathy declines consistent across research designs?

As shown in Table 3, the studies differed in respect of the empathy measures and sample sizes. In the fourteen studies using the JSE-S four reported empathy declines or statistically significant lower empathy scores in older students (Hojat et al. 2009; Lim et al. 2013; Nunes et al. 2011; Youssef et al. 2014), four reported non-significant score variations/differences (Duarte et al., 2015; Hegazi and Wilson 2013; Montilva et al. 2015; Silva et al. 2014), three reported increases (Kataoka et al. 2009; Magalhães et al. 2011; Wen et al. 2013) and three presented mixed results (Chen et al. 2012; Costa et al. 2013; Loureiro et al. 2011). Studies using the IRI, on the other hand, reported either increases (Handford et al. 2013; Toto et al. 2015) or non-significant score variations/differences (Bratek et al. 2015; Quince et al. 2011) of empathy through medical school. Six exhibited non-significant score variations/differences (Bratek et al. 2015; Duarte et al. 2015; Hegazi and Wilson 2013; Montilva et al. 2015; Quince et al. 2011; Silva et al. 2014) and four out of the fourteen studies with large sample sizes ($n > 150$) reported empathy increases (Kataoka et al. 2009; Magalhães et al. 2011; Toto et al. 2015; Wen et al. 2013). Only two of these studies presented either empathy declines (Nunes et al. 2011; Youssef et al. 2014), and other two reported mixed results (Chen et al. 2012; Esquerda et al. 2016). Most studies with small

Table 3 Summary of results by design, measure, samples size, presence/absence of transition, and country (number of studies with the indicated result)

	Empathy increase or is higher in older students	n.s.	Empathy declines or is lower in older students	Mixed results	Total number of studies
Empathy measure					
JPSE-S	3	4	4	3	14
IRI	2	2	–	–	4
Other	1	1	1	1	4
Sample size					
<150	1	1	2	2	6
>150	4	6	2	2	14
Type of study					
Longitudinal	–	2	2	3	7
Cross-section	5	5	2	1	13
Transition					
With	4	5	2	3	14
Without	–	1	2	1	4

Results from Handford et al. (2013) were here considered empathy increase as a statistically significant increase was found for two out of three empathy measures. Results from Quince et al. (2011) were here considered n.s. as a statistically significant decline was found only for males in the IRI-EC. Results from Youssef et al. (2014) were here considered as empathy declines as a statistically significant decrease was found for two out of three empathy measures

samples, in contrast, reported either declines (two out of six studies; Lim et al. 2013; Hojat et al. 2009) or mixed results (two studies; Costa et al. 2013; Loureiro et al. 2011).

The predominant trend in cross-sectional studies was either of significantly higher empathy scores in later years (five out of 13 studies; Kataoka et al. 2009; Magalhães et al. 2011; Handford et al. 2013; Toto et al. 2015; Wen et al. 2013) or similar empathy scores across cohorts (five studies; Bratek et al. 2015; Duarte et al. 2015; Hegazi and Wilson 2013; Montilva et al. 2015; Silva et al. 2014). The two studies of Trinidad and Tobago (Nunes et al. 2011; Youssef et al. 2014) had opposite findings (i.e. lower empathy scores in later years), while one study reported mixed-results (Esquerda et al. 2016). Yet, evidence from cross-sectional studies is conditioned by cohort differences in terms of participants and sometimes in terms of formal educational experiences.

The predominant longitudinal trend (3 out of 7 studies) was of mixed-results (Chen et al. 2012; Costa et al. 2013; Loureiro et al. 2011). There were only two studies reporting declines in empathy scores over time (Hojat et al. 2009; Lim et al. 2013), whereas the remaining two reported non-statistically significant variations of empathy scores over time (Quince et al. 2011; Thomazi et al. 2014). A detailed analysis of these studies follows.

Figure 2 plots results of longitudinal studies, in reference to the transition moments from the pre-clinical to the clinical period of medical education. The four studies comprising the preclinical/clinical transition did not show a coinciding sharp decline, but rather either a small decline (Chen et al. 2012; Hojat et al. 2009), or non-statistically significant

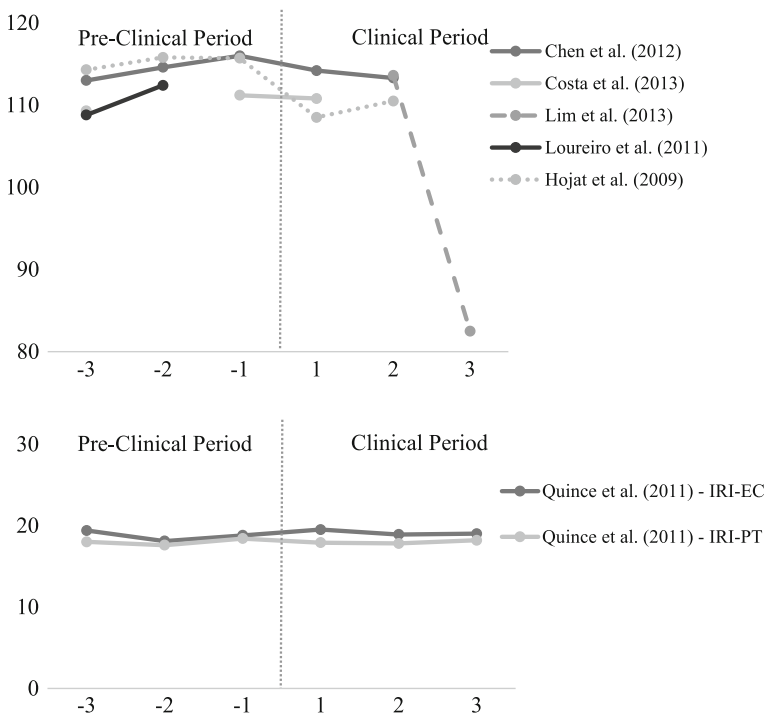


Fig. 2 Empathy growth for longitudinal studies. *Note:* the transition from pre-clinical to clinical period is taken as reference; results are presented for years before (*negative values* in the x-axis) and after (*positive values* in the x-axis); Thomazi et al. (2014) was not included due to insufficient information

variations (Costa et al. 2013; Quince et al. 2011) within that period. Pre-clinical studies showed a consistent trend of small empathy growth of participants (Chen et al. 2012; Costa et al. 2013; Loureiro et al. 2011). The study developed exclusively in the clinical period (Lim et al. 2013) reported a marked decline in empathy dissonant from all other studies.

Are empathy declines a general trend across countries in the scoped studies?

The studies conducted in Trinidad and Tobago identified declines in student empathy scores over time (Nunes et al. 2011), and a statistically significant lower empathy in third year students compared to first and second year students (Youssef et al. 2014). In New Zealand there were significant declines over time (Lim et al. 2013).

In the USA, Australia, Portugal and Spain there were contradictory results. In the USA, declines were observed in the two longitudinal studies using the JSE-S (Chen et al. 2012; Hojat et al. 2009). Yet, empathy scores increased slightly during pre-clinical training, and gradually declined after. In one cross-sectional study, however, older students reported higher empathy scores than younger ones (Toto et al. 2015). In Australia, participants' empathic concern, perspective taking and overall empathy (IRI and EQ-60 measures) were higher in years five and six as compared to one and two (Handford et al. 2013). However differences between groups were n.s., as measured by RMET (Handford et al. 2013), and by the JSE-S (Hegazi and Wilson 2013). In Portugal, studies pointed, overall, to a trend of non-significant variations/differences (Costa et al. 2013; Duarte et al. 2015) or slight increases of empathy scores (Loureiro et al. 2011; Magalhães et al. 2011). The study from Spain (Esquerda et al. 2016) evidenced mixed-results, with n.s. differences having been found in the total sample and among male students, while older female students reported higher empathy scores than younger ones.

Non-significant score variations were reported in studies from Brazil (Thomazi et al. 2014), the Dominican Republic (Silva et al. 2014), Venezuela (Montilva et al. 2015) and Poland (Bratek et al. 2015), and for cognitive empathy. in the UK (Quince et al. 2011). In the latter study, there was a statistically significant, albeit extremely small, decline of male students' affective component of empathy.

Finally, studies from Japan and China found increases in empathy scores (Kataoka et al. 2009; Wen et al. 2013). In China, JSE-S empathy scores of fourth year were higher than first year students. Similarly in Japan, JSE-S scores of first year UMS were the lowest, with the exception of scores of fourth year students.

The overall picture is very heterogeneous, and includes non-significant variations/differences, along with statistically significant positive or negative variations in empathy scores. The calculated effect sizes were small ($r^2 < 0.1$) for 16 studies suggesting little practical effects. Medium and very large size effects were detected only for the studies of Magalhães et al. (2011) and Lim et al. (2013). Overall, the predominant variations were small and with no practical significance.

Discussion

This international scoping review suggests that there is not a generalized trend in changes in student empathy throughout medical school. Many studies presented evidence suggestive of empathy non-significant variations, others of positive variations, but a minority of studies reported significant declines in empathy scores. In what concerns the pre-

clinical/clinical transition, even though most longitudinal studies failed to include such period, there was a dominant trend of small declines or of non-statistically significant variation in empathy. This scoping review highlighted large heterogeneities in study designs which are disadvantageous for the clarification of the research question. Globally, the adoption of common methodological frameworks and the implementation of comparable study designs in future research would be beneficial.

However our scoping review is limited to 20 self-report questionnaire-based studies, which might limit our inferences on if self-report measures scores reflect actual variations/differences in behaviors at the bedside (Colliver et al. 2010) and its generalizability, some tentative conclusions on empathy change over time or differences in empathy scores between cohorts might be drawn.

Firstly, in respect of research designs, all studies were based on questionnaires, using predominantly the JSE-S or the IRI. However, studies varied in key aspects such as sample sizes, type of study—cross-sectional or longitudinal—and timing of collection of measures. Few studies described the socio-demography of participants or the curriculum in terms of length or characterized the pre-clinical to clinical transition period. Such characterizations would be important to answer key questions concerning empathy development, in this case, related to the identification of influential key components in individual and institutional variables (Costa et al. 2014). To facilitate inter-study comparison, we suggest that future studies should report the following elements: socio-demographic characteristics of participants (e.g. age, sex and socio-economic status), year of study, characterization of the curriculum and measurement moments (with reference to the preclinical/clinical transition), descriptive statistics for empathy scores for each cohort and group by gender, (non)parametric tests of hypothesis (including effect sizes), and, for longitudinal studies, correlation coefficients between measures. Such characterization would facilitate the possibility of replication of longitudinal research designs across different institutions.

Secondly, this scoping review did not confirm the existence of a general international trend of a declines in student empathy during medical school, which has been assumed in previous studies (e.g. Hojat et al. 2002, 2004; Shariat and Habibi 2013; Stansfield et al. 2015). The studies in our final selection reported variations in empathy covering the whole spectrum—negative, positive and non-statistically significant change—regardless of the instrument applied, sample sizes or type of study. Identical findings apply to the pre-clinical/clinical transition. For example, there was an apparent negative effect in two studies conducted in the USA (Chen et al. 2012; Hojat et al. 2009), but non-statistically significant changes in Portugal (Costa et al. 2013) and in the UK (Quince et al. 2011).

The initial years in medical school seemed to be beneficial for empathy growth, which might be influenced by a ubiquitous presence of courses related to physician-patient interactions. For example, a study specifically mentioned that the growth was related to the attendance of a Medical Psychology course (Loureiro et al. 2011). Because the studies mostly use a scale that captures the cognitive components of attitudes towards empathy, it is unclear how the other dimensions and empathy might change. In contrast, three out of five longitudinal studies covering the clinical period evidenced a decline in empathy across time, with one study reporting a large effect size (Lim et al. 2013) suggestive of practical significance. The authors attributed the decline to the “the addition of clinical responsibility, higher patient contact and the implementation of diagnostic skills and management decisions are the major differences” (Lim et al. 2013, p. 949). This study does not exclude the possibility that cognitive component increases during the early years might be transient, eventually reach a plateau, and then decline later on in clerkships and beyond.

This review confirmed the existence of statistically significant differences between empathy measures, between cohorts and between assessment moments in the same study. However, the effect sizes were small for all but two studies (Lim et al. 2013; Magalhães et al. 2011). As a result, the practical significance of these differences is difficult to ascertain. Interestingly, none of the reviewed studies aimed at providing a fine grained view for the causality of empathy variations, but rather explored the timings—"when"—of empathy changes. We would argue that empathy research should increasingly focus on the causes—"why"—that underpin such changes. In this regard, it is imperative to that studies' consider the learning contexts, including the structure of curricula—in particular the presence and timing of courses related to physician-patient relationship—but also the real experience of the student—in particular how empathy is learned through supervised student-patient contact. The analysis of such variables could clarify inconsistencies in findings, as noticed by Hojat et al. (2009).

Thirdly, the present study could not draw conclusions about regional discrepancies in empathy variations. In effect, despite targeting literature in 4 languages, the pooled studies were restricted to 20 self-report questionnaire-based studies, which are not representative of the world cultures. Within these studies, those conducted in countries in the USA, the UK, Australia, New Zealand and Trinidad and Tobago revealed either decreases (Chen et al. 2012; Hojat et al. 2009; Lim et al. 2013; Nunes et al. 2011; Youssef et al. 2014) or n.s. (Handford et al. 2013; Hegazi and Wilson 2013; Quince et al. 2011) changes in empathy. On the contrary, the trends in studies conducted in China and Japan, were of empathy increases throughout medical school (Kataoka et al. 2009; Wen et al. 2013), while in South European and Latin South American countries, as well as Poland, present a trend of no variations or increase in empathy during medical school (Bravek et al. 2015; Costa et al. 2013; Duarte et al. 2015; Esquerda et al. 2016; Loureiro et al. 2011; Magalhães et al. 2011; Montilva et al. 2015; Silva et al. 2014; Thomazi et al. 2014).

The study's limitations include the non-systematic nature of scoping reviews, and the lack of an analysis of the quality of the research reports (Schwellnus and Carnahan 2014). Also, this scoping review addresses exclusively patterns of empathy development revealed by self-report questionnaires, which may not be the best proxies of patient's perceptions of students' empathy and of students' actual behavior when interacting with patients. Furthermore, we acknowledge the existence of empathy studies published in journals missed out by this study's inclusion criteria—for example, studies published in other languages. Nevertheless we consider that the option of screening literature in four idioms is a strong point of this study. Further positive aspects, were that decisions about literature selection involved three reviewers and that a scoping review methodology was carefully put in place. The adoption of a scoping review methodology was effective in terms of summarizing the research findings, identifying limitations in studies' methodologies and findings and provided a more rigorous vision of the international stat of the art.

Conclusions

Our findings, from different countries and institutions, questions the prevailing view that empathy declines occurs during medical school and is a generalizable trend across different countries. It represents a unique contribution which considers literature usually not considered in previous reviews. However, firm conclusions are dogged by uncertainty over the impact of culture and context, study design and measures, as well as practical significance

of measured change. To overcome the inherent limitation of scoping reviews, a systematic review with meta-analysis examining the practical significance of previous findings would enable more definitive conclusions on the existence or absence of a consistent trend in empathy growth throughout medical education. Such a review is warranted in the near future. As directions for future research, our study highlights the importance of a common framework.

Acknowledgments The authors thank Thelma Quince and Mohammadreza Hojat for critically reviewing the text.

References

- Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. doi:[10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616).
- Austin, E. J., Evans, P., Magnus, B., & O'Hanlon, K. (2007). A preliminary study of empathy, emotional intelligence and examination performance in MBChB students. *Medical Education*, 41(7), 684–689. doi:[10.1111/j.1365-2923.2007.02795.x](https://doi.org/10.1111/j.1365-2923.2007.02795.x).
- Bratek, A., Bulska, W., Bonk, M., Seweryn, M., & Krysta, K. (2015). Empathy among physicians, medical students and candidates. *Psychiatria Danubina*, 27(Suppl 1), S48–S52.
- Chen, D., Lew, R., Hershman, W., & Orlander, J. (2007). A cross-sectional measurement of medical student empathy. *Journal of General Internal Medicine*, 22(10), 1434–1438. doi:[10.1007/s11606-007-0298-x](https://doi.org/10.1007/s11606-007-0298-x).
- Chen, D. C. R., Kirshenbaum, D. S., Yan, J., Kirshenbaum, E., & Aseltine, R. H. (2012). Characterizing changes in student empathy throughout medical school. *Medical Teacher*, 34(4), 305–311. doi:[10.3109/0142159X.2012.644600](https://doi.org/10.3109/0142159X.2012.644600).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.
- Colliver, J. A., Conlee, M. J., Verhulst, S. J., & Dorsey, J. K. (2010). Reports of the decline of empathy during medical education are greatly exaggerated: A reexamination of the research. *Academic Medicine*, 85(4), 588–593.
- Costa, P., Alves, R., Neto, I., Marvão, P., Portela, M., & Costa, M. J. (2014). Associations between medical student empathy and personality: A multi-institutional study. *PLoS One*, 9(3), e89254. doi:[10.1371/journal.pone.0089254](https://doi.org/10.1371/journal.pone.0089254).
- Costa, P., Magalhães, E., & Costa, M. J. (2013). A latent growth model suggests that empathy of medical students does not decline over time. *Advances in Health Sciences Education*, 18(3), 509–522. doi:[10.1007/s10459-012-9390-z](https://doi.org/10.1007/s10459-012-9390-z).
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126. doi:[10.1037/0022-3514.44.1.113](https://doi.org/10.1037/0022-3514.44.1.113).
- Duarte, M. I., Branco, M. C., Raposo, M. L., & Rodrigues, P. J. (2015). Empathy in medical students as related to gender, year of medical school and specialty interest. *South East Asian Journal of Medical Education*, 9(1), 51.
- Esquerda, M., Yuguero, O., Viñas, J., & Pifarré, J. (2016). La empatía médica, ¿nace o se hace? Evolución de la empatía en estudiantes de medicina. *Atención Primaria*, 48(1), 8–14. doi:[10.1016/j.aprim.2014.12.012](https://doi.org/10.1016/j.aprim.2014.12.012).
- Handford, C., Lemon, J., Grimm, M. C., & Vollmer-Conna, U. (2013). Empathy as a function of clinical exposure-reading emotion in the eyes. *PLoS One*, 8(6), e65159. doi:[10.1371/journal.pone.0065159](https://doi.org/10.1371/journal.pone.0065159).
- Hegazi, I., & Wilson, I. (2013). Maintaining empathy in medical school: It is possible. *Medical Teacher*, 35(12), 1002–1008. doi:[10.3109/0142159X.2013.802296](https://doi.org/10.3109/0142159X.2013.802296).
- Hojat, M., Gonnella, J. S., Mangione, S., Nasca, T. J., Veloski, J. J., Erdmann, J. B., et al. (2002). Empathy in medical students as related to academic performance, clinical competence and gender. *Medical Education*, 36(6), 522–527. doi:[10.1046/j.1365-2923.2002.01234.x](https://doi.org/10.1046/j.1365-2923.2002.01234.x).
- Hojat, M., Mangione, S., Nasca, T. J., Rattner, S., Erdmann, J. B., Gonnella, J. S., et al. (2004). An empirical study of decline in empathy in medical school. *Medical Education*, 38(9), 934–941. doi:[10.1111/j.1365-2929.2004.01911.x](https://doi.org/10.1111/j.1365-2929.2004.01911.x).

- Hojat, M., Vergare, M. J., Maxwell, K., Brainard, G., Herrine, S. K., Isenberg, G. A., et al. (2009). The devil is in the third year: A longitudinal study of erosion of empathy in medical school. *Academic Medicine*, *84*(9), 1182–1191.
- Kataoka, H. U., Koide, N., Ochi, K., Hojat, M., & Gonnella, J. S. (2009). Measurement of empathy among Japanese medical students: Psychometrics and score differences by gender and level of medical education. *Academic Medicine*, *84*(9), 1192–1197.
- Kliszcz, J., Hebanowski, M., & Rembowski, J. (1998). Emotional and cognitive empathy in medical schools. *Academic Medicine*, *73*, 541.
- Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*(1), 69. doi:[10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69).
- Lim, B. T., Moriarty, H., Huthwaite, M., Gray, L., Pullon, S., & Gallagher, P. (2013). How well do medical students rate and communicate clinical empathy? *Medical Teacher*, *35*(2), e946–e951. doi:[10.3109/0142159X.2012.715783](https://doi.org/10.3109/0142159X.2012.715783).
- Loureiro, J., Gonçalves-Pereira, M., Trancas, B., Caldas-de-Almeida, J. M., & Castro-Caldas, A. (2011). Empathy in the doctor-patient relationship as viewed by first-year medical students: Data on validity and sensibility to change of the Jefferson measure in Portugal. *Acta Médica Portuguesa*, *24*, 431–442.
- Magalhães, E., Salgueira, A. P., Costa, P., & Costa, M. J. (2011). Empathy in senior year and first year medical students: A cross-sectional study. *BMC Medical Education*, *11*(1), 52.
- Mays, N., Roberts, E., & Popay, J. (2001). Synthesising research evidence. In N. Fulop, P. Allen, A. Clarke, & N. Black (Eds.), *Studying the organisation and delivery of health services: research methods* (pp. 188–220). London, UK: Routledge.
- Montilva, M., García, M., Torres, A., Puertas, M., & Zapata, E. (2015). Empatía según la escala de Jefferson en estudiantes de Medicina y Enfermería en Venezuela. *Investigación en Educación Médica*, *4*(16), 223–228. doi:[10.1016/j.riem.2015.04.006](https://doi.org/10.1016/j.riem.2015.04.006).
- Newton, B. W., Barber, L., Clardy, J., Cleveland, E., & O'Sullivan, P. (2008). Is there hardening of the heart during medical school? *Academic Medicine*, *83*(3), 244–249. doi:[10.1097/ACM.0b013e3181637837](https://doi.org/10.1097/ACM.0b013e3181637837).
- Nunes, P., Williams, S., Sa, B., & Stevenson, K. (2011). A study of empathy decline in students from five health disciplines during their first year of training. *International Journal of Medical Education*, *2*, 12–17. doi:[10.5116/ijme.4d47.ddb0](https://doi.org/10.5116/ijme.4d47.ddb0).
- Pedersen, R. (2009). Empirical research on empathy in medicine—A critical review. *Patient Education and Counseling*, *76*(3), 307–322. doi:[10.1016/j.pec.2009.06.012](https://doi.org/10.1016/j.pec.2009.06.012).
- Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods*, *5*(4), 371–385. doi:[10.1002/jrsm.1123](https://doi.org/10.1002/jrsm.1123).
- Quince, T. A., Parker, R. A., Wood, D. F., & Benson, J. A. (2011). Stability of empathy among undergraduate medical students: A longitudinal study at one UK medical school. *BMC Medical Education*, *11*(1), 90.
- Roff, S. (2015). Reconsidering the “decline” of medical student empathy as reported in studies using the Jefferson Scale of Physician Empathy-Student version (JSPE-S). *Medical Teacher*. doi:[10.3109/0142159X.2015.1009022](https://doi.org/10.3109/0142159X.2015.1009022).
- Roh, M.-S., Hahm, B.-J., Lee, D. H., & Suh, D. H. (2010). Evaluation of empathy among Korean medical students: A cross-sectional study using the Korean version of the Jefferson scale of physician empathy. *Teaching and Learning in Medicine*, *22*(3), 167–171. doi:[10.1080/10401334.2010.488191](https://doi.org/10.1080/10401334.2010.488191).
- Schweltnus, H., & Carnahan, H. (2014). Peer-coaching with health care professionals: What is the current status of the literature and what are the key components necessary in peer-coaching? *A scoping review. Medical Teacher*, *36*(1), 38–46. doi:[10.3109/0142159X.2013.836269](https://doi.org/10.3109/0142159X.2013.836269).
- Shariat, S. V., & Habibi, M. (2013). Empathy in Iranian medical students: Measurement model of the Jefferson scale of empathy. *Medical Teacher*, *35*(1), e913–e918. doi:[10.3109/0142159X.2012.714881](https://doi.org/10.3109/0142159X.2012.714881).
- Shashikumar, R., Chaudhary, R., Ryali, V. S. S. R., Bhat, P. S., Srivastava, K., Prakash, J., & Basannar, D. (2014). Cross sectional assessment of empathy among undergraduates from a medical college. *Medical Journal Armed Forces India*, *70*(2), 179–185. doi:[10.1016/j.mjafi.2014.02.005](https://doi.org/10.1016/j.mjafi.2014.02.005).
- Silva, M. G., Arboleda Castillo, J., & Díaz Narváez, V. P. (2014). Orientación empática en estudiantes de Medicina en una universidad de República Dominicana. *Educación Médica Superior*, *28*(1), 74–83.
- Stansfield, R. B., Schwartz, A., O'Brien, C. L., Dekhtyar, M., Dunham, L., & Quirk, M. (2015). 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*. doi:[10.1007/s10459-015-9605-1](https://doi.org/10.1007/s10459-015-9605-1).
- Thomazi, L., Moreira, F. G., & Marco, M. A. (2014). Avaliação da evolução da empatia em alunos do quarto ano da graduação em medicina da Unifesp em 2012. *Revista Brasileira de Educação Médica*, *38*(1), 87–93.

- Todres, M., Tsimtsiou, Z., Stephenson, A., & Jones, R. (2010). The emotional intelligence of medical students: An exploratory cross-sectional study. *Medical Teacher*, 32(1), e42–e48. doi:[10.3109/01421590903199668](https://doi.org/10.3109/01421590903199668).
- Toto, R. L., Man, L., Blatt, B., Simmens, S. J., & Greenberg, L. (2015). Do empathy, perspective-taking, sense of power and personality differ across undergraduate education and are they inter-related? *Advances in Health Sciences Education: Theory and Practice*, 20(1), 23–31. doi:[10.1007/s10459-014-9502-z](https://doi.org/10.1007/s10459-014-9502-z).
- Wen, D., Ma, X., Li, H., Liu, Z., Xian, B., & Liu, Y. (2013). Empathy in Chinese medical students: Psychometric characteristics and differences by gender and year of medical education. *BMC Medical Education*, 13(1), 130.
- Youssef, F. F., Nunes, P., Sa, B., & Williams, S. (2014). An exploration of changes in cognitive and emotional empathy among medical students in the Caribbean. *International Journal of Medical Education*, 5, 185–192. doi:[10.5116/ijme.5412.e641](https://doi.org/10.5116/ijme.5412.e641).