# The effectiveness of teaching clinical empathy to medical students: a systematic review and meta-analysis of randomized controlled trials

**Konstantinos C. Fragkos, MBBS, MSc, MA, PhD, and Paul E.S. Crampton, MSc, PhD**
*K.C. Fragkos is a Clinical Fellow in Gastroenterology at University College London Hospitals NHS Foundation Trust, London, UK. ORCID: https://orcid.org/0000-0002-7677-7989.*
*P.E.S. Crampton is a Lecturer at the Health Professions Education Unit, Hull York Medical School, York, UK; adjunct Research Fellow at University College London (UCL) Medical School, London, UK; and adjunct Research Fellow at Monash Centre for Scholarship in Health Education, Monash University, Victoria, Australia. ORCID: https://orcid.org/0000-0001-8744-930X.*

Correspondence should be addressed to Konstantinos C. Fragkos, University College London Hospitals NHS Foundation Trust, 250 Euston Road, London NW1 2PG, United Kingdom; e-mail: constantinos.frangos.09@ucl.ac.uk; Tel. +44 (0)7960 340489, Fax: +44 (0)20 344 79217; Twitter: @KCFragkos

## Abstract

Purpose: Clinical empathy is highlighted as a prerequisite for medical professionalism, despite being variously constructed and measured, and recently there has been an influx of randomized controlled studies investigating undergraduate interventions. The aim of the study was to examine whether undergraduate empathy interventions are effective and what factors serve as potential moderators.
Method: A systematic review was performed between 1948-2018 using database searching, citation tracking and hand-searching relevant journals. Key inclusion criterion was randomized controlled studies examining empathy intervention in medical students. Meta-analysis was performed with a random effects model to produce a pooled estimate of the standardized mean difference (SMD) followed by subgroup analyses.
Results: The search revealed 380 studies which after applying the inclusion criteria were reduced to 16 studies included in the meta-analysis ($n$=1,736). Quality assessment indicated the possibility of response and reporting bias. The pooled SMD was 0.68 [95% CI (0.43, 0.93)] indicating a moderately positive effect of developing empathy after the educational intervention compared to controls. There was no evidence of publication bias but heterogeneity was significantly high ($I^2$ = 88.5%). Subgroup analyses indicated that significant moderators for developing empathy were age, country, empathy measurement scope, type of empathy intervention, and rehearsal. However, moderators with limited evidence included gender, quality of studies and intervention characteristics.
Conclusions: Despite the described heterogeneity and biases, undergraduate empathy educational interventions are effective. The findings reinforce the current literature but adds considerable rigor as we performed a meta-analysis. A conceptual model is proposed for educationalists to consider when designing undergraduate empathy interventions.

# Introduction

Clinical empathy is commonly recognized as necessary to provide effective patient care,[1-3] yet there is an acknowledgement that physicians are often too detached and their approach to patients can be dispassionate, lacking empathy.[4,5] There are ongoing debates about definitions of clinical empathy, generally incorporating one or more of the following three features: thinking (cognitive), feeling (affective) and acting (behavioral).[6,7] Thinking and acting are most frequently cited, with just above ten percent of articles using all three features.[6] A multidimensional approach to empathy encompasses cognitive, affective and behavioral features.

Educational interventions to develop empathy are frequently implemented throughout medical education.[1] Designs vary with most intervention studies being non-controlled pre-post comparisons.[8,9] The interventions include experiential training, didactic methods, skills training, role-playing, mixed methods[10] as well as communication skills training with behavior-based workshops.[1] In terms of effectiveness, evidence although limited and heterogeneous, suggests that interventions improve empathy.[1,9,10] Limiting factors are variable samples, intervention conditions, empathy assessment and experimental design, which all result in a heterogeneous landscape which will now be addressed.[1,9,10]

Here we discuss a conceptual framework examining possible factors that affect the development of empathy namely demographics,[11-15] educational and intervention characteristics,[10,16,17] and study quality.[9,10] Sex has been described as a moderator of empathy in many studies, with females presenting generally a higher degree of empathy.[18-20] Age and ethnicity also have an effect on empathy.[18,19,21] White undergraduate medical students have scored higher empathy scores than white Asian Americans,[19] while male black/African American students had the lowest scores of empathy, but the authors noted a gender/ethnicity interaction, suggesting possible ethnicity and gender biases.[18] In terms of age, younger people (less than 30 years old or premedical students) have frequently exhibited higher empathic scores than older ones (over 50 years old or fourth year medical students).[22] This is consistent with the frequently cited empathy decline during medical education.[23,24]

In terms of intervention characteristics, empathy training belongs to the larger field of workplace learning and development[16] and follows the principles of behavior modelling training.[17] Such training commonly defines distinct behaviors (skills) to be learned, provides examples/models displaying effective use of those behaviors, allows opportunities to practice and feedback, and supports learners to transfer behaviors to practice.[25,26] Within this framework, length of training and time after which the effects of an empathy intervention are measured, are important variables.[16] A meta-analytic review suggested that training knowledge appeared to diminish post-training, but newly learned skills were maintained or even increased over time.[17] Other authors have found no association of time and empathy effectiveness between baseline and post-test measurements.[10] Compensation on the other hand is generally considered to affect participation because of the presence of participation bias.[10] Finally, empathy studies have not shown a difference in empathy depending on whether there was an active control group or waiting list.[10] The ideal control group is one where outcome expectations between intervention and control group are equivalent. Although an arguably qualitative difference, this is not the case with a waiting list, thus impeding genuine causal inferences between training and outcomes.[27]

Study quality examining empathy interventions varies with limitations including lack of a control group, non-randomized design, conducted at a single institution, lack of pre-intervention or baseline measurement, and measurement of attitudes rather than skills or patient outcomes.[1,9,10,28] This limits the ability of the literature to produce clear implications for whether empathy increases with training intervention.

Hence, the aim of the current study is to examine whether empathy interventions among medical students are effective and how do confounding factors potentially moderate this effect. A meta-analysis of randomized controlled studies will determine how effective empathy interventions are and how demographics, educational and intervention characteristics, and study quality impact this effect.

# Methods

A systematic review and meta-analysis was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[29] Electronic database searches were conducted between 1 January 1948 and 31 January 2018. Inclusion criteria were randomized controlled study designs only, which examined empathy interventions in medical students (Figure 1A). Quality of studies was assessed using the Medical Education Research Study Quality Instrument (MERSQI).[30] For the meta-analysis, standardized mean differences (SMDs) were extracted. A random effects model was used to produce a pooled estimate of the SMDs. Statistical heterogeneity was assessed with the $I^2$ statistic[31] and further investigated with subgroup analysis and meta-regression. Publication bias was assessed using funnel plots, Egger's test, Begg's test, Rosenthal's number and the trim-and-fill method.[32-36] Meta-analysis was performed with Stata 15.0 (Stata Corp., College Station, Texas) and R 3.5.1 (R Core Team, Vienna, Austria). Detailed methods are described in Supplemental Digital Appendix 1.

# Results

## Study characteristics

The initial search revealed 380 studies which after reviewing by title/abstract and applying the inclusion criteria were reduced to 16 studies[37-52] included in the meta-analysis. The flowchart is shown in Figure 1B and study information for included studies is in Supplemental Digital Appendix 2 (see also Supplemental Appendix 3 for the whole dataset). All studies were randomized controlled trials that were published between 2008 and 2018 (six were published in 2017). Over 80% of studies (*n*=14) were performed at one institution[37,39-46,48-52] rather than multiple (*n*=2).[38,47] Most studies (*n*=5, 31.3%) were performed in the USA.[37,40,46,47,52] The total number of participants was 1,736 (range 13-299, 876: intervention groups and 860: controls) with a mean age of 23.6 years. Females comprised 59% of the total sample (*n*=705/1187). Most students were in their third (*n*=635, 36.6%) or fourth year (*n*=479, 27.6%) (Supplemental Digital Appendix 4). To account for international differences, pre-clinical students were considered years one/two/three; clinical students were considered years four/five. Only one study[42] reported long-term follow-up (2 years) post publication of their original study.[53]

**A**

**Study eligibility criteria**

*Inclusion criteria*

1. Randomized controlled study designs examining empathy interventions in medical students
2. English language papers (1 January 1948 till 31 January 2018)
3. Sufficient data for effect measure estimation
4. Clinical empathy considered as a component driven from medical students towards patients
   General concept and not its sole separated concepts (e.g. solely cognitive or affective or behavioral empathy)

*Exclusion criteria*

1. Examining other types of empathy [empathy from professionals (doctors, nurses, allied healthcare professionals, etc.) towards patients; empathy from students not from healthcare; empathy from teachers toward students (teacher empathy); perceived empathy of patients from their caring professionals; empathy among adolescents and high school students; cross-cultural empathy; multicultural empathy; empathy towards particular medical conditions from non-medical students]
2. Other study types: ross sectional studies, case studies, pre-post experimental designs, qualitative studies, reviews, non-English studies

**Search strategy and terms**

*Keywords*: empathy, caring, humanism, cognitive, emotional, healthcare, medical students, compassion, care, randomized

*Electronic databases:* PubMed/Medline, Scopus, EMBASE, Google Scholar, ERIC, publisher databases (ScienceDirect, Springer Link, Wiley Online Library, Taylor & Francis Online, and Oxford Academic)

*Hand-searching:* Included manuscripts' bibliographies and healthcare education journals (Academic Medicine, Advances in Health Sciences Education, BMC Medical Education, International Journal of Medical Education, Medical Education, Medical Education Online, Medical Teacher, Teaching and Learning in Medicine, Perspectives on Medical Education)

**Study selection, data extraction and quality assessment**

Both authors screened abstracts/full text and performed data extraction

*Descriptive information* [study aim, country, sample size, mean age, gender distribution, student distribution per year of study, journal Impact Factor (2017), article impact ratio]

*Empathy features* [Subtypes (Cognitive, affective, behavioral, or multidimensional); Measure scope (narrow, broad); Self-reported vs objective measure; Scale characteristics (name, reliability, number of items)]

*Empathy intervention* [Type (experiential, didactic, skills training, mixed methods); Behavior skills training (modelling, instructions, rehearsal, feedback); Compensation; Control (waiting list, active); Intervention length (hours trained); Months pre-post measurement; Effect size

*Quality assessment* [Medical Education Research Study Quality Instrument (MERSQI)]

**B**

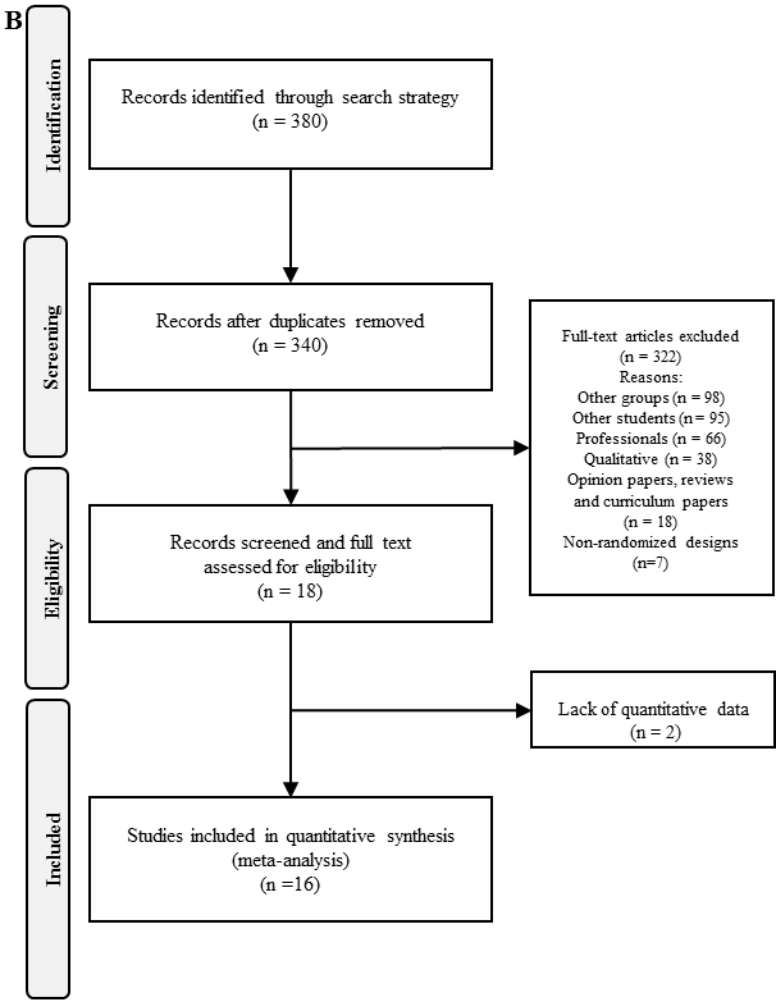



**Figure 1.** A. Eligibility criteria, search strategy and data extraction for current study. B. Flow diagram of the study.

## Study quality

The median MERSQI score was 13.0 [Interquartile range (IQR) 12.0-15.5] and the risk of bias graph is shown in Figure 2 (see Supplemental Digital Appendix 3 for the assessment of each individual study). Eleven studies (69%) had good response rates including over 75% of their samples.[38-41,43,45-47,49,51,52] Self-reported measures of empathy were present in 56% of studies (*n*=9),[37,41,42,44-48,52] the rest of the studies assessed empathy using more objective measures (experts or standardized patients).[38-40,43,49-51] The internal structure was not described in most studies, but their content and relationship to other variables was high. There is a possibility of reporting bias[54] due to under-reporting of internal structure and relationships to other variables, and response bias due to self-reported data. Finally, most studies assessed knowledge, skills or behaviors but no study assessed patient outcomes (Figure 2).

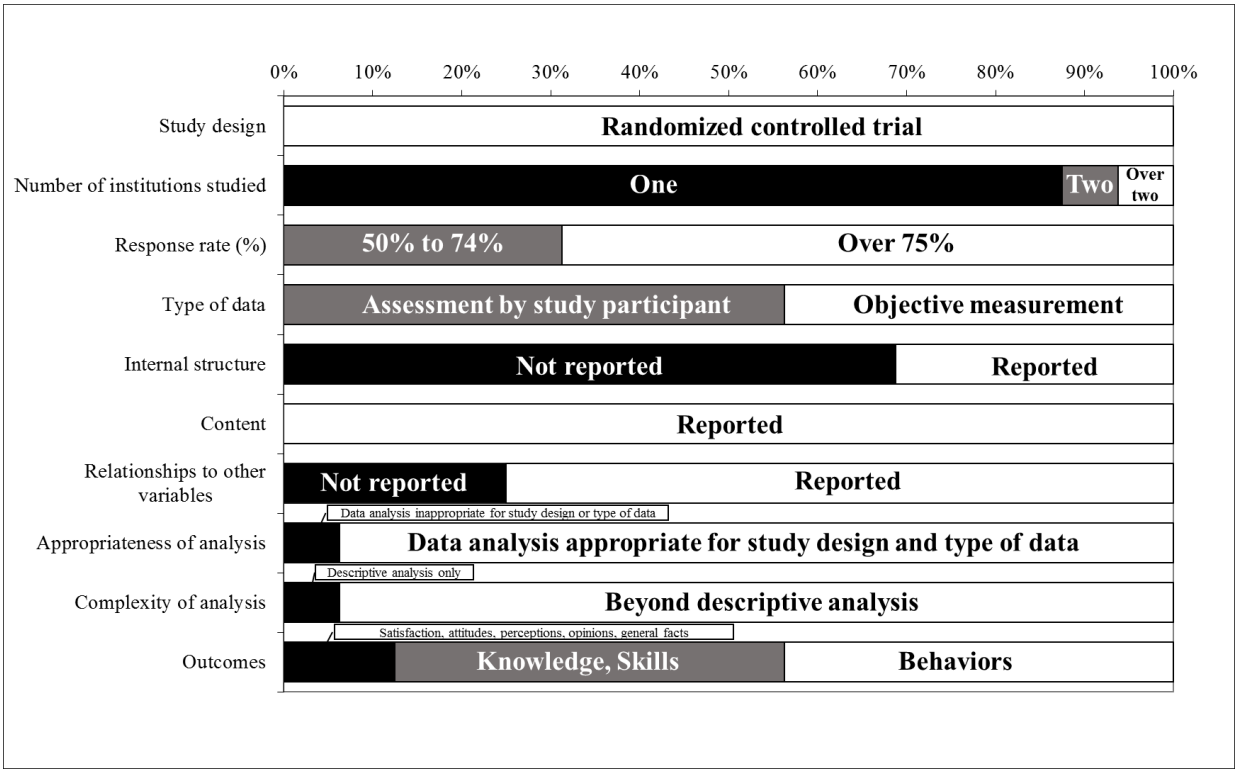



**Figure 2.** Risk of bias graph. Review author judgements about each risk of bias item presented as percentages across all included studies (all studies). Black indicates the lowest weight for that domain, grey a middle weight and white a higher weight.

## How effective are empathy interventions?

Empathy interventions were typically assessed by multiple methods. Six studies used more than one method to assess empathy.[38-40,43,45,51] The random-effects meta-analysis of SMDs at the study level produced a pooled effect of 0.68 [95% Confidence Interval (CI) (0.43, 0.93), range (-0.30, 2.77)], which indicates a moderately positive effect of developing empathy after an intervention compared to controls (Figure 3A). Moreover, there was no evidence of publication bias [funnel plot symmetry (Supplemental Digital Appendix 5), Egger's test $p$=0.66, Begg's test $p$=0.56, Fail-safe $N$=1,258] but heterogeneity was significantly high ($I^2$ = 88.5%, $p$<0.01). Cumulative meta-analysis by year showed that essentially after the study by Singh et al.,[41] the effect size remained steadily above 0.60 (Supplemental Digital Appendix 6). Also, a sensitivity analysis indicated that omission of single studies did not change substantially the main effect, apart from the omission of the study by Singh et al.,[41] which reduced the effect to 0.55 [95% CI (0.35-0.76)] (Supplemental Digital Appendix 7).

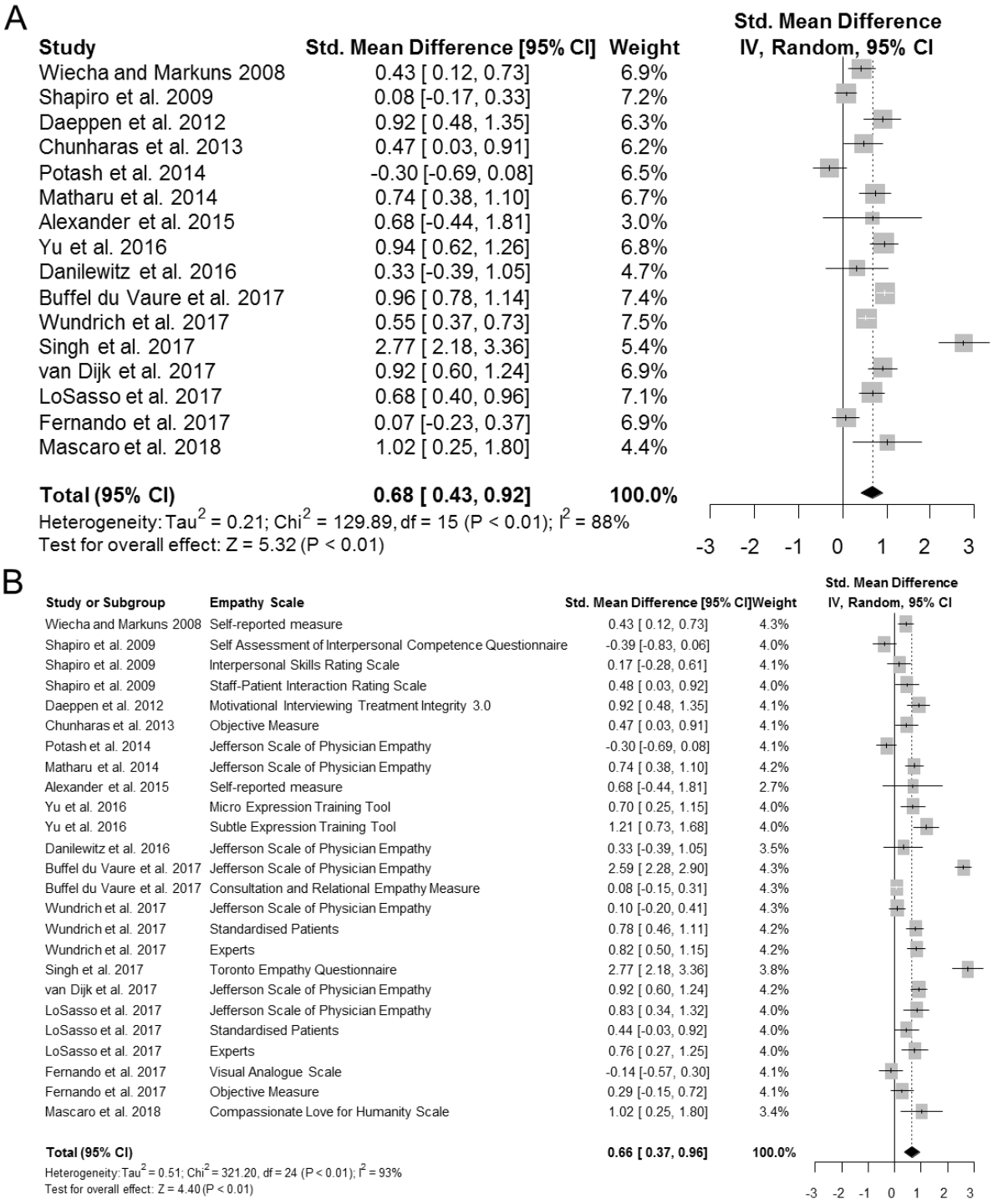



**Figure 3.** A. Forest plot of SMD for empathy development. B. Meta-analysis of all SMDs (*k*=25) for each empathy measure described (some studies have measured empathy with more than one measures). IV: Inverse Variance.

## How do demographic characteristics impact on empathy interventions' effect?

The USA studies[37,40,46,47,52] had an overall effect of 0.63 [95% CI (0.46, 0.80)] with no heterogeneity ($I^2$=0.0%). Meta-analysis indicated that the SMD was significantly different between countries ($p$<0.01). When examined by continent, the largest effect was noted in Europe[38,42,43,50] [SMD=0.82, 95% CI (0.58, 1.06), $I^2$=72.8%], with no significant difference between continents nevertheless. Age seemed to effect outcome ($p$<0.01), with students over the mean having an effect of 0.52 compared to 0.07. Sex distribution was not a significant moderator of the outcome ($p$=0.36). The journal's impact factor had no significant effect on the SMD. Studies from journals with no impact factor had heterogeneity 0.3%,[44,47] possibly reflecting the use of the Jefferson Scale of Physician Empathy (JSPE) (its effect is discussed below) (Table 1, Supplemental Digital Appendices 8, 9 and 10).

**Table 1**. Subgroup analyses of all categorical moderators.

| Variable | k | N | Effect Size | | Heterogeneity | |
|---|---|---|---|---|---|---|
| | | | SMD (95% CI) | *p* | $I^2$ | *p* |
| ***Demographics*** | | | | | | |
| *Country[Q(10) = 126.28, p < 0.01]* | | | | | | |
| Canada | 2 | 109 | 0.11 (-0.13, 0.35) | 0.38 | 0.0% | 0.52 |
| France | 1 | 299 | 0.96 (0.78, 1.15) | <0.01 | -- | -- |
| Germany | 1 | 158 | 0.55 (0.37, 0.73) | <0.01 | -- | -- |
| Hong Kong | 1 | 106 | -0.30 (-0.69, 0.08) | 0.13 | -- | -- |
| India | 1 | 93 | 2.77 (2.18, 3.36) | <0.01 | -- | -- |
| Netherlands | 1 | 167 | 0.92 (0.60, 1.24) | <0.01 | -- | -- |
| New Zealand | 1 | 83 | 0.07 (-0.24, 0.38) | 0.65 | -- | -- |
| South Korea | 1 | 82 | 0.94 (0.62, 1.27) | <0.01 | -- | -- |
| Switzerland | 1 | 91 | 0.92 (0.48, 1.35) | <0.01 | -- | -- |
| Thailand | 1 | 89 | 0.47 (0.03, 0.91) | 0.036 | -- | -- |
| USA | 5 | 459 | 0.63 (0.46, 0.80) | <0.01 | 0.0% | 0.53 |
| *Continent [Q(2) = 2.97, p= 0.23]* | | | | | | |
| Asia and Oceania | 5 | 453 | 0.77 (-0.05, 1.58) | 0.07 | 95.5% | <0.01 |
| Europe | 4 | 714 | 0.82 (0.58, 1.06) | <0.01 | 72.8% | 0.01 |
| North America | 7 | 668 | 0.51 (0.25, 0.77) | <0.01 | 61.6% | 0.02 |
| *Age [Q(1) = 27.85, p<0.01]* | | | | | | |
| < 23.5 years old | 4 | 357 | 0.07 (-0.19, 0.32) | 0.60 | 55.6% | 0.08 |
| ≥ 23.5 years old | 5 | 501 | 0.52 (0.20, 0.84) | <0.01 | 0.0% | 0.92 |
| *Impact Factor [Q(2) = 0.06, p=0.97]* | | | | | | |
| No Impact Factor | 2 | 159 | 0.66 (0.34, 0.98) | <0.01 | 0.3% | 0.32 |
| Low Impact Factor | 5 | 506 | 0.70 (-0.12, 1.53) | 0.09 | 95.1% | <0.01 |
| High Impact Factor | 9 | 1071 | 0.71 (0.50, 0.91) | <0.01 | 75.9% | <0.01 |
| ***Empathy Characteristics*** | | | | | | |
| *Cognitive Empathy* | | | | | | |
| Yes | 16 | 1,736 | 0.68 (0.43, 0.93) | <0.01 | 88.5% | <0.01 |
| No | -- | -- | -- | -- | -- | -- |
| *Affective Empathy [Q(1) = 0.16, p=0.69]* | | | | | | |
| Yes | 13 | 1,436 | 0.67 (0.37, 0.97) | <0.01 | 90.4% | <0.01 |
| No | 3 | 300 | 0.75 (0.49, 1.01) | <0.01 | 30.6% | 0.24 |
| *Behavioral Empathy [Q(1) = 0.45, p=0.50]* | | | | | | |
| Yes | 12 | 1,321 | 0.73 (0.45, 1.01) | <0.01 | 89.4% | <0.01 |
| No | 4 | 415 | 0.49 (-0.14, 1.13) | 0.13 | 88.0% | <0.01 |
| *Multidimensional Empathy [Q(1) = 0.38, p=0.54]* | | | | | | |
| Yes | 10 | 1,150 | 0.74 (0.41, 1.07) | <0.01 | 91.0% | <0.01 |
| No | 6 | 586 | 0.57 (0.161, 0.98) | <0.01 | 83.4% | <0.01 |
| *Empathy Measure [Q(2) = 0.98, p =0.61]* | | | | | | |

| Variable | k | N | Effect Size | | Heterogeneity | |
|---|---|---|---|---|---|---|
| | | | SMD (95% CI) | *p* | $I^2$ | *p* |
| JSPE | 4 | 432 | 0.43 (-0.16, 1.03) | 0.15 | 88.1% | <0.01 |
| JSPE + Other measure | 3 | 527 | 0.73 (0.46, 1.00) | <0.01 | 79.8% | <0.01 |
| Other | 9 | 777 | 0.79 (0.34, 1.24) | <0.01 | 90.9% | <0.01 |
| *Self-Report vs Objective Measure Used [Q(2) = 1.53, p = 0.47]* | | | | | | |
| Mixed | 5 | 689 | 0.48 (0.14, 0.81) | <0.01 | 90.5% | <0.01 |
| Self-reported | 9 | 867 | 0.83 (0.35, 1.30) | <0.01 | 90.3% | <0.01 |
| Objective | 2 | 180 | 0.69 (0.26, 1.13) | <0.01 | 50.2% | 0.16 |
| *Empathy Measure Scope [Q(1) = 5.77, p = 0.02]* | | | | | | |
| Broad | 12 | 1,336 | 0.77 (0.48, 1.07) | <0.01 | 90.3% | <0.01 |
| Narrow | 4 | 400 | 0.31 (0.08, 0.54) | <0.01 | 22.0% | 0.28 |
| ***Empathy Training Characteristics*** | | | | | | |
| *Type of Empathy Training [Q(3) = 11.49, p <0.01]* | | | | | | |
| Experiential Training | 6 | 870 | 0.53 (0.14, 0.92) | <0.01 | 87.0% | <0.0001 |
| Mixed | 7 | 612 | 0.92 (0.46, 1.39) | <0.01 | 91.9% | <0.0001 |
| Didactic | 1 | 83 | 0.07 (-0.24, 0.38) | 0.65 | -- | -- |
| Skills Training | 2 | 171 | 0.73 (0.27, 1.19) | <0.01 | 65.0% | 0.09 |
| *Control Type [Q(1) = 0.13, p = 0.71]* | | | | | | |
| Waiting List | 4 | 440 | 0.59 (-0.003, 1.17) | 0.05 | 90.5% | <0.01 |
| Active | 12 | 1,296 | 0.71 (0.43, 0.93) | <0.01 | 88.8% | <0.01 |
| *Compensation [Q(1) = 0.38, p = 0.54]* | | | | | | |
| Yes | 4 | 282 | 0.78 (0.43, 0.93) | <0.01 | 0.0% | 0.51 |
| No | 12 | 1,454 | 0.66 (0.36, 0.95) | <0.01 | 91.3% | <0.01 |
| *Use of Four Components of Training [Q(1) =0.54, p = 0.46]* | | | | | | |
| Yes | 7 | 738 | 0.80 (0.43, 0.93) | <0.01 | 91.6% | <0.01 |
| No | 9 | 998 | 0.60 (0.30, 0.90) | <0.01 | 85.6% | <0.01 |
| Modelling [Q(1) = 0.02, p = 0.88] | | | | | | |
| Yes | 14 | 1,577 | 0.69 (0.42, 0.97) | <0.01 | 89.9% | <0.01 |
| No | 2 | 159 | 0.66 (0.34, 0.98) | <0.01 | 0.3% | 0.32 |
| Instruction | | | | | | |
| Yes | 16 | 1,736 | 0.68 (0.43, 0.93) | <0.01 | 88.5% | <0.01 |
| No | -- | -- | -- | -- | -- | -- |
| Rehearsal [Q(1) = 3.24, p = 0.07] | | | | | | |
| Yes | 14 | 1,415 | 0.77 (0.51, 1.03) | <0.01 | 87.6% | <0.01 |
| No | 2 | 321 | 0.07 (-0.64, 0.78) | 0.84 | 88.1% | 0.04 |
| Feedback [Q(1) = 0.24, p = 0.62] | | | | | | |
| Yes | 8 | 968 | 0.75 (0.32, 1.18) | <0.01 | 90.2% | <0.01 |
| No | 8 | 768 | 0.62 (0.30, 0.93) | <0.01 | 87.2% | <0.01 |

Abbreviations: k: number of studies; N: number of participants; SMD: standardized mean difference; CI: confidence Interval; P: p-value; $I^2$: I-squared heterogeneity statistic

## What is the effect of year of study on empathy interventions' outcome?

Meta-regression was performed with the number of students per year. Only the number of year two students seemed to moderate the effect on empathy, with higher effects being exhibited when more year two students were present (p<0.01) (Supplemental Digital Appendix 10). There was no effect of the number of pre-clinical and clinical year students on empathy (Supplemental Digital Appendix 8).

## Which assessment methods (scale, definition, scope) affect empathy interventions' effect?

The operating definitions for each study are presented in Supplemental Digital Appendix 11. The included studies reflect a general debate that empathy and similar concepts (e.g. compassion, humanism, communication) are linked through their definition but also educational and professional consequences. Examining their structure, exemplary phases and their assessment scale, the elements of empathy are readily recognized (one or more of them) in all studies. However, it is essential to understand how many studies used empathy's different elements. The cognitive element was considered by all studies, the affective element by 13 studies, the behavioral element by 12 studies and multidimensional empathy (cognitive, affective, and behavioral) was considered by 10 studies (Table 1). When examining the effect by each element considered, there were no significant differences between having an element or not ($p$>0.05 for all elements, Supplemental Digital Appendix 8, Supplemental Digital Appendix 12).

When the SMD was examined by the empathy measure used (JSPE, JSPE+other, other), there was no significant difference ($p$=0.61); nevertheless, four studies[42,44,47,48] that measured empathy with JSPE alone had the lowest SMD=0.43 [95% CI (-0.16, 1.03)] (Supplemental Digital Figure 7). When performing meta-analysis of the SMDs with regards to each empathy measure (hence $k$=25, because a few studies used more than one way to measure empathy, as mentioned previously), the overall effect is 0.66 [95% CI (0.37-0.96)] (Figure 2B), which is very close to the previously reported SMD=0.69. Cumulative meta-analysis by years, shows much clearer here that the overall effect doesn't change essentially after Buffel du Vaure et al.[38] (Supplemental Digital Appendix 13, figure panel A). Self-reported measures of empathy exhibit the largest effect [SMD=0.83, 95% CI (0.35, 1.30)] but are not significantly different from objective measures or papers that included both methods (mixed) ($p$=0.47) (Table 1, Supplemental Digital Appendix 13, figure panel B). Finally, narrow-scope empathy measures (including one item) exhibit a lower effect [SMD=0.31, 95% CI (0.08, 0.54)] than broad-scope measures, but they are more homogenous ($I^2$=22.0%) (Table 1, Supplemental Digital Appendix 13, figure panel C).

## How do intervention characteristics (design, length, duration of effect, compensation) impact on empathy interventions' effect?

Type of empathy intervention was a significant moderator of the overall effect on empathy ($p$<0.01) (Table 1, Supplemental Digital Appendix 14). Didactic methods were not significant in producing an effect and the largest effect was exhibited by a combination of methods (mixed) [(SMD=0.92, 95% CI (0.46, 1.39)]; however, heterogeneity was present at 91.9%. Skill training had the second largest effect followed by experiential training. The type of control did not affect the outcome ($p$=0.71) and neither did compensation for participation in the educational intervention ($p$=0.54). The use of the four components of behavioral training was not a significant moderator overall, but only rehearsal seemed to produce a significant effect on stronger empathy development ($SMD_{rehearsal}$=0.77 vs $SMD_{no\ rehearsal}$=0.07). The duration of the education intervention and the months pre-post assessment were also not significant moderators of the effect size ($p$>0.05 for both).

## What is the impact of study quality on empathy interventions' effect?

Subgroup analysis was performed for each MERSQI item and differences were not statistically different between the options of each domain ($p$>0.05 for all items; Supplemental Digital Appendix 15). Meta-regression of the sum of the MERSQI items taken as a continuous score showed that it did not affect significantly the effect sizes of educational interventions for empathy development (Supplemental Digital Appendix 10). Although single center studies frequently inflate the overall effect size, interestingly, the SMD for more than one institutions was larger (SMD=0.91)[38,47] than the one from single-center studies (SMD=0.65),[37,39-46,48-52] supporting that no small study effects were present.

# Discussion

To our knowledge, we have performed the first systematic review and meta-analysis of randomized controlled studies of clinical empathy educational interventions amongst medical students. Previously there has been only one meta-analysis examining empathy interventions however this included multiple types of participants (professionals, students, citizens, children)[10] whilst other systematic reviews did not include meta-analysis limiting their generalisability.[1,8,9,28] The current study is strengthened by including only randomized controlled trials,[55] and an above average MERSQI median study score of 13.0, which was larger than a previously reported 11.3.[56]

In summary, the meta-analysis shows that educational interventions had a significant moderate positive effect on increasing empathy, in agreement with other systematic reviews.[1,8,9,28] It is important, however, to consider this effect within the context of measuring clinical empathy. Although a broader empathy measurement tool may be useful and desirable from a theoretical standpoint, they exhibit significant heterogeneity,[57] as shown in the present meta-analysis. Narrow scope assessment tools had less heterogeneity but possibly underestimated the effect of empathy interventions, due to their narrow focus. Another aspect of assessing empathy concerns whether participants filled in self-reported questionnaires or were assessed by experts or standardized patients. In the present meta-analysis, objective vs self-reported questionnaires effect sizes were not significant, however, self-reported instruments had a tendency towards higher effect sizes, possibly explained by participants' desire to respond in such a way to avoid criticism. Empathy is generally considered as a positive personality trait, especially for doctors.[24] Self-reported empathy has correlated with social desirability,[58] which in turn has been noted to be inversely associated with empathic concerns in medical students.[59] However, this association has not been universally described[60] and also simply attributing empathy changes to social desirability tendencies is not adequate to explain empathy decline which has been correlated with emotional intelligence or moral judgement competence as well.[24,61] Finally, there seems to be a gender effect regarding social desirability bias, with female responses correlating with social desirability.[62] In our study, females were 59% (705/1187) of the sample, which could have affected the results.

Moving onto empathy assessment tools, the variety of measures has been described as an existing issue in the medical literature.[6,63] The use of empathy measurement tools is also closely connected to the definition of empathy used. Most studies in the present review followed a multidimensional empathy definition, constructively aligned with the complexity of clinical empathy. The need for consistency and broad scope of empathy measurement tools has been noted in other fields as well.[64,65] Although most tools are valid and are based on rigorous psychometric testing, there is debate as to how such heterogeneous instruments can be used in clinical care, medical student education and medical school admissions.[63] Our results showed no superiority of any tool against another in terms of effect but heterogeneity was present throughout studies, hence stressing the need for a more consistent less heterogeneous instrument.

Regarding the behavior modelling training aspects of the education interventions examined, they were more effective in developing empathy when rehearsal was present and where mixed training techniques (experiential, skilled, didactic) were used. This has been replicated in other studies as well.[9,10] It is generally expected that mixed model training techniques will increase the production of the intended behaviors and skills[17,66] and rehearsal will enhance retention of the outcomes of the education interventions.[67]

Next, one interesting finding here is that older age (≥ 23.5 years old) was associated with a stronger effect of the empathy interventions examined. Usually older ages correspond to a higher year of student and hence more senior year students might respond better to empathy intervention. However, medical students' admissions age is generally higher in the USA compared to other countries.[68,69] Also, empathy decline in senior medical students is not mutually exclusive with the effect of an empathy intervention being larger in senior students. Hence, trying to interpret this finding, we need to consider the growth curve of adult intellectual development.[70] The participants analyzed belong to the early adult group during where occupational knowledge (in our case, clinical empathy) increases until it reaches a

plateau at middle adult age.[70] Social desirability response bias should also be considered as a possible confounding factor for this finding since it has been suggested to increase with age.[71] Also, through the possibility of response shift bias, older students might be aware of empathy and its importance, therefore more receptive to training.

Contextually, there was a higher effect of empathy interventions in studies from Europe followed by Asia/Oceania and North America. It is usual that between different countries, design, measurement methods, and psychometric properties of measurement scales are existent.[72] These cross-cultural differences have also been suggested to be due to nonverbal expressions of empathy, some of which are culturally specific.[73]

The non-significant predictors of empathy intervention effectiveness [gender, quality of studies as assessed by the MERSQI, journal impact factor, and intervention characteristics (length, control type, duration of effect, compensation)] could be attributed to the described heterogeneity between studies, which could cause underestimation or non-significance. The lack of establishing an effect for quality of studies is possibly due to the higher median MERSQI of the studies included in the meta-analysis, which could be a result of the inclusion criteria.

Based on the findings of the present study, we propose a model highlighting the implications which is shown in Figure 4. This model suggests educationalists should design empathy interventions which include rehearsal, are delivered using a mixture of pedagogic techniques (experiential, didactic and skills training) and are aimed primarily at senior medical students towards the end of medical school. Moreover, empathy should be assessed objectively by experts or standardized patients, and the definition of empathy should be broad so that it encompasses multidimensional elements (cognitive, affective, and behavioral).

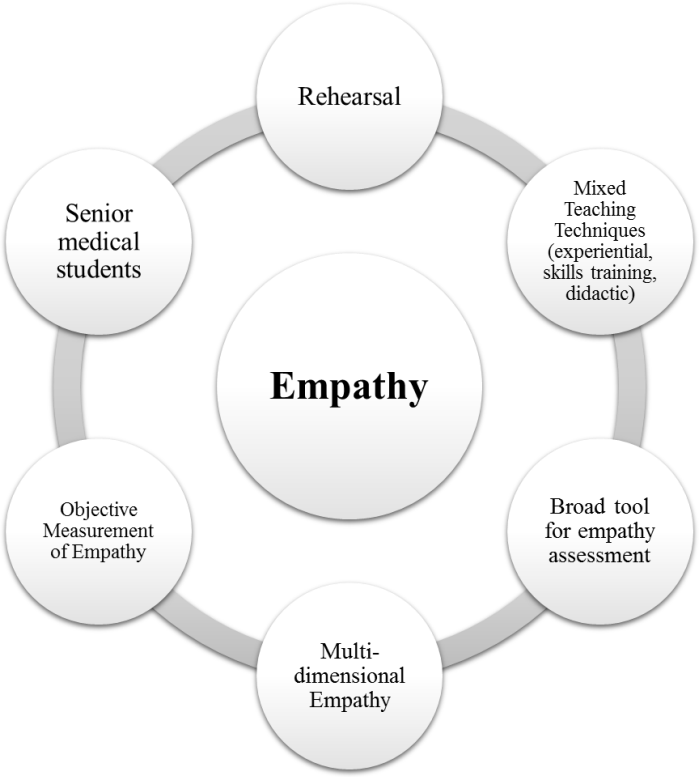



**Figure 4.** Suggested model for empathy development in medical students.

## Limitations

This literature review included only randomized controlled trials which is both a strength and a weakness.[55] In education, it has been argued that randomized controlled trials have confused and trivial results, with possibly zero practical application to the field of human affairs.[74-76] Empathy is a spuriously difficult construct to define and hence any standardization is limited by the tools which are implemented to test, measure and research its properties. Additionally, blinding is not possible in education which can introduce a further limitation,[77] while lack of long term follow up data does not provide insight into whether these behavioral or cognitive changes are maintained. The results of the present meta-analysis had increased heterogeneity addressed by using a random effects model, whilst subgroup analyses were not adjusted for multiple testing. Finally, the risks of design biases, response shift bias, and social desirability response bias were possibly present. Response biases occur when individuals offer biased estimates of self-assessed behavior (e.g. due to misunderstanding or social desirability) which can be augmented during intervention studies.[78] These biases are supported by the meta-analysis performed which indicated that SMDs were higher with studies that used self-assessment instruments and did not report relationships to other variables (Table 1). Finally, most studies assessed knowledge, skills or behaviors but no study assessed patient outcomes (Figure 2). This limits the practical applications of these educational interventions for patients, but interestingly the highest SMD was noted in studies that aimed to assess behavioral changes (0.86) compared to knowledge development (0.56) or satisfaction and attitudes (0.44).

## Conclusion

The results of the meta-analysis indicate that undergraduate empathy educational interventions significantly increase student empathy compared to controls. A range of moderating variables impacted on the effectiveness of empathy intervention, which included age, country, empathy measurement scope, type of empathy intervention and rehearsal. However, moderators with limited evidence included gender, quality of studies assessed by the MERSQI, journal impact factor and intervention characteristics. Based on the findings we propose a model highlighting the implications of the study.

The study had limitations that stemmed from biases inherent to the design of each trial which possibly added up during the meta-analysis. Hence, future research should focus on eliminating design biases during empathy measurement or development studies. Currently, it is unclear whether these educational interventions should be compulsory or not in undergraduate in medical education. Although there is a regulatory and public need for clinical empathy,[79-82] its development has also been shown to be a cultural trait with different weights depending on circumstances.[73,83] Nevertheless, improving empathy should be an essential aim of undergraduate medical education,[84] since empathy has been linked to positive outcomes[6,10,85,86] and addressing empathy education may also reduce the empathy decline documented in medical education.[24]


# Acknowledgements

We thank Dr Asta Medisauskaite and Dr Shah-Jalal Sarker for providing constructive feedback during the writing of the present study.

# Funding/Support

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.


# Other Disclosures

The authors declare no conflict of interest.

## Previous Presentations
This work was completed as partial fulfilment of a Masters in Clinical Education at UCL Medical School.

## Ethical Approval
Not applicable for this study.

## Disclaimers
None.

## Data
Not applicable for this study.

## Data Sharing
All data generated or analyzed during this study are included in this published article (Supplemental Digital Appendix 3).

# References


**1.** Kelm Z, Womer J, Walter JK, Feudtner C. Interventions to cultivate physician empathy: a systematic review. BMC Med Educ. 2014;14:219.

**2.** Hardy C. Empathizing with patients: the role of interaction and narratives in providing better patient care. Med Health Care Philos. 2017;20(2):237-248.

**3.** Weiss T, Swede MJ. Transforming Preprofessional Health Education Through Relationship-Centered Care and Narrative Medicine. Teach Learn Med. 2016:1-12.

**4.** Halpern J. From detached concern to empathy: humanizing medical practice. Oxford; New York: Oxford University Press; 2001.

**5.** Underman K, Hirshfield LE. Detached concern?: Emotional socialization in twenty-first century medical education. Soc Sci Med. 2016;160:94-101.

**6.** Sulzer SH, Feinstein NW, Wendland CL. Assessing empathy development in medical education: a systematic review. Med Educ. 2016;50(3):300-310.

**7.** Dohrenwend AM. Defining Empathy to Better Teach, Measure, and Understand its Impact. Acad Med. 2018;93(12):1754-1756.

**8.** Stepien KA, Baernstein A. Educating for empathy. A review. J Gen Intern Med. 2006;21(5):524-530.

**9.** Batt-Rawden SA, Chisolm MS, Anton B, Flickinger TE. Teaching empathy to medical students: an updated, systematic review. Acad Med. 2013;88(8):1171-1177.

**10.** Teding van Berkhout E, Malouff JM. The efficacy of empathy training: A meta-analysis of randomized controlled trials. J Couns Psychol. 2016;63(1):32-41.

**11.** 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;35(2):e946-951.

**12.** Neumann M, Scheffer C, Tauschel D, Lutz G, Wirtz M, Edelhauser F. Physician empathy: definition, outcome-relevance and its measurement in patient care and medical education. GMS Z Med Ausbild. 2012;29(1):Doc11.

**13.** Chan JC, Hamamura T. Nursing students' assessment of pain and decision of triage for different ethnic groups: An experimental study. Nurse Educ Today. 2015;35(8):921-925.

**14.** Choudhury S, Blakemore SJ, Charman T. Social cognitive development during adolescence. Soc Cogn Affect Neurosci. 2006;1(3):165-174.

**15.** Gorman BK. Birth weight and cognitive development in adolescence: causal relationship or social selection? Soc Biol. 2002;49(1-2):13-34.

**16.** Cole N. How long should a training program be? A field study of “rules-of-thumb”. J Workplace Learn. 2008;20(1):54-70.

**17.** Taylor PJ, Russ-Eft DF, Chan DWL. A Meta-Analytic Review of Behavior Modeling Training. J Appl Psychol. 2005;90(4):692-709.

**18.** Berg K, Blatt B, Lopreiato J, et al. Standardized patient assessment of medical student empathy: ethnicity and gender effects in a multi-institutional study. Acad Med. 2015;90(1):105-111.

**19.** Berg K, Majdan JF, Berg D, Veloski J, Hojat M. Medical students' self-reported empathy and simulated patients' assessments of student empathy: an analysis by gender and ethnicity. Acad Med. 2011;86(8):984-988.

**20.** Christov-Moore L, Simpson EA, Coude G, Grigaityte K, Iacoboni M, Ferrari PF. Empathy: gender effects in brain and behavior. Neurosci Biobehav Rev. 2014;46 Pt 4:604-627.

**21.** Soto JA, Levenson RW. Emotion recognition across cultures: the influence of ethnicity on empathic accuracy and physiological linkage. Emotion. 2009;9(6):874-884.

**22.** DiLalla LF, Hull SK, Dorsey JK. Effect of gender, age, and relevant course work on attitudes toward empathy, patient spirituality, and physician wellness. Teach Learn Med. 2004;16(2):165-170.

**23.** Ferreira-Valente A, Monteiro JS, Barbosa RM, Salgueira A, Costa P, Costa MJ. Clarifying changes in student empathy throughout medical school: a scoping review. Adv Health Sci Educ Theory Pract. 2016.

**24.** Neumann M, Edelhauser F, Tauschel D, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med. 2011;86(8):996-1009.

**25.** Bandura A. Social learning theory. Englewood Cliffs, NJ: Prentice Hall; 1977.

**26.** Decker PJ, Nathan BR. Behavior modeling training: Principles and applications. New York: Praeger; 1985.

**27.** Boot WR, Simons DJ, Stothart C, Stutts C. The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. Perspect Psychol Sci. 2013;8(4):445-454.

**28.** Bearman M, Palermo C, Allen LM, Williams B. Learning Empathy Through Simulation: A Systematic Literature Review. Simul Healthc. 2015;10(5):308-319.

**29.** Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-269, W264.

**30.** Cook DA, Levinson AJ, Garside S. Method and reporting quality in health professions education research: a systematic review. Med Educ. 2011;45(3):227-238.

**31.** Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-560.

**32.** Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. Biometrics. 1994;50(4):1088-1101.

**33.** Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-634.

**34.** Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.

**35.** Fragkos KC, Tsagris M, Frangos CC. Publication Bias in Meta-Analysis: Confidence Intervals for Rosenthal's Fail-Safe Number. Int Sch Res Notices. 2014;2014:825383.

**36.** Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455-463.

**37.** Mascaro JS, Kelley S, Darcher A, et al. Meditation buffers medical student compassion from the deleterious effects of depression. J Posit Psychol. 2018;13(2):133-142.

**38.** Buffel du Vaure C, Lemogne C, Bunge L, et al. Promoting empathy among medical students: A two-site randomized controlled study. J Psychosom Res. 2017;103:102-107.

**39.** Fernando AT, Skinner K, Consedine NS. Increasing Compassion in Medical Decision-Making: Can a Brief Mindfulness Intervention Help? Mindfulness. 2017;8(2):276-285.

**40.** LoSasso AA, Lamberton CE, Sammon M, et al. Enhancing Student Empathetic Engagement, History-Taking, and Communication Skills During Electronic Medical Record Use in Patient Care. Acad Med. 2017;92(7):1022-1027.

**41.** Singh SP, Modi CM, Patel CP, Pathak AG. Low-fidelity simulation to enhance understanding of infection control among undergraduate medical students. Natl Med J India. 2017;30(4):215-218.

**42.** van Dijk I, Lucassen P, Akkermans RP, van Engelen BGM, van Weel C, Speckens AEM. Effects of Mindfulness-Based Stress Reduction on the Mental Health of Clinical Clerkship Students: A Cluster-Randomized Controlled Trial. Acad Med. 2017;92(7):1012-1021.

**43.** Wundrich M, Schwartz C, Feige B, Lemper D, Nissen C, Voderholzer U. Empathy training in medical students - a randomized controlled trial. Med Teach. 2017;39(10):1096-1098.

**44.** Danilewitz M, Bradwejn J, Koszycki D. A pilot feasibility study of a peer-led mindfulness program for medical students. Can Med Educ J. 2016;7(1):e31-37.

**45.** Yu EH, Choi EJ, Lee SY, Im SJ, Yune SJ, Baek SY. Effects of micro- and subtle-expression reading skill training in medical students: A randomized trial. Patient Educ Couns. 2016;99(10):1670-1675.

**46.** Alexander C, Sheeler RD, Rasmussen NH, Hayden L. Teaching an experiential mind-body method to medical students to increase interpersonal skills: a pilot study. Acad Psychiatry. 2015;39(3):316-319.

**47.** Matharu K, Shapiro JF, Hammer RR, Kravitz RL, Wilson MD, Fitzgerald FT. Reducing obesity prejudice in medical education. Educ Health (Abingdon). 2014;27(3):231-237.

**48.** Potash JS, Chen JY, Lam CL, Chau VT. Art-making in a family medicine clerkship: how does it affect medical student empathy? BMC Med Educ. 2014;14:247.

**49.** Chunharas A, Hetrakul P, Boonyobol R, Udomkitti T, Tassanapitikul T, Wattanasirichaigoon D. Medical students themselves as surrogate patients increased satisfaction, confidence, and performance in practicing injection skill. Med Teach. 2013;35(4):308-313.

**50.** Daeppen JB, Fortini C, Bertholet N, et al. Training medical students to conduct motivational interviewing: a randomized controlled trial. Patient Educ Couns. 2012;87(3):313-318.

**51.** Shapiro SM, Lancee WJ, Richards-Bentley CM. Evaluation of a communication skills program for first-year medical students at the University of Toronto. BMC Med Educ. 2009;9:11.

**52.** Wiecha JM, Markuns JF. Promoting medical humanism: design and evaluation of an online curriculum. Fam Med. 2008;40(9):617-619.

**53.** van Dijk I. I was always trying to do everything at once: mindfulness-based stress reduction for medical clerkship students (PhD Thesis): Radboud University; 2018.

**54.** Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ. 2010;340:c365.

**55.** Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? BMJ. 1998;316(7126):201.

**56.** Cook DA, Reed DA. Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. Acad Med. 2015;90(8):1067-1076.

**57.** Fragkos KC, Sotiropoulos I, Frangos CC. Empathy assessment in healthcare students is highly heterogeneous: A systematic review and meta-analysis (2012-2016). World J Meta-Anal. 2019;7(1):1-17.

**58.** Kämpfe N, Penzhorn J, Schikora J, Dünzl J, Schneidenbach J. Empathy and social desirability: a comparison of delinquent and non-delinquent participants using direct and indirect measures. Psychol Crime Law. 2009;15(1):1-17.

**59.** Merrill JM, Laux LF, Lorimor RJ, Thornby JI, Vallbona C. Measuring Social Desirability among Senior Medical Students. Psychol Rep. 1995;77(3):859-864.

**60.** Hojat M, Zuckerman M, Magee M, et al. Empathy in medical students as related to specialty interest, personality, and perceptions of mother and father. Pers Individ Dif. 2005;39(7):1205-1215.

**61.** Jena AB, Press VG, Arora VM. Social Desirability Bias in Self-rated Presenteeism Among Resident Physicians—Reply. JAMA Intern Med. 2013;173(2):166.

**62.** Hebert JR, Ma Y, Clemow L, et al. Gender Differences in Social Desirability and Social Approval Bias in Dietary Self-report. Am J Epidemiol. 1997;146(12):1046-1055.

**63.** Hemmerdinger JM, Stoddart SD, Lilford RJ. A systematic review of tests of empathy in medicine. BMC Med Educ. 2007;7:24.

**64.** Yu J, Kirk M. Measurement of empathy in nursing research: systematic review. J Adv Nurs. 2008;64(5):440-454.

**65.** Yu J, Kirk M. Evaluation of empathy measurement tools in nursing: systematic review. J Adv Nurs. 2009;65(9):1790-1806.

**66.** Lauzier M, Haccoun RR. The Interactive Effect of Modeling Strategies and Goal Orientations on Affective, Motivational, and Behavioral Training Outcomes. Perform Improv Q. 2014;27(2):83-102.

**67.** Decker PJ. The Enhancement of Behavior Modeling Training of Supervisory Skills by the Inclusion of Retention Processes. Pers Psychol. 1982;35(2):323-332.

**68.** Parry J, Mathers J, Stevens A, et al. Admissions processes for five year medical courses at English schools: review. BMJ. 2006;332(7548):1005-1009.

**69.** Monroe A, Quinn E, Samuelson W, Dunleavy DM, Dowd KW. An overview of the medical school admission process and use of applicant data in decision making: what has changed since the 1980s? Acad Med. 2013;88(5):672-681.

**70.** Ackerman PL. A theory of adult intellectual development: Process, personality, interests, and knowledge. Intelligence. 1996;22(2):227-257.

**71.** Soubelet A, Salthouse TA. Influence of Social Desirability on Age Differences in Self-Reports of Mood and Personality. J Pers. 2011;79(4):741-762.

**72.** Jorgensen JG. Cross-Cultural Comparisons. Annu Rev Anthropol. 1979;8(1):309-331.

**73.** Lorie A, Reinero DA, Phillips M, Zhang L, Riess H. Culture and nonverbal expressions of empathy in clinical settings: A systematic review. Patient Educ Couns. 2017;100(3):411-424.

**74.** Thomas G. After the Gold Rush: Questioning the "Gold Standard" and Reappraising the Status of Experiment and Randomized Controlled Trials in Education. Harv Educ Rev. 2016;86(3):390-411.

**75.** Morrison K. Randomised Controlled Trials for Evidence-based Education: Some Problems in Judging 'What Works'. Eval Res Educ. 2001;15(2):69-83.

**76.** Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med. 2018;210:2-21.

**77.** Ainsworth H, Hewitt CE, Higgins S, Wiggins A, Torgerson DJ, Torgerson CJ. Sources of bias in outcome assessment in randomised controlled trials: a case study. Educ Res Eval. 2014;21(1):3-14.

**78.** Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. Int J Behav Healthc Res. 2011;2(4):320-332.

**79.** Horton T. The importance of empathy. 2016; https://www.health.org.uk/blog/importance-empathy. Accessed 18 July, 2018.

**80.** Nelson K. Empathy in Heatlhcare: 7 Benefits. 2017; https://www.wegohealth.com/2017/11/14/empathy-in-healthcare/. Accessed 18 July, 2018.

**81.** Ross S, Bennett L. Addressing the emotional realities of the job: compassionate care for the NHS workforce. 2018; https://www.pointofcarefoundation.org.uk/blog/addressing-emotional-realities-job-compassionate-care-nhs-workforce/. Accessed 18 July, 2018.

**82.** General Medical Council (UK). Good medical practice 2014.

**83.** Morling B, Uchida Y, Frentrup S. Social Support in Two Cultures: Everyday Transactions in the U.S. and Empathic Assurance in Japan. PLoS One. 2015;10(6):e0127737.

**84.** Jeffrey D, Downie R. Empathy - can it be taught? J R Coll Physicians Edinb. 2016;46(2):107-112.

**85.** Rieger KL, Chernomas WM, McMillan DE, Morin FL, Demczuk L. Effectiveness and experience of arts-based pedagogy among undergraduate nursing students: a mixed methods systematic review. JBI Database System Rev Implement Rep. 2016;14(11):139-239.

**86.** Riley R, Weiss MC. A qualitative thematic review: emotional labour in healthcare settings. J Adv Nurs. 2016;72(1):6-17.

# Supplemental Digital Content

## Supplemental Digital Appendix 1. Detailed methods of the present manuscript.

### Study eligibility criteria

The inclusion criteria related to study quality, language context and relevance. Only randomized controlled study designs were included which examined empathy interventions in medical students. Randomized controlled designs are often considered the highest quality of studies since they are based on the existence of a control group and the location to each arm in random,[1] maximizing statistical power and minimizing selection and allocation bias.[2] There was a restriction to English language papers published from 1 January 1948 until 31 January 2018. Furthermore, papers had to provide sufficient data to produce an effect measure for the meta-analysis. Clinical empathy was considered as a component driven from medical students towards patients. It was considered as a general concept and not as its sole separated concepts (e.g. cognitive or affective empathy). Studies were excluded when other types of empathy were examined: empathy from professionals (doctors, nurses, allied healthcare professionals, etc.) towards patients, empathy from students not from healthcare, empathy from teachers toward students (teacher empathy), perceived empathy of patients from their caring professionals, and empathy among adolescents and high school students. Also specific types of empathy were not included, such as cross-cultural empathy, multicultural empathy, and empathy towards particular medical conditions from non-medical students. Other exclusion criteria were cross sectional studies, case studies, pre-post experimental designs, qualitative studies, reviews and non-English studies.

### Search strategy and terms

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for systematic reviews and meta-analyses were used.[3] The keywords for searching were: empathy, caring, humanism, cognitive, emotional, healthcare, medical students, compassion, care and randomized. Electronic database searches using appropriate variations of the search terms were conducted in PubMed/Medline, Scopus, EMBASE, Google Scholar and ERIC. Publisher databases were also searched (ScienceDirect, Springer Link, Wiley Online Library, Taylor & Francis Online, and Oxford Academic). The bibliographies from all included manuscripts and hand searching of relevant healthcare education journals were used to identify further references. Hand searching of healthcare education journals included the following: Academic Medicine, Advances in Health Sciences Education, BMC Medical Education, International Journal of Medical Education, Medical Education, Medical Education Online, Medical Teacher, Teaching and Learning in Medicine, Perspectives on Medical Education.

### Study selection, data extraction and quality assessment

The resulting studies (in abstract form) were assessed against the inclusion criteria. Both authors screened the abstracts. A random 10% of the initial search was selected to check the reliability in applying the inclusion criteria. The kappa coefficient for interrater agreement regarding study selection between the two authors was 0.47 (moderate) and the decision on the disagreed articles was settled with a consensus meeting. When there was insufficient information available in the abstract, the full text was reviewed. Data extraction was performed by both authors. Based on the literature review and hypotheses, the extracted data from the selected studies were:

#### Descriptive information

Author, year of publication, study aim, country, sample size, mean age, gender distribution, distribution of students per year of study, journal name, journal Impact Factor (2017),[4] and impact ratio of each article. Papers were divided into high-impact or low-impact journals if they were respectively higher or lower than the 2017 aggregate impact factor for Education, Scientific Disciplines (1.800).[4] The impact ratio of each article was chosen to assess the relative impact of each article, since not every article will necessary reflect

impact which is equivalent to the impact of the journal it was published in. The impact ratio was calculated as the number of citations each paper received from Google Scholar divided by the years since publication of the article (up until March 2019).[5]

**Empathy features**
*Subtypes*: cognitive, affective, behavioral, or multidimensional. These subtypes of empathy are broadly based on available accepted definitions and comprehensive reviews.[6-17]
*Measure scope*: narrow (assessed by one item only) or broad (assessed by multiple items or a validated scale).
*Self-reported vs objective measure for empathy*.
*Scale characteristics*: name, reliability, number of items.

**Empathy intervention**
*Type of empathy training*: experiential: instructors provide "experiences" such as games and role-play; didactic, which refers mainly to lecture based; skills training, which includes lectures, demonstrations and practice; and mixed methods, which includes combinations of the above.[11,14,17]
*Whether there was use of behavior skills training four components* (modelling, instructions, rehearsal, feedback).[17,18]
*Compensation*: yes or no.
*Type of control group*: waiting list, active control.
*Intervention length* (hours trained).
*Months pre-post measurement*.
*Effect size*

**Quality assessment** was performed with the Medical Education Research Study Quality Instrument (MERSQI), which has the highest interrater agreement score and correlates satisfactorily with other similar instruments.[19] The full instrument is shown in Table 1 of the current Supplemental Digital Appendix 1.

## Statistical analysis

Meta-analysis was performed with Stata 15.0 (Stata Corp., College Station, Texas) and R 3.4.1 (R Core Team, Vienna, Austria). Standardized mean differences (SMDs) were extracted from studies when available. The strength of association was categorized as following: small, SMD=0.2; medium, SMD=0.5; and large, SMD=0.8. A random effects model was used to produce a pooled estimate of the SMDs. Statistical heterogeneity was assessed using Cochran's $Q$ test and quantified with the $I^2$ statistic,[20] while statistical significance for heterogeneity was set as $p \leq 0.10$. Heterogeneity was further investigated with subgroup analysis and meta-regression. Publication bias was assessed using funnel plots, Egger's test, Begg's test, Rosenthal's number and the trim and fill method.[21-25]

Table 1. MERSQI items and scoring.[26]

| Domain | Categories | Score |
|---|---|---|
| **Study design** | *Study design* | |
| | Single group cross-sectional or single group post-test only | 1 |
| | Single group pre-test and post-test | 1.5 |
| | Nonrandomized, 2 group | 2 |
| | Randomized controlled trial | 3 |
| **Sampling** | *No of institutions studied* | |
| | 1 | 0.5 |
| | 2 | 1 |
| | >2 | 1.5 |
| | *Response rate, %* | |
| | Not applicable | |
| | < 50 or not reported | 0.5 |
| | 50-74 | 1 |
| | ≥75 | 1.5 |
| **Type of data** | *Type of Data* | |
| | Assessment by study participant | 1 |
| | Objective measurement | 3 |
| **Validity of evaluation instrument** | *Internal structure* | |
| | Not applicable | |
| | Not reported | 0 |
| | Reported | 1 |
| | *Content* | |
| | Not applicable | |
| | Not reported | 0 |
| | Reported | 1 |
| | *Relationships to other variables* | |
| | Not applicable | |
| | Not reported | 0 |
| | Reported | 1 |
| **Data Analysis** | *Appropriateness of analysis* | |
| | Data analysis inappropriate for study design or type of data | 0 |
| | Data analysis appropriate for study design and type of data | 1 |
| | *Complexity of analysis* | |
| | Descriptive analysis only | 1 |
| | Beyond descriptive analysis | 2 |
| **Outcomes** | *Outcomes* | |
| | Satisfaction, attitudes, perceptions, opinions, general facts | 1 |
| | Knowledge, Skills | 1.5 |
| | Behaviors | 2 |
| | Patient/Health care outcome | 3 |

## References


**1.** Sullivan GM. Getting off the "gold standard": randomized controlled trials and education research. J Grad Med Educ. 2011;3(3):285-289.
**2.** Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? BMJ. 1998;316(7126):201.
**3.** Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264-269, W264.
**4.** 2017 Journal Citation Reports: Clarivate Analytics; 2018.
**5.** Valvi AC, Fragkos KC. Critical review of the e-loyalty literature: a purchase-centred framework. Electron Commerc Res. 2012;12(3):331-378.
**6.** Hemmerdinger JM, Stoddart SD, Lilford RJ. A systematic review of tests of empathy in medicine. BMC Med Educ. 2007;7:24.
**7.** Yu J, Kirk M. Measurement of empathy in nursing research: systematic review. J Adv Nurs. 2008;64(5):440-454.
**8.** Yu J, Kirk M. Evaluation of empathy measurement tools in nursing: systematic review. J Adv Nurs. 2009;65(9):1790-1806.
**9.** Neumann M, Edelhauser F, Tauschel D, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med. 2011;86(8):996-1009.
**10.** Lelorain S, Bredart A, Dolbeault S, Sultan S. A systematic review of the associations between empathy measures and patient outcomes in cancer care. Psychooncology. 2012;21(12):1255-1264.
**11.** Batt-Rawden SA, Chisolm MS, Anton B, Flickinger TE. Teaching empathy to medical students: an updated, systematic review. Acad Med. 2013;88(8):1171-1177.
**12.** Derksen F, Bensing J, Lagro-Janssen A. Effectiveness of empathy in general practice: a systematic review. Br J Gen Pract. 2013;63(606):e76-84.
**13.** Kelm Z, Womer J, Walter JK, Feudtner C. Interventions to cultivate physician empathy: a systematic review. BMC Med Educ. 2014;14:219.
**14.** Bearman M, Palermo C, Allen LM, Williams B. Learning Empathy Through Simulation: A Systematic Literature Review. Simul Healthc. 2015;10(5):308-319.
**15.** Liu X, Rohrer W, Luo A, Fang Z, He T, Xie W. Doctor-patient communication skills training in mainland China: a systematic review of the literature. Patient Educ Couns. 2015;98(1):3-14.
**16.** Sulzer SH, Feinstein NW, Wendland CL. Assessing empathy development in medical education: a systematic review. Med Educ. 2016;50(3):300-310.
**17.** Teding van Berkhout E, Malouff JM. The efficacy of empathy training: A meta-analysis of randomized controlled trials. J Couns Psychol. 2016;63(1):32-41.
**18.** Bandura A. Social learning theory. Englewood Cliffs, NJ: Prentice Hall; 1977.
**19.** Cook DA, Levinson AJ, Garside S. Method and reporting quality in health professions education research: a systematic review. Med Educ. 2011;45(3):227-238.
**20.** Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-560.
**21.** Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. Biometrics. 1994;50(4):1088-1101.
**22.** Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-634.
**23.** Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.
**24.** Fragkos KC, Tsagris M, Frangos CC. Publication Bias in Meta-Analysis: Confidence Intervals for Rosenthal's Fail-Safe Number. Int Sch Res Notices. 2014;2014:825383.
**25.** Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455-463.
**26.** Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. JAMA. 2007;298(9):1002-1009.

Supplemental Digital Appendix 2. Table with main study characteristics for the included studies of the review.

Table 1. Studies included in the review and meta-analysis. This table presents basic demographic information, characteristics of the empathy educational intervention, the journal's impact and the article's impact ratio. All studies have medical students as participants.

| Study | Country, N (Male %, Age in years) | Aim | Main results | Educational Intervention | | | | | | Journal Impact Factor | Article Impact ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Name | Used four components of behavior skills training | Type | Type of control group | Intervention Length (hours trained) | Months pre-post | | |
| Mascaro et al.[1] | USA, 32 (62.5%, 25) | To investigate the feasibility of cognitively-based compassion training and to test whether it can decrease depression, enhance compassion, and improve daily functioning | • The intervention increased compassion and decreased loneliness and depression<br>• Change in compassion was associated with depression | Cognitively-Based Compassion Training | No | Experiential /mixed | Waiting list | 15 | 2.5 | 2.594 (high impact journal) | 4.97 |
| Buffel du Vaure et al.[2] | France, 299 (40.5%, Not Reported) | To assess the effects of Balint groups on empathy | • Significant difference in the Consultation and Relational Empathy Measure scale score at follow-up between the two groups<br>• The intervention group displayed significantly higher Jefferson Scale of Physician Empathy score at follow-up than the control group | Balint Groups | No | Experiential | Waiting list | 10.5 | 0.25 | 2.947 (high impact journal) | 6.00 |
| Fernando et al.[3] | New Zealand, 83 (54.2%, 21.4) | To evaluate whether a brief mindfulness induction increased compassionate responding to difficult patients | • Results showed that mindfulness predicted greater patient "liking" and "caring" but only among persons lower in self-compassion<br>• The mindfulness intervention predicted greater helping behavior, mainly in those with higher self-compassion | Mindfulness-Based Exercise | No | Didactic | Active | 2 | 0 | 3.024 (high impact journal) | 3.48 |
| LoSasso et al.[4] | USA, 70 (Not Reported, Not Reported) | To examine whether an intervention on proper use of electronic medical records could help improve medical students' empathic engagement | • Faculty mean ratings on the Jefferson Scale of Physician Empathy were higher for the intervention group than the control group<br>• Both groups' Jefferson Scale of Physician Empathy mean scores increased pre-test to post-test, but changes were not significant<br>• Intervention group's post-test Jefferson Scale of Physician Empathy score was higher than the control group's (p>0.05) | Training in Electronic Medical Record-Specific Communication | No | Mixed | Active | 1 | 1.5 | 4.801 (high impact journal) | 2.40 |
| Singh et al.[5] | India, 93 (Not Reported, Not Reported) | To test emotional sensitization using low-fidelity techniques to enhance its effectiveness | • No significant difference in the pre-test Toronto Empathy Questionnaire score (p=0.87)<br>• Significant difference in the post-test Toronto Empathy Questionnaire (p = 0.026) | Low-Fidelity Simulation Techniques (Case Discussions and a Video Show) | Yes | Mixed | Active | 4.5 | 1 | 0.786 (low impact journal) | 0.00 |
| van Dijk et al.[6] | Netherlands, 167 (21.5%, 23.5) | To examine the effect of mindfulness-based stress reduction training on the mental health of medical students during clinical clerkships | • The intervention group reported a small reduction of psychological distress and dysfunctional cognitions and a moderate increase of positive mental health, life satisfaction, and mindfulness skills but no significant effect on physician empathy | Mindfulness-Based Stress Reduction | No | Mixed | Active | 16 | 12 | 4.801 (high impact journal) | 6.00 |
| Wundrich et al.[7] | Germany, 158 (Not Reported, Not Reported) | To examine whether empathy in medical students can be improved by specific training | • Participants of the intervention group showed significantly higher levels of empathy when rated by standardized patients and experts than the control group<br>• No significant group differences were observed in self-rated empathy | Empathy Skills Training and Teaching with Simulated Patients | Yes | Mixed | Active | 4.5 | 0.75 | 2.450 (high impact journal) | 13.26 |
| Danilewitz et al.[8] | Canada, 30 (26.7%, Not Reported) | To evaluate the feasibility and benefits of a peer-led mindfulness meditation program | • The intervention decreased levels of stress and enhanced mindfulness, self-compassion and altruism from baseline to post-study<br>• Changes were not significant for the wait-list condition | Mindfulness Based Stress Reduction Program | No | Experiential | Waiting list | 12 | 2 | No impact factor available | 2.74 |

| Study | Country, N (Male %, Age in years) | Aim | Main results | Educational Intervention | | | | | | Journal Impact Factor | Article Impact ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Name | Used four components of behavior skills training | Type | Type of control group | Intervention Length (hours trained) | Months pre-post | | |
| Yu et al.[9] | South Korea, 82 (53.7%, 26.1) | To investigate the effectiveness of the Micro Expression Training Tool and Subtle Expression Training Tool to help improve the non-verbal communication skills of medical students | • Micro Expression Training Tool pre-test scores were positively correlated with female gender, agreeableness<br>• Subtle Expression Training Tool pre-test scores were negatively correlated with age and positively correlated with female gender<br>• Increases in both test scores in the interventional group were significantly higher than in the control group | Micro- and Subtle-Expression Reading Skill Training | No | Skills training | Active | 1 | 0.25 | 2.785 (high impact journal) | 0.71 |
| Alexander et al.[10] | USA, 13 (Not Reported, Not Reported) | To investigate whether inner relationship focusing increases self-awareness and empathic listening in medical students | • The intervention group showed improvement in all areas compared to the control group<br>• Improvement in comfort talking to patients about how recurring symptoms might relate to issues in their lives was significant | Inner Relationship Focusing | Yes | Mixed | Active | 20 | 5 | 1.603 (low impact journal) | 0.63 |
| Matharu et al.[11] | USA, 129 (29.5%, 25.2) | To determine whether reading a play about obesity could diminish obesity prejudice | • Significant increase in empathy for those in both the theatre ($p = 0.007$) and lecture group ($p = 0.02$) | Reading a Play about Obesity | No | Experiential | Active | 1 | 4 | No impact factor available | 4.75 |
| Potash et al.[12] | Hong Kong, 106 (59.4%, 21.2) | To evaluate the impact of an arts-making workshop on medical student empathy | • The level of empathy declined in both groups over time, but with no statistically significant differences between groups<br>• For Jefferson Scale of Physician Empathy items relating to emotional influence on medical decision making, participants in the arts-making workshop changed more than those in the problem-solving workshop | Arts-Making Workshop | No | Experiential | Active | 3 | 2.5 | 1.511 (low impact journal) | 5.41 |
| Chunharas et al.[13] | Thailand, 89 (43.8%, 23.0) | To evaluate the satisfaction, perceptions of confidence and feeling of empathy toward patients using manikin only compared to additional training using themselves as surrogate patients | • The intervention group reported significantly higher satisfaction, confidence and empathy | Injection Skill Using Themselves as Surrogate Patients | Yes | Skills training | Active | 2 | 0 | 2.450 (high impact journal) | 2.88 |
| Daeppen et al.[14] | Switzerland, 91 (40.7%, 24.7) | To examine the effectiveness of motivational interviewing training among medical students | • Students in the intervention group demonstrated significantly higher scores for empathy motivational interviewing spirit | Motivational Interviewing Training | Yes | Experiential | Active | 8 | 0.25 | 2.785 (high impact journal) | 5.44 |
| Shapiro et al.[15] | Canada, 79 (38.9%, 23.0) | To evaluate the effectiveness of the University of Toronto's Therapeutic Communication Program at improving first-year medical students' communication skills | • In terms of external rating, the intervention improved students' communication skills | University Of Toronto's Therapeutic Communication Program | Yes | Mixed | Waiting list | 16 | 4 | 1.511 (low impact journal) | 5.30 |
| Wiecha and Markuns[16] | USA. 215 (Not Reported, Not Reported) | To evaluate an online clerkship program which promoted student confidence in three areas of humanistic practice (cultural competence, empathy, and assessing non-adherent patients non-judgmentally) | • Students in the online group showed a greater increase, from before to after, in self-reported ability in each of the three areas examined | Online Clerkship | Yes | Experiential | Active | 26 | 1.5 | 1.140 (low impact journal) | 2.11 |

## References


**1.** Mascaro JS, Kelley S, Darcher A, et al. Meditation buffers medical student compassion from the deleterious effects of depression. J Posit Psychol. 2018;13(2):133-142.

**2.** Buffel du Vaure C, Lemogne C, Bunge L, et al. Promoting empathy among medical students: A two-site randomized controlled study. J Psychosom Res. 2017;103:102-107.

**3.** Fernando AT, Skinner K, Consedine NS. Increasing Compassion in Medical Decision-Making: Can a Brief Mindfulness Intervention Help? Mindfulness. 2017;8(2):276-285.

**4.** LoSasso AA, Lamberton CE, Sammon M, et al. Enhancing Student Empathetic Engagement, History-Taking, and Communication Skills During Electronic Medical Record Use in Patient Care. Acad Med. 2017;92(7):1022-1027.

**5.** Singh SP, Modi CM, Patel CP, Pathak AG. Low-fidelity simulation to enhance understanding of infection control among undergraduate medical students. Natl Med J India. 2017;30(4):215-218.

**6.** van Dijk I, Lucassen P, Akkermans RP, van Engelen BGM, van Weel C, Speckens AEM. Effects of Mindfulness-Based Stress Reduction on the Mental Health of Clinical Clerkship Students: A Cluster-Randomized Controlled Trial. Acad Med. 2017;92(7):1012-1021.

**7.** Wundrich M, Schwartz C, Feige B, Lemper D, Nissen C, Voderholzer U. Empathy training in medical students - a randomized controlled trial. Med Teach. 2017;39(10):1096-1098.

**8.** Danilewitz M, Bradwejn J, Koszycki D. A pilot feasibility study of a peer-led mindfulness program for medical students. Can Med Educ J. 2016;7(1):e31-37.

**9.** Yu EH, Choi EJ, Lee SY, Im SJ, Yune SJ, Baek SY. Effects of micro- and subtle-expression reading skill training in medical students: A randomized trial. Patient Educ Couns. 2016;99(10):1670-1675.

**10.** Alexander C, Sheeler RD, Rasmussen NH, Hayden L. Teaching an experiential mind-body method to medical students to increase interpersonal skills: a pilot study. Acad Psychiatry. 2015;39(3):316-319.

**11.** Matharu K, Shapiro JF, Hammer RR, Kravitz RL, Wilson MD, Fitzgerald FT. Reducing obesity prejudice in medical education. Educ Health (Abingdon). 2014;27(3):231-237.

**12.** Potash JS, Chen JY, Lam CL, Chau VT. Art-making in a family medicine clerkship: how does it affect medical student empathy? BMC Med Educ. 2014;14:247.

**13.** Chunharas A, Hetrakul P, Boonyobol R, Udomkitti T, Tassanapitikul T, Wattanasirichaigoon D. Medical students themselves as surrogate patients increased satisfaction, confidence, and performance in practicing injection skill. Med Teach. 2013;35(4):308-313.

**14.** Daeppen JB, Fortini C, Bertholet N, et al. Training medical students to conduct motivational interviewing: a randomized controlled trial. Patient Educ Couns. 2012;87(3):313-318.

**15.** Shapiro SM, Lancee WJ, Richards-Bentley CM. Evaluation of a communication skills program for first-year medical students at the University of Toronto. BMC Med Educ. 2009;9:11.

**16.** Wiecha JM, Markuns JF. Promoting medical humanism: design and evaluation of an online curriculum. Fam Med. 2008;40(9):617-619.

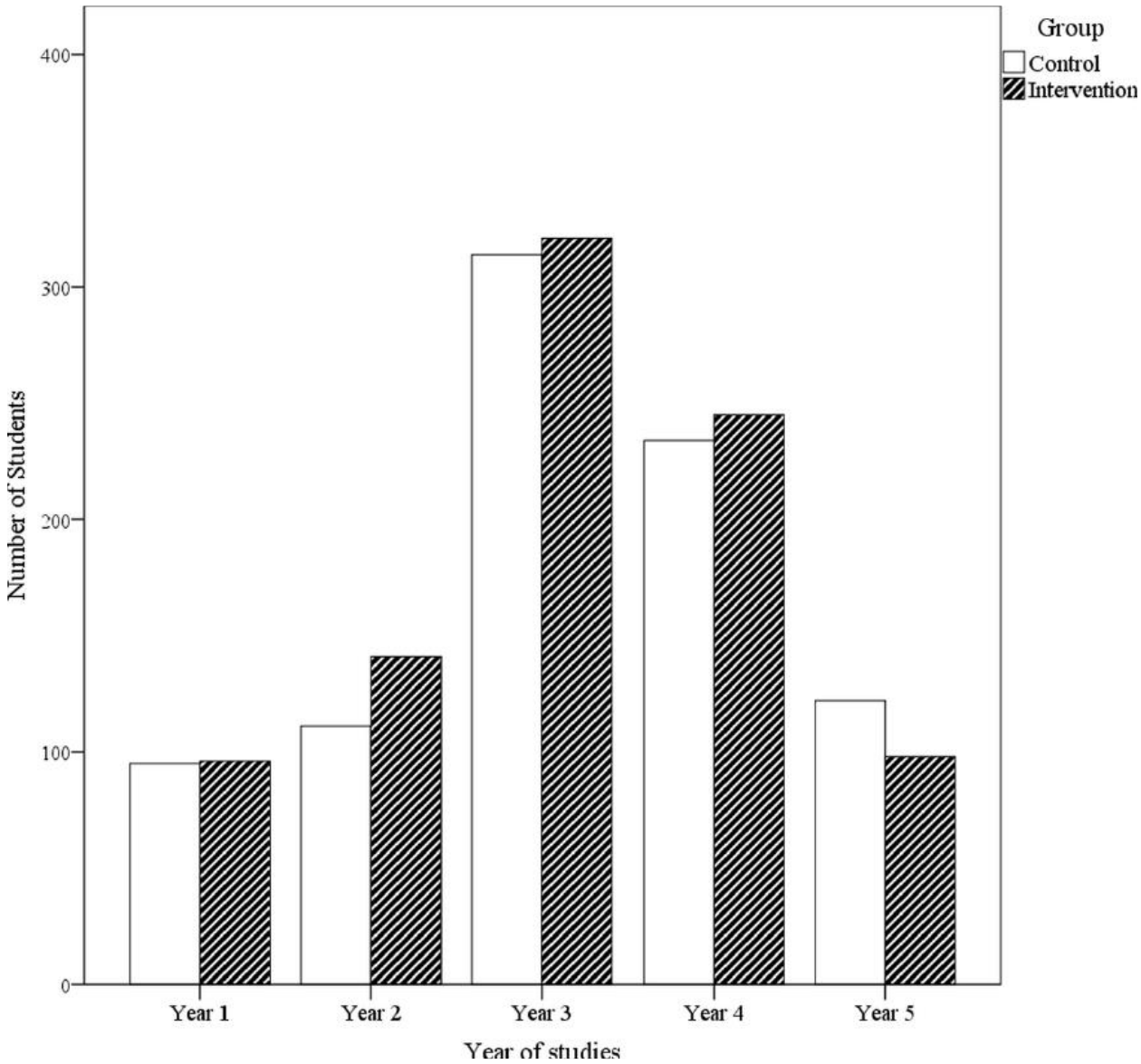
Supplemental Digital Appendix 3. Table with the dataset analyzed in the present study.

| | A | B | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF | AG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Study Details | | | | Journal Impact | | | | | | | Study Demographics | | | | | | Scale used | | | Distribution of students per year | | | | | | | | | |
| 2 | | | | | | | | | | | | Control | | | Intervention | | | | | | Medical school year (Control) | | | | | | Medical school year (Intervention) | | | |
| 3 | ID | Study | Country | Total Sample Size | Impact factor (2017) | SCImago Journal Rank (2017) | CiteScore (2017) | SNIP (2017) | Citations in Google Scholar in March 2019 | Impact ratio | Journal | Sample Size | Mean Age | Male % | Sample Size | Mean Age | Male % | Scale | Reliability | Items | n_year_1 | n_year_2 | n_year_3 | n_year_4 | n_year_5 | n_year_6 | n_year_1 | n_year_2 | n_year_3 | n_year_4 |
| 4 | 0001 | Buffel du Vaure et al. 2017 | France | 299 | 2.947 | 1.344 | 3.02 | 1.323 | 8 | 6.00 | Journal of Psychosomatic Research | 144 | | 34.7 | 155 | | 45.8 | Jefferson Scale of Physician Empathy / Consultation and Relational Empathy Measure | | | 0 | 0 | 0 | 144 | 0 | 0 | 0 | 0 | 0 | 155 |
| 5 | 0001 | Buffel du Vaure et al. 2017 | | | | | | | | | | | | | | | | Jefferson Scale of Physician Empathy | 0.785 | 20 | | | | | | | | | | |
| 6 | 0001 | Buffel du Vaure et al. 2017 | | | | | | | | | | | | | | | | Consultation and Relational Empathy Measure | 0.95 | 10 | | | | | | | | | | |
| 7 | 0002 | Wundrich et al. 2017 | Germany | 158 | 2.45 | 1.195 | 1.44 | 1.575 | 21 | 13.26 | Medical Teacher | 79 | | | 79 | | | Jefferson Scale of Physician Empathy / Standardised patient / Experts | | | 0 | 0 | 79 | 0 | 0 | 0 | 0 | 0 | 79 | 0 |
| 8 | 0002 | Wundrich et al. 2017 | | | | | | | | | | | | | | | | Jefferson Scale of Physician Empathy | | 20 | | | | | | | | | | |
| 9 | 0002 | Wundrich et al. 2017 | | | | | | | | | | | | | | | | Standardised patient | 0.94 | 11 | | | | | | | | | | |
| 10 | 0002 | Wundrich et al. 2017 | | | | | | | | | | | | | | | | Experts | 0.94 | 11 | | | | | | | | | | |
| 11 | 0003 | Singh et al. 2017 | India | 93 | 0.786 | 0.188 | 0.3 | 0.43 | 0 | 0.00 | National Medical Journal of India | 31 | | | 62 | | | Toronto Empathy Questionnaire | | 16 | 0 | 31 | 0 | 0 | 0 | 0 | 0 | 63 | 0 | 0 |
| 12 | 0004 | van Dijk et al. 2017 | Netherlands | 167 | 4.801 | 2.53 | 2.37 | 2.248 | 10 | 6.00 | Academic Medicine | 84 | 23.3 | 15 | 83 | 23.7 | 28 | Jefferson Scale of Physician Empathy | | 20 | 0 | 0 | 0 | 84 | 0 | 0 | 0 | 0 | 0 | 83 |
| 13 | 0005 | LoSasso et al. 2017 | USA | 70 | 4.801 | 2.53 | 2.37 | 2.248 | 4 | 2.40 | Academic Medicine | 32 | | | 38 | | | Jefferson Scale of Physician Empathy / Standardised patient / Experts | | | 0 | 0 | 32 | 0 | 0 | 0 | 0 | 0 | 38 | 0 |
| 14 | 0005 | LoSasso et al. 2017 | | | | | | | | | | | | | | | | Jefferson Scale of Physician Empathy | | 20 | | | | | | | | | | |
| 15 | 0005 | LoSasso et al. 2017 | | | | | | | | | | | | | | | | Standardised patient | | 6 | | | | | | | | | | |
| 16 | 0005 | LoSasso et al. 2017 | | | | | | | | | | | | | | | | Experts | | 6 | | | | | | | | | | |
| 17 | 0006 | Fernando et al. 2017 | New Zealand | 83 | 3.024 | 1.132 | 2.73 | 1.214 | 9 | 3.48 | Mindfulness | 42 | 21.62 | 50 | 41 | 21.2 | 58.5 | Visual Analogue Scale / Objective measure | | | 0 | 0 | 42 | 0 | 0 | 0 | 0 | 0 | 41 | 0 |
| 18 | 0006 | Fernando et al. 2017 | | | | | | | | | | | | | | | | Visual Analogue Scale | | 3 | | | | | | | | | | |
| 19 | 0006 | Fernando et al. 2017 | | | | | | | | | | | | | | | | Objective Measure | | 1 | | | | | | | | | | |
| 20 | 0007 | Mascaro et al. 2018 | USA | 32 | 2.594 | 1.226 | 3.11 | 1.374 | 12 | 4.97 | Journal of Positive Psychology | 11 | 24.4 | 63.6 | 21 | 25.3 | 61.9 | Compassionate Love for Humanity Scale | | 21 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 0 |
| 21 | 0008 | Yu et al. 2016 | South Korea | 82 | 2.785 | 1.38 | 2.83 | 1.503 | 2 | 0.71 | Patient Education and Counseling | 41 | 26.2 | 56.1 | 41 | 26 | 51.2 | Micro Expression Training Tool / Subtle Expression Training Tool | | | 17 | 24 | 0 | 0 | 0 | 0 | 23 | 18 | 0 | 0 |
| 22 | 0008 | Yu et al. 2016 | | | | | | | | | | | | | | | | Micro Expression Training Tool | | | | | | | | | | | | |
| 23 | 0008 | Yu et al. 2016 | | | | | | | | | | | | | | | | Subtle Expression Training Tool | | | | | | | | | | | | |
| 24 | 0009 | Danilewitz et al. 2016 | Canada | 30 | | | | | 8 | 2.74 | Canadian Medical Educational Journal | 15 | | 26.7 | 15 | | 26.7 | Jefferson Scale of Physician Empathy | | 20 | 7 | 8 | 0 | 0 | 0 | 0 | 7 | 8 | 0 | 0 |
| 25 | 0010 | Alexander et al. 2015 | USA | 13 | 1.603 | 0.53 | 0.84 | 0.674 | 3 | 0.63 | Academic Psychiatry | 7 | | | 6 | | | Self-reported measure | | 4 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 |
| 26 | 0011 | Potash et al. 2014 | Hong Kong | 106 | 1.511 | 0.765 | 1.71 | 1.169 | 23 | 5.41 | BMC Medical Education | 58 | 21.1 | 58.6 | 48 | 21.4 | 60.4 | Jefferson Scale of Physician Empathy | | 20 | 0 | 0 | 58 | 0 | 0 | 0 | 0 | 0 | 48 | 0 |
| 27 | 0012 | Matharu et al. 2014 | USA | 129 | | 0.242 | 0.37 | 0.389 | 19 | 4.75 | Education for Health | 66 | 25.2 | 25.8 | 63 | 25.1 | 33.3 | Jefferson Scale of Physician Empathy | 0.76 | 20 | 30 | 30 | 0 | 6 | 0 | 0 | 28 | 25 | 3 | 7 |
| 28 | 0013 | Chunharas et al. 2013 | Thailand | 89 | 2.45 | 1.195 | 1.44 | 1.575 | 18 | 2.88 | Medical Teacher | 57 | 23 | 36.8 | 32 | 23 | 56.3 | Objective Measure | | 1 | 0 | 0 | 0 | 0 | 57 | 0 | 0 | 0 | 0 | 0 |
| 29 | 0014 | Daeppen et al. 2012 | Switzerland | 91 | 2.785 | 1.38 | 2.83 | 1.503 | 39 | 5.44 | Patient Education and Counseling | 49 | 24.7 | 40.7 | 42 | 24.7 | 40.7 | Motivational Interviewing Treatment Integrity 3.0 | 0.85 | 12 | 0 | 0 | 0 | 0 | 65 | 0 | 0 | 0 | 0 | 0 |
| 30 | 0015 | Shapiro et al. 2009 | Canada | 79 | 1.511 | 0.765 | 1.71 | 1.169 | 53 | 5.30 | BMC Medical Education | 41 | 23 | 37 | 38 | 23.1 | 41 | Self-Assessment of Interpersonal Competence Questionnaire / Interpersonal Skills Rating Scale / Staff-Patient Interaction Rating Scale | | | 41 | 0 | 0 | 0 | 0 | 0 | 38 | 0 | 0 | 0 |
| 31 | 0015 | Shapiro et al. 2009 | | | | | | | | | | | | | | | | Self Assessment of Interpersonal Competence Questionnaire | | 40 | | | | | | | | | | |
| 32 | 0015 | Shapiro et al. 2009 | | | | | | | | | | | | | | | | Interpersonal Skills Rating Scale | | 7 | | | | | | | | | | |
| 33 | 0015 | Shapiro et al. 2009 | | | | | | | | | | | | | | | | Staff-Patient Interaction Rating Scale | | 24 | | | | | | | | | | |
| 34 | 0016 | Wiecha and Markuns 2008 | USA | 215 | 1.14 | 0.567 | 0.74 | 0.95 | 22 | 2.11 | Family Medicine | 103 | | | 112 | | | Self-reported measure | | 3 | 0 | 0 | 103 | 0 | 0 | 0 | 0 | 0 | 112 | 0 |

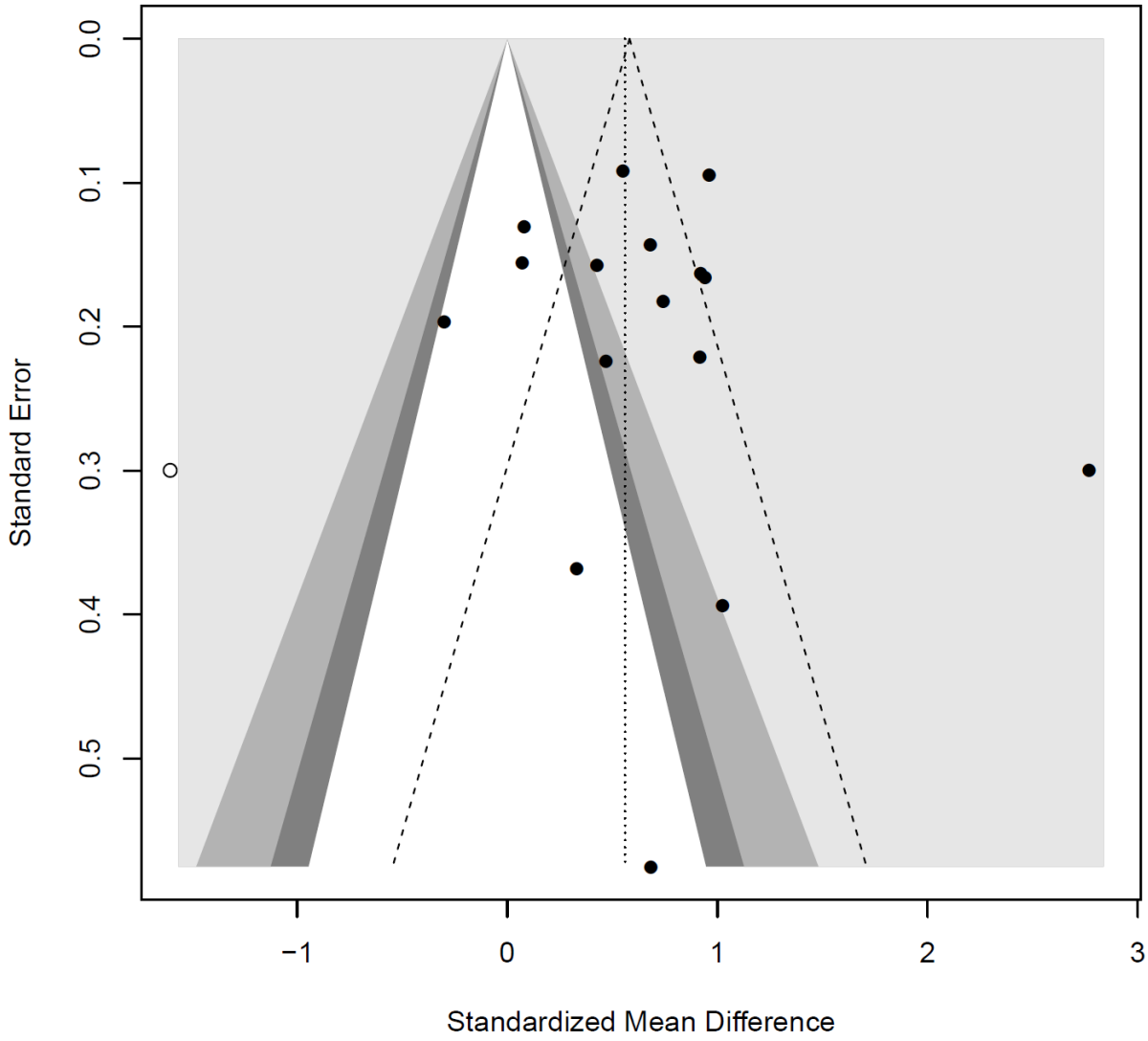
| | AH | AI | AJ | AK | AL | AM | AN | AO | AP | AQ | AR | AS | AT | AU | AV | AW | AX | AY | AZ | BA | BB | BC | BD | BE | BF | BG | BH | BI | BJ | BK | BL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | Empathy Intervention characteristics | | | | | | | | | | | | | | | Effect Size | | | | | | | | | | | | | |
| 2 | | | Teaching Methods | | | | | | | | | | | | | | | Control (post-pre) | | | Training (post-pre) | | | | SMD | | | | | Quality Assessment | |
| 3 | n_year_5 | n_year_6 | Name of intervention | Type of empathy trained | Used four components of behavior skills training | Type | Modelling | instructions | rehearsal | feedback | Compensation provided | Type of control group | Empathy measure used | Self-report vs. objective measure used | Scope of empathy | Intervention Length (hours trained) | Months pre-post | n_c | m_c | sd_c | n_t | m_ | sd | p | smd | var | LCI | UCI | SE | MERSQI Study design | MERSQI Sampling Institutions |
| 4 | 0 | 0 | Balint groups | Cognitive, affective and behavioral | No | Experiential | Yes | Yes | Yes | No | No | Waiting list | Jefferson Scale of Physician Empathy / Consultation and Relational Empathy Measure | Mixed | Broad | 10.5 | 0.25 | 144 | | | 155 | | | | 0.96 | 0.0089 | 0.78 | 1.15 | 0.0944 | 3 | 1 |
| 5 | | | | | | | | | | | | | Jefferson Scale of Physician Empathy | Self-report | Broad | | | 144 | -1.8 | 1.44 | 155 | 1.6 | 1.18 | 0.006 | 2.592 | 0.02471 | 2.28392 | 2.90008 | 0.15718 | | |
| 6 | | | | | | | | | | | | | Consultation and Relational Empathy Measure | Objective | Broad | | | 144 | 62.2 | 13.6 | 155 | 63.4 | 16.5 | 0.488 | 0.079 | 0.01341 | -0.14794 | 0.30594 | 0.11579 | | |
| 7 | 0 | 0 | empathy skills training and teaching with simulated patients | Cognitive, affective and behavioral | Yes | Mixed | Yes | Yes | Yes | Yes | No | Active | Jefferson Scale of Physician Empathy / Standardised patient / Experts | Mixed | Broad | 4.5 | 0.75 | 79 | | | 79 | | | | 0.55 | 0.0084 | 0.37 | 0.73 | 0.0918 | 3 | 0.5 |
| 8 | | | | | | | | | | | | | Jefferson Scale of Physician Empathy | Self-report | Broad | | | 79 | | | 79 | | | >0.05 | 0.105 | 0.02399 | -0.19855 | 0.40855 | 0.15487 | | |
| 9 | | | | | | | | | | | | | Standardised patient | Objective | Broad | | | 79 | 4.1 | 0.73 | 79 | 4.65 | 0.67 | <0.001 | 0.785 | 0.02729 | 0.46121 | 1.10879 | 0.1652 | | |
| 10 | | | | | | | | | | | | | Experts | Objective | Broad | | | 79 | 3.9 | 0.79 | 79 | 4.55 | 0.79 | <0.001 | 0.823 | 0.02749 | 0.49805 | 1.14795 | 0.16579 | | |
| 11 | 0 | 0 | low-fidelity simulation techniques (case discussions and a video show) | Cognitive, affective and behavioral | Yes | Mixed | Yes | Yes | Yes | Yes | No | Active | Toronto Empathy Questionnaire | Self-report | Broad | 4.5 | 1 | 31 | -1.08 | 1.8 | 63 | 2.94 | 1.25 | <0.001 | 2.768 | 0.0898 | 2.1807 | 3.3553 | 0.2996 | 3 | 0.5 |
| 12 | 0 | 0 | Mindfulness-Based Stress Reduction | Cognitive, affective | No | Mixed | Yes | Yes | Yes | No | No | Active | Jefferson Scale of Physician Empathy | Self-report | Broad | 16 | 12 | 84 | -1.2 | 1.49 | 83 | 0.26 | 1.68 | <0.05 | 0.92 | 0.0265 | 0.6008 | 1.2392 | 0.1628 | 3 | 0.5 |
| 13 | 0 | 0 | training in electronic medical record-specific communication | Cognitive, affective and behavioral | No | Mixed | Yes | Yes | Yes | No | No | Active | Jefferson Scale of Physician Empathy / Standardised patient / Experts | Mixed | Broad | 1 | 1.5 | | | | | | | | 0.68 | 0.0204 | 0.4 | 0.96 | 0.1429 | 3 | 0.5 |
| 14 | | | | | | | | | | | | | Jefferson Scale of Physician Empathy | Self-report | Broad | | | 32 | -1.62 | 3.58 | 38 | 0.82 | 2.24 | <0.05 | 0.833 | 0.06267 | 0.34234 | 1.32366 | 0.25034 | | |
| 15 | | | | | | | | | | | | | Standardised patient | Objective | Broad | | | 32 | 3.1 | 0.9 | 38 | 3.5 | 0.9 | 0.07 | 0.444 | 0.05902 | -0.03214 | 0.92014 | 0.24293 | | |
| 16 | | | | | | | | | | | | | Experts | Objective | Broad | | | 32 | 3.9 | 0.6 | 38 | 4.4 | 0.7 | 0.02 | 0.762 | 0.06184 | 0.27461 | 1.24939 | 0.24867 | | |
| 17 | 0 | 0 | Mindfulness-Based Exercise | | No | Didactic | Yes | Yes | Yes | No | No | Active | Visual Analogue Scale / Objective measure | Mixed | Narrow | 2 | 0 | | | | | | | | 0.07 | 0.0242 | -0.23 | 0.38 | 0.1556 | 3 | 0.5 |
| 18 | | | | | | | | | | | | | Visual Analogue Scale | Self-report | Narrow | | | 42 | 46.84 | 11.72 | 41 | 45.12 | 13.73 | | -0.135 | 0.04831 | -0.56581 | 0.29581 | 0.2198 | | |
| 19 | | | | | | | | | | | | | Objective Measure | Objective | Narrow | | | 42 | 1.52 | 0.69 | 41 | 1.72 | 0.71 | | 0.286 | 0.0487 | -0.14656 | 0.71856 | 0.22069 | | |
| 20 | 0 | 0 | cognitively-based compassion training | Cognitive, affective and behavioral | No | Experiential/mixed | Yes | Yes | Yes | No | Yes | Waiting list | Compassionate Love for Humanity Scale | Self-report | Broad | 15 | 2.5 | 11 | | | 21 | | | 0.01 | 1.0235 | 0.155 | 0.2521 | 1.7949 | 0.3937 | 3 | 0.5 |
| 21 | 0 | 0 | micro- and subtle-expression reading skill training | Cognitive and behavioral | No | skills training | Yes | Yes | Yes | No | No | Active | Micro Expression Training Tool / Subtle Expression Training Tool | Self-report | Broad | 1 | 0.25 | 41 | | | 41 | | | | 0.94 | 0.0275 | 0.61 | 1.26 | 0.1658 | 3 | 0.5 |
| 22 | | | | | | | | | | | | | Micro Expression Training Tool | Self-report | Broad | | | 41 | 1.8 | 5.2 | 41 | 5.8 | 6.2 | 0.003 | 0.699 | 0.05183 | 0.25276 | 1.14524 | 0.22767 | | |
| 23 | | | | | | | | | | | | | Subtle Expression Training Tool | Self-report | Broad | | | 41 | -0.3 | 3.5 | 41 | 4.9 | 5 | <0.001 | 1.205 | 0.05786 | 0.73356 | 1.67644 | 0.24053 | | |
| 24 | 0 | 0 | mindfulness based stress reduction (MBSR) program | Cognitive, affective and behavioral | No | Experiential | No | Yes | Yes | Yes | Yes | Waiting list | Jefferson Scale of Physician Empathy | Self-report | Broad | 12 | 2 | 15 | -1.5 | 10.6 | 15 | 1.7 | 8.7 | >0.05 | 0.33 | 0.1353 | -0.391 | 1.0509 | 0.3678 | 3 | 0.5 |
| 25 | 0 | 0 | inner relationship focusing | Cognitive, affective | Yes | Mixed | Yes | Yes | Yes | Yes | No | Active | Self-reported measure | Self-report | Narrow | 20 | 5 | 7 | -0.25 | 0.67 | 6 | 0.54 | 1.55 | | 0.683 | 0.3307 | -0.444 | 1.8102 | 0.5751 | 3 | 0.5 |
| 26 | 0 | 0 | arts-making workshop | Cognitive, affective | No | Experiential | Yes | Yes | No | No | No | Active | Jefferson Scale of Physician Empathy | Self-report | Broad | 3 | 2.5 | 58 | -0.6 | 2.45 | 48 | -4.6 | 2.73 | 0.13 | -0.301 | 0.0385 | -0.686 | 0.0836 | 0.1962 | 3 | 0.5 |
| 27 | 0 | 0 | play-reading | Cognitive | No | Experiential | No | Yes | Yes | No | Yes | Active | Jefferson Scale of Physician Empathy | Self-report | Broad | 1 | 4 | 66 | 2.8 | 2.2 | 63 | 4.7 | 2.9 | <0.05 | 0.741 | 0.0332 | 0.3839 | 1.0981 | 0.1822 | 3 | 1.5 |
| 28 | 32 | 0 | injection skill using themselves as surrogate patients | Behavioral | Yes | skills training | Yes | Yes | Yes | Yes | No | Active | Objective Measure | Objective | Narrow | 2 | 0 | 57 | -0.17 | 1.9 | 32 | 0.72 | 1.89 | 0.036 | 0.469 | 0.0501 | 0.0305 | 0.9075 | 0.2237 | 3 | 0.5 |
| 29 | 66 | 0 | Motivational interviewing training | Cognitive, affective and behavioral | Yes | Experiential | Yes | Yes | Yes | Yes | Yes | Active | Motivational Interviewing Treatment Integrity 3.0 | Objective | Broad | 8 | 0.25 | 49 | 3.4 | 0.7 | 42 | 4 | 0.6 | <0.001 | 0.915 | 0.0489 | 0.4815 | 1.3485 | 0.2212 | 3 | 0.5 |
| 30 | 0 | 0 | University of Toronto's Therapeutic Communication Program | Cognitive, affective and behavioral | Yes | Mixed | Yes | Yes | Yes | Yes | No | Waiting list | Self-Assessment of Interpersonal Competence Questionnaire / Interpersonal Skills Rating Scale / Staff Patient Interaction Rating Scale | Mixed | Broad | 16 | 4 | 41 | | | 38 | | | | 0.08 | 0.0169 | -0.17 | 0.34 | 0.1301 | 3 | 0.5 |
| 31 | | | | | | | | | | | | | Self Assessment of Interpersonal Competence Questionnaire | Self-report | Broad | | | 41 | 0.02 | 0.28 | 38 | -0.09 | 0.29 | >0.05 | -0.386 | 0.0517 | -0.832 | 0.0595 | 0.2273 | | |
| 32 | | | | | | | | | | | | | Interpersonal Skills Rating Scale | Objective | Broad | | | 41 | 2.9 | 16.8 | 38 | 5.7 | 16.8 | >0.05 | 0.167 | 0.0509 | -0.275 | 0.6091 | 0.2256 | | |
| 33 | | | | | | | | | | | | | Staff-Patient Interaction Rating Scale | Objective | Broad | | | 41 | -0.68 | 6.2 | 38 | 2.29 | 6.3 | 0.038 | 0.475 | 0.0522 | 0.0273 | 0.9227 | 0.2284 | | |
| 34 | 0 | 0 | Online clerkship | Cognitive, affective and behavioral | Yes | Experiential | Yes | Yes | No | Yes | No | Active | Self-reported measure | Self-report | Narrow | 26 | 1.5 | 103 | | | 112 | | | | 0.426 | 0.0246 | 0.1186 | 0.7334 | 0.1568 | 3 | 0.5 |

| | BM | BN | BO | BP | BQ | BR | BS | BT | BU | BV | BW | BX | BY | BZ | CA | CB | CC | CD | CE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | | | | | | | |
| 3 | Sampling: response rate | Type of data | Internal structure | Content | Relationships to other variables | Appropriateness of analysis | Complexity of analysis | Outcomes | Sum | Study design | Sampling Institions | Sampling: response rate | Type of data | Internal structure | Content | Relationships to other variables | Appropriateness of analysis | Complexity of analysis | Outcomes |
| 4 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 2 | 16.5 | Randomized controlled trial | 2 | ≥75 | Objective measurement | Reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Behaviours |
| 5 | | | | | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | | | | | |
| 7 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 2 | 16 | Randomized controlled trial | 1 | ≥75 | Objective measurement | Reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Behaviours |
| 8 | | | | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | | | | | | |
| 11 | 1.5 | 1 | 0 | 1 | 0 | 1 | 2 | 2 | 12 | Randomized controlled trial | 1 | ≥75 | Assessment by study participant | Not reported | Reported | Not reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Behaviours |
| 12 | 1 | 1 | 0 | 1 | 0 | 1 | 2 | 1.5 | 11 | Randomized controlled trial | 1 | 50-74 | Assessment by study participant | Not reported | Reported | Reported | Data analysis inappropriate for study design or type of data | Beyond descriptive analysis | Knowledge, Skills |
| 13 | 1.5 | 3 | 1 | 1 | 1 | 1 | 2 | 1.5 | 15.5 | Randomized controlled trial | 1 | ≥75 | Objective measurement | Reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Knowledge, Skills |
| 14 | | | | | | | | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | | | | | | | | |
| 17 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 2 | 15 | Randomized controlled trial | 1 | ≥75 | Objective measurement | Not reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Behaviours |
| 18 | | | | | | | | | | | | | | | | | | | |
| 19 | | | | | | | | | | | | | | | | | | | |
| 20 | 1 | 1 | 0 | 1 | 1 | 1 | 2 | 2 | 12.5 | Randomized controlled trial | 1 | 50-74 | Assessment by study participant | Not reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Behaviours |
| 21 | 1.5 | 1 | 0 | 1 | 1 | 1 | 2 | 1.5 | 12.5 | Randomized controlled trial | 1 | ≥75 | Assessment by study participant | Not reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Knowledge, Skills |
| 22 | | | | | | | | | | | | | | | | | | | |
| 23 | | | | | | | | | | | | | | | | | | | |
| 24 | 1 | 1 | 0 | 1 | 1 | 1 | 2 | 1.5 | 12 | Randomized controlled trial | 1 | 50-74 | Assessment by study participant | Not reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Knowledge, Skills |
| 25 | 1.5 | 1 | 0 | 1 | 1 | 1 | 2 | 1 | 12 | Randomized controlled trial | 1 | ≥75 | Assessment by study participant | Not reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Satisfaction, attitudes, perceptions, opinions, general facts |
| 26 | 1 | 1 | 0 | 1 | 1 | 1 | 2 | 1.5 | 12 | Randomized controlled trial | 1 | 50-74 | Assessment by study participant | Not reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Knowledge, Skills |
| 27 | 1.5 | 1 | 1 | 1 | 1 | 1 | 2 | 1.5 | 14.5 | Randomized controlled trial | >2 | ≥75 | Assessment by study participant | Reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Knowledge, Skills |
| 28 | 1.5 | 3 | 0 | 1 | 0 | 1 | 2 | 1.5 | 13.5 | Randomized controlled trial | 1 | ≥75 | Objective measurement | Not reported | Reported | Not reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Knowledge, Skills |
| 29 | 1 | 3 | 1 | 1 | 1 | 1 | 2 | 2 | 15.5 | Randomized controlled trial | 1 | 50-74 | Objective measurement | Reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Behaviours |
| 30 | 1.5 | 3 | 0 | 1 | 1 | 1 | 2 | 2 | 15 | Randomized controlled trial | 1 | ≥75 | Objective measurement | Not reported | Reported | Reported | Data analysis appropriate for study design and type of data | Beyond descriptive analysis | Behaviours |
| 31 | | | | | | | | | | | | | | | | | | | |
| 32 | | | | | | | | | | | | | | | | | | | |
| 33 | | | | | | | | | | | | | | | | | | | |
| 34 | 1.5 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 9 | Randomized controlled trial | 1 | ≥75 | Assessment by study participant | Not reported | Reported | Not reported | Data analysis inappropriate for study design or type of data | Descriptive analysis only | Satisfaction, attitudes, perceptions, opinions, general facts |

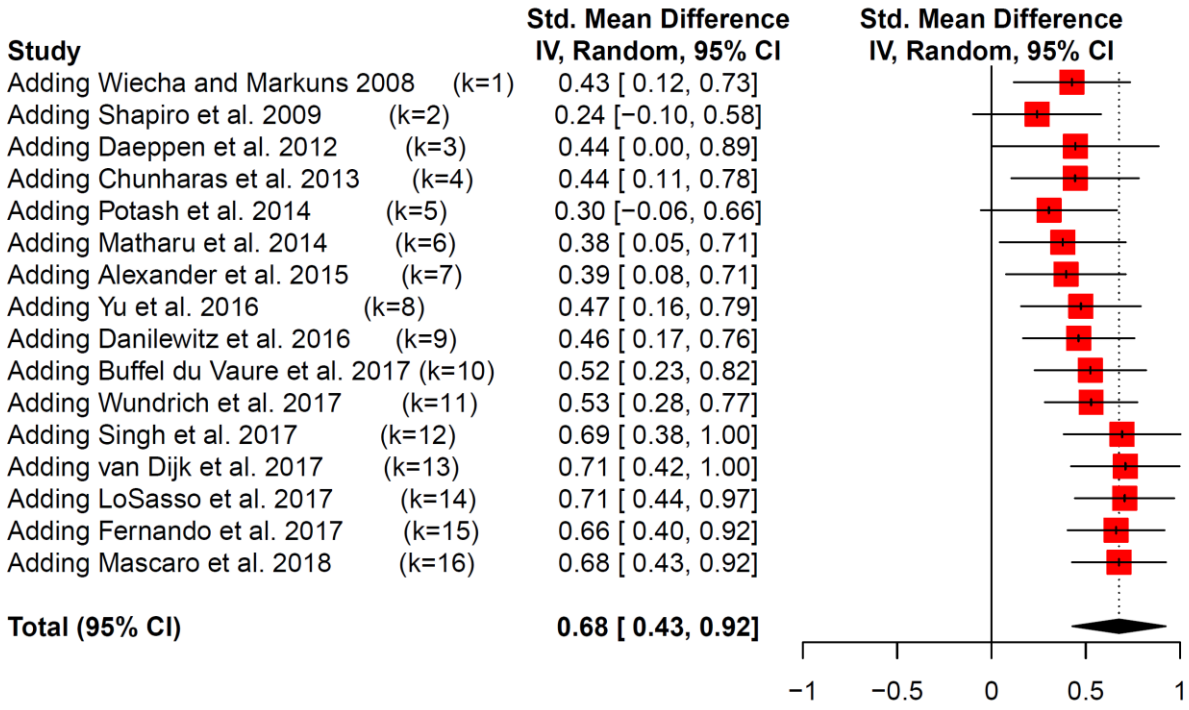
Supplemental Digital Appendix 4. Figure showing the distribution of students by year of medical school.



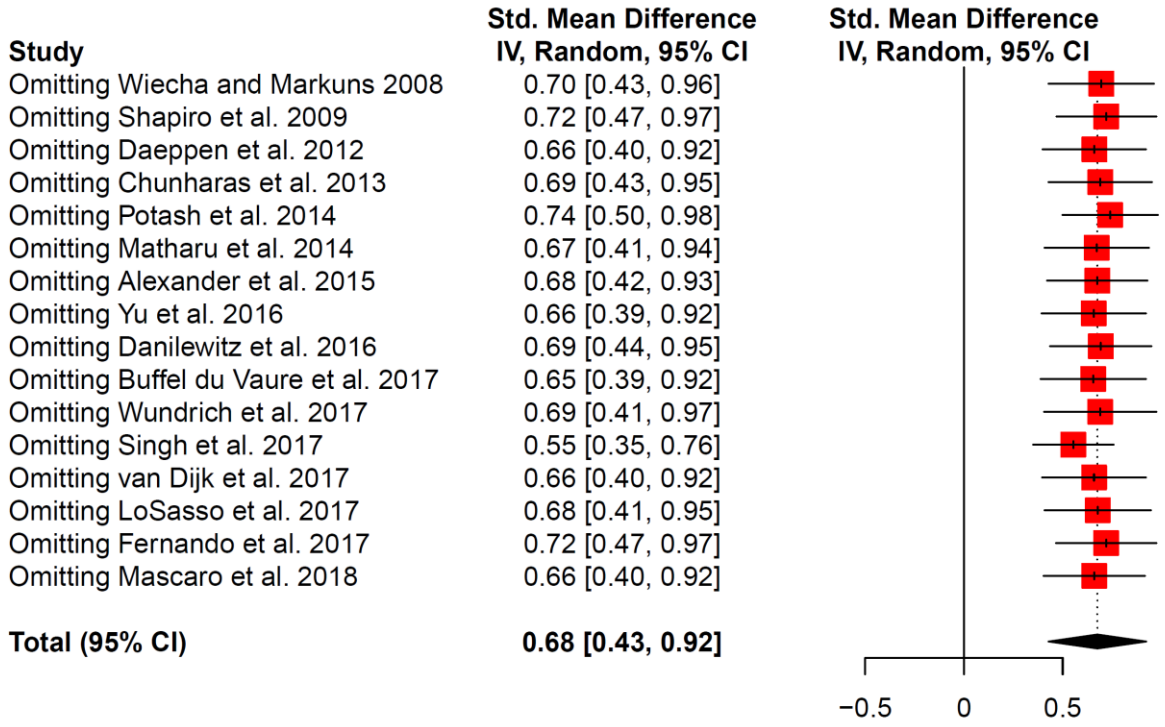
Supplemental Digital Appendix 5. Funnel plot for publication bias. Black dots indicate studies of the review and the empty dot is the imputed study according to the method of Duval and Tweedie.36 No asymmetry seen.



Supplemental Digital Appendix 6. Figure showing the cumulative meta-analysis of SMD for empathy.



Supplemental Digital Appendix 7. Figure showing the sensitivity analysis of SMD for empathy.

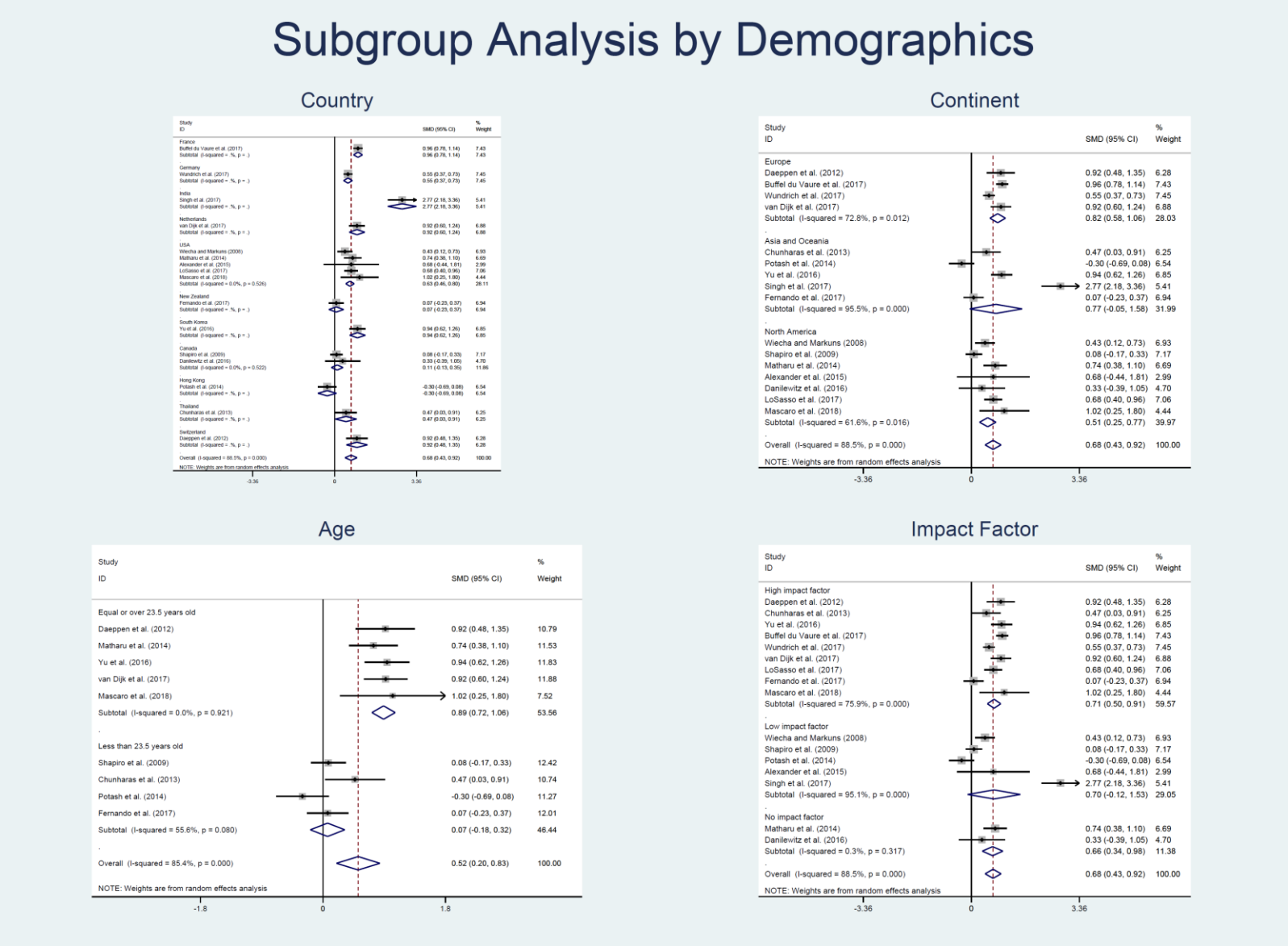


Supplemental Digital Appendix 8. Table showing meta-regression results of continuous variables on the effect size.

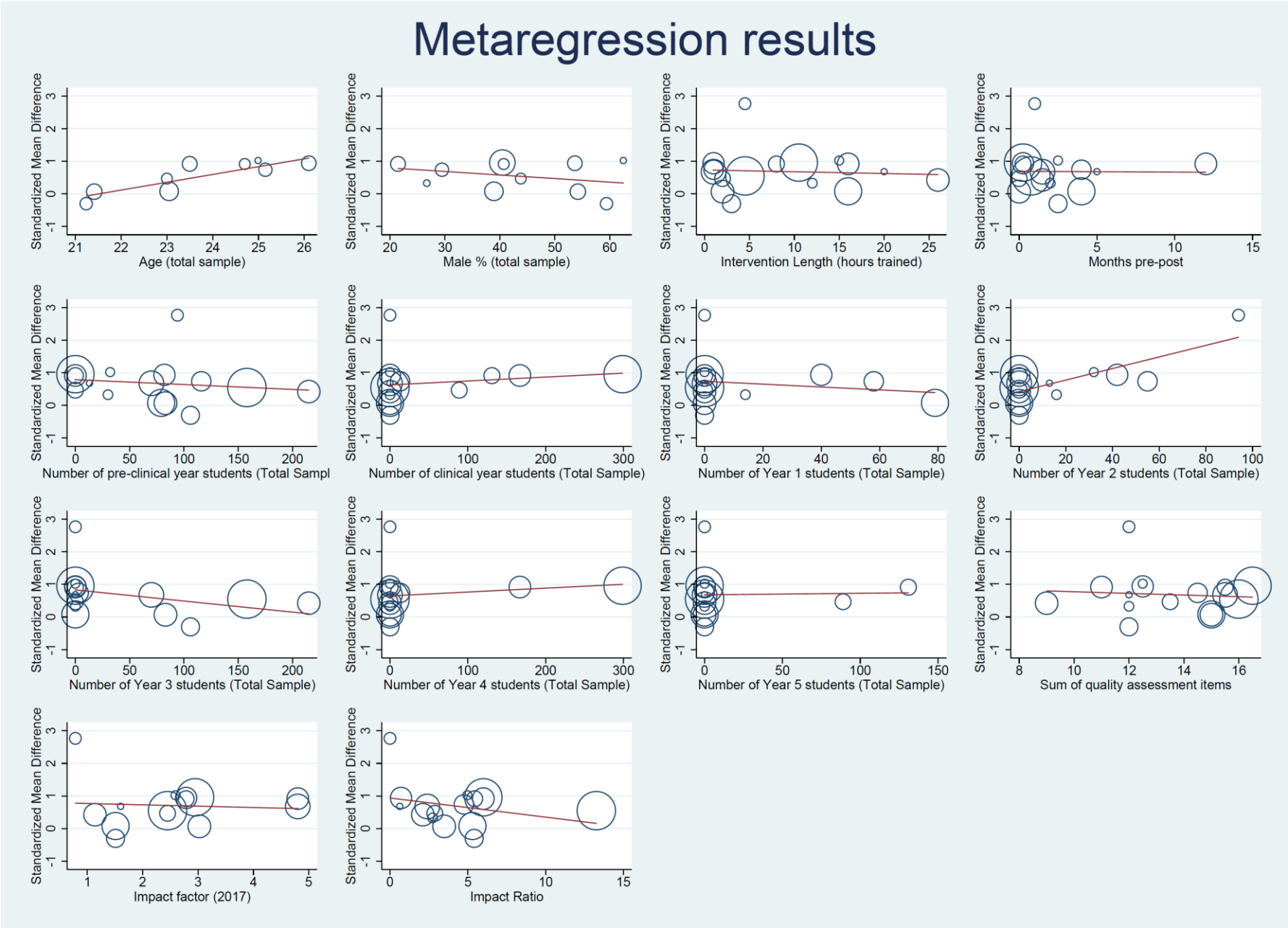
Table 1. Meta-regression results of continuous variables on the effect size.

| **Moderators** | **Univariate models** | | | **Multivariate model Adjusted $R^2$ = 77.2%, p = 0.02** | |
|---|---|---|---|---|---|
| | ***β*** | **p** | **Adjusted $R^2$** | ***β*** | **p** |
| *Age* | 0.24 | <0.01 | 77.1% | 0.31 | 0.01 |
| *Male %* | -0.01 | 0.36 | 5.0% | -- | -- |
| *No. of Year 1 students* | -0.004 | 0.53 | 0.0% | -- | -- |
| *No. of Year 2 students* | 0.02 | <0.01 | 54.9% | -0.01 | 0.33 |
| *No. of Year 3 students* | -0.003 | 0.17 | 8.3% | -- | -- |
| *No. of Year 4 students* | 0.001 | 0.56 | 0.0% | -- | -- |
| *No. of Year 5 students* | 0.0004 | 0.92 | 0.0% | -- | -- |
| *No. of pre-clinical students* | -0.002 | 0.59 | 0.0% | -- | -- |
| *No. of clinical students* | 0.001 | 0.54 | 0.0% | -- | -- |
| *Intervention Length (hours)* | -0.01 | 0.81 | 0.0% | -- | -- |
| *Months pre-post assessment* | -0.002 | 0.97 | 0.0% | -- | -- |
| *Medical Education Research Study Quality Instrument (MERSQI) score* | -0.03 | 0.76 | 0.0% | -- | -- |
| *Impact factor* | -0.04 | 0.81 | 0.0% | -- | -- |
| *Impact ratio* | -0.06 | 0.28 | 0.3% | -- | -- |

Supplemental Digital Appendix 9. Figure showing subgroup analyses of SMD for empathy by demographics (country, continent, age, and journal impact factor).



Supplemental Digital Appendix 10. Figure showing the univariate meta-regression graphs by moderator.



Supplemental Digital Appendix 11. Table showing empathy characteristics and operationalizations by study.

Table 1. Empathy characteristics and operating definitions used throughout the studies.

| **Study** | **Empathy** | | | | **Operating definition and scale** |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| Mascaro et al.[1] | Cognitive, affective, behavioral | Compassionate Love for Humanity Scale | Self-report | Broad | • **Reason for inclusion:** The authors recognize cognitive elements of empathy are closely related to compassion. When the scale used is closely examined, it has "caring" or "concern" as well as behavioral aspects, and hence it can be deduced that their approach is indeed very similar to empathy and hence was included. Their **definition** on empathy is (verbatim):<br><br>*"Researchers have long recognized that imprecise construct definitions have hampered research on prosocial emotions and behaviors,[2] and noted the apparent disconnect between research in the domains of medical education, on the one hand, and in the domains of social psychology and social cognitive neuroscience on the other.[3] Studies of medical education emphasize a more cognitive definition of empathy, for example, as a 'cognitive attribute that involves an ability to understand the patient's inner experiences and perspective and a capability to communicate this understanding' (p. 1564).[4] Coupled with this approach is the idea that emotionality and affective sharing may be detrimental to both the physician's professional objectivity and his or her own well-being. In contrast, researchers in psychology and neuroscience often operationalize empathy as containing both cognitive and affective components, which allow one to experience the emotions of another while also identifying those emotions as belonging to the other.[5]*<br>*While additional studies regarding the exact nature of optimal caretaker empathy are urgently needed, the current study sets aside this question to focus on a construct closely related to empathy, namely compassion, defined as the deep wish that another be free from suffering, coupled with the motivation to alleviate such suffering.[6-8] Because studies have shown that enhancing compassion increases prosocial behavior[9] and suggest that compassion may be an interpersonal stance more closely tied to well-being than is empathy,[7] it is arguably more predictive of behavior than the aforementioned process of empathy and hence more salient for clinicians-in-training.[10]" (pp. 133-134).*<br><br>• **Assessment**: The authors subsequently used the Compassionate Love for Humanity Scale,[11] which is a 21-item scale designed to measure an attitude of concern, caring, and support for humanity that involves a motivation to understand and help others, including strangers, when they are most in need. |

| Study | Empathy | | | | Operating definition and scale |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| Buffel du Vaure et al.[12] | Cognitive, affective, behavioral | Jefferson Scale of Physician Empathy – Medical Student, Consultation and Relational Empathy Measure | Mixed | Broad | • **Reason for inclusion**: The authors **define** empathy as<br><br>*"the ability to share and/or understand others' emotional state without confusion between self and others. Clinical empathy, i.e. empathy within the context of a doctor-patient relationship, is considered as a core feature of the doctor-patient relationship,[13,14] as acknowledged by most of medical schools. For instance, the Association of American Medical Colleges states that "physicians must be compassionate and empathetic in caring for patients".[15]"*<br><br>• **Assessment**: Participants had to complete the self-rated Jefferson Scale of Physician Empathy – Medical Student. This scale encompasses 20 items Likert-type items, rated from 1 (strongly disagree) to 7 (strongly agree) (e.g. "Patients feel better when their physicians understand their feelings"), leading to a summed score ranging from 20 to 140 with higher score indicating higher levels of empathy.[16]<br>• Standardized patients fulfilled the Consultation and Relational Empathy Measure scale immediately after each Objective Structured Clinical Examination (i.e. two measures). The Consultation and Relational Empathy Measure scale is a 10-item patient-rated questionnaire of physician empathy [e.g. "How good was the practitioner at showing care and compassion (seeming genuinely concerned, connecting with you on a human level; not being indifferent or detached)?"].[17] Each item can be scored on a Likert scale from 1 (poor) to 5 (excellent), with a 'does not apply' option, leading to a summed score ranging from 10 to 50 with higher score indicating higher levels of empathy. |

| Study | Empathy | | | | Operating definition and scale |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| Fernando et al.[18] | Cognitive, affective, behavioral | Visual Analogue Scale, Objective Measure | Mixed | Narrow | • **Reason for inclusion**: The authors recognize that cognitive elements of empathy are closely related to compassion. When the clinical vignettes are examined carefully, it includes elements such as caring and concern and a behavioral response. Hence, it can be deduced that their approach is indeed very similar to empathy, leading to its inclusion. Their **definition** is:<br><br>*Compassion is an essential component of medical practice. It is expected by patients,[19-25] regulatory bodies[26,27] and doctors themselves.[28] More than being an expectation, however, compassionate care matters. Compassion predicts greater patient satisfaction, better patient-physician relationships, and better patient health outcomes.[20,29-31] In medicine as elsewhere, compassion is often confused with empathy. However, while empathy refers to the cognitive and emotional processes involved in adopting another's perspective,[32] compassion involves the additional step of wanting to relieve suffering.[6,7,33] However, sustaining compassion in medicine can be difficult[34] and compassion fatigue impacts between 20 and 70 % of physicians;[35-38] nearly half of patients and doctors report that compassionate care is missing in the health care system.[28] Developing educational approaches and interventions that enhance or sustain compassion have become serious challenges in modern medicine and medical training.[39]*<br><br>• **Assessment**: Participants underwent clinical vignettes describing initial interactions with a series of patients with challenging personal and clinical characteristics. In a standard order, participants rated how they felt towards each patient and what they would do during consultations, two important and separate aspects of physician response.[40] Ratings were made regarding how much participants liked the patient, wanted to help, and felt caring towards the patient. Because of their high sensitivity, naturalistic nature and superior numerical properties, ratings were made on 10 cm visual analogue scale.[41,42] A final rating regarding subjective closeness to patients was made using the Inclusion of Other in Self Scale, a method that asks participants to choose among a series of 7 images in which two circles overlap to varied degrees; greater overlap indicates greater feelings of closeness to the hypothetical patient.[43] Other exploratory ratings regarding possible referrals, patient responsibility, and acceding to patient demands were difficult to interpret and are not considered here. Two additional measures assessed more objective aspects of compassionate responding. First, participants were required to allocate consultation time to each patient within a constrained period (totaling 60 min across patients); allocating and balancing time is a clinically important process and differences may reflect aspects of compassion. Finally, a covert behavioral measure assessed compassionate behavior directly. Having been told the study was finished, participants were asked to help the research assistant with an unrelated administrative task (45 min of help with questionnaire compilation work that was outstanding because of a personal commitment). Given skew in the raw data, responses were scored as 0 (No help), 1 (willing to help but less than 45 min), or 2 (willing to help for 45 min or more if needed) for analytic purposes. |

| Study | Empathy | | | | Operating definition and scale |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| LoSasso et al.[44] | Cognitive, affective, behavioral | Jefferson Scale of Physician Empathy, Standardized Patients, Experts | Mixed | Broad | • **Reason for inclusion**: The authors use validated measures of empathy.<br>• **Assessment**:<br>*Jefferson Scale of Physician Empathy*. Developed via literature reviews and pilot studies in response to a gap in instruments available to measure empathy specific to medical education, the Jefferson Scale of Physician Empathy serves as a student self-report of attitudes and orientation toward empathic engagement in patient care, based on a definition of empathy as a primarily cognitive attribute.[45,46] The scale consists of 20 items answered on a seven-point Likert-type scale (1 = strongly disagree to 7 = strongly agree), with a higher score indicating a more empathetic attitude toward patient care (possible range of scores: 20-140). Sample items are "It is difficult for a physician to view things from patients' perspectives" and "Patients feel better when their physicians understand their feelings." In a series of studies, the Jefferson Scale of Physician Empathy has demonstrated construct validity,[4,47] criterion-related validity,[47,48] predictive validity,[49] internal consistency reliability,[4,47] and test-retest reliability.[4]<br>*Jefferson Scale of Patient Perceptions of Physician Empathy*. The Jefferson Scale of Patient Perceptions of Physician Empathy was designed to assess a patient's or a standardized patient's perception of a physician's or medical student's empathy.[50,51] The original scale 23 consists of five items. The six-item version of the scale was used, which was previously used by Berg et al.[45] in their study of standardized patient assessment of medical student empathy. In their study, items were answered on a five-point Likert-type scale (poor = 1, fair = 2, good = 3, very good = 4, excellent = 5), with a higher score indicating more empathic engagement (possible range of scores: 1-5). An example item is, "Did the student seem concerned about me and my family?" Psychometric evidence supporting the reliability and validity of the scale has been shown in studies of internal medicine residents[50] and family medicine residents.[51] |
| Singh et al.[52] | Cognitive, affective, behavioral | Toronto Empathy Questionnaire | Self-report | Broad | • **Reason for inclusion**: The authors use a validated measure of empathy and during their themes' discussion in the video show group and case discussion group, they used phrases which had elements of cognitive and affective empathy.<br>• **Assessment**:<br>*Group discussions*: In both the video show group and case discussion groups, the following discussion themes were used: Put yourself in place of X and feel what would be your reaction when you come to know about the incident? What would be the impact of this incident on your life when you are detected with healthcare-associated infections? Was this a preventable tragedy? Are you in a position to save such damages? What are the promises you make to yourself as a future doctor to prevent such incidents? These themes were meant to help students empathize with patients and realize the importance of compliance to infection control practices.<br>*The Toronto Empathy Questionnaire* contains 16 questions that encompass a wide range of attributes associated with theoretical facets of empathy such as emotional contagion, emotional comprehension, sympathetic physiological arousal, higher-order empathic responding, such as pro-social helping behaviors and altruism. Scoring of the items depends upon the nature of items. Positively worded items (No. 1, 3, 5, 6, 8, 9, 13, 16) were scored as Never 0; Rarely 1; Sometimes 2; Often 3; Always 4. The negatively worded items (No. 2, 4, 7, 10, 11, 12, 14, 15) were reverse scored. Scores were summed to derive the total for the Toronto empathy questionnaire score.[53] |

| Study | Empathy | | | | Operating definition and scale |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| van Dijk et al.[54] | Cognitive, affective | Jefferson Scale of Physician Empathy | Self-report | Broad | • **Reason for inclusion**: The authors use a validated measure of empathy.<br>• **Assessment**:<br>*Jefferson Scale of Physician Empathy*. The 20-item Jefferson Scale of Physician Empathy measures empathy in the physician-patient relationship. It contains statements such as "Patients feel better when their physicians understand their feelings." Items are scored on a seven-point Likert-type scale, ranging from 1 = "strongly disagree" to 7 = "strongly agree." The total score may range from 20 to 140, and higher scores indicate a higher level of empathy. Validity and reliability of the Jefferson Scale of Physician have been demonstrated to be high for physicians ([alpha] = 0.81-0.85) and medical students ([alpha] = 0.89).[4,47,55] |
| Wundrich et al.[56] | Cognitive, affective, behavioral | Jefferson Scale of Physician Empathy, Standardized Patients, Experts | Mixed | Broad | • **Reason for inclusion**: The authors use validated measures of empathy. They also use mainstream definitions of empathy. They subsequently engage students in training which gives the definition of empathy and asks students to reflect on their own experience as a patient. Specifically they **define** empathy with:<br>*A widely used model differentiates four dimensions of empathy:[57] a moral, cognitive, emotive, and behavioral dimension. In an extension, empathy has been proposed to develop as a process over time.[58] This process includes the following: (i) an inner process of listening, reasoning and understanding, (ii) the communication of this awareness by the empathizing person, and (iii) the perception of being understood by the counterpart.*<br>• **Assessment**:<br>During the objective clinical structured examination, empathy was rated by blinded experts and Standardized Patients who had been trained with a five-hour session that explained the empathy construct and the rating instrument.<br>Experts and Standardized Patients used a questionnaire representing the 11 aspects of empathy-related communication skills as well as general interview techniques: Active listening; Understanding the situation; Understanding the problems; Understanding feelings; Explanation (of the illness, drugs, and so on); Shared decision-making; Communicating hope; Being competent; Verbal expression; Non-verbal expression; Degree of coherence in the interview.<br>The first 10 main dimensions were defined by four experienced psychotherapists in a critical review process of the aforementioned empathy models. Each dimension of this construct is rated on a Likert scale from 1 ("not fulfilled at all") to 5 ("completely fulfilled").<br>Student Self-assessment: Students filled out the German student version of the Jefferson Scale of Physician Empathy,[59] that represents the assessment of the attitudinal dimension of empathy. |
| Danilewitz et al.[60] | Cognitive, affective, behavioral | Jefferson Scale of Physician Empathy | Self-report | Broad | • **Reason for inclusion**: The authors use a validated measure of empathy.<br>• **Assessment**:<br>*Jefferson Scale of Physician Empathy:* 20-item questionnaire assessing health care professional empathy in patient care settings |

| Study | Empathy | | | | Operating definition and scale |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| Yu et al.[61] | Cognitive, behavioral | Micro Expression Training Tool, Subtle Expression Training Tool | Self-report | Broad | • **Reason for inclusion**: The authors provide a definition of empathy and subsequently point out how communication skills are needed for empathy. Hence they have focused on the cognitive aspect of empathy. Their **definition** is:<br><br>*Empathy is the ability to identify and understand the feelings and emotions of others, and in physicians, empathy-building improves patient outcomes and reduces complaints about medical services.[62,63] Communication skill training and education aimed at improving empathy are indispensable for preparing medical students to be good doctors.[64,65] There are two basic categories of communication: verbal and non-verbal, and non-verbal skills may have greater effects on patient trust and satisfaction with medical services.[66] A patient-doctor communication skills model that includes understanding of patient's perspective has been developed,[67] but the time allowed within medical curricula for teaching and reinforcing non-verbal communication skills is insufficient.*<br><br>• **Assessment**:<br>*Micro Expression Training Tool* is a training tool designed to improve one's ability to spot micro expressions, that is, very brief facial expressions, which last <1/2 s. Micro Expression Training Tool consisted of anger, fear, sadness, disgust, contempt, surprise and happiness as universal emotions, which can be easily recognized by different racial-culture groups.[68] For test consistency, 42 randomly selected pictures of facial micro expressions showing these seven emotions were pre- or post-tested in the training and control groups.<br>*Subtle Expression Training Tool is* a training tool that improves one's ability to recognize small facial expressions that sometimes register in only a portion of the face, and usually last between 1/15 and 1/25 of a second. For test consistency, 41 randomly selected facial subtle expressions were pre- or post-tested in both groups. |

| Study | Empathy | | | | Operating definition and scale |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| Alexander et al.[69] | Cognitive, affective | Self-reported Measure | Self-report | Narrow | • **Reason for inclusion**: The authors provide a definition of empathy and use narrow scope items which reflect empathy. Their **definition** is:<br><br>*The medical literature suggests that physician empathy is based primarily on a combination of reflective listening and mindfulness. In a large study on empathy, Winseman et al.[70] identified listening as among the top factors helping medical students develop a deeper understanding of patients. Mindfulness-based interventions also promote physician empathy.[64] Thus, we decided to offer a novel elective on mind-body awareness that included an empathic listening technique called inner relationship focusing… Describing bodily felt experiences can awaken emotions that may not be available through intellectual or linguistic pathways. Gendlin collaborator, Carl Rogers, emphasized empathic listening as the core component of trust and healing. Once a strong alliance, such as between client and clinician, is established through empathy, it becomes a foundation for behavioral change.[71] Empathy and listening are central to the physician-patient relationship. To the extent that inner relationship focusing teaches these skill sets, it could potentially be of value to physicians in many fields.*<br><br>• **Assessment**:<br>Items which were associated to empathy and answered on a Likert scale: Relating recurrent physical/emotional symptoms to patients' life issues; Accessing one's own experience; Being a fully present listener; Reflective listening to help patients de-identify with problem |
| Matharu et al.[72] | Cognitive | Jefferson Scale of Physician Empathy | Self-report | Broad | • **Reason for inclusion**: The authors use a validated measure of empathy.<br>• **Assessment**:<br>*The Jefferson Scale of Physician Empathy*, with score ranges of 20-180, was included for assessment of empathy and has been utilized in medical student and resident studies on professionalism and humility.[55] Cronbach's alpha values for the empathy scale in the study was 0.76. |
| Potash et al.[73] | Cognitive, affective | Jefferson Scale of Physician Empathy | Self-report | Broad | • **Reason for inclusion**: The authors use a validated measure of empathy.<br>• **Assessment**:<br>*Jefferson Scale of Physician Empathy.* Students completed the Revised Jefferson Scale of Physician Empathy – Student Version[4,74] during the initial orientation briefing and at the conclusion of each block. The Jefferson Scale of Physician Empathy consisted of 20 statements rated on a 7-point Likert scale (1 = Strongly Disagree,…, 7 = Strongly Agree). A higher Jefferson Scale of Physician Empathy score equated to a higher level of empathy. This measure was developed and validated in a US medical school[74] and has been validated in the Asian context among medical and nursing students.[75-77] |

| Study | Empathy | | | | Operating definition and scale |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| Chunharas et al.[78] | Cognitive, behavioral | Objective Measure | Objective | Narrow | • **Reason for inclusion**: The authors use a self-defined item to measure empathy. The base their concept of empathy on the ability to reflect on their own experience as a patient. Specifically they **define** empathy with the phrase:<br><br>*Anecdotally, our retired staffs informed us that they learned by performing each other with some technical skills such as nasogastric tube insertion, finger puncture, and venipuncture. Subsequently, they became more confident and developed a higher level of empathy before performing procedure to their patients.*<br><br>• **Assessment**: The students were asked to complete a four point Likert-type scales developed by the one of the authors: the feeling of empathy to the children who were injected by the students comparing before and after completion of the injection. |
| Daeppen et al.[79] | Cognitive, affective, behavioral | Motivational Interviewing Treatment Integrity 3.0 | Objective | Broad | • **Reason for inclusion**: The authors use a validated measure of empathy.<br>• **Assessment**: The Motivational Interviewing Treatment Integrity 3.0 is a reliable and valid behavioral coding system[80,81] that provides information on the use of Motivational Interviewing by practitioners, as a treatment integrity measure for Motivational Interviewing clinical trials. Its components are: (1) global scores, which require the coder to assign a single number from a five-point scale to characterize the entire interaction and are meant to capture the global impression, or overall judgment, of the dimension by the rater. They are Empathy, Direction and Motivational Interviewing Spirit (itself comprised of three subscales, i.e., evocation, collaboration, and autonomy/support), which are averaged to give a single score; and (2) behavior counts, where the coder is asked to tally (but not judge the quality or overall adequacy of) instances of certain interviewer behaviors. |
| Shapiro et al.[82] | Cognitive, affective, behavioral | Self-Assessment of Interpersonal Competence Questionnaire, Interpersonal Skills Rating Scale, Staff-Patient Interaction Rating Scale | Mixed | Broad | • **Reason for inclusion**: The authors use validated measures of empathy and consider empathy in terms of communications.<br>• **Assessment**:<br>1. The participants completed the Self Assessment of Interpersonal Competence Questionnaire,[83] a 40-item scale, with a Likert scale of 1 (I'm poor at this) to 5 (I'm extremely good at this). Examples of items are: "**Carrying on conversations with someone new whom you think you might like to get to know**" and "**Turning down a request by a companion that is unreasonable**".<br>(b) Participants interviewed two standardized patients,[84] actors portraying patients with psychosocial problems, trained and tested for reliability by the Department of Family and Community Medicine Standardized Patient Program of the University of Toronto. The interviews were rated by the standardized patients using the Interpersonal Skills Rating Scale.[85] There are 7 items, with a Likert scale of 1 (strongly disagree) to 7 (strongly agree). Examples are: "The doctor wanted to understand how I saw things" and "The doctor just took no notice of some things that I thought or felt".<br>(c) The Staff-Patient Interaction Rating Scale,[86,87] a reliable (test-retest r = 0.79) and valid (φ = 0.67-0.78) instrument designed to assess participants' **expressed empathy** based on their written open-ended responses to a series of 24 statements made by hypothetical patients, was administered. Examples of statements are: "Why do I have to keep on seeing you?" and "I just want to do nothing and stay in bed". The responses to these items were rated by trained external raters according to a manual that describes in detail how to classify responses into disengaging and engaging sets. The score for each of the 24 items on the scale ranges from -1 for a disengaging response, 0 for a neutral response, to +1 for an engaging response. |

| Study | Empathy | | | | Operating definition and scale |
|---|---|---|---|---|---|
| | **Features** | **Measurement tool** | **Self-report vs. objective measure** | **Scope** | |
| Wiecha and Markuns[88] | Cognitive, affective, behavioral | Self-reported Measure | Self-report | Narrow | • **Reason for inclusion**: The authors define empathy in terms as being very similar to humanism. They include in their educational intervention teachings related to empathy and their assessment includes issues of caring and consideration, hence issues of cognitive, affective and behavioral empathy. They **define** as:<br><br>*Medical humanism has been described as fostering relationships with patients that are compassionate and empathetic and includes attitudes and behaviors that are sensitive to the values, autonomy, and cultural and ethnic backgrounds of others.[89] It has been suggested that a hidden curriculum in medicine is responsible for a decline in empathy over the course of medical training.[90-93]*<br><br>• **Assessment**:<br>1. During week 4 of their intervention, students read a narrative case study on empathy[94] and reflect on the reading and how it applies to the students' experiences, followed by a posting addressing these issues.<br>2. During evaluation, competence was self-assessed using a 5-point Likert scale on items "Integrating patient's cultural beliefs about health into your care of that patient" and "Eliciting how a patient has been emotionally impacted by an illness" |

## References


1. Mascaro JS, Kelley S, Darcher A, et al. Meditation buffers medical student compassion from the deleterious effects of depression. J Posit Psychol. 2018;13(2):133-142.
2. Stepien KA, Baernstein A. Educating for empathy. A review. J Gen Intern Med. 2006;21(5):524-530.
3. Preusche I, Lamm C. Reflections on empathy in medical education: What can we learn from social neurosciences? Adv Health Sci Educ Theory Pract. 2016;21(1):235-249.
4. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: definition, components, measurement, and relationship to gender and specialty. Am J Psychiatry. 2002;159(9):1563-1569.
5. de Vignemont F, Singer T. The empathic brain: how, when and why? Trends Cogn Sci. 2006;10(10):435-441.
6. Goetz JL, Keltner D, Simon-Thomas E. Compassion: an evolutionary analysis and empirical review. Psychol Bull. 2010;136(3):351-374.
7. Klimecki OM, Leiberg S, Lamm C, Singer T. Functional neural plasticity and associated changes in positive affect after compassion training. Cereb Cortex. 2013;23(7):1552-1561.
8. Kim JW, Kim SE, Kim JJ, et al. Compassionate attitude towards others' suffering activates the mesolimbic neural system. Neuropsychologia. 2009;47(10):2073-2081.
9. Leiberg S, Klimecki O, Singer T. Short-term compassion training increases prosocial behavior in a newly developed prosocial game. PLoS One. 2011;6(3):e17798.
10. Mascaro JS, Darcher A, Negi LT, Raison CL. The neural mediators of kindness-based meditation: a theoretical model. Front Psychol. 2015;6:109.
11. Sprecher S, Fehr B. Compassionate love for close others and humanity. J Soc Pers Relat. 2016;22(5):629-651.
12. Buffel du Vaure C, Lemogne C, Bunge L, et al. Promoting empathy among medical students: A two-site randomized controlled study. J Psychosom Res. 2017;103:102-107.
13. Kelm Z, Womer J, Walter JK, Feudtner C. Interventions to cultivate physician empathy: a systematic review. BMC Med Educ. 2014;14:219.
14. Hemmerdinger JM, Stoddart SD, Lilford RJ. A systematic review of tests of empathy in medicine. BMC Med Educ. 2007;7:24.
15. Learning objectives for medical student education--guidelines for medical schools: report I of the Medical School Objectives Project. Acad Med. 1999;74(1):13-18.
16. Bitoun A. Validation du score JSPE-Medical Student© en langue française dans l'évaluation de l'empathie des étudiants en médecine (MD dissertation), 2017.
17. Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Fam Pract. 2004;21(6):699-705.
18. Fernando AT, Skinner K, Consedine NS. Increasing Compassion in Medical Decision-Making: Can a Brief Mindfulness Intervention Help? Mindfulness. 2017;8(2):276-285.
19. Emanuel EJ, Dubler NN. Preserving the physician-patient relationship in the era of managed care. JAMA. 1995;273(4):323-329.
20. Fogarty LA, Curbow BA, Wingard JR, McDonnell K, Somerfield MR. Can 40 seconds of compassion reduce patient anxiety? J Clin Oncol. 1999;17(1):371-379.
21. Meagher G. What Can We Expect from Paid Carers? Polit Soc. 2016;34(1):33-54.
22. Old A, Adams B, Foley P, White HD. Society's expectation of the role of the doctor in New Zealand: results of a national survey. N Z Med J. 2011;124(1342):10-22.
23. Wiggins MN, Coker K, Hicks EK. Patient perceptions of professionalism: implications for residency education. Med Educ. 2009;43(1):28-33.
24. Youngson R. Compassion in healthcare—the missing dimension of healthcare reform. In (Ed.), (p. 37). . In: Renzenbrink I. Caregiver Stress and Staff Support in Illness, Dying, and Bereavement. Oxford: Oxford University Press; 2011.
25. Youngson R. Time to care: how to love your patients and your job. New Zealand: Rebelheart Publishers; 2012.
26. American Medical Association. AMA's Code of Medical Ethics. 2012; http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/principles-medical-ethics.page. Accessed 24 October, 2012.
27. New Zealand Medical Association. NZMA Code of Ethics. 2012; http://www.nzma.org.nz/about/ethics.html. Accessed 24 October, 2012.
28. Lown BA, Rosen J, Marttila J. An agenda for improving compassionate care: a survey shows about half of patients say such care is missing. Health Aff (Millwood). 2011;30(9):1772-1778.
29. Del Canale S, Louis DZ, Maio V, et al. The relationship between physician empathy and disease complications: an empirical study of primary care physicians and their diabetic patients in Parma, Italy. Acad Med. 2012;87(9):1243-1249.
30. Lelorain S, Bredart A, Dolbeault S, Sultan S. A systematic review of the associations between empathy measures and patient outcomes in cancer care. Psychooncology. 2012;21(12):1255-1264.

**31.** Steinhausen S, Ommen O, Thum S, et al. Physician empathy and subjective evaluation of medical treatment outcome in trauma surgery patients. Patient Educ Couns. 2014;95(1):53-60.

**32.** Engelen E-M, Röttger-Rössler B. Current Disciplinary and Interdisciplinary Debates on Empathy. Emot Rev. 2012;4(1):3-8.

**33.** Jazaieri H, McGonigal K, Jinpa T, Doty JR, Gross JJ, Goldin PR. A randomized controlled trial of compassion cultivation training: Effects on mindfulness, affect, and emotion regulation. Motiv Emot. 2013;38(1):23-35.

**34.** Hojat M, Vergare MJ, Maxwell K, et al. The devil is in the third year: a longitudinal study of erosion of empathy in medical school. Acad Med. 2009;84(9):1182-1191.

**35.** Benson S, Sammour T, Neuhaus SJ, Findlay B, Hill AG. Burnout in Australasian Younger Fellows. ANZ J Surg. 2009;79(9):590-597.

**36.** Lee FJ, Stewart M, Brown JB. Stress, burnout, and strategies for reducing them: what's the situation among Canadian family physicians? Can Fam Physician. 2008;54(2):234-235.

**37.** Markwell AL, Wainer Z. The health and wellbeing of junior doctors: insights from a national survey. Med J Aust. 2009;191(8):441-444.

**38.** Shanafelt TD, Bradley KA, Wipf JE, Back AL. Burnout and self-reported patient care in an internal medicine residency program. Ann Intern Med. 2002;136(5):358-367.

**39.** Fernando AT, 3rd, Consedine NS. Beyond compassion fatigue: the transactional model of physician compassion. J Pain Symptom Manage. 2014;48(2):289-298.

**40.** Greenhalgh T, Chowdhury M, Wood GW. Story-based scales: development and validation of questionnaires to measure subjective health status and cultural adherence in British Bangladeshis with diabetes. Psychol Health Med. 2006;11(4):432-448.

**41.** Erisman SM, Roemer L. A preliminary investigation of the effects of experimentally induced mindfulness on emotional responding to film clips. Emotion. 2010;10(1):72-82.

**42.** Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain. 1983;17(1):45-56.

**43.** Bakker AB, Schaufeli WB, Sixma HJ, Bosveld W, Van Dierendonck D. Patient demands, lack of reciprocity, and burnout: a five-year longitudinal study among general practitioners. 2000;21(4):425-441.

**44.** LoSasso AA, Lamberton CE, Sammon M, et al. Enhancing Student Empathetic Engagement, History-Taking, and Communication Skills During Electronic Medical Record Use in Patient Care. Acad Med. 2017;92(7):1022-1027.

**45.** Berg K, Majdan JF, Berg D, Veloski J, Hojat M. Medical students' self-reported empathy and simulated patients' assessments of student empathy: an analysis by gender and ethnicity. Acad Med. 2011;86(8):984-988.

**46.** Hojat M. Empathy in Patient Care: Antecedents, Development, Measurement, and Outcomes. New York, NY: Springer; 2007.

**47.** Hojat M, Mangione S, Nasca TJ, et al. The Jefferson Scale of Physician Empathy: Development and Preliminary Psychometric Data. Educ Psychol Meas. 2001;61(2):349-365.

**48.** Hojat M, Gonnella JS, Mangione S, et al. Empathy in medical students as related to academic performance, clinical competence and gender. Med Educ. 2002;36(6):522-527.

**49.** Hojat M, Mangione S, Nasca TJ, Gonnella JS, Magee M. Empathy scores in medical school and ratings of empathic behavior in residency training 3 years later. J Soc Psychol. 2005;145(6):663-672.

**50.** Kane GC, Gotto JL, Mangione S, West S, Hojat M. Jefferson Scale of Patient's Perceptions of Physician Empathy: preliminary psychometric data. Croat Med J. 2007;48(1):81-86.

**51.** Glaser KM, Markham FW, Adler HM, McManus PR, Hojat M. Relationships between scores on the Jefferson Scale of physician empathy, patient perceptions of physician empathy, and humanistic approaches to patient care: a validity study. Med Sci Monit. 2007;13(7):CR291-294.

**52.** Singh SP, Modi CM, Patel CP, Pathak AG. Low-fidelity simulation to enhance understanding of infection control among undergraduate medical students. Natl Med J India. 2017;30(4):215-218.

**53.** Spreng RN, McKinnon MC, Mar RA, Levine B. The Toronto Empathy Questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. J Pers Assess. 2009;91(1):62-71.

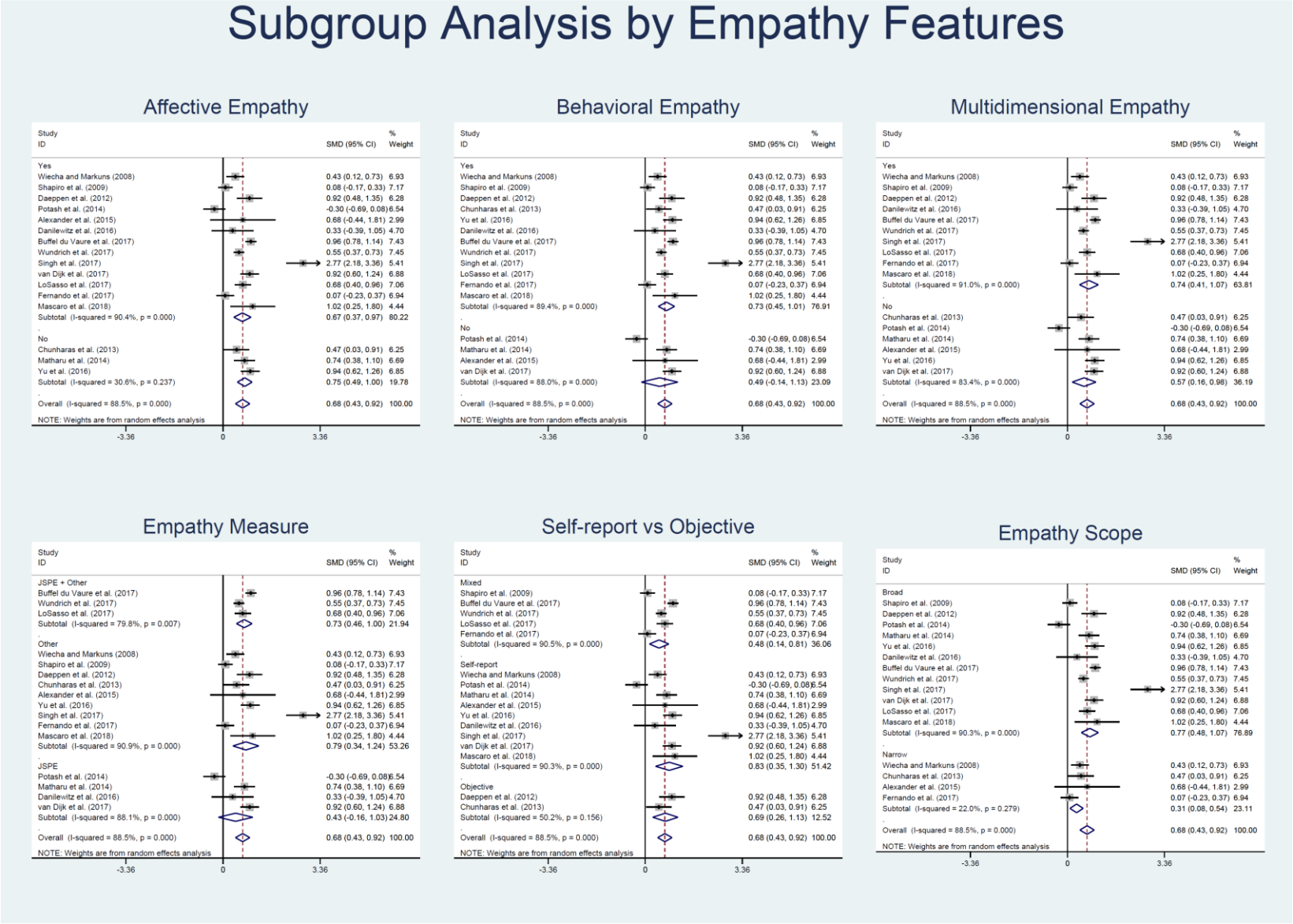
**54.** van Dijk I, Lucassen P, Akkermans RP, van Engelen BGM, van Weel C, Speckens AEM. Effects of Mindfulness-Based Stress Reduction on the Mental Health of Clinical Clerkship Students: A Cluster-Randomized Controlled Trial. Acad Med. 2017;92(7):1012-1021.

**55.** Hojat M, Gonnella JS, Nasca TJ, Mangione S, Veloksi JJ, Magee M. The Jefferson Scale of Physician Empathy: further psychometric data and differences by gender and specialty at item level. Acad Med. 2002;77(10 Suppl):S58-60.

**56.** Wundrich M, Schwartz C, Feige B, Lemper D, Nissen C, Voderholzer U. Empathy training in medical students - a randomized controlled trial. Med Teach. 2017;39(10):1096-1098.
**57.** Morse JM, Anderson G, Bottorff JL, et al. Exploring empathy: a conceptual fit for nursing practice? Image J Nurs Sch. 1992;24(4):273-280.
**58.** Barrett-Lennard GT. The phases and focus of empathy. Br J Med Psychol. 1993;66 ( Pt 1):3-14.
**59.** 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 Theory Pract. 2013;18(4):573-587.
**60.** Danilewitz M, Bradwejn J, Koszycki D. A pilot feasibility study of a peer-led mindfulness program for medical students. Can Med Educ J. 2016;7(1):e31-37.
**61.** Yu EH, Choi EJ, Lee SY, Im SJ, Yune SJ, Baek SY. Effects of micro- and subtle-expression reading skill training in medical students: A randomized trial. Patient Educ Couns. 2016;99(10):1670-1675.
**62.** Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians' empathy and clinical outcomes for diabetic patients. Acad Med. 2011;86(3):359-364.
**63.** Jangland E, Gunningberg L, Carlsson M. Patients' and relatives' complaints about encounters and communication in health care: evidence for quality improvement. Patient Educ Couns. 2009;75(2):199-204.
**64.** Krasner MS, Epstein RM, Beckman H, et al. Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians. JAMA. 2009;302(12):1284-1293.
**65.** An JH, Kwon I, Lee SN, Han JJ, Jeong JE. Study on the Medical Humanities and Social Sciences Curriculum in Korean Medical School: Current Teaching Status and Learning Subjects. Korean J Med Educ. 2008;20(2):133-144.
**66.** Endres J, Laidlaw A. Micro-expression recognition training in medical students: a pilot study. BMC Med Educ. 2009;9:47.
**67.** Bombeke K, Van Roosbroeck S, De Winter B, et al. Medical students trained in communication skills show a decline in patient-centred attitudes: an observational study comparing two cohorts during clinical clerkships. Patient Educ Couns. 2011;84(3):310-318.
**68.** Ekman P. The argument and evidence about universals in facial expressions of emotion. In: Wagner H, Manstead A. Handbook of Social Psychophysiology. Chichester, England: Wiley; 1989.
**69.** Alexander C, Sheeler RD, Rasmussen NH, Hayden L. Teaching an experiential mind-body method to medical students to increase interpersonal skills: a pilot study. Acad Psychiatry. 2015;39(3):316-319.
**70.** Winseman J, Malik A, Morison J, Balkoski V. Students' views on factors affecting empathy in medical education. Acad Psychiatry. 2009;33(6):484-491.
**71.** Rogers CG. The therapeutic relationship with schizophrenics. Madison: The University of Wisconsin Press; 1967.
**72.** Matharu K, Shapiro JF, Hammer RR, Kravitz RL, Wilson MD, Fitzgerald FT. Reducing obesity prejudice in medical education. Educ Health (Abingdon). 2014;27(3):231-237.
**73.** Potash JS, Chen JY, Lam CL, Chau VT. Art-making in a family medicine clerkship: how does it affect medical student empathy? BMC Med Educ. 2014;14:247.
**74.** Fields SK, Mahan P, Tillman P, Harris J, Maxwell K, Hojat M. Measuring empathy in healthcare profession students using the Jefferson Scale of Physician Empathy: health provider--student version. J Interprof Care. 2011;25(4):287-293.
**75.** Roh MS, Hahm BJ, Lee DH, Suh DH. Evaluation of empathy among Korean medical students: a cross-sectional study using the Korean Version of the Jefferson Scale of Physician Empathy. Teach Learn Med. 2010;22(3):167-171.
**76.** Kataoka HU, Koide N, Ochi K, Hojat M, Gonnella JS. Measurement of empathy among Japanese medical students: psychometrics and score differences by gender and level of medical education. Acad Med. 2009;84(9):1192-1197.
**77.** Hsiao CY, Tsai YF, Kao YC. Psychometric properties of a Chinese version of the Jefferson Scale of Empathy-Health Profession Students. J Psychiatr Ment Health Nurs. 2013;20(10):866-873.
**78.** Chunharas A, Hetrakul P, Boonyobol R, Udomkitti T, Tassanapitikul T, Wattanasirichaigoon D. Medical students themselves as surrogate patients increased satisfaction, confidence, and performance in practicing injection skill. Med Teach. 2013;35(4):308-313.
**79.** Daeppen JB, Fortini C, Bertholet N, et al. Training medical students to conduct motivational interviewing: a randomized controlled trial. Patient Educ Couns. 2012;87(3):313-318.
**80.** Moyers TB, Martin T, Manuel JK, Hendrickson SM, Miller WR. Assessing competence in the use of motivational interviewing. J Subst Abuse Treat. 2005;28(1):19-26.
**81.** Moyers TB, Martin T, Manuel JK, Miller WR, Ernst D. Motivational Interviewing Treatment Integrity 3.0 (MITI 3.0). University of New Mexico, Center on Alcoholism, Substance Abuse, and Addictions. 2007; http://casaa.unm.edu/download/miti3.pdf. Accessed 15 March, 2011.

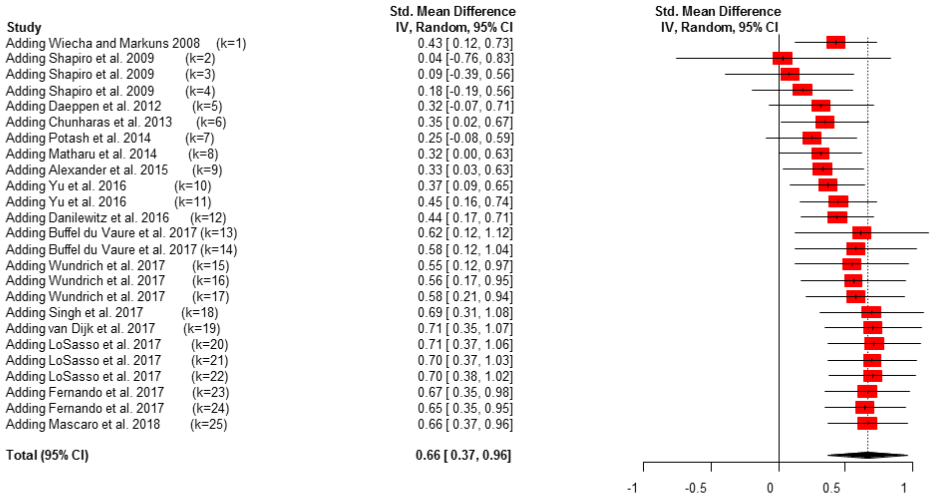
**82.** Shapiro SM, Lancee WJ, Richards-Bentley CM. Evaluation of a communication skills program for first-year medical students at the University of Toronto. BMC Med Educ. 2009;9:11.
**83.** Buhrmester D, Furman W, Wittenberg MT, Reis HT. Five domains of interpersonal competence in peer relationships. J Pers Soc Psychol. 1988;55(6):991-1008.
**84.** Lehmann F, Cote L, Bourque A, Fontaine D. Physician-Patient Interaction: A Reliable and Valid Check-list of Quality. Can Fam Physician. 1990;36:1711-1716.
**85.** Schnabl GK, Hassard TH, Kopelow ML. The assessment of interpersonal skills using standardized patients. Acad Med. 1991;66(9 Suppl):S34-36.
**86.** Gallop R, Lancee WJ, Garfinkel P. How nursing staff respond to the label "borderline personality disorder". Hosp Community Psychiatry. 1989;40(8):815-819.
**87.** Gallop R, Lancee WJ, Garfinkel PE. The empathic process and its mediators. A heuristic model. J Nerv Ment Dis. 1990;178(10):649-654.
**88.** Wiecha JM, Markuns JF. Promoting medical humanism: design and evaluation of an online curriculum. Fam Med. 2008;40(9):617-619.
**89.** The Arnold P. Gold Foundation. . 2008; http://humanism-in-medicine.org. Accessed 15 May, 2008.
**90.** Hojat M, Mangione S, Nasca TJ, et al. An empirical study of decline in empathy in medical school. Med Educ. 2004;38(9):934-941.
**91.** Chen D, Lew R, Hershman W, Orlander J. A cross-sectional measurement of medical student empathy. J Gen Intern Med. 2007;22(10):1434-1438.
**92.** Diseker RA, Michielutte R. An analysis of empathy in medical students before and following clinical experience. J Med Educ. 1981;56(12):1004-1010.
**93.** Hafferty FW. Beyond curriculum reform: confronting medicine's hidden curriculum. Acad Med. 1998;73(4):403-407.
**94.** Horn MO. The other side of the bed rail. Ann Intern Med. 1999;130(11):940-941.

Supplemental Digital Appendix 12. Figure showing subgroup analyses of SMD for empathy by empathy features.

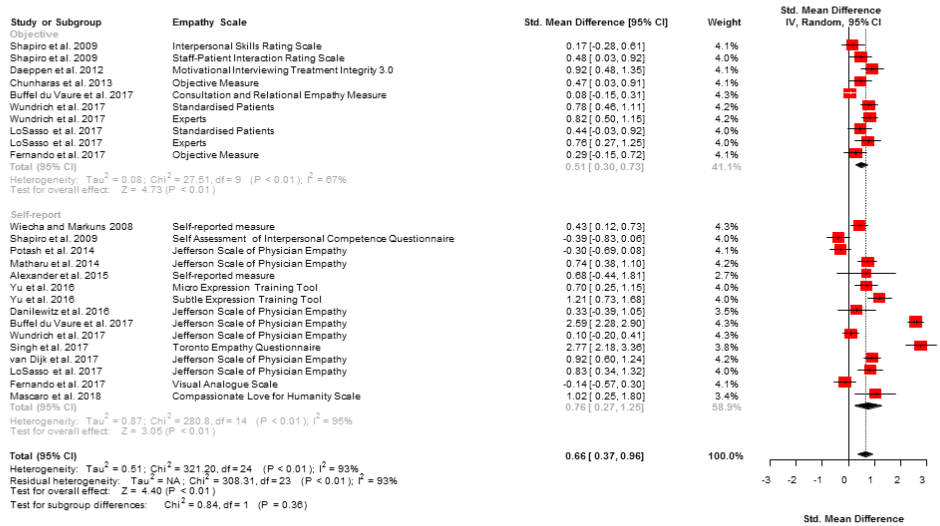


Supplemental Digital Appendix 13. Figure showing A. Cumulative meta-analysis for SMDs for all empathy measures at scale level. B. Subgroup meta-analysis for SMDs for all empathy measures by self-report or objective at scale level. C. Subgroup meta-analysis for SMDs for all empathy measures by scope (broad or narrow).

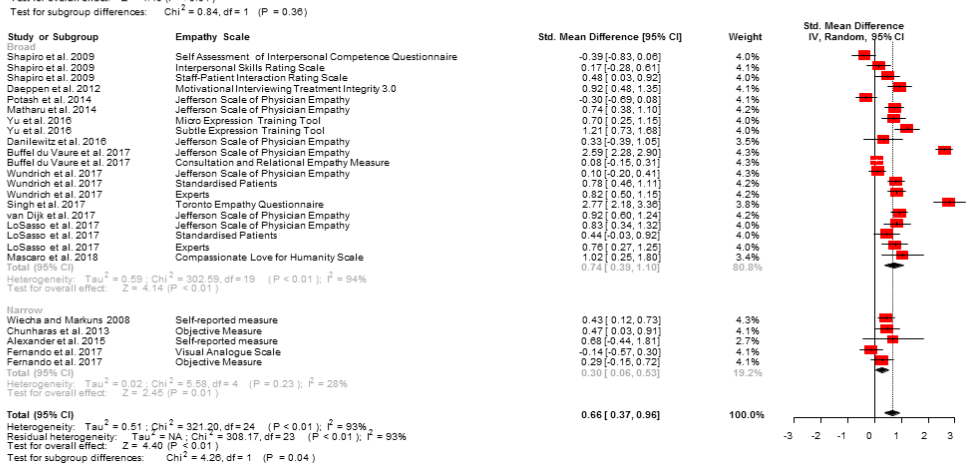
A



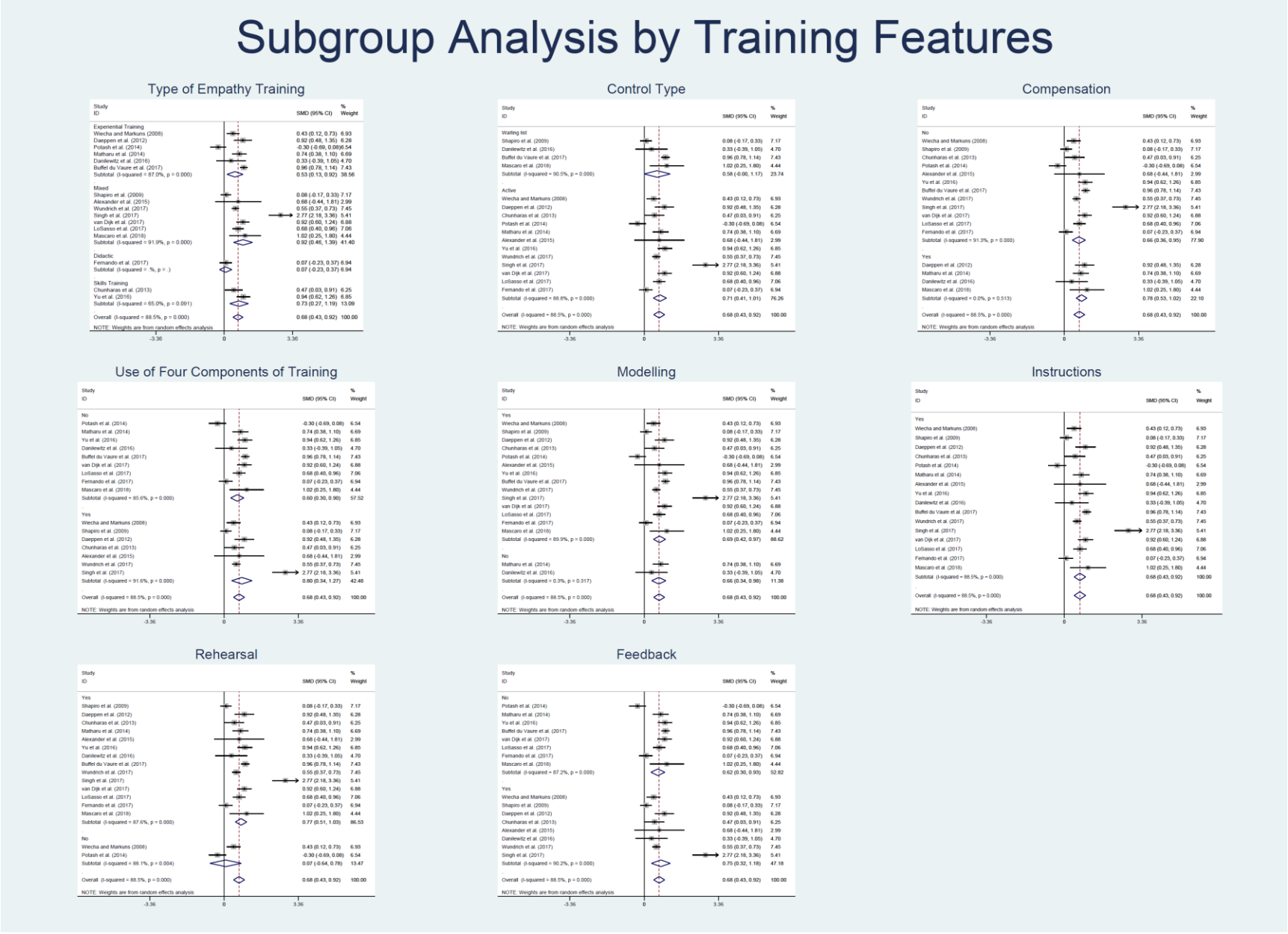

B




C



Supplemental Digital Appendix 14. Figure showing subgroup analyses of SMD for empathy by training features.



Supplemental Digital Appendix 15. Table and Figure showing subgroup analyses by MERSQI items.

Table 1. Subgroup analysis by MERSQI items.

| Variable | k | N | Effect Size | | Heterogeneity | |
|---|---|---|---|---|---|---|
| | | | SMD (95% CI) | *p* | $I^2$ | *p* |
| ***MERSQI Items*** | | | | | | |
| *Study Design* | | | | | | |
| Single group cross-sectional or single group post-test only | -- | -- | -- | -- | -- | -- |
| Single group pre-test and post-test | -- | -- | -- | -- | -- | -- |
| Nonrandomized, 2 group | -- | -- | -- | -- | -- | -- |
| Randomized controlled trial | 16 | 1,736 | 0.68 (0.43, 0.93) | <0.01 | 88.5% | <0.01 |
| *No of institutions studied [Q(1) = 2.19, p = 0.14]* | | | | | | |
| One | 14 | 1,308 | 0.65 (0.37, 0.93) | <0.01 | 88.5% | <0.01 |
| More than one | 2 | 428 | 0.91 (0.72, 1.09) | <0.01 | 12.2% | 0.29 |
| *Response rate, % [Q(1) = 0.23, p = 0.63]* | | | | | | |
| < 50 or not reported | -- | -- | -- | -- | -- | -- |
| 50-74 | 5 | 426 | 0.57 (-0.001, 1.13) | 0.05 | 86.0% | <0.01 |
| ≥75 | 11 | 1,310 | 0.72 (0.43, 1.01) | <0.01 | 90.1% | <0.01 |
| *Type of Data [Q(1) = 1.12, p = 0.29]* | | | | | | |
| Assessment by study participant | 9 | 867 | 0.83 (0.35, 1.30) | <0.01 | 90.3% | <0.01 |
| Objective measurement | 7 | 869 | 0.53 (0.25, 0.81) | <0.01 | 86.6% | <0.01 |
| *Internal structure [Q(1) = 0.24, p=0.62]* | | | | | | |
| Not reported | 11 | 989 | 0.65 (0.25, 1.05) | <0.01 | 90.7% | <0.01 |
| Reported | 5 | 747 | 0.76 (0.57, 0.95) | <0.01 | 61.9% | 0.03 |
| *Content* | | | | | | |
| Not reported | -- | -- | -- | -- | -- | -- |
| Reported | 16 | 1,736 | 0.68 (0.43, 0.93) | <0.01 | 88.5% | <0.01 |
| *Relationships to other variables [Q(1) = 1.79, p = 0.18]* | | | | | | |
| Not reported | 4 | 564 | 1.12 (0.31, 1.92) | <0.01 | 94.2% | <0.01 |
| Reported | 12 | 1,172 | 0.54 (0.29, 0.79) | <0.01 | 84.7% | <0.01 |
| *Appropriateness of analysis [Q(1) = 1.69, p = 0.19]* | | | | | | |
| Data analysis inappropriate for study design or type of data | 1 | 215 | 0.43 (0.12, 0.73) | <0.01 | -- | -- |
| Data analysis appropriate for study design and type of data | 15 | 1,521 | 0.70 (0.43, 0.96) | <0.01 | 89.1% | <0.01 |
| *Complexity of analysis [Q(1) = 1.69, p = 0.19]* | | | | | | |
| Descriptive analysis only | 1 | 215 | 0.43 (0.12, 0.73) | <0.01 | -- | -- |
| Beyond descriptive analysis | 15 | 1,521 | 0.70 (0.43, 0.96) | <0.01 | 89.1% | <0.01 |

| Variable | k | N | Effect Size | | Heterogeneity | |
|---|---|---|---|---|---|---|
| | | | **SMD (95% CI)** | ***p*** | ***I²*** | ***p*** |
| *Outcomes [Q(2) = 2.24, p = 0.33]* | | | | | | |
| Satisfaction, attitudes, perceptions, opinions, general facts | 2 | 228 | 0.44 (0.43, 0.93) | <0.01 | 0.0% | 0.67 |
| Knowledge, Skills | 7 | 673 | 0.56 (0.24, 0.88) | <0.01 | 80.5% | <0.01 |
| Behaviours | 7 | 835 | 0.86 (0.41, 1.31) | <0.01 | 93.8% | <0.01 |
| Patient/Healthcare outcome | -- | -- | -- | -- | -- | -- |

k: number of studies; N: number of participants; SMD: standardized mean difference; P: p-value; $I^2$: heterogeneity statistics

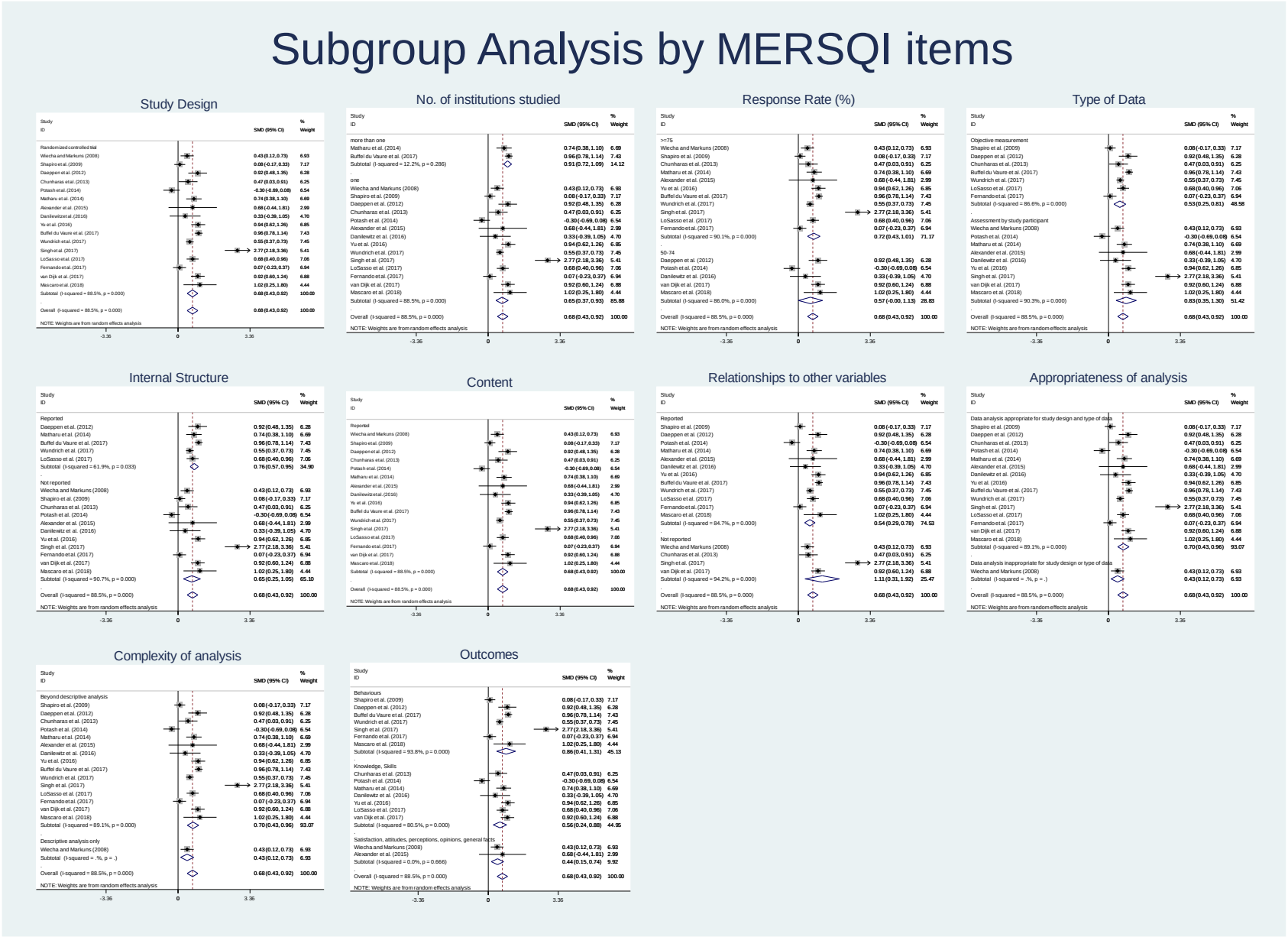



Figure 1. Subgroup analysis of SMD for empathy by MERSQI items.