BMJ Open

Assessing the effect of empathy-enhancing interventions in health education and training: A systematic review of randomised controlled trials

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-036471
Article Type:	Original research
Date Submitted by the Author:	17-Dec-2019
Complete List of Authors:	Winter, Rachel; University of Leicester, Health Sciences Issa, Eyad; University of Leicester, Health Sciences Roberts, Nia; University of Oxford, UK, Bodleian Health Care Libraries, Norman, Robert; University of Leicester, Health Sciences Howick, Jeremy; University of Oxford, Faculty of Philosophy
Keywords:	MEDICAL EDUCATION & TRAINING, EDUCATION & TRAINING (see Medical Education & Training), STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Assessing the effect of empathy-enhancing interventions in health education and training: A systematic review of randomised controlled trials

Rachel Winter (0000-0002-8775-2461), Eyad Isa, Nia Roberts, Robert I Norman, Jeremy Howick

Rachel Winter

Academic Clinical Lecturer

College of Life Sciences, University of Leicester, George Davies Centre, Leicester, England LE1 7RH

Eyad Isa

Academic Clinical Lecturer

College of Life Sciences, University of Leicester, George Davies Centre, Leicester, England LE1 7RH

Robert I Norman

Director of Learning and Teaching

College of Life Sciences, University of Leicester, George Davies Centre, Leicester, England LE1 7RH

Dr Nia Roberts

Librarian

Bodleian Health Care Libraries, University of Oxford, Oxford, England

Dr Jeremy Howick

Director, Oxford Empathy Programme

Faculty of Philosophy, University of Oxford, Oxford, England

Correspondence to: Rachel Winter rw205@le.ac.uk, College of Life Sciences, George Davies Centre, University Road, Leicester, England LE1 7RH

Word Count: 3786

ABSTRACT

Objective: To estimate the effect of empathy interventions in health education and training from randomised controlled trials (RCTs).

Methods: MEDLINE, PsycINFO, EMBASE, CINAHL and Cochrane databases were searched from inception to June 2019 for RCTs investigating the effect of empathy-enhancing interventions in medical and healthcare students and professionals. Studies measuring any aspect of 'clinical empathy' as a primary or secondary outcome were included. Two reviewers extracted data and assessed risk of bias of eligible studies using the Cochrane Risk of Bias Tool. Random effects meta-analyses of the impact of empathy training on participants' empathy levels were performed.

Results: Twenty-six trials were included, with 22 providing adequate data for meta-analysis. An overall moderate effect on participant empathy post-intervention (standardised mean difference 0.52, 95% confidence interval 0.36 to 0.67) was found. Heterogeneity across trial results was substantial (I2=63%). Data on sustainability of effect was provided by 11 trials and found a moderate effect size for improved empathy up until 12 weeks (0.69 95% confidence interval 0.23 to 1.15), and a small but statistically significant effect size for sustainability at 12 weeks and beyond (standardised mean difference 0.34 95% confidence interval 0.11 to 0.57). In total 15 studies were considered to be either unclear or high risk of bias. The quality of evidence of included studies was low.

Conclusions: Findings suggest empathy-enhancing interventions can be effective at cultivating and sustaining empathy with intervention specifics contributing to effectiveness. This review focuses on an important, growing area of medical education, and provides guidance to those looking to develop effective interventions to enhance empathy in the

healthcare setting. Further high quality trials are needed that include patient-led outcome assessments and further evaluate the long-term sustainability of empathy training.

Protocol registration: PROSPERO registration number (CRD42019126843).

Strengths and limitations of this study

- This is an up-to-date review that excludes non-randomised studies, follows a pre-published protocol, and measures the longer term effects of empathy training.
- The quality of the review was limited by the reporting quality of some of the included studies.
- The studies in our review were heterogeneous, which we anticipated.
- We found only four studies that followed-up participants for at least three months,

INTRODUCTION

Rationale

Clinical empathy has multiple benefits for patient care[1-4] and practitioner health.[5, 6] Indeed, person-centred and empathic care are central to all professional healthcare education.[7] Empathy in the clinical setting has been defined in various ways[8] and can be considered as a multidimensional construct incorporating affective, cognitive, behavioural and moral components.[9] A widely accepted definition of clinical empathy involves the ability to understand the patient's situation, perspective and feelings, communicate that understanding to them, and act on it in a helpful and therapeutic way.[10] Although contested by some,[11,12] there is evidence that empathy in medical and health science

students declines during undergraduate education.[13-15] Researchers also agree that empathetic skills can be taught.[16-19] Despite this, no standard empathy-curriculum for healthcare training exists and empathy-based training does not appear routinely in medical or healthcare education.[13]

Three systematic reviews of empathy-promoting interventions have been conducted.[16,19,20] Kelm et al[16] conducted a qualitative synthesis of empathy-cultivating interventions for medical students or physicians. Their findings support the hypothesis that interventions can increase physician and medical student empathy. However, they also identified a lack of rigorous study design in most studies (such as lack of control groups and a failure to use random assignment). More recently, Vassilios et al[19] published a systematic review of randomised control trials (RCTs) of empathy-promoting interventions for health professionals. However, only two out of 17 included reported change in empathy as a primary outcome, focusing instead on general communication skills. Hence, the review did not provide robust evidence of empathy-enhancing interventions. In 2019, Patel et al[20] reviewed educational interventions aimed at enhancing both empathy and/or compassion. They included observational as well as randomised studies and looked only at physicians and physicians-in-training. They were not able to pool their results statistically and did not investigate whether potential benefits of empathy were sustained over time.

The problems listed above present barriers for medical educators looking to implement empathy training into their curriculum. It is not currently known how large the effect size of effective empathy training is; whether the effect is sustained over time; or how best to train

students and continuing learners. It is important to measure the effect of empathy training over time, because such an effect could increase rather than decrease. Arthur et al. [21] found no effect of empathy training immediately after the training, but significant improvement 12 weeks after the end of the training. A delayed improvement in empathy could potentially be accounted for by participants only recognising the benefits of training once they are putting any lessons learnt into action.

In this systematic review and meta-analysis we addressed these gaps, with an up-to-date synthesis of RCTs of interventions aimed at promoting empathy, delivered to both medical and healthcare students and professionals.

Objectives

The overarching objective of this systematic review and meta-analysis is to combine data from all available randomised controlled trials of empathy-enhancing educational interventions in health education and training. This was achieved with two subsidiary objectives:

(1) to assess the effectiveness of interventions aiming to enhance empathy in undergraduate and postgraduate health education and training; and

(2) to assess any lasting effect of empathy training.

We also had three secondary aims:

(1) to identify the intervention type (e.g. communication skills training) that is most effective at enhancing empathy;

(2) to identify the duration of training that is most effective; and

(3) to identify the tools used to measure empathy levels in participants and therefore the effectiveness of the intervention.

METHODS

Protocol and registration

In accordance with the Cochrane Handbook for systematic reviews of interventions,[22] we published a protocol for this systematic review,[23] registered with PROSPERO international prospective register of systematic reviews (registration number CRD42019126843). We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[24]

Eligibility criteria

Randomised controlled trials (RCTs), including cluster RCTs, which investigated the effect of empathy-enhancing interventions on medical and other healthcare students and professionals' empathy levels as a primary or secondary outcome were eligible for inclusion. We included studies with students and trainees at any level and qualified practitioners from any medical profession (including medicine, dentistry, nursing, pharmacy, midwifery and allied healthcare professions). Studies measuring any aspect of 'clinical empathy' were eligible for inclusion. In addition, terminology and measures used in each study were assessed to ensure that outcomes reported under different terms but using the same definitions (for example, reporting on compassion taken to mean empathy) would be captured. Empathy scores reported via self- and/or observer-reported outcome measures post-intervention were included.

Information sources and search strategies

The following databases were searched from inception to 6 June 2019: MEDLINE, PsycINFO, EMBASE, CINAHL and Cochrane. Search strategies are detailed within eTable 1 in the Supplement. Electronic searches were supplemented by hand-searching the references of retrieved papers.

Study selection

All studies retrieved through the search strategy were stored using EndNote with duplicates removed. Two authors (RW and EI) reviewed all titles and abstracts to identify those that met the inclusion criteria. Full text manuscripts were retrieved for potentially relevant articles, the reference lists of which were also hand-searched for further potential studies. If the decision to include or exclude a study was unclear, the study was discussed with a third author (JH) to reach a consensus. Seven papers were discussed with the third author. A PRISMA flow chart was used to record the screening and selection process.

Data collection

One reviewer (RW) extracted, summarised and recorded data to assess quality and to synthesise evidence from included studies. A second author independently extracted a random sample (10%) of studies to ensure agreement on the information extracted and summarised (JH). Data was extracted about: general demographics of the study (first author, date published, country of origin, whether empathy is defined); study design (participants and recruitment, inclusion/exclusion criteria, study duration, control conditions); description of the intervention (setting, duration and frequency); outcome

measures (type of measure, whether measure is validated); results (sample size, completeness of outcome data, data that can be used to calculate an effect size); risk of bias and funding source. If data was not reported, study authors were contacted.

Risk of bias in individual studies

Risk of bias was assessed using the Cochrane Collaboration's Tool for assessing the risk of bias in clinical trials. This recommends the explicit reporting of each individual element of an RCT: random sequence generation and allocation concealment (selection bias); blinding of participants and blinding of outcome assessment (detection bias); incomplete outcome data (attrition bias); and selective reporting (reporting bias). Using the criteria provided by Higgins (2011)[22], each item was scored as high, low or unclear risk of bias, and evidence from the study was used to justify each score given. For cluster RCTs, an additional domain was assessed: selective recruitment of cluster participants. Given that evidence increasingly suggests that sequence generation and allocation concealment are of particular importance in determining the overall risk of bias,[22] a study was classed as being at high risk of bias if it scored as high or unclear risk on either of these domains.

Synthesis of results

We calculated the overall effect size of empathy interventions using the standardised mean difference (SMD) and 95% confidence intervals (CI) based on the data provided in the studies: post-intervention sample size, mean and standard deviation (SD) for experimental versus control group (except where only mean difference and SD between pre- and post-intervention for the experimental and control groups were provided). We used a random

effects model (REM) to allow for likely different (though related) intervention effects. If a study had more than one intervention arm, we used the results for the most comprehensive training intervention. If a study provided measures of empathy using different scales, the primary outcome measure of empathy was used. If it was unclear which was the primary outcome measure, we used the first reported measure of empathy.

Heterogeneity was anticipated between studies and assessed using Cochran's Q Statistic (statistically significant for p0.01) and quantified using the I^2 statistic, with an I^2 value of 50% or more being considered to represent levels of heterogeneity.

Primary analysis included all studies providing the data needed to calculate the mean and SD (or standard error (SE)) of the post-intervention control and intervention groups. Where studies provided more than one point for outcome assessment, the data closest to the end-point of the intervention was used. Studies that provided no numerical data on empathy-related outcomes or data from which it was not possible to calculate mean values and SD were excluded from the meta-analysis.

Additional analyses

We performed a sensitivity analysis excluding studies that were considered to be at high risk of bias (scoring unclear or high risk of bias for either sequence generation or allocation concealment, with evidence suggesting these domains are of particular importance in establishing risk of bias).[22]

We conducted separate meta-analyses to look at: sustainability of the effects of the intervention; the intervention type that is most effective; the duration of intervention that is most effective; the outcome assessment tools (comparing objective and subjective outcome measures); and participant populations (effectiveness of interventions aimed at student populations compared with those aimed at professional populations). To assess for sustainability, studies that provided follow-up measurements of the impact of an empathy intervention were grouped into measurements taken before 12 weeks, and at 12 weeks or later. To evaluate the type of intervention most effective at cultivating empathy, we divided interventions into communication skills-based training interventions, perspective-taking interventions, empathy skills-based training, psychotherapy-focused training, arts and humanities-focused interventions, stress management-focused training, serious gaming interventions, and mixed educational programmes. Interventions were categorised based on the descriptions given of the training programmes in each individual study. Where an intervention could not be put into one or other category, it was allocated to the 'mixed educational programme' category. To assess impact of duration on cultivating empathy, interventions were divided on the basis of the length of time participants spent engaging with the intervention.

Risk of bias across studies

Reporting bias was assessed qualitatively based on inspection of the characteristics of the studies included. A funnel plot was produced to investigate small study effects, which may indicate the presence of publication bias. The GRADE system was used to evaluate the overall quality of evidence for the primary outcome.[25]

RESULTS

Study selection

The literature search resulted in 4,904 citations with duplicates removed. Figure 1 provides an overview of the selection process (see eResults in the Supplement for further details). Seventy-two articles were retrieved for full-text review. Forty-six studies were excluded (eTable 2 in the Supplement). Twenty-six trials were included.[25-50] (n=2,900) Table 1 provides a summary of characteristics (eTable 3 in the Supplement gives further details).

Table 1. Summary of characteristics of included studies

Study	Year	County	No. participants	Participant type	Intervention type	Duration of intervention (hours)	Outcome assessor	Outcome measure	Effect of intervention
Alhassan	2019	Ghana	210	Nursing and midwifery students	Communication skills training	12	Self	JSE	No significant effect found
Arthur	2017	UK	112	Health care assistants	Perspective-taking training	12	Self	JSE	No significant effect found
Blair Irvine	2012	USA	172	Health care professionals	Mixed	4	Self	VST	Significant effect found
Buffel Du Vaure	2017	France	352	Medical students	Balint group	10.5	Self Observer	JSE CARE	Mixed. No significant effect for JSE, significant effect for CARE
Butow	2007	Australia	30	Physicians	Communication skills training	15	Observer	CRP	No significant effect found
Collins	2017	USA	25	Student pharmacists	Literature intervention	2	Self	JSE	No significant effect found
Daniels	1998	Canada	53	Nursing students	Communication skills training	18	Self Self	ECRS CIC	Significant effect found
Foster	2016	USA	70	Medical students	Communication skills training	NE	Observer	ECCS	Significant effect found
Gholamzadeh	2018	Iran	63	Nursing students	Empathy skills training	8	Self	JSE	
Gould	2017	UK	249	Nursing staff and healthcare assistants	Mixed	NE	Self	JSE	No significant effect found

Hastings	2018	UK	236	Qualified care staff	Mixed	3	Self	SECBQ	No significant effect found
Hattink	2015	Netherlands and UK	142	Qualified care staff	Mixed	NE	Self	IRI	Significant effect found
Larti	2014	Iran	82	Nursing students	Communication skills training	12	Self	JSE	Significant effect found
Lobchuck	2018	Canada	44	Nursing staff and students	Perspective-taking training	2.66	Observer Self	CARE CARE (modified)	Mixed. No significant effect found for CARE. Significant effect found on modified CARE
Lor	2014	USA	40	Student pharmacists	Perspective-taking training	18	Self	JSE	Significant effect found
LoSasso	2017	USA	70	Medical students	Communication skills training	1	Self	JSE	No significant effect found
Mueller	2001	USA	37	Physical therapy students	Mixed	11	Self	JSE	Significant effect found
Reiss	2012	USA	99	Physicians	Empathy skills training	4	Observer Self	CARE JSE BEES EFDT	Mixed No significant effect found for CARE, JSE, BEES. Significant effect for EFDT
Shapiro	1998	USA	78	Medical students	Mindfulness training	17.5	Self	ECRS	Significant effect found
Sripada	2010	USA	12	Physicians	Psychotherapy intervention	NE	Observer	BLRI	Significant effect found
Sterkenburg	2018	Netherlands	224	Qualified care staff	Serious game	0.33	Self	SQ	Significant effect found
Tulsky	2011	USA	48	Physicians	Communication skills training	NE	Observer	ES EO PE	Significant effect found
Vaghee	2018	Iran	127	Nursing students	Perspective-taking training	3	Self	JSE	Significant effect found
Wolf	1987	Canada	134	Medical students	Communication skills training	12	Self	HRI MCI	Significant effect found
Wundrich	2017	Germany	158	Medical students	Empathy skills training	6	Self Observer	JSE OSCE	Mixed. No significant effect found for JSE. Significant effect found on OSCE scores
Yang	2018	China	177	Nursing students	Narrative medicine intervention	42	Self	JSE	Significant effect found

Study characteristics

Study publication dates ranged from 1987 to 2019, with 15 out of 26 trials published in the last five years.[21,26,28,30,32-36,38,40,45,47,49,50] Thirteen were carried out in the USA and Canada,[27,30-32,38-41,43,44,46,48] seven in Europe,[21,28,34-36,45,49] three in Iran,[33,37,47] and one each in Australia,[29] Ghana[26] and China.[50] Fourteen studies provided a definition of empathy.[28,30,32-35,38,41-45,49,50]

Study design

Sample size ranged from 12 to 352 participants (median of 90.5; interquartile range (IQR) 49.25-154). Twelve studies had 100 or more participants.[21,26-27,34-36,45,47-49] Seven had fewer than 50 participants.[29,30,38,39,41,44,46] Fifteen studies evaluated empathy interventions for student populations,[26,28,30-33,37,39,40,41,43,47-50] including seven which looked at medical students,[28,32,33,40,43,48,49] five with nursing students,[31,37,38,48,50] two with student pharmacists,[30,39] one with physiotherapy students,[41] and one with a mixed nursing and midwifery student population.[26] Ten trials used professional/qualified populations,[21,27,29,34-36,42,44-46] with four of these focusing on physicians,[29,42,44,46] one on nurses,[34] and five with qualified care staff, including healthcare assistants.[21,27,35,36,45] One study used a mixed student and professional population (nursing students and nurse practitioners).[38]

Five trials used multiple sites,[21,28,34,35,38] and five were cluster RCTs.[21,34,35,47,50] Ten studies defined both inclusion and exclusion criteria for the study.[21,26-27,33,35,37,39,47,50] Thirteen defined inclusion criteria only[28-31,34,36,38,40,41,43-

45,48] and in three studies inclusion/exclusion criteria were either not given or were not clear.[32,46,49]

Study interventions

While the aims of eligible trials in this review were to enhance empathy through an educational intervention, a range of intervention types were employed. The most commonly used approach was a communication skills-based training intervention, with eight[26,29,31.32,37,40,46,48] studies using this. Four studies used perspective-taking training,[21,38,39,47] two used interventions with a psychotherapy focus,[28,44] three used empathy skills-based training sessions,[33,42,49] two used an arts and humanities approach,[30,50] one used mindfulness-based training,[43] and one a serious gaming intervention.[45] Five studies could not be classified and were described as 'mixed' interventions, using various elements of theoretical knowledge teaching and experiential learning sessions.[27,34-36,41] Seventeen of the 26 interventions had been specifically designed to foster empathy[21,30,32-35,37-42,44-46,49,50] while the remaining studies used interventions not specifically designed to improve empathy but with the hypothesis that they would. For example, Buffel Du Vaure et al[28] explored the impact of a psychotherapy-focused 'Balint Group' intervention on medical student empathy.

The most frequently used mode of delivery was face-to-face, with eighteen interventions delivered this way.[21,26,28,29,31,33-35,38-40,42,43,44,47-50] Six interventions were delivered online,[27,32,36,37,41,45] one employed a self-directed mode of delivery,[30] and one used a CD-ROM to deliver the intervention.[46]

Studies ranged in duration of intervention (total time spent participating in the intervention) from 20 minutes to 18 hours. The mean duration was 10.2 hours (SD 8.8). Five studies did not explicitly state duration.[32,34,36,44,46] Training packages in six studies were considered to be 'short duration', lasting three hours or less;[30,35,38,40,45,47] ten were considered 'medium duration', lasting between four and 12 hours;[21,26-28,33,37,41,42,48,49] and five were considered 'long duration', lasting more than 12 hours.[29,31,39,43,50]

Timespan of the interventions ranged from one to 120 days, with a mean length of 38.5 days (SD 40.2).

Outcome measures

Studies used either self-report or other (objective)-report measures to assess a change in participants' empathy. Objective measures included those completed by patients or experts (for example faculty staff or trained actors playing simulated patients). The majority of studies (18) used only self-report measures.[21,26,27,30,31,33-37,39-41,43,45,46-48,50] Four studies used objective measures[29,32,44,46] (of which only Tulsky et al[46] used patients rather than simulated patients or experts as the outcome assessors). Four studies used a combination of self- and objective-report tools to measure empathy.[28,38,42,49]

The Jefferson Scale of Empathy (JSE)[51] or a version of it was the most frequently used self-reported outcome measurement tool, with 13 studies employing it.[21,26,27,30,33,34,37,39-41,47,49,50] Other self-report tools used included the Balanced Emotional Empathy Scale (BEES),[52] the Ekman Facial Decoding test,[53] and the Toronto

Empathy Questionnaire (TEQ).[54] The Consultation and Relational Empathy Scale (CARE)[55] was the most frequently used objective measure of empathy, with three studies employing it.[28,38,42] Other objective outcome measures of empathy included the Carkhuff Empathy Rating Scale.[56] In addition, some studies developed their own measures of empathy, for example Tulksy et al[46] used a Likert scale with ten items to assess perceived oncologist empathy. Butow et al[29] created a manual to code transcripts of videoed patient interactions to assess empathic behaviour, in addition to using the CARE scale.[55] All studies except three[27,29,46] employed a validated tool to measure empathy.

Outcome assessment strategy

Timeframes for measuring outcomes varied across studies. Fifteen studies did not specify a timeframe for post-intervention measurements or were unclear.[31-33,35,36-38,40,41,43-46,47-50] For example, Hastings et al[35] reported measuring empathy six-weeks post-randomisation but were not clear how long after the intervention had ended that this measurement was taken. For studies that were explicit, post-intervention measures varied between two days and six months, with the majority of measures taken within two weeks of the intervention.[21,26,28,39,46,27,30] Eleven studies measured the effects of the intervention at one or more follow-up points (in addition to the post-intervention measurement),[21,26,27,29,31,33,35,37,39,47,50] which varied between four weeks and 18 months.

Risk of bias within studies

In total, 11 studies[21,26,29,34-37,41,42,45,46] were considered to be at low risk of bias overall (with a low risk of bias for sequence generation and allocation concealment).[22] Thirteen studies were considered to be low risk for random sequence generation[21,26,29,33-38,41,45,46] and 11 were low risk for allocation concealment.[21,26,29,34-37,41,42,45,46] Blinding was not possible in the majority of studies due to the nature of the interventions and the outcome assessment (self-report) strategy frequently employed. Full details of the risk of bias assessment are reported in the eResults of the Supplement and eFigure 1 illustrates the overall findings.

Results of individual studies

The majority of studies (19/26) found that the tested intervention significantly improved empathy on at least one outcome measure.[27,28,31-33,36-39,41-50] Seven studies did not find any significant increase in empathy.[21,26,29,30,34,35,40] Of the studies that reported a significant improvement in empathy on at least one outcome measure, 11 were aimed at student populations (representing approximately 73% of student population studies)[28,31-33,39,41,43,47-50] and seven were aimed at professionals (representing 70% of professional population studies).[27,36,37,44,45,46,42] Fifteen studies reported a significant improvement in empathy using a self-rated outcome measure (this represents 68% of studies (15/22) using a self-report outcome tool).[27,28,31,33,36-39,41,43-45,47,48,50] Four studies reported an increase in empathy when using an objective measure (representing 50% (4/8) of studies using an objective outcome measure).[32,42,46,49] Seventeen studies employed an educational intervention that had been specifically designed to foster empathy.[21,30,32-35,37-42,44-46,49,50] Of these, 12 (70%) were

successful.[32,33,37-42,44-46,49,50] Four out of five studies that were classed as 'long duration' (lasting >12 hours) reported a significant improvement in empathy post intervention;[31,39,43,50] 50% of 'medium duration' studies (between 3 and 12 hours) reported a significant increase in empathy;[27,33,37,48,49] and 33% of 'short duration' studies (<3 hours) reported a significant improvement.[45,47]

Synthesis of results

Of the 26 studies included in this review, four were excluded from meta-analysis as they did not provide adequate data from which to calculate the SMD and SD.[29,34,44,49] For the studies that were excluded from the primary analysis, Butow et al[29] reported a positive but not statistically significant effect and Gould et al[34] found no significant difference between control and intervention groups. Wundrich[49] reported no significant influence of the intervention as measured by the JSE (student version) but did report a positive and statistically significant effect on the observer-assessed outcome. Sripada et al[44] also reported a statistically significant positive effect. Of the 22 studies that had adequate data for pooling, all but one (Arthur et al[21]) showed a benefit of intervention. The primary analysis identified that the overall effect of empathy interventions in terms of improving participant empathy was statistically significant (SMD 0.52, 95% CI 0.36 to 0.67) (figure 2). The Q value indicated significant heterogeneity, with p equal to 0.0001 and I^2 equal to 63%. A summary of findings is presented in table 2.

Table 2. Summary of effect sizes for studies included in meta-analyses

	Standardised mean difference (95% confidence interval)	Heterogeneity (I^2)	References
Overall effect of empathy interventions	0.52 (0.36-0.67)	63%	21,26-28,30-33,35-43,45-48,50
Effect of intervention with least risk of bias	0.44 (0.19-0.69)	66%	21,26,35-37,41,42,45,46
Sustainability of effect			
- Follow-up measurement before 12 weeks	0.69 (0.23-1.15)	84%	26,27,33,35,37,47
- Follow-up measurement at 12 weeks or later	0.34 (0.11-0.57)	0%	21,35,39,50
Effect by type of intervention			
- Communication skills training	0.69 (0.32-1.06)	78%	26,31,32,37,40,46,48
- Perspective-taking training	0.60 (0.17-1.04)	55%	21,38,39,47
- Mixed educational programmes	0.39 (0.18-0.61)	0%	27,35,36,41
- Empathy skills training	0.60 (-0.02-1.21)	71%	33,42
- Arts/humanities interventions	0.38 (0.03-0.73)	0%	30,50
Effect by duration of intervention			
- Short (3 hours or less)	0.44 (0.25-0.63)	23%	30,35,38,40,45,47
- Medium (4 to 12 hours)	0.46 (0.15-0.77)	82%	21,26,27,28,33,37,41,42,48
- Long (more than 12 hours)	0.57 (0.32-0.82)	0%	31,39,43,50
Effect by participant population			
- Student population	0.62 (0.38-0.85)	74%	26,28,30-33,37-41,43,47,48,50
- Professional/qualified population	0.33 (0.18-0.47)	0%	21,27,35,36,42,45,46
Effect by outcome assessor			
- Self-assessment	0.52 (0.37-0.68)	58%	21,26-28,30,31,33,35-43,45,47,48.50
- Observer-assessment	0.28 (-0.18-0.75)	81%	28,32,38,42,46

Additional analyses

Sensitivity analysis

For the sensitivity analysis of the least biased studies (table 2), 11 were judged to have low risk of bias for random allocation or allocation concealment[21,26,29,34-37,41,42,45,46] and nine of these provided sufficient data to be included in a meta-analysis (figure 3).[21,26,35-37,41,42,45,46]

Sustainability of improved empathy analysis

Eleven studies provided follow-up data assessing sustainability of changes to empathy, in addition to post-intervention measurement.[21,26,27,29,31,33,35,37,39,47,50] Eight were eligible for inclusion in a sub-group analysis (see eResults in the Supplement for further details).[21,27,33,35,37,39,47,50] Meta-analysis found a moderate effect size for improved

empathy until 12 weeks and a small but statistically significant effect size for sustainability at 12 weeks and later (figure 4 and table 2).

Type of intervention analysis

A meta-analysis comparing sub-groups of different types of intervention (eFigure 2 in the Supplement and eResults for further details) found the greatest effect was with empathy training that was communication skills-based (table 2). The smallest effect reported was for interventions that were described as 'mixed educational programmes' and ones based in the arts and humanities (table 2).

Duration of intervention analysis

Interventions of medium and longer duration (eFigure 3 in the Supplement) were most effective. Interventions of short duration had the smallest effect size (table 2).

Participant population analysis

Studies using healthcare student participant populations appeared to have a larger effect size than those directed at professional/qualified participant populations (eFigure 4 in the Supplement). Studies included in a sub-analysis of interventions for students showed a moderate effect size of training, compared to a smaller but still significant effect size for training directed at professional/qualified populations (table 2).

Outcome assessor analysis

Studies using a self-assessment outcome scale showed a moderate and significant benefit to empathy for the intervention tested (eFigure 5), compared to a small and statistically not

significant effect size for the observer-assessed outcome studies (table 2).

Risk of bias across studies

A funnel plot of trials used in the primary meta-analysis (22 studies) did not reveal evidence of publication bias (figure 5). An evaluation of evidence using GRADE software found that the quality of evidence was low (eTable 4). This was due to a high or uncertain risk of bias based on random sequence generation and/or allocation concealment in a number of studies and a high degree of heterogeneity across studies.

DISCUSSION

Summary of evidence

We found that empathy interventions delivered to healthcare students and practitioners consistently improved empathy levels in participants by a moderate amount, and that this effect was sustained over time. The quality of the evidence was low due to lack of blinding and allocation concealment.

Comparison with other evidence

Other systematic reviews have found benefits of empathy training[16,19,20] and that practitioner empathy training makes a difference to patients.[57] Our study adds to this evidence by providing an estimate of empathy training from higher quality (randomised) trials, and by showing that the effect lasts well beyond the intervention.

Strengths and limitations

This is an up-to-date review that excludes non-randomised studies, follows a pre-published protocol, and measures the longer term effects of empathy training. The quality of the review was limited by the reporting quality of some of the included studies. In studies where it was unclear which was the primary measure of empathy, we choose to use the first reported measure of empathy. This might have been biased, as authors may have chosen to report the most positive outcome first. However, we found that this was not necessarily the

case. For example, the first outcome reported by Buffel du Vaure et al [28] (who did not specify which was primary) had a smaller effect than the second.

The studies in our review were also heterogeneous, which we anticipated. Finally, we only found four studies that followed participants for at least three months. However, the ones that did do so found a positive effect.

Implications for research and practice

An area for future research highlighted through this review is the lack of more objective, patient-evaluated empathy interventions. The results of this review have implications for practice and may be useful to those involved in education and training. With competition for time and space in both undergraduate and postgraduate healthcare curriculums, robust evidence when considering how best to develop effective interventions to improve empathy in students and practitioners and ultimately to improve patient care is vital.

CONCLUSION

Empathy-enhancing interventions for healthcare students and professionals can be effective at cultivating and sustaining empathy. Designers of future trials of empathy training for medical students can use the results of this review as a guide to their intervention development.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

CONTRIBUTORSHIP STATEMENT

RW had a lead role in the planning, conduct and reporting of the work described in this article. EI had an equal role in the conduct of the work described in this article. NR had an equal role in the conduct of the work described. RI had a supporting role in the planning and reporting of the work described in this article. JW had an equal role in the planning, conduct and reporting of the work described in this article. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

COMPETING INTEREST STATEMENT

All authors have completed the Unified Competing Interest form and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

TRANSPARENCY DECLARATION

This manuscript is an honest, accurate and transparent account of the studies being reported. No important aspects of the review have been omitted.

ETHICAL APPROVAL

Ethical approval was not required.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sections.

PATIENT AND PUBLIC INVOLVEMENT

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

DATA AVAILABILITY STATEMENT

All data relevant to the study are included in the article or uploaded as supplementary information.

FIGURES LEGEND

Figure 1. PRISMA flow diagram

Figure 2. Meta-analysis of eligible studies providing adequate data to calculate standardised mean difference with 95% confidence interval

Figure 3. Meta-analysis of eligible studies, excluding those considered to be at high risk of bias

Figure 4. Meta-analysis of studies that provided follow-up observation points to determine long-term effectiveness of intervention

Figure 5. Funnel plot of effect sizes and standard errors.

REFERENCES

1. Kim SS, Kaplowitz S, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. Eval Health Prof. 2004;27:237-251
2. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther. 2001;26(5):331-342
3. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijen J. Influence of context effects on health outcomes: A systematic review. Lancet. 2001;357(9258):757-62
4. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Reiss H. The influences of the patient-clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomised controlled trials. PLoS One. 2014;9(4):e94207
5. Thomas MR, Dyrbye LN, Huntington JL, Lawson KL, Novotny PJ, Sloan JA, Shanafelt TD. How do distress and well-being relate to medical student empathy? A multicentre study. J Gen Intern Med. 2007;22(2):177-183
6. 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

7. The NHS Constitution for England. Health Education England. Department of Health. 2015. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/480482/NHS_Constituti on_WEB.pdf Accessed 5th February 2019
8. Cummins J and Bennett V. Compassion in Practice: Nursing, Midwifery and Care Staff, our Vision and Strategy. Leeds: NHS Commissioning Board, 2012. Available at: https://www.england.nhs.uk/wp-content/uploads/2012/12/compassion-in-practice.pdf. Accessed 5th February 2019
9. Francis R. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Summary. London: The Stationery Office, 2013. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/279124/0947.pdf. Accessed 5th February 2019
10. Mercer SW, Reynolds W. Empathy and quality of care. Br J Gen Pract. 2002; 52(Suppl):S9-12
11. Colliver JA, Conlee MJ, Verhulst SJ, Dorsey JK. Reports of the decline of empathy during medical education are greatly exaggerated: A re-examination of the research. Acad Med. 2010;85(4):588–593
12. Smith KE, Norman GJ, Decety J. The complexity of empathy during medical school training: Evidence for positive changes. Med Educ, 2017;51(11):1146-1159
13. 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 J Assoc Am Med Coll. 2011;86(8):996-1009

14. 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 J Assoc Am Med Coll. 2009;84(9):1182-1191

15. Nunes P, Williams S, Sa B, Stevenson K. A study of empathy decline in students from five health disciplines during their first year of training. Int J Med Educ. 2011:2;12-17

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

17. Chen D, Lew R, Hershman W, Orlander J. A cross-sectional measurement of medical student empathy. J Gen Intern Med. 2007;22(10):1434-8

18. Batt-Rawden SA, Chisholm MD, Anton B, Flickinger TE. Teaching empathy to medical students: An updated, systematic review. Acad Med. 2013;88(8):1171-7

19. Kiosses VN, Karathanos VT, Tatsioni A. Empathy promoting interventions for health professionals: A systematic review of RCTs. Journal of Compassionate Health Care. 2016;3:7

20. Patel S, Pelletier-Bui A, Smith S, Roberts MB, Kilgannon H, Trzeciak S, Roberts BW. Curricula for empathy and compassion training in medical education: A systematic review. PLoS One. 2019;14(8):e0221412

21. Arthur A, Maben J, Wharrad H, Aldus C, Sarre S, Schneider J et al. Can Healthcare Assistant Training (CHAT) improve the relational care of older people? Study protocol for a pilot cluster randomised controlled trial. Trials. 2015;16:559

22. Higgins JJ and Green S. The Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Chichester. The Cochrane Collaboration, 2011

23. Winter R, Roberts N, Howick J. Assessing the effect of empathy-enhancing interventions in health education and training: A systematic review of randomised controlled trials. PROSPERO: International prospective register of systematic reviews. 2019

24. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses:* The PRISMA Statement. BMJ 2009;339:b2535

25. Schunemann H, Brozek J and Oxman A. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendation. Hamilton: Group TGW, 2008

26. Alhassan M. Effect of a 2-day communication skills training on nursing and midwifery students' empathy: a randomised controlled trial. BMJ Open 2019;9(3):e023666

27. Blair Irvine A, Billow MB, Eberhage MG, Seeley JR, McMahon E, Bourgeois M. Mental illness training for licensed staff in long-term care. Issues Ment Health Nurs 2012;33(3):181-194

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

29. Butow P, Cockburn J, Girgis A, Bowman D, Schofield P, D'Este C et al. Increasing oncologists' skills in eliciting and responding to emotional cues: Evaluation of a communication skills training program. Psycho-oncology. 2008;17(3):209-218.

30. Collins KL, Zweber A, Irwin AN et al. Impact of a fictional reading intervention on empathy development in student pharmacists. Curr Pharm Teach Learn 2017;9(3): 498-503

31. Daniels TG, Denny A, Andrews D. Using microcounseling to teach RN nursing students skills of therapeutic communication. Journal of Nursing Education 1988;27(6):246-252

32. Foster A, Chaudhary N, Kim T, Waller JL, Wong J, Borish M et al. Using Virtual Patients to Teach Empathy: A Randomized Controlled Study to Enhance Medical Students' Empathic Communication. Simulation in Healthcare: The Journal of The Society for Medical Simulation 2016;11(3):181-189

33. Gholamzadeh S, Khastavaneh M, Khademian Z, Ghadakpour S. The effects of empathy skills training on nursing students' empathy and attitudes toward elderly people. BMC Med Educ 2018;18(1):198

34. Gould LJ, Griffiths P, Barker HR, Libberton P, Mesa-Eguiagaray I, Pickering R et al. Compassionate care intervention for hospital nursing teams caring for older people: A pilot cluster randomised controlled trial. BMJ Open 2018;8(2):e018563

35. Hastings RP, Gillespie D, Flynn S, McNamara R, Taylor Z, Knight R et al. Who's challenging who training for staff empathy towards adults with challenging behaviour: cluster randomised controlled trial. JIDR 2018;62(9):798-813

36. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J et al. Web-Based STAR E-Learning Course Increases Empathy and Understanding in Dementia Caregivers: Results from a Randomized Controlled Trial in the Netherlands and the United Kingdom. J Med Internet Res 2015;17(10):e241

37. Larti N, Ashouri E, Aarabi A. The effect of an empathy role-play program for operating room nursing students. J Educ Eval Health Prof 2018;15:29

38. Lobchuk M, Hoplock L, Halas G, West C, Dika C, Schroeder W et al. Heart health whispering: A randomized, controlled pilot study to promote nursing student perspective-taking on carers' health risk behaviors. BMC Nurs 2018;17:21

39. Lor KB, Truong JTT, Ip EJ, Barnett MJ. A randomized prospective study on outcomes of an empathy intervention among second-year student pharmacists. Am J Pharm Educ 2015;79(2):18s

40. LoSasso AA, Lamberton CE, Sammon M, Berg K, Caruso JW, Cass J 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. Mueller K, Prins R, de Heer H. An Online Intervention Increases Empathy, Resilience, and Work Engagement Among Physical Therapy Students. Journal of Allied Health 2018; 47(3):196-203

42. Reiss H, Kelley JM, Bailey RW, Dunn EJ, Phillips M. Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum. J Gen Intern Med 2012;27(10):1280-1286

43. Shapiro SL, Schwartz GE, Bonner G. Effects of mindfulness-based stress reduction on medical and premedical students. Journal of Behavioral Medicine 1998;21(6):581-599

44. Sripada BN, Henry DB, Jobe TH, Winer JA, Schoeny ME, Gibbons RD. A randomized controlled trial of a feedback method for improving empathic accuracy in psychotherapy. Psychol Psychother 2011;84(2):113-127

45. Sterkenburg PS, Vacaru VS. The effectiveness of a serious game to enhance empathy for care workers for people with disabilities: A parallel randomized controlled trial. Disability and Health J 2018;11(4):576-582

46. Tulsky JA, Arnold RM, Alexander SC, Olsen MK, Jeffreys AS, Rodriguez KL. Enhancing communication between oncologists and patients with a computer-based training program: A randomized trial. Annals of Internal Medicine 2011;155(9):593-601

47. Vaghee S, Loffabadi MK, Salarhaji A, Vaghei N, Hashemi BM. Comparing the Effects of Contact-Based Education and Acceptance and Commitment-Based Training on Empathy toward Mental Illnesses among Nursing Students. Iranian Journal of Psychiatry 2018;13(2):119-127

48. Wolf FM, Woolliscroft JO, Calhourn JG, Boxer GJ. A controlled experiment in teaching students to respond to patients' emotional concerns. J Med Educ 1987;62(1):25-3

49. 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

50. Yang N, Xiao H, Cao Y, Li S, Yan H, Wang Y. Does narrative medicine education improve nursing students' empathic abilities and academic achievement? A randomised controlled trial. J Int Med Res 2018;46(8):3306-3317

51. Hojat M, Gonnella JS, Nasca TJ, *et al*. Physician empathy: Definition, components, measurement, and relationship to gender and specialty. Am J Psychiatry 2002;159:1563–9

52. Mehrabian A, Epstein N. A measure of emotional empathy. J Pers. 1972;40:525–43.

53. Ekman P. Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. New York: Holt; 2004

54. 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;91:62–71

55. Mercer SW, Maxwell M, Heaney D, Watt GCM. 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

56. Carkhuff RR. Helping and human relations: A primer for lay and professional helpers. Montreal-Toronto: Holt, Rinehart and Winston Inc

57. Howick J, Moscrop A, Mebius A, Fanshawe T, Lewith G, Bishop FL et al. Effects of empathic and positive communication in healthcare consultations: a systematic review and meta-analysis. J R Soc Med. 2018;111(70:240-252

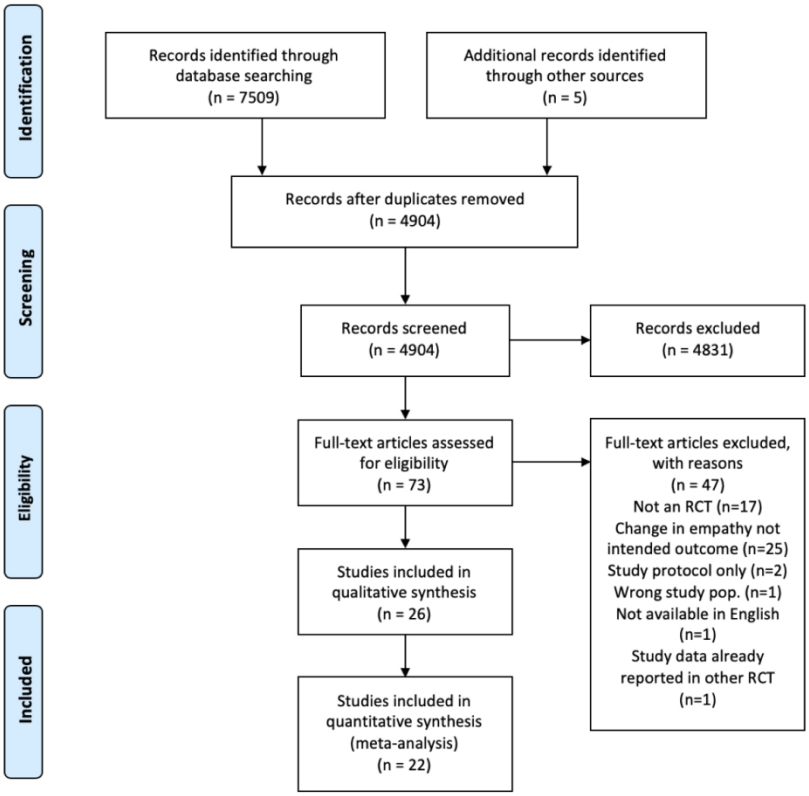


Figure 1. PRISMA flow diagram

209x297mm (150 x 150 DPI)

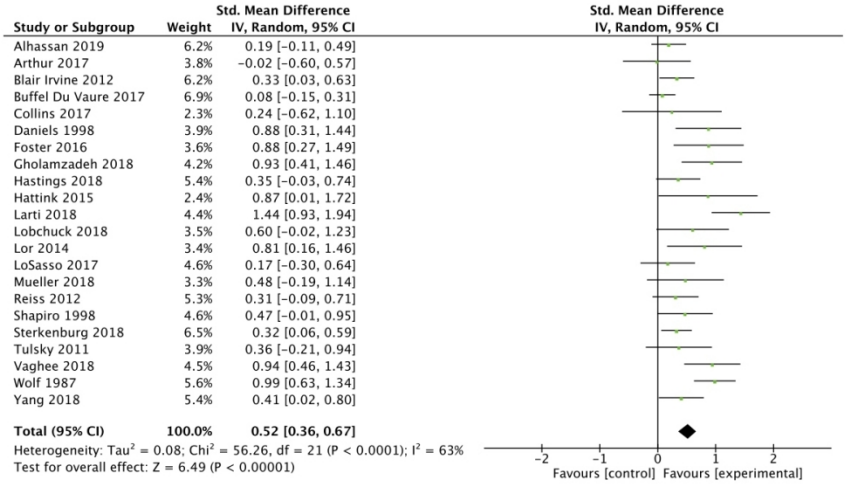


Figure 2. Meta-analysis of eligible studies providing adequate data to calculate standardised mean difference with 95% confidence interval

215x279mm (150 x 150 DPI)

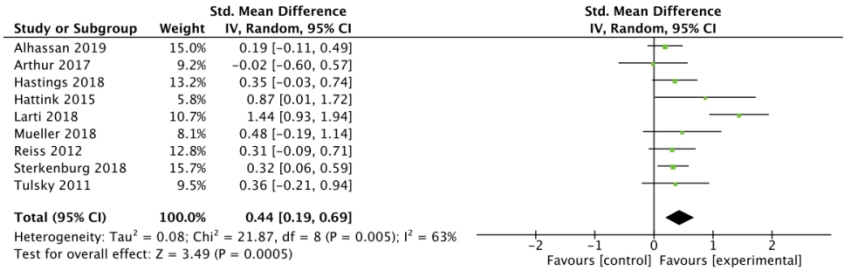


Figure 3. Meta-analysis of eligible studies, excluding those considered to be at high risk of bias

215x279mm (150 x 150 DPI)

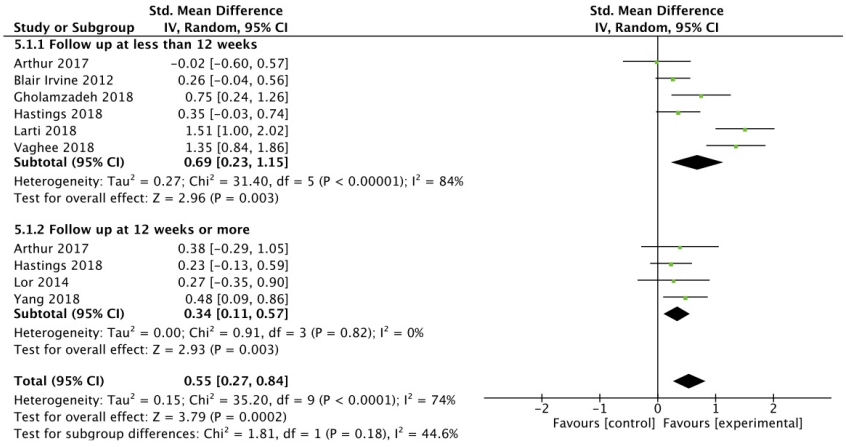


Figure 4. Meta-analysis of studies that provided follow-up observation points to determine long-term effectiveness of intervention

215x279mm (150 x 150 DPI)

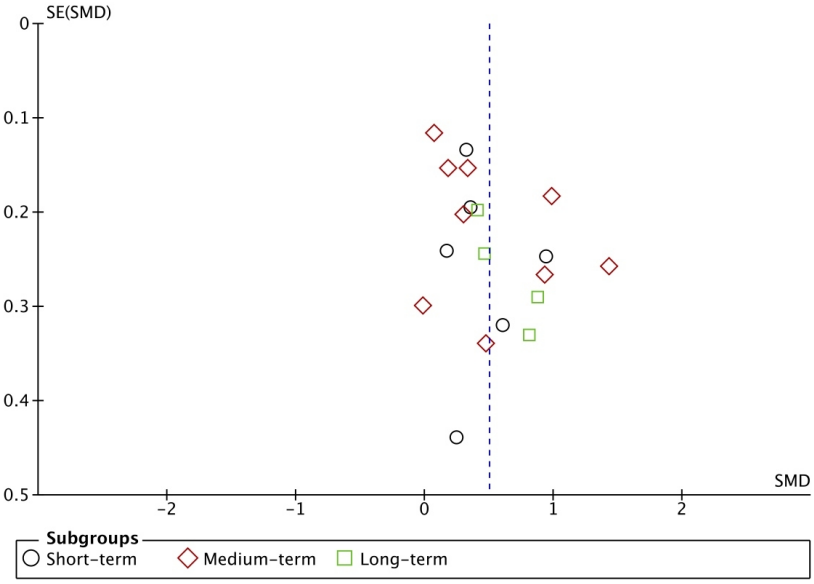


Figure 5. Funnel plot of effect sizes and standard errors

215x279mm (150 x 150 DPI)

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

SUPPLEMENT

Additional results (eResults)

Study selection

The literature search resulted in 7,509 citations. EMBASE included 2,754, PsychINFO 1767, CINAHL 381, MEDLINE 2441 and Cochrane 346. An additional five records were identified through other sources. After duplications were removed 4904 citations remained. 4831 citations were excluded after screening abstracts. Seventy-two articles were retrieved for full-text review. Forty-six studies were excluded (eTable 2). The total number of eligible papers included in this review was 26[25-50] (n=2,900). See eTable3 for descriptive characteristics.

Risk of bias within studies

Allocation

Thirteen studies were considered to be low risk for random sequence generation,[25,26,29,33-38,41,45,46] of which seven employed some form of computer randomisation,[26,34,35,36,38,42,45] one used the minimisation method,[46] one used a random numbers table[29] and three used a low-tech method[25,37,41] (for example a shuffled pack of cards). Thirteen trials were considered to have an unclear risk[27,28,30-32,39,40.43,44,47-50] with 12 of these stating that participants were randomly assigned but not describing the method.[27,30-32,39,40.43,44,47-50] One trial used participants from two different sites, using computer randomisation at one site but not describing the method

of randomisation at the other.[28] The risk of bias for allocation concealment was considered low for 11 studies[25,26,29,34-37,41,42,45,46] and was well described in each of these. Fifteen studies did not describe or clearly describe allocation concealment and so were considered unclear in terms of risk.[27,28,30-33,38,39,40,43,44,47-50]

Blinding

Whilst blinding of participants was not possible in the majority of the trials, due to the nature of the interventions, one study did blind participants.[45] This was achieved by using an online package to deliver either a 'serious game'(experimental) intervention or a 'digital reading' (control) intervention. Participants were unaware of which was the control and which was the experimental intervention so were unaware which they were participating in once they had been randomly allocated to one or the other. In two trials it was unclear whether participants had been adequately blinded.[27,32] Similarly, blinding of outcome assessors was not always possible due to the self-reported nature of outcome assessments used by many studies. However three studies reported blinding of outcome assessors[32,45,46] three were unclear if blinding had occurred[27,29,44] and 15 were rated as high risk as no blinding of outcome assessment had occurred.[25,26,30,31,33-37,39,41,43,47,48,50] Five studies reported a 'mixed' picture with blinding of the outcome assessment reported for some outcome measures and not for others.[28,38,40,42,49] For example Reiss et al [42] used the observer rated CARE scale, blinding the assessors to physician randomisation and three non-blinded self-rated scales to measure empathy.

Incomplete outcome data

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Incomplete outcome data was considered to be 'low risk' in 19 studies,[25,27-30,32,33,37-47,50] with attrition rates ranging from 0-16%. The risk was unclear in three studies[31,48,49]and considered high in four.[26,34,35,36]

Selective reporting

Eighteen trials described all pre-specified outcomes as stated in the methodology.[25-30,32,35-41,45,46] One trial presented an 'unclear risk' (Daniels et al[31] described dropping all males from the analysis) and seven studies were high risk for selective reporting.[33,34,43,47-49] Gould et al[34] for example did not report the data associated with the JSE questionnaire which was one of the specified outcomes.

Other potential sources of bias

Five trials were cluster RCTs,[26,34,35,47,50] of which three were considered low risk for recruitment bias[26,34,35] and two were identified as either unclear or high risk.[47,50] Eight studies were identified to be at either a high risk or unclear risk from 'other potential sources of bias.[27,29,31,34,38,44,48,49] For example Butow et al[29] reported differences between the study groups in baseline characteristics and six other studies did not report baseline demographics and/or empathy measurements at baseline.

Sustainability of improved empathy analysis

Eleven studies provided follow-up data assessing sustainability of changes to empathy, in addition to post intervention measurement.[25-27,29,31,33,35,37,39,47,50] Eight were eligible for inclusion in a sub-group analysis.[26,27,33,35,37,39,47,50] One was excluded

from all meta-analyses due to lack of data,[29] one was excluded from this meta-analysis as the empathy-intervention was delivered to the control group prior to the follow-up measures being taken,[25] and one was excluded as the follow-up data was not reported.[31] Studies were divided into two groups; those reporting follow up measures at less than 12 weeks and those reporting follow up at 12 weeks or later (figure 4). Arthur et al[26] and Hastings et al[35] provided multiple follow up data at time points that could be included in both groups (at 8 weeks and 12 weeks, and at 6 weeks and 20 weeks respectively). Meta-analysis found a moderate effect size for improved empathy until 12 weeks (effect size 0.69 95% CI 0.23-1.15) and a small but statistically significant effect size for sustainability at 12 weeks and later (effect size 0.34 95% CI 0.11 to 0.57).

Type of intervention analysis

A meta-analysis comparing sub-groups of different types of intervention (eFigure 2) found the greatest effect was with empathy training that was communication skills-based (effect size 0.69 [95% confidence interval 0.32 to 1.06]). The smallest effect reported was for interventions that were described as 'mixed educational programmes' and ones based in the arts and humanities (effect size 0.39 [95% confidence interval 0.18 to 0.61] and 0.38 [95% confidence interval 0.03 to 0.73] respectively). Interventions labelled as 'empathy skills-based training' had a positive but not statistically significant overall effect (0.60, 95% confidence interval -0.02 to 1.21).

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

eTable 1. Search strategies

MEDLINE		
# ▲	Searches	Results
1	exp Students/	116946
2	student?.ti,ab.	254787
3	(physician? or doctor? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff*).ti.	295930
4	exp Health Personnel/	481003
5	1 or 2 or 3 or 4	906748
6	exp Education/	767285
7	ed.fs.	264737
8	((intervention? or program*) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	137613
9	(train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula).ti.	369134
10	(intervention or program*).ti.	260613
11	6 or 7 or 8 or 9 or 10	1249776
12	5 and 11	335534
13	((physician? or doctor? or surgeon? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff* or practitioner? or student?) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	137434
14	12 or 13	393662
15	Empathy/	17455
16	(empath* or compassion*).ti,ab.	21716
17	15 or 16	31561
18	randomized controlled trial.pt.	481154
19	controlled clinical trial.pt.	93050
20	randomized.ab.	441413
21	placebo.ab.	197236
22	drug therapy.fs.	2104120
23	randomly.ab.	309893
24	trial.ab.	461528
25	groups.ab.	1906393

26	multicenter study.pt.	249476
27	pragmatic clinical trial.pt.	1037
28	(multicenter or multi center or multicentre or multi centre).ti.	47574
29	(intervention? or effect? or impact? or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab.	8937416
30	18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29	11030368
31	14 and 17 and 30	2441

EMBASE		
# ▲	Searches	Results
1	*student/ or exp *health student/	68463
2	student?.ti,ab.	326421
3	exp *health care personnel/	479224
4	(physician? or doctor? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff*).ti.	301997
5	1 or 2 or 3 or 4	941482
6	education/ or continuing education/ or curriculum/ or education program/ or in service training/ or lifelong learning/ or exp medical education/ or exp paramedical education/ or postgraduate education/	736812
7	((intervention? or program*) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	184005
8	(train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula).ti.	399259
9	(intervention or program*).ti.	318923
10	6 or 7 or 8 or 9	1266300
11	5 and 10	281380
12	((physician? or doctor? or surgeon? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff* or practitioner? or student?) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	179470
13	11 or 12	369015

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

14	Empathy/	23785
15	(empath* or compassion*).ti,ab.	28390
16	14 or 15	39458
17	13 and 16	4903
18	randomized controlled trial/	545326
19	single blind procedure/ or double blind procedure/	192596
20	crossover procedure/	58851
21	random*.tw.	1400168
22	(((singl* or doubl*) adj (blind* or mask*)) or crossover or cross over or factorial* or latin square or assign* or allocat* or volunteer*).ti,ab.	983905
23	pragmatic trial/ or multicenter study/	213866
24	intervention study/	40085
25	(multicenter or multi center or multicentre or multi centre).ti.	74011
26	(intervention? or effect? or impact? or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab.	11312699
27	18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26	12032330
28	(exp animals/ or nonhuman/) not human/	6212385
29	27 not 28	9294426
30	17 and 29	2574

PsychINFO		
# ▲	Searches	Results
1	students/ or medical students/	35317
2	student?.ti,ab.	481295
3	exp health personnel/	128154
4	(physician? or doctor? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff*).ti.	47232
5	1 or 2 or 3 or 4	616902
6	education/ or exp curriculum/ or distance education/ or nursing education/ or paraprofessional education/ or exp personnel training/ or exp medical education/	186066
7	((intervention? or program*) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	100952

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

8	(train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula).ti.	207043
9	(intervention or program*).ti.	121597
10	6 or 7 or 8 or 9	455304
11	5 and 10	166574
12	((physician? or doctor? or surgeon? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff* or practitioner? or student?) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	98357
13	11 or 12	209818
14	Empathy/	12489
15	(empath* or compassion*).ti,ab.	37254
16	14 or 15	38291
17	13 and 16	3043
18	random*.ti,ab,hw,id.	187448
19	trial*.ti,ab,hw,id.	172104
20	controlled stud*.ti,ab,hw,id.	11726
21	placebo*.ti,ab,hw,id.	38934
22	((singl* or doubl* or trebl* or tripl*) and (blind* or mask*)).ti,ab,hw,id.	27892
23	(cross over or crossover or factorial* or latin square).ti,ab,hw,id.	28819
24	(assign* or allocat* or volunteer*).ti,ab,hw,id.	156473
25	treatment effectiveness evaluation/	22860
26	mental health program evaluation/	2062
27	exp experimental design/	54976
28	(clinical trial or treatment outcome).md.	41809
29	intervention/	58790
30	(multicenter or multi center or multicentre or multi centre).ti.	2788
31	(intervention? or effect? or impact? or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab.	1834258
32	18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31	2026090
33	17 and 32	1767

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

CINAHL		
#	Query	Results
S17	S13 AND S16	381
S16	S14 NOT S15	556,315
S15	(MH animals+ OR MH (animal studies) OR TI (animal model*)) NOT MH (human)	154,114
S14	MH randomized controlled trials OR MH double-blind studies OR MH single-blind studies OR MH random assignment OR MH pretest-posttest design OR MH cluster sample OR TI (randomised OR randomized) OR AB (random*) OR TI (trial) OR (MH (sample size) AND AB (assigned OR allocated OR control)) OR MH (placebos) OR PT (randomized controlled trial) OR AB (control W5 group) OR MH (crossover design) OR MH (comparative studies) OR AB (cluster W3 RCT)	579,579
S13	S9 AND S12	2,335
S12	S10 OR S11	17,823
S11	TI (empath* or compassion*) OR AB (empath* or compassion*)	13,814
S10	(MH "Empathy")	8,360
S9	S7 OR S8	188,626
S8	TI ((physician? or doctor? or intern? or internship or resident? or residency or nurse? or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*") N5 (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula)) OR AB ((physician? or doctor? or intern? or internship or resident? or residency or nurse? or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*") N5 (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula))	55,142
S7	(S3 AND S6)	158,577

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

S6	S4 OR S5	550,634
S5	TI (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula) OR AB (((intervention? or program*) N5 (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula))) OR TI(intervention? or program*)	349,186
S4	(MH "Curriculum+") OR (MH "Education, Clinical+") OR (MH "Education, Health Sciences+") OR (MH "Staff Development") OR (MH "Education")	294,559
S3	S1 OR S2	663,254
S2	TI student? OR AB student? OR TI (physician? or doctor? or intern? or internship or resident? or residency or nurse? or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*")	226,699
S1	(MH "Students, Health Occupations+") OR (MH "Health Personnel+")	529,459

COCHRANE	
ID	Search
#1	MeSH descriptor: [Students] explode all trees
#2	(student*):ti,ab,kw
#3	MeSH descriptor: [Health Personnel] explode all trees
#4	(physician* or doctor* or intern or interns or internship or resident* or residency or nurse* or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*"):ti
#5	#1 or #2 or #3 or #4
#6	MeSH descriptor: [Education] explode all trees
#7	(train* or educat* or course* or workshop* or "staff development" or "professional development" or curriculum or curricula):ti OR (intervention* or program*):ti OR (((intervention8 or program*) N5 (train* or educat* or course* or workshop* or "staff development" or "professional development" or curriculum or curricula))):ti,ab,kw
#8	#6 or #7
#9	#5 and #8

#10	((physician* or doctor* or intern or interns or internship or resident* or residency or nurse* or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*") NEAR/5 (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula)):ti,ab,kw
#11	#9 or #10
#12	MeSH descriptor: [Empathy] explode all trees
#13	(empath* or compassion*):ti,ab,kw
#14	#11 and #13

eTable 2. Characteristics of excluded studies

Study	**Reason for exclusions**
Arthur 2015	Study protocol.
Bonvicini 2008	Observational data taken from an RCT. Intervention not specifically designed with outcome of change in empathy. Secondary analysis of data to see if there is an impact on empathy.
Bosse 2012	Change in empathy not a specified outcome of study
Bruera 2007	Change in empathy not measured or intended outcome.
Chen 2016	Not an RCT. Quasi-experimental design, not randomised.
Chunharas 2013	Not an RCT
Daeppen 2012	Change in empathy is not an intended outcome
Danucalov 2017	Empathy is not an intended outcome of the study. Participants not healthcare students or professionals.
Delvaux 2005	Change in empathy not an intended outcome and not measured
Downar 2016	Change in empathy not an intended outcome
Downar 2017	Change in empathy is not an intended outcome of the study.
Dundas 2017	Participants are not healthcare students/professionals.
Fallowfield 2002	Empathy is not directly measured
Fine 1977	Not an RCT
Gibon 2013	Change in empathy not an intended outcome
Gorniewicz 2016	Change in empathy not an intended outcome and is not measured
Hojat 2013	Not an RCT. Experimental control groups without randomisation.
Jaury 2018	Analysis of data already reported in RCT
Johnson 2013	Not an RCT. Controls selected from a waitlist group and intervention participants from a group who were due to undergo training in a set time-period.
Kahriman 2016	Change in empathy is not intended outcome
Klein 1999	Change in empathy is not measured
Liao 2016	Not an RCT. Quasi-experimental design
Lienard 2010	Change in empathy not an intended outcome
Lim 2011	Change in empathy not an intended outcome
Little 2015	Change in empathy not intended outcome of study and not specifically measured
Misra-Herbert 2012	Not an RCT
Nasr Esfahani 2014	No control arm, comparison between wo groups receiving same training, one as distant learning, one as attendants on course.
Nixon 2018	Not an RCT. Quasi-experimental design "partial randomisation was conducted" with participants designated to their preference group
Oz 2001	Not an RCT.
Perula de Torres 2019	Study protocol only
Potash 2014	No control arm "mixed-methods quantitative-qualitative study"
Rask 2009	Empathy not measured as an outcome
Razavi 2002	Change in empathy is not an intended outcome

Razavi 2003	Empathy not explicitly measured as an outcome
Rosenzweig 2016	Not an RCT
Roter 1995	Unclear whether intervention is looking to cultivate empathy and whether change in empathy is an intended outcome
Schroeder 2018	Change in empathy is not an intended outcome of the study
Shapiro 2004	Not an RCT
Shapiro 2009	Not an RCT
Shapiro 2011	Change in empathy is not an intended outcome
Smith 1995	Change in empathy is not intended outcome
Tamuma 2017	Only available in Japanese
Van Dijk 2017	Change in empathy is not an intended aim of the study
Van Vilet 2017	Not an RCT. Exploratory, controlled, quasi-experimental study using students not on a specific course as control group
Weatherdale 2018	Correspondence and not research study
West 2014	Change in empathy is not an intended outcome.

eTable 3. Characteristics of included studies

Alhassan 2019

Methods	Randomised controlled trial
Participants	The country of origin was Ghana. 104 students were randomised to the intervention group and 106 to the control group. The inclusion criteria were nursing and midwifery students in their second year of training, above age 18 and available for follow-up data collection after 6 months. The exclusion criteria included students not studying at Tamale Nursing and Midwifery College
Interventions	Communication Skills Training (CST) developed by author (MA) using 'Four Habits Model' and 'PCNF' (person-centred nursing framework). The mode of delivery were small group discussions, brainstorming, personal experience from participants, group reports, roleplaying, questions and answers, videos and summaries. The duration was 2 days and frequency was one off.
Outcomes	The outcome was empathy measured with JSE HPS version Outcome assessment 2 days post intervention and 6 months post intervention
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"NMS were separated before random assignment to ensure that both professions were approximately equally represented in the groups" "The researcher (MA) and research assistants conducted this by allowing participants to pick numbers written on papers, which had been randomly shuffled in a box."
Allocation concealment (selection bias)	Low risk	"There was allocation concealment to the researcher, research assistants and the participants. The researcher (MA) and research assistants conducted this by allowing participants to pick numbers written on papers, which had been randomly shuffled in a box."
Blinding of participants and personnel (performance bias)	High risk	"The participants were made aware of empathy being an outcome of this study and since JSE is self-reported, it may have impacted their self-report."
Blinding of outcome assessment (detection bias)	High risk	"The participants were made aware of empathy being an outcome of this study and since JES is self-reported, it may have impacted their self-report." "The data was analysed by the author (MA) without blinding."
Incomplete outcome data (attrition bias)	Low risk	11 participants in intervention group and 26 in control were excluded from analysis due to incomplete data or outcome measures not returned.

Selective reporting (reporting bias)	Low risk	Outcomes reported as pre-determined
Other bias	Low risk	No other bias detected

Arthur 2017

Methods	Pilot cluster randomised controlled trial
Participants	The country of origin was UK. Clusters were wards within three acute hospital trusts in England. General medical, stroke or care of the elderly/older people wards were eligible. Specialist dementia wards and medical admissions units were excluded. Health Care Assistants (HCAs) employed full or part time within enrolled wards were eligible to enter trial. Bank staff and not part of the named staff on ward roster were ineligible. In total 59 Health Care Assistants were randomised to the intervention group and 53 to the control group.
Interventions	'Older People's Shoes' training intervention that focuses on relational care of older people. The mode of delivery was small group teaching led by nurses who had received full training in content and delivery of the intervention from a member of the research team. The setting was the hospital, the duration of the intervention was 2 weeks and frequency was 1 half day session for 2 consecutive days followed by a weeks break and then repeated.
Outcomes	HCA outcomes were empathy, as measured by The Toronto Empathy Questionnaire (TEQ) at baseline and post intervention at 8 and 12 weeks post randomisation.
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Stratified by NHS hospital trust, wards were randomly allocated by the Norwich Clinical Trials Unit. Each ward had an equal chance of receiving either Older People's Shoes training for HCAs or TAU. Random allocation was generated via computer-written code using block sizes of four"
Allocation concealment (selection bias)	Low risk	"To conceal allocation from those responsible for recruitment, randomisation took place immediately after baseline measures were completed and 4 weeks ahead of the start of the intervention (set-up period) to allow appropriate arrangements, including HCA staffing cover to be arranged."
Blinding of participants and personnel (performance bias)	High risk	"At a number of ward-based meetings during the 4-week baseline period, HCAs were given information about the study"
Blinding of outcome assessment (detection bias)	High risk	Not described. Outcome measure is self-reported
Incomplete outcome data (attrition bias)	High risk	"For HCAs, completion of questionnaires was 72 out of 112 (64.2%) at baseline, 52 out of 112 (46.4%) at the first follow-up and 40 out of 112 (35.7%) at the second follow-up."
Selective reporting (reporting bias)	Low risk	Outcomes are reported as per methodology
Other bias	Low risk	Recruitment bias considered to be low risk: "Each ward had an equal chance of receiving either Older People's Shoes training for HCAs or TAU".

Blair Irvine 2012

Methods	Randomised controlled trial
Participants	The country of origin was the USA. 84 healthcare professionals were randomised to the intervention group and 88 to the control group. Eligibility criteria included: identification of professional license from a pre-determined list, working in nursing home and assisted living settings Exclusion criteria included: Working as Certified Nursing Assistant, Nursing Assistant, and Home Health Aide, working in a psychiatric/Alzheimer's care units and hospitals, working less than 20 hours per week, a 'moderate' or 'a lot' of self-reported level of mental illness,

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

	'extremely confident' self-reported confidence to deal with resident behaviours associated with mental illness
Interventions	Online training designed to develop skills and confidence to deal with symptoms of whatever mental illness was causing a particular behaviour. The mental illness training approach included video modelling vignettes, right-way and wrong-way exemplars, testimonials and narration supplemented by short on-screen text designed to create empathy for residents with mental illness. A minimum 'viewing time' for all online courses was 4 hours with two online 'visits' one week apart.
Outcomes	Video situational testing (VST) was used to assess participant reactions to short video vignettes of resident behaviour. Four items in VST were used to assess participant empathy towards a resident.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No detail given on how randomisation occurred
Allocation concealment (selection bias)	Unclear risk	No detail given on allocation of participant
Blinding of participants and personnel (performance bias)	High risk	"After submitting the baseline assessment, treatment participants were e-mailed login information to the Internet training program for Visit 1. One week after logging on to the Visit 1 courses, each participant was sent a second e-mail with log-in information for Visit 2."
Blinding of outcome assessment (detection bias)	Unclear risk	No detail given on how/who assessed video situational vignettes and whether outcome assessors were blinded
Incomplete outcome data (attrition bias)	Low risk	"Of the 172 study participants 91% completed all three assessment surveys, 6% completed two surveys, and 3% completed one survey Participants who completed all three surveys were compared to those who completed one or two surveys on study condition, demographic characteristics, and all baseline outcome measures. Attrition was not significantly related to any of the measures, which suggests that dropping out of the study did not bias results."
Selective reporting (reporting bias)	Low risk	Outcomes reported as stated in methodology
Other bias	Unclear risk	"our measures of empathy and stigma did not provide an in-depth assessment of these constructs, nor is it clear what elements of the training were influential"

Buffel Du Vaure 2017

Methods	Two site parallel group randomised controlled trial
Participants	The country of origin was France 176 fourth year medical students were randomised to the intervention group and 176 to the control group from two medical schools. No exclusion criteria were stated.
Interventions	Balint group training was the intervention with control conditions as 'teaching as usual'. The intervention was delivered in small group discussions held at the university. The duration of the intervention was 10.5 hours delivered in 1.5-hour weekly sessions over 7 weeks.
Outcomes	Empathy was assessed using the observer-rated CARE scale post intervention and JSPE student version self-rated scale pre and post intervention.
Notes	-

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Students from Paris Diderot were randomized with a simple randomization using computer generated random numbers" "students from Paris Descartes, we took advantage of the randomization routinely performed each year by university staff to allocate each student to one of three groups, each corresponding to a particular order of the three mandatory 3-month programs of the fourth-year curriculum"
Allocation concealment (selection bias)	Unclear risk	"students from Paris Descartes, we took advantage of the randomization routinely performed each year by university staff to allocate each student to one of three groups, each corresponding to a particular order of the three mandatory 3-month programs of the fourth-year curriculum"
Blinding of participants and personnel (performance bias)	High risk	"Participants in the intervention group received a training of 7 sessions of 1.5 hour Balint groups, over 3 months"
Blinding of outcome assessment (detection bias)	Unclear risk	Outcome assessed both by observer and self. "Whereas students and facilitators were aware of the allocated group, standardized patients, OSCE's observers and data analysts were kept blinded to the allocation". Self-assessment for JSPE so unable to blind outcome assessors (students themselves)
Incomplete outcome data (attrition bias)	Low risk	52 lost to follow up but study over recruited to ensure significance level of 5% and power of 80%. 14.7% attrition (21 intervention and 32 controls)
Selective reporting (reporting bias)	Low risk	Primary and secondary outcomes reported as stated in the methods
Other bias	Low risk	No other bias detected

Butow 2007

Methods	Randomised controlled trial
Participants	The country of origin was Australia. 16 medical and radiation oncologists were randomised to the intervention group and 14 to the control group. All medical and radiation oncologists from six tertiary care hospitals in six Australian cities which incorporated oncology outpatient clinics were invited to participate in the study No exclusion criteria stated
Interventions	Communication skills training was an intensive face-to-face workshop incorporating presentation of principles, a DVD modelling ideal behaviour and role-play practice, followed by four 1.5 hour monthly video-conferences incorporating role-play of doctor-generated scenarios.
Outcomes	The outcome was a change in doctor behaviour in eliciting and responding to emotional cues in patients and was measured via coding of a transcript from a filmed role-play at baseline, after completing the training and at 12 months post intervention.
Notes	No funding source stated

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"oncologists individually randomised immediately after giving consent and baseline data collection, to receive the training or not. Oncologists were stratified by hospital to ensure approximately equal numbers in the control and intervention arms within each institution, and then randomised within permuted blocks of size 6 constructed by the central research team using a random number table"

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Allocation concealment (selection bias)	Low risk	"oncologists individually randomised immediately after giving consent and baseline data collection, to receive the training or not."
Blinding of participants and personnel (performance bias)	High risk	"Control group doctors were offered training at the completion of the study." "It is possible that intervention doctors shared some study materials with control doctors although they were strictly instructed not to do so" "all doctors were aware that they were being assessed, which likely motivated them to be on 'their best behaviour"
Blinding of outcome assessment (detection bias)	Unclear risk	Does not state whether assessors were blinded
Incomplete outcome data (attrition bias)	Unclear risk	Two controls and two intervention participants lost to follow-up. 11.4% overall attrition
Selective reporting (reporting bias)	Low risk	Outcomes reported as stated in methodology
Other bias	High risk	Baseline imbalance: "EE and DP scores were significantly higher in the intervention group compared to the control group at baseline".

Collins 2017

Methods	Randomised controlled trial
Participants	The country of origin was USA 13 student pharmacists were randomised to the intervention group and 12 to the control group. First through to third year pharmacist students invited to participate. No exclusion criteria stated
Interventions	Students randomized to the literature intervention group were then sent a weekly email that included the reading assignment. Reading assignments were divided into three segments (approximately three to five minutes apiece), and students were requested to complete the readings in three separate sittings throughout the week. The intervention duration was 8 weeks with weekly sessions.
Outcomes	A change in empathy was measured using the JSE-HPS two weeks post end of the intervention.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Participants were randomized into either an intervention or control group." No detail of how randomisation occurred
Allocation concealment (selection bias)	Unclear risk	No details given
Blinding of participants and personnel (performance bias)	High risk	"The announcement was then followed by an email further explaining the study and inviting students to participate."
Blinding of outcome assessment (detection bias)	Unclear risk	No details given. However, outcome assessment is self-assessed by participants and participants not blinded.
Incomplete outcome data (attrition bias)	Low risk	Overall attrition rate 16%. (15.4% for intervention group, 16.7% for control group dropout rate)
Selective reporting (reporting bias)	Low risk	Outcomes reported as stated in results
Other bias	Low risk	No other bias detected

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Daniels 1998

Methods	Randomised controlled trial
Participants	The country of origin was Canada 53 full-time second year nursing students were randomly allocated to either the intervention or control group. Full-time second year female students in a two-year, eight-month registered nurse (RN) diploma program. Males not excluded from study randomisation but were excluded from analysis.
Interventions	Micro-counselling training divided into six segments with one micro-skill taught per segment including attending behaviour, questioning, minimal encouragers, paraphrasing, reflection of feeling and summarizing. The intervention was delivered face-to-face and training was divided into 6 segments of 3-5 hours with a minimum of 18 hours training.
Outcomes	The Empathy Construct Rating Scale and The Carkhuff Index of Communication (Empathy) self-rated scales were administered to assess changes in empathy post intervention.
Notes	No details on funding source given.

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"Subjects were randomly assigned to either an experimental group or a non-attention control group." No details of how random sequence generated
Allocation concealment (selection bias)	Unclear risk	"Subjects were randomly assigned to either an experimental group or a non-attention control group." No details on allocation of students to experimental/control
Blinding of participants and personnel (performance bias)	High risk	"During the period of micro-counselling training of the experimental subjects, the control subjects were non-attended. Essentially, the control subjects spent this period of time entirely on their own and received no supervision or structured training experience of any kind."
Blinding of outcome assessment (detection bias)	High risk	No details given of blinding outcome assessors however outcome assessment is self-assessment
Incomplete outcome data (attrition bias)	Unclear risk	"The sample consists of all full-time second year female students (n=60). In all, there are 56 females and 4 males. The males were dropped from the analysis and there was a further attrition of three subjects."
Selective reporting (reporting bias)	High risk	The males were dropped from the analysis and there was a further attrition of three subjects
Other bias	Unclear risk	No results tables/figures published for the 9-month follow-up data ("At the nine-month follow-up period, the experimental group performed better on all the dependent measures than the control group. However, these differences failed to reach statistical significance")

Foster 2016

Methods	Randomised controlled trial
Participants	The country of origin was USA. 35 and 18 medical students were allocated to 2 intervention arms and 17 to a control arm.
Interventions	Student engagement with a virtual patient (VP). Students interacted with VP online test-based interface. They conducted interviews as they would with live patients, but typed what they wanted to say rather than speaking. The three arms to the study consisted of: -The empathy-feedback VP: Human-assisted empathy feedback is a technique where human 'assessors' anonymously follow online the trainee's interaction with the VP in real

	time. The assessors' feedback about opportunities to express empathy was available to students for review at the end of the VP interaction -The Backstory VP: Combines embodied conversational agents and narrative video vignettes. When specific questions are asked of the VP, noninteractive video vignettes are presented which show scenes of the VP illustrating their condition. -Control VP: Provides typed interaction with VP without empathy feedback or patient backstory.
Outcomes	The primary outcome was to assess students' verbal responses to all the opportunities to show empathy presented to them by the simulated patients. The Empathic Communication Encoding System (ECCS) (developed to code empathic opportunities, defined as an explicit, clear and direct statement of emotion, progress or challenge by the patient) was used to assess empathy.
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	High risk	"Students were randomized into one of three groups." No detail on random sequence generation given.
Allocation concealment (selection bias)	Unclear risk	No detail on allocation given.
Blinding of participants and personnel (performance bias)	Unclear risk	"The (VP) assessors were not aware of the students' identity or study group assignment and could not see the students, and the students were not aware of the assessors' presence"
Blinding of outcome assessment (detection bias)	Low risk	"The (VP) assessors were not aware of the students' identity or study group assignment and could not see the students, and the students were not aware of the assessors' presence." "Measures were taken to label the transcripts (of SP interactions) in each study group such that the source of the transcript was not identifiable to the assessors" "The SPs (standardised patients) were blinded to students' study group assignment."
Incomplete outcome data (attrition bias)	Low risk	No attrition reported. N=70 randomised and n=70 analysed
Selective reporting (reporting bias)	Unclear risk	Study outcomes reported as stated in methodology
Other bias	Low risk	No other bias detected

Gholamzadeh 2018

Methods	Quasi-experimental randomised controlled design
Participants	The country of origin was Iran 63 third and fourth year medical students were allocated to either the control or intervention group. The inclusion criteria of the study were willingness to participate, being a third- or fourth-year nursing student, and not having taken any empathy courses in the past 6 months. In case the students were unwilling to continue participation in the study or were participating in another educational program at the same time, they were excluded.
Interventions	Workshop on empathy skills including self-awareness, and definition and examples of empathy towards patients. The intervention consisted of an 8-hour workshop on empathy skills that was held at the college for 2 days. The content of the workshop was designed by the researchers and reviewed and revised by some of the college professors. The workshop was mainly based on constructivist learning theory.
Outcomes	The JSE-HP self-rating scale was used to examine the effects of empathy skills training immediately and 2 months after the intervention.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"the 70 students were randomly divided into a control and an intervention group through block randomization."
Allocation concealment (selection bias)	Unclear risk	"the 70 students were randomly divided into a control and an intervention group through block randomization." No details of allocation to groups post randomisation.
Blinding of participants and personnel (performance bias)	High risk	"All students in the intervention group participated in the same workshop. The students were informed about the date of the workshop in advance."
Blinding of outcome assessment (detection bias)	High risk	Self-rated questionnaire (outcome assessor is participant)
Incomplete outcome data (attrition bias)	Low risk	All participants randomised completed the study
Selective reporting (reporting bias)	High risk	Outcomes not specifically stated in methodology.
Other bias	Low risk	No other bias detected

Gould 2017

Methods	Multi-site pilot randomised controlled trial (as part of a wider feasibility study)
Participants	Six ward teams were randomised to either intervention or control groups with a total of 168 nursing staff randomised to the intervention group and 81 to the control group. Medical and surgical wards with high proportion of older patients were eligible.
Interventions	The Creating Learning Environments for Compassionate Care (CLECC): educational programme focused on developing manager and team practices at a group level that create an expansive learning environment, theorised to enhance team capacity to provide compassionate care
Outcomes	Nurses' self-reported empathy was measured using the Jefferson Scale of Empathy (JSE) (Physician/HP version).
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomisation of clusters was undertaken using the ralloc command in Stata (Release 12, StataCorp) by the team statistician (IM-E) blinded to hospital and ward information other than ward specialty."
Allocation concealment (selection bias)	Low risk	"Procedures for allocation concealment and blinding proceeded as planned, with the exception of two researcher observers at follow-up reporting that they learnt of ward allocation from ward staff."
Blinding of participants and personnel (performance bias)	High risk	"It was not possible to conceal allocation from ward team nursing staff. Patients were not informed of allocation."
Blinding of outcome assessment (detection bias)	High risk	Empathy measurement is self-rated questionnaire so unable to blind outcome assessor Researchers gathering questionnaire data were aware of ward allocation.
Incomplete outcome data (attrition bias)	High risk	No attrition of wards during the study
Selective reporting (reporting bias)	High risk	No data reported on JSE other than: "There was no significant difference between groups (P=0.800)"
Other bias	Unclear risk	Baseline demographic and baseline measurement difference not fully reported for JSE.

		Recruitment bias low risk: Six wards in two NHS hospital Trusts in England were enrolled and allocated to intervention (n=4) or control (n=2). The number of clusters was determined by funding availability and the plan to run the study in at least two hospital organisations, and at least two ward specialties. Randomisation of clusters was undertaken using the ralloc command in Stata (Release 12, StataCorp) by the team statistician (IM-E) blinded to hospital and ward information other than ward specialty.

Hastings 2018

Methods	Cluster randomised controlled trail
Participants	118 residential care settings for people with intellectual disability (with a total of 236 staff) were randomised to either the intervention or control group. Residential settings were eligible for inclusion if: they were based in a community setting, provided services via publicly funded contracts, supported between one and 10 people with ID, employed staff who provided at least some 24-h support, provided care for at least one person with ID who displayed aggressive CB, could identify one manager/lead staff member and one other support staff member who could attend WCW training together. Staff were eligible for inclusion if: they were either a manager (or lead staff member as defined by the service provider organisation) or a direct support worker whose roles were no more than 50% administrative/management. Staff who worked less than 70% of full-time equivalent were also ineligible.
Interventions	WCW (Who's challenging who) training course for support staff in ID context covering communication, frustrations of people with CB (challenging behaviours), experience of being physically restrained, medication, feeling excluded and unhelpful attitudes and behaviour or support staff). The intervention was delivered in small group facilitated learning sessions by trained trainers. It was delivered in a one off half day session.
Outcomes	The Staff Empathy for people with Challenging Behaviour Questionnaire (SECBQ) was used to measure staff self-reported empathy at baseline and at 6 weeks and 20 weeks post randomisation.
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Randomisation occurred at one point in time for each phase, was carried out by a study-independent statistician from the Centre for Trials Research and used a dynamic balancing algorithm specifically designed for cluster randomised trials"
Allocation concealment (selection bias)	Low risk	The trial statistician remained blind to allocation up until the point of data analysis.
Blinding of participants and personnel (performance bias)	High risk	"Settings, and staff members within them, could not be masked to the intervention but were recruited prior to randomisation."
Blinding of outcome assessment (detection bias)	High risk	Self-reported outcomes to measure empathy
Incomplete outcome data (attrition bias)	High risk	Intervention group: 77% received intervention 6 week follow up 44.1% 20 week follow up 48.3%
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	Recruitment bias low: Randomisation occurred at one point in time for each phase, was carried out by a study-independent statistician from the Centre for Trials Research and used a dynamic balancing algorithm specifically designed for cluster randomised trials

		No evidence that further residential settings were added to the trial following randomisation.

Hattink 2015

Methods	Randomised controlled trial
Participants	The countries of origin were UK and the Netherlands. 142 care givers (informal or professional) were randomised to the intervention or control group. 24 were professional care givers. Participants who fulfilled the following criteria were recruited for the evaluation study: (1) were sufficiently computer literate to utilize the STAR website and (2) were currently an informal caregiver for someone with dementia living in the community, or a volunteer working with people with dementia with direct contact with community-dwelling people with dementia, or a professional caregiver for people with dementia with direct contact with community-dwelling people with dementia.
Interventions	STAR training portal, a Web-based portal consisting of 8 modules, 2 of which had a basic level and 6 additional modules at intermediate and advanced levels about dementia care. In addition, users had access to online peer and expert communities for support and information exchange. Up to 4 months to complete on-line training modules at participants own pace.
Outcomes	The Interpersonal Reactivity Index (IRI) was used to measure empathy pre and post intervention (empathy was measured as a secondary outcome) with changes to knowledge about dementia and attitudes to it being primary outcomes.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomization software was used to classify participants into either the experimental or control group."
Allocation concealment (selection bias)	Low risk	"Randomization software was used to classify participants into either the experimental or control group"
Blinding of participants and personnel (performance bias)	High risk	"Participants in the experimental group received a link to the STAR registration" "People in the control group were informed that they were assigned to the group that could follow the course free of charge after post-test measurements 4 months later."
Blinding of outcome assessment (detection bias)	High risk	Self-rated instrument used to measure empathy
Incomplete outcome data (attrition bias)	High risk	"During the pilot, 59 participants dropped out. The total response at post-test was 61%. Reasons for dropouts in the Netherlands (n=29) were no time (n=4) or unknown (n=25; no response to repeated emails of researchers to remind them of filling in the questionnaires). Reasons for dropouts in the United Kingdom (n=30) were no time (n=1), no computer at home (n=1), or unknown (n=28; no response to repeated requests by researchers to fill in the questionnaires)."
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Larti 2018

Methods	Comparative study with random allocation to control and intervention groups.
Participants	The country of origin was Iran 82 operating room nursing students were randomised to either the intervention or control group.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

	Inclusion criteria: second-semester or higher students who had entered the stage of clinical practice, had experience with communicating with patients, had not been diagnosed with any psychological conditions, and had no history of participation in communication or patient empathy workshops The exclusion criteria included incomplete responses to questionnaires, absence at any of the training sessions, and withdrawal from continuation of the study.
Interventions	Training programme for empathetic communication with patients in the operating room, mainly during the perioperative phase, using role-playing technique. The training was delivered face-to-face by the researchers with assistance from psychologists specialising in running empathy workshops. The duration of training was 12 hours delivered in 3 x 4 hour sessions with weekly sessions over 3 weeks.
Outcomes	The purpose of this study was to investigate the effects of a role-playing training program for empathetic communication with patients on the empathy scores of operating room nursing students. The JSE-HPS was used to measure self-rated empathy pre and one month post intervention.
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"A number was then randomly assigned to each of the students, and the numbers were poured into a bowl. The first paper drawn out of the bowl was for the experimental group, the second paper was for the control group, and this procedure was continued to select students from all years of study"
Allocation concealment (selection bias)	Low risk	"A number was then randomly assigned to each of the students, and the numbers were poured into a bowl. The first paper drawn out of the bowl was for the experimental group, the second paper was for the control group, and this procedure was continued to select students from all years of study"
Blinding of participants and personnel (performance bias)	High risk	"The objectives of the training program were then explained"
Blinding of outcome assessment (detection bias)	High risk	Self-assessment so no blinding of outcome assessor
Incomplete outcome data (attrition bias)	Low risk	Low attrition rate (6%)
Selective reporting (reporting bias)	Low risk	No other bias detected
Other bias	Unclear risk	

Lobchuck 2018

Methods	Two centre randomised controlled pilot study
Participants	The country of origin was Canada 25 nursing students were allocated to the intervention group and 19 to the control group. Students at: (a) the end of the second year or in the third year of a three-year accelerated baccalaureate program at the college or (b) the end of the second year or in the third or fourth year of a four-year baccalaureate program at the university were included. No exclusion criteria listed.
Interventions	Heart Health Whispering intervention was delivered as a novel person-cantered approach for counselling and health promotion. The training programme on perspective taking involved 4 phases. Phase 1 – individual teaching on perspective taking followed by 2 week period and instructions to practice skills. Phase 2 10 minute videoed conversation with actor. Phase 3, researcher and actor watch video and 'video-tag' thoughts and feelings actor remembered having experienced, shared, displayed etc. Phase 4 exit interviews
Outcomes	Empathy post intervention was assessed using the CARE scale completed by observer An adapted version of the CARE scale was also completed by the participant to capture their inference of the actors response to his or her clinical empathy.
Notes	-

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The Research Assistant (RA) conducted a computerized randomization process to assign students to Group I (n=24) or Group PI (n=18)"
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias)	High risk	"Due to practical reasons, students, the interventionist (JL), and interviewers (ML and LH) were not blinded"
Blinding of outcome assessment (detection bias)	Unclear risk	Mixed High – self reported measure of empathy (JSE) Low – observer reported - actor was blinded to group assignment.
Incomplete outcome data (attrition bias)	Low risk	Low attrition rate 5%
Selective reporting (reporting bias)	Unclear risk	Outcomes reported as per methodology
Other bias	Unclear risk	Baseline demographic differences not reported

Lor 2014

Methods	Randomised controlled trial
Participants	The country of origin was USA 40 student pharmacists were randomised to either the intervention or the control group. Students with pre-existing medical conditions were asked not to participate, and students with any self-reported medical conditions were automatically excluded.
Interventions	A 3 day simulation with each day including a designed activity with loss of the dominant hand usage, vision and speech. Simulations were followed by small group discussions regarding the daily activity, which covered its purpose, their feelings about the activity, items they learned, key take-away points, and how the items would affect their practice as future health care providers. This was followed by a large group discussion
Outcomes	The purpose of this study was to determine the immediate and sustained impact of a single, 3-day empathy intervention on empathy levels among students. The JSE-HPS was used to measure self-reported empathy at baseline, 7 days post-intervention and 90 days post-intervention.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Forty student pharmacists who volunteered and provided informed consent were then randomly assigned to either the intervention or control group" No information provided on random sequence generation
Allocation concealment (selection bias)	Unclear risk	"Subjects were randomized to an intervention group (n520) or control group (n520) and completed the JSE-HPS at baseline, 7 days postintervention, and 90 days postintervention." No information provided on allocation of students
Blinding of participants and personnel (performance bias)	High risk	"The purpose of this study was to determine the immediate and sustained impact of a single, 3-day empathy intervention on empathy levels among students and to address the lack of a control group by using a randomized, non-blinded, quasi-controlled design"

Blinding of outcome assessment (detection bias)	High risk	"The Jefferson Scale of Empathy-Health Profession Students version (JSE-HPS) was administered to the intervention and control groups at baseline, 7 days following the intervention (as post-test 1), and 90 days following the intervention (as post-test 2)."
Incomplete outcome data (attrition bias)	Low risk	No attrition from randomisation to reporting
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

LoSasso 2017

Methods	Randomised controlled trial
Participants	The country of origin was USA. 70 medical students were randomised to either the intervention or control groups. Third-year students were eligible to participate in the study while on their regularly scheduled six-week paediatric clerkship if their outpatient assignment was at a site using the Epic EMR system
Interventions	Training session on EMR (electronic medical records) specific communication skills, including discussion of EMR use, the SALTED (set-up, ask, listen, type, exceptions, documentation) mnemonic and technique and role-play.
Outcomes	Empathy was measured pre and post intervention using the self-rated JSE questionnaire. In addition an observer rating of empathy was taken using the JSPPPE (Jefferson Scale of Patient Perception of Physician Empathy).
Notes	No funding source reported.

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Participants in each six-week clerkship block were randomly assigned to the intervention group (n = 38) or to the control group (n = 32)." Not stated how randomisation occurred
Allocation concealment (selection bias)	Unclear risk	Details on allocation process not given
Blinding of participants and personnel (performance bias)	High risk	"In consenting for the study, students in both groups were made aware that the study examined how the training may improve empathy, which could have led to some bias."
Blinding of outcome assessment (detection bias)	Unclear risk	The SP and faculty raters' were blinded to whether students were in the intervention or control group – and completed the observer-rated scale JSPPPE (low risk) Self-reported scale JSE outcome assessors not blinded (high risk)
Incomplete outcome data (attrition bias)	Low risk	No attrition from randomisation to analysis
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Mueller 2018

Methods	Randomised controlled trial.
Participants	The country of origin was USA. 19 physical therapy students were randomised to the intervention group and 18 to the control group (which was a 'delayed' intervention group). All students entering the third year were approached. No exclusion criteria listed.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Interventions	On-line Called to Care curriculum used to improve patient outcomes through the development of optimal physical therapist behaviours. (employs film clips, quidded questions, research articles and other readings to promote the clinical application of educational concepts. Participants post and respond via a discussion board for each of the 11 modules.
Outcomes	The JSE-HP was used to measure a change in empathy pre and post intervention.
Notes	

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Participants were randomly assigned (via a blinded shuffle of cards) to an immediate intervention group or a delayed intervention group."
Allocation concealment (selection bias)	Low risk	"Participants were randomly assigned (via a blinded shuffle of cards) to an immediate intervention group or a delayed intervention group. The deck included only the numbered cards (to ensure an event 50/50 split) and group assignment based on events or odds)."
Blinding of participants and personnel (performance bias)	High risk	An orientation to the Called to Care curriculum was provided to all participants at the end of the spring 2015 semester. The participants were informed of their designation into the immediate or delayed intervention group.
Blinding of outcome assessment (detection bias)	High risk	Self- reported scale
Incomplete outcome data (attrition bias)	Low risk	Of the 37 participants 1 withdraw due to pregnancy-related delay in her internship (2.7%)
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	High risk	No other bias detected

Reiss 2012

Methods	Randomised controlled trial
Participants	The country of origin was USA. 54 residents and fellows were randomised to the intervention group and 45 to the control group. Residents and fellows were eligible if they (1) were currently in training, (2) were available to attend all three training modules, and (3) had clinical interactions with adult outpatients or inpatients able to complete physician rating surveys. Trainees on clinical rotations outside MEEI or MGH were excluded. Trainees on night float, paediatrics, ICU or research rotations were excluded unless they had a clinic with adult patients.
Interventions	Empathy and relational skills training protocol developed by first author and previously tested in a pilot study. Aims of training (1) scientific foundation of empathy, (2) increase awareness of physiology of emotions, (3) improve skills in decoding facial expressions of emotion, (4) teach empathic responses. Training was delivered any a trained physician in both the inpatient and outpatient setting. The duration of intervention was 4 hours and was delivered in 60 minute modules spaced over 4 weeks.
Outcomes	Change in empathy was assessed by patients using the CARE measure as the primary outcome. As secondary outcomes the following was measured: Physician skill at decoding facial expression (The Ekman Facial Decoding Test). Self-rated physician attitude about empathy (JSPSE, validated scale). Self-rated general empathic responsiveness in personal life (The Balanced Emotional Empathy Scale, BEES)
Notes	-

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Group assignment was determined by a computer-generated random number sequence"
Allocation concealment (selection bias)	Low risk	"Participating physicians were randomly assigned in a 1:1 allocation ratio to either the training intervention or to standard residency or fellowship training"
Blinding of participants and personnel (performance bias)	High risk	"Participating physicians were randomly assigned in a 1:1 allocation ratio to either the training intervention or to standard residency or fellowship training." "The training was comprised of three 60-minute modules spaced over 4 weeks"
Blinding of outcome assessment (detection bias)	Unclear risk	"Patients were blind to physician randomization, and physicians were blinded to which patients completed the surveys" "The primary outcome measure was change in empathetic and relational skills as assessed by patients blinded to physician randomization" Secondary outcomes – self rated scales of empathy so unable to blind outcome assessor
Incomplete outcome data (attrition bias)	Low risk	Overall attrition rate 7.5% (4 participants lost in control group, 1 participant lost in intervention group).
Selective reporting (reporting bias)	Low risk	Primary and secondary outcomes reported as stated in methods.
Other bias	Low risk	No other bias detected

Shapiro 1998

Methods	Matched randomised experiment with wait-list controls.
Participants	78 premedical and medical students were randomised to either the intervention or control groups. Inclusion criteria: first- and second-year medical students, the premedical honours society, and the Fostering and Achieving Cultural Equity and Sensitivity (FACES) premedical student group. Only those students willing to be randomly assigned to either the intervention or control group were included in the study.
Interventions	Elective module in Stress Reduction and Relaxation. The core of the program focused on training the students in mindfulness. Participants received training in: "Sitting Meditation", "Body Scan" and "Hatha Yoga". Emphasis on mindful breathing, "lovingkindness" and "forgiveness". In addition, students participated in experiential exercises designed to cultivate mindful listening skills and empathy. The training was delivered via a mixture of didactic teaching and small group sessions. The duration was approximately 18 hours delivered in 2.5 hour weekly sessions over 8 weeks.
Outcomes	Empathy was measured using an adapted version (half of the original version of 84 items) of The Empathy Construct Rating Scale (ECRS).
Notes	No funding source reported

Risk of bias table

Bias	Authors' judgement	Support for judgement

Random sequence generation (selection bias)	Unclear risk	"The design was a matched randomized experiment in which participants were assigned to a 7-week mindfulness-based intervention or a wait-list control group." Random sequence generation not reported
Allocation concealment (selection bias)	Unclear risk	Details of allocation concealment not stated
Blinding of participants and personnel (performance bias)	High risk	"The design was a matched randomized experiment in which participants were assigned to a 7-week mindfulness-based intervention or a wait-list control group"
Blinding of outcome assessment (detection bias)	Unclear risk	"all assessment measures were self-report psychological questionnaires which are intrinsically limited and open to response bias."
Incomplete outcome data (attrition bias)	Low risk	"One student did not complete the intervention due to severe medical problems for which she was hospitalized. Four of the participants in the control group did not complete the post-measures. The final count of participants was 73, consisting of 32 males and 41 females, 35 premedical students and 38 medical students."
Selective reporting (reporting bias)	High risk	"Outcomes reported as a cohort in general."
Other bias	Low risk	No other bias detected

Sripada 2010

Methods	Pilot randomised controlled trial
Participants	The country of origin was USA. 12 psychiatry residents were randomised to either the intervention or control group. All second- through fourth-year psychiatry residents treating out-patients at the University of Illinois College of Medicine during the academic years 2002–2005 were eligible to participate in this study. Patients were eligible if they were between the ages of 18 and 65, were in treatment for an Axis I psychiatric disorder, had no intellectual disability, and were not suicidal or psychotic.
Interventions	A feedback intervention designed to increase therapist empathic understanding and improve patient outcomes in psychotherapy was delivered. The feedback intervention condition involved completing the empathy measure along with other measures, and engaging in the feedback intervention which involved: At the end of each therapy session, patients and therapists recorded their views of the patient's GAF and predicted the GAF ratings of the other. In the intervention condition, at the beginning of the next session, therapists and patients exchanged ratings from the preceding session, providing an opportunity to discuss their respective views. The average number of sessions completed by each therapist–patient pair was 14.1 The average duration of patient participation in the study was 13.75 (±7.0) sessions or 183.87 (± 111.1) days. The average duration of therapist participation was 195.8 (± 117.4) days.
Outcomes	The Barrett-Lennard Relationship Inventory - 6-item scale designed to assess patients' ratings of therapist empathy as well as therapists' self-ratings of empathy.
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"Patient-therapist pairs were randomly assigned by the first author to the intervention or control group by flipping a coin." However how therapists were assigned to intervention or control not reported.
Allocation concealment (selection bias)	Unclear risk	Allocation to intervention/control not descried
Blinding of participants and personnel (performance bias)	High risk	"Patients were blind to intervention condition, but therapists were not, as they administered the intervention".

Blinding of outcome assessment (detection bias)	Unclear risk	"A more methodological limitation of this study is the potential for contamination that existed because a single therapist treated five patients, three of whom were assigned to control, and two of whom were assigned to intervention."
Incomplete outcome data (attrition bias)	Low risk	Methodology states: "Additionally, at the end of the 1st, 5th, 10th, 15th, and 20th sessions, patient and therapist subjects in both groups completed their respective forms of the BLRI (Barrett-Lennard, 1976). Only patient scores reported in results"
Selective reporting (reporting bias)	High risk	Data not explicitly reported for each group
Other bias	Unclear risk	difference in baseline demographics of therapists and patients not reported

Sterkenburg 2018

Methods	Parallel randomised controlled trial
Participants	The country of origin was the Netherlands. 111 care workers were randomised to the intervention group and 113 to the control group. Inclusion: Care workers working with people with disabilities
Interventions	Playing a computer-based serious game "The World of EMPA", aimed at enhancing empathy towards people with disabilities. The game illustrates characters with several types of disability, with six levels in which players have to respond to multiple-choice questions. The intervention was delivered online and took 20 minutes to complete. It was a one-off intervention.
Outcomes	The Empathy Quotient (EQ) short version self-rating questionnaire was administered to assess changes in empathy at baseline and immediately following the intervention.
Notes	Funding source not stated.

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"Upon completion of the pre-test phase, participants were automatically randomized via a computerized random assignment to one of the two conditions, based on the Mersenne Twister pseudorandom number generator (PRNG)"
Allocation concealment (selection bias)	Low risk	"The automatic computer-based randomization was implemented in the programming script of the experiment, resulting in the concealed allocation of the participants into one of the two intervention arms"
Blinding of participants and personnel (performance bias)	Low risk	"The participants were also unaware whether the condition they were allocated to was the experimental or control condition"
Blinding of outcome assessment (detection bias)	Low risk	"The researcher was blind to condition once participants started the computer program".
Incomplete outcome data (attrition bias)	Low risk	a total of 224 care workers working with people with disabilities were recruited, and 223 completed the study
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Tulsky 2011

Methods	Parallel randomised controlled trial
Participants	The country of origin was USA. 24 medical, gynaecological and radiation oncologists were randomised to the intervention group and 24 to the control group. Inclusion and exclusion criteria were not stated.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Interventions	A communication lecture (1 hour) was delivered to all intervention and control students. An interactive CD-ROM about responding to patients' negative emotions was then given to intervention participants. The CD-ROM included tailored feedback on the oncologists own recorded conversations. Participants had up to one month to view the CD-ROM.
Outcomes	Empathic statements - Post-intervention audio recordings were used to identify the number of empathic statements and responses to patients' expressions of negative emotion. Perceived empathy - 10 Likert scale items was used to assess perceived oncologist empathy (as assessed by patient)
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	"The oncologists were then randomly assigned by using the minimization method"
Allocation concealment (selection bias)	Low risk	"The oncologists were stratified by balanced randomization in a 1:1 ratio by site (Durham or Pittsburgh), sex (men or women), and specialty (medical oncology, solid and liquid tumours; medical oncology, solid tumours only; malignant haematology, liquid tumours only; gynaecologic oncology; or radiation oncology)."
Blinding of participants and personnel (performance bias)	High risk	"All of the oncologists viewed a 1-hour lecture on communication skills delivered by one of the investigators. In addition, oncologists in the intervention group received a CD-ROM training program on communication skills that was tailored with exemplars from their own audio-recorded clinic visits."
Blinding of outcome assessment (detection bias)	Low risk	"Two independent, blinded coders were trained over 6 weeks"
Incomplete outcome data (attrition bias)	Low risk	No attrition from randomisation to analysis
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Vaghee 2018

Methods	Cluster randomised controlled trial
Participants	The country of origin was Iran. Nursing faculties training mental health clerkship in Ibne-Sina psychiatric hospital were invited to attend in the study, and accordingly, 12 faculties accepted the invitation, and 4 faculties were randomly selected. 127 nursing students were randomised to one of three groups: two intervention groups or a control group. Inclusion criteria were no work experience in psychiatric wards, no psychological disorders, and no mental illness in their first and second degree relatives. Exclusion criteria were reluctance to continue the study, absence of the post-test, and being absent or lack of participation in 1 or more intervention sessions.
Interventions	The two intervention groups were: Contact based education: In contact-based education, 3 patients with improved disorders who were working daily for 4 hours as a connector between different wards of the hospital were selected. They had schizophrenia, bipolar type I, and major depression. The patients were asked to talk about their experiences and personal life with students Acceptance and commitment education: According to Steven Hayse protocol (1986), ACT with the content of mental illnesses stigma was held as a workshop by one master of clinical psychology and 2 masters of psychiatric nursing,
Outcomes	The study aimed at comparing the effects of contact-based education and commitment and acceptance-based training on empathy toward mental illnesses among nursing students. The JSE was used as a self-rating measure of empathy pre and post intervention.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Two groups of male and female students were randomly selected (according to clerkship division group) from each university by quota sampling based on gender distribution. Finally, each group was separately divided into 3 groups of contact-based education, ACT, and control." No details on random sequence generation
Allocation concealment (selection bias)	Unclear risk	No details on allocation concealment reported
Blinding of participants and personnel (performance bias)	High risk	"The patients were asked to talk about their experiences and personal life with students"
Blinding of outcome assessment (detection bias)	High risk	Self-reported outcome measures
Incomplete outcome data (attrition bias)	Low risk	Low attrition rate (12.5%)
Selective reporting (reporting bias)	High risk	Outcomes are not clearly stated in methodology
Other bias	Unclear risk	Recruitment bias: Random cluster and quota sampling methods were used. Nursing faculties training mental health clerkship in Ibne-Sina psychiatric hospital were invited to attend in the study, and accordingly, 12 faculties accepted the invitation, and 4 faculties were randomly selected.

Wolf 1987

Methods	Randomised controlled trial
Participants	The county of origin was Canada 65 medical students were randomised to the intervention group and 69 to the control group. Part of course was conducted in community nursing homes, so not all students could be scheduled to participate in it at the same time. Therefore, some of the students participated in the main part of the study. The remaining (excluded) students participated in the course after the study was completed.
Interventions	Programme in medical interviewing and history taking that integrates humanistic principles and medical content. The course is designed to use community resources and maximise efficient use of faculty members' time. Consists of set of large group lectures and then small group teaching sessions which included discussing strategies for responding empathically to patients. The teaching was delivered in small group sessions by social workers and educational psychologists. It consisted of 3 x 4 hour sessions and was delivered weekly.
Outcomes	The Medical Communication Index (MCI) served as the dependent variable to measure the students' responses to patients' emotional concerns The Helping relationship Inventory (HRI) served to measure the dependent variable to measure the students' preferences for responses that expressed empathy or understanding.
Notes	No funding source stated

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"All students in both the intervention and control groups attended these large group lectures. Following this instruction, the students were randomly assigned to an intervention or control group" Details of random sequence generation not reported

Allocation concealment (selection bias)	Unclear risk	"Part of course conducted in community nursing homes, not all students could be scheduled to participate in it at the same time. Therefore, only 134 of these students participated in the main part of the study. The remaining (excluded) students participated in the course after the study was completed." Allocation concealment not reported
Blinding of participants and personnel (performance bias)	High risk	"The 69 students in the control group received no other instruction in communication skills during the study. The 65 students in the intervention group were divided into four smaller groups. Each group met for four weekly, three-hour sessions."
Blinding of outcome assessment (detection bias)	High risk	Self-rated outcome assessment
Incomplete outcome data (attrition bias)	Unclear risk	24 lost to follow up (not clearly stated) on analysis of MCI). Not explicitly stated on what number of students' basis analysis carried out, how many lost to follow up or reasons
Selective reporting (reporting bias)	High risk	Outcomes not clearly stated in methodology.
Other bias	Unclear risk	no baseline demographics reported so cannot comment on baseline differences

Wundrich 2017

Methods	Randomised controlled trial.
Participants	The country of origin was Germany. 158 third year medical students were randomised to either an intervention or control group. No inclusion or exclusion criteria were stated.
Interventions	A three week training course with focus on empathy: The empathy skills training consisted of an introduction course on empathy and empathy skills training with simulated patients. The duration of the intervention was 6 hours delivered over 3 weeks.
Outcomes	The self-rated JSPE (student version) was used to measure empathy in addition to an empathy-related communications skills questionnaire completed by an observer.
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"A total of 158 3rd year medical students at the University of Freiburg Medical Centre were assigned into an intervention group receiving an empathy training and a control group" Details of random sequence generation not reported
Allocation concealment (selection bias)	Unclear risk	Allocation concealment not stated
Blinding of participants and personnel (performance bias)	High risk	"The intervention group participated in an empathy skills training with simulated patients (SPs). The control group participated in a history course."
Blinding of outcome assessment (detection bias)	Unclear risk	Experts and SPs were blinded to the students' group membership - low risk for observer rated outcome. Self-rated outcome high risk
Incomplete outcome data (attrition bias)	Unclear risk	Number analysed not reported. Missing data not reported
Selective reporting (reporting bias)	High risk	Number analysed not reported. Missing data not reported
Other bias	Unclear risk	no baseline demographics reported so cannot comment on baseline differences

Yang 2018

Methods	Cluster randomised controlled trial
Participants	The country of origin was China.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

	59 'grade 3' nursing students each were randomised to two intervention arms and 59 to a control arm of the study. Exclusion criteria: students who were taking doctor–patient communication- related courses and students who were planning to take those courses during the study.
Interventions	The intervention was a narrative medicine programme. Two intervention groups: One group received the theoretical education part of the programme and one intervention group received both theoretical teaching and clinical experience. The theoretical component was delivered by a teacher 'well trained in narrative medicine'. The clinical component was delivered by teaching nurses who had been trained in narrative medicine.
Outcomes	The JSE (Chinese version) was administered to students at baseline and then at various follow up points post intervention: T1: January 2015 (pre-intervention), T2: July 2015 (post-step 1 intervention) T3: January 2016 (post-step 2 intervention), T4: July 2016 (0.5 years after the intervention), T5: January 2017 (1 year after the intervention), and T6: July 2017 (1.5 years after the intervention).
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Random sequence generation not stated. "the six classes were randomly divided into three groups"
Allocation concealment (selection bias)	Unclear risk	"Of the sixteen classes, six (30 students per class) were randomly selected to participate in this study." "Taking each class as a unit, the six classes were randomly divided into three groups: one observation group (Group 1) and two experimental groups (Groups 2 and 3)." Method of allocation not stated.
Blinding of participants and personnel (performance bias)	High risk	No blinding of participants or personnel
Blinding of outcome assessment (detection bias)	High risk	Outcome assessors were not blinded.
Incomplete outcome data (attrition bias)	Low risk	5 participants from intervention groups and 7 controls lost to follow up. Attrition 6.6%
Selective reporting (reporting bias)	Low risk	Outcomes reported as stated in methods.
Other bias	Unclear risk	Recruitment bias: Method of randomisation not described "six [classes] were randomly selected" According to methodology, no participants were recruited after the clusters had been randomised.

eTable 4 Empathy effect summary of findings

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Summary of findings:

Empathy training compared to Control for Healthcare students and professionals

Patient or population: Healthcare students and professionals
Setting: University, primary care settings, secondary care settings
Intervention: Empathy training
Comparison: Control

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Control	Risk with Empathy training				
empathy	-	SMD **0.52 SD more** (0.36 more to 0.37 more)	-	2024 (22 RCTs)	⨁⨁◯◯ LOW [a,b]	Empathy training may increase empathy.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

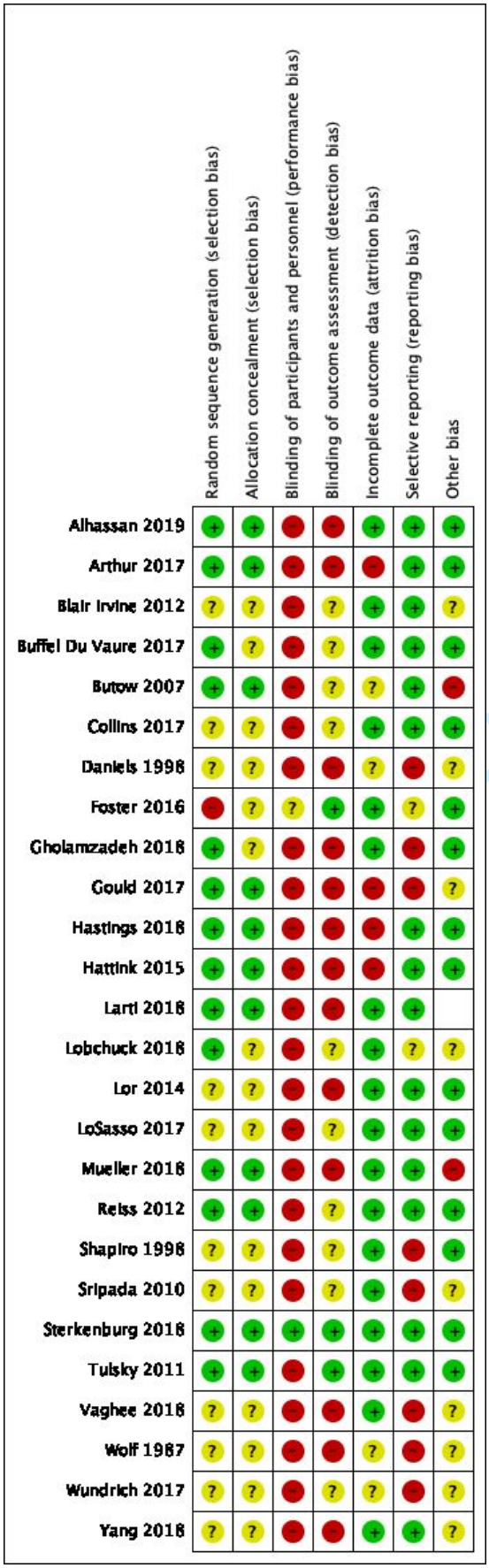
CI: Confidence interval; **SMD:** Standardised mean difference

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Footnotes
a High risk of bias suspected in 11 studies (with a high or unclear risk of bias for sequence generation and allocation concealment)
b There was variation across all studies with type of intervention and population studied

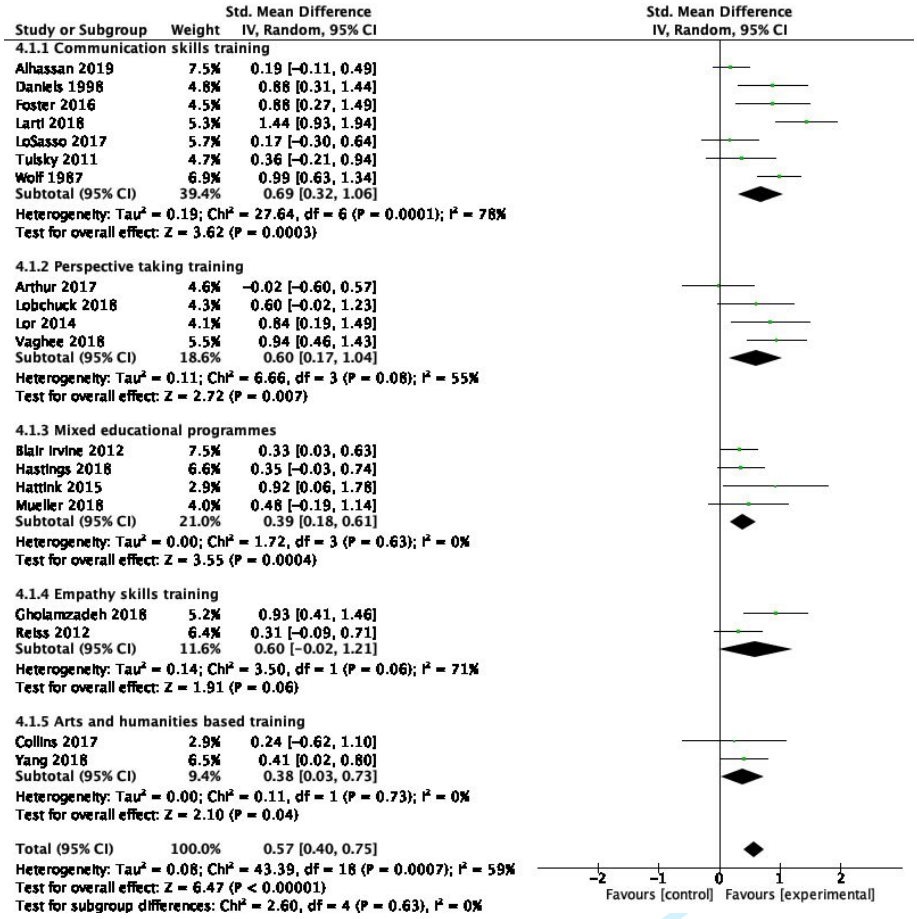
eFigure 1. Risk of bias assessment

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019



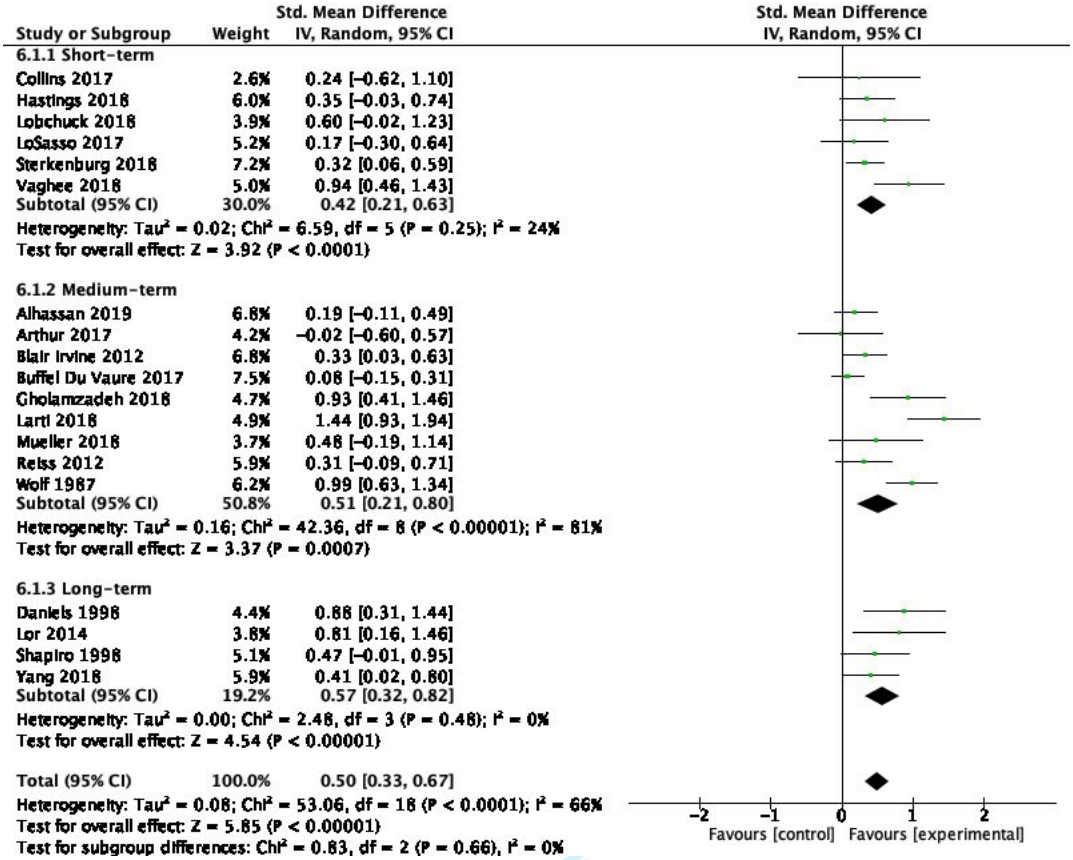
Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

eFigure 2. **Meta-analysis of sub-groups according to type of intervention**

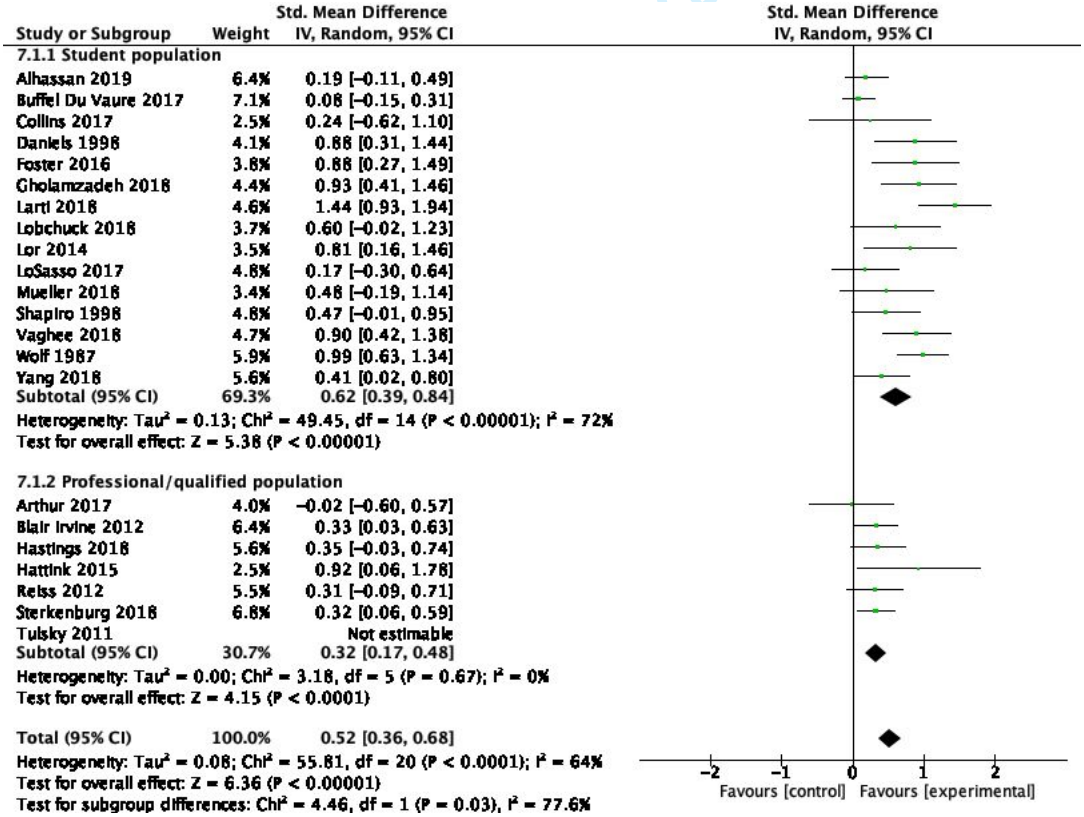


eFigure 3. **Meta-analysis of subgroups according to duration of intervention**

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

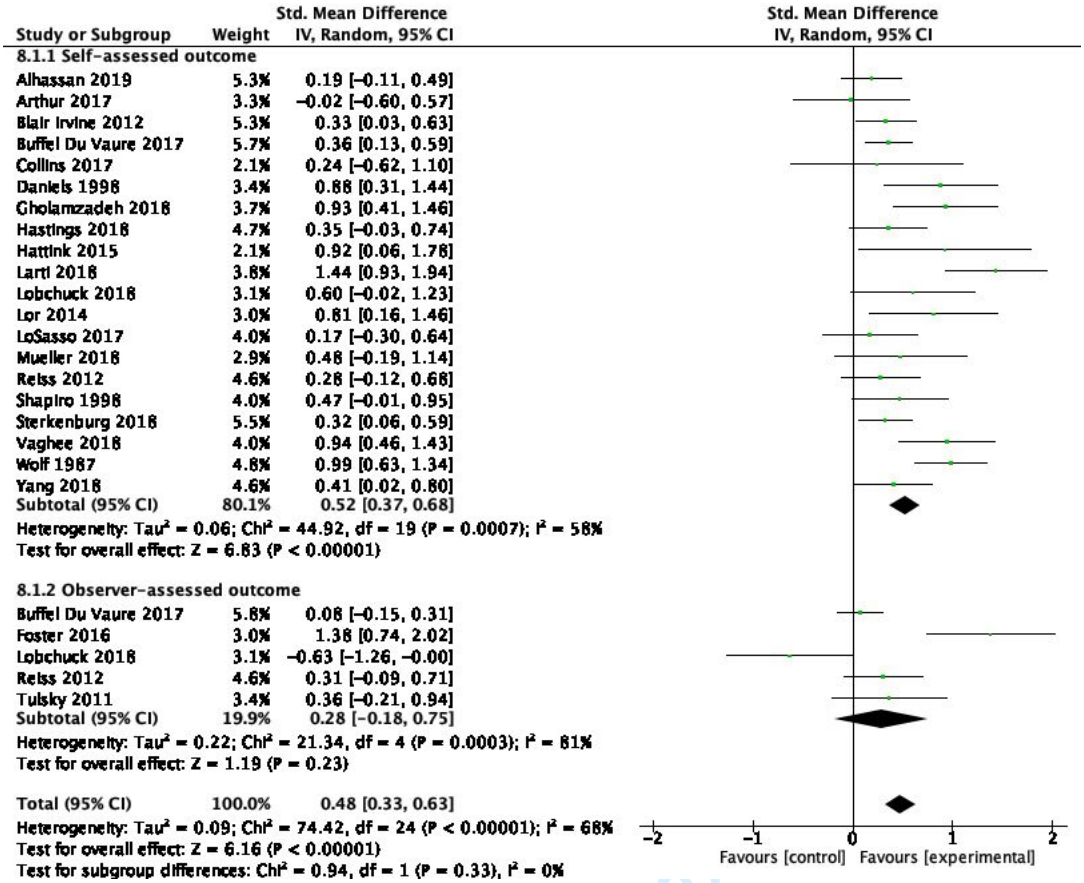


eFigure 4 **Meta-analysis of subgroups according to participant population**



Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

eFigure 5 **Meta-analysis of subgroups according to outcome assessor**



PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	7
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	8
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	8

PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	9
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	11
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	12
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	16
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	16
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	17
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	18
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	21
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	21
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	22
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

For more information, visit: **www.prisma-statement.org**.

BMJ Open

Assessing the effect of empathy-enhancing interventions in health education and training: A systematic review of randomised controlled trials

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-036471.R1
Article Type:	Original research
Date Submitted by the Author:	13-Jul-2020
Complete List of Authors:	Winter, Rachel; University of Leicester, Health Sciences Issa, Eyad; University of Leicester, Health Sciences Roberts, Nia; University of Oxford, UK, Bodleian Health Care Libraries, Norman, Robert; University of Leicester, Health Sciences Howick, Jeremy; University of Oxford, Faculty of Philosophy
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Patient-centred medicine
Keywords:	MEDICAL EDUCATION & TRAINING, EDUCATION & TRAINING (see Medical Education & Training), STATISTICS & RESEARCH METHODS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Assessing the effect of empathy-enhancing interventions in health education and training: A systematic review of randomised controlled trials

Rachel Winter (0000-0002-8775-2461), Eyad Issa, Nia Roberts, Robert I Norman, Jeremy Howick

Rachel Winter

Academic Clinical Lecturer

College of Life Sciences, University of Leicester, George Davies Centre, Leicester, England LE1 7RH

Eyad Issa

Academic Clinical Lecturer

College of Life Sciences, University of Leicester, George Davies Centre, Leicester, England LE1 7RH

Robert I Norman

Director of Learning and Teaching

College of Life Sciences, University of Leicester, George Davies Centre, Leicester, England LE1 7RH

Nia Roberts

Librarian

Bodleian Health Care Libraries, University of Oxford, Oxford, England

Jeremy Howick

Director, Oxford Empathy Programme

Faculty of Philosophy, University of Oxford, Oxford, England

Correspondence to: Rachel Winter rw205@le.ac.uk, College of Life Sciences, George Davies Centre, University Road, Leicester, England LE1 7RH

Word Count: 4000

ABSTRACT

Objective: To estimate the effect of empathy interventions in health education and training from randomised controlled trials (RCTs).

Methods: MEDLINE, PsycINFO, EMBASE, CINAHL and Cochrane databases were searched from inception to June 2019 for RCTs investigating the effect of empathy-enhancing interventions in medical and healthcare students and professionals. Studies measuring any aspect of 'clinical empathy' as a primary or secondary outcome were included. Two reviewers extracted data and assessed risk of bias of eligible studies using the Cochrane Risk of Bias Tool. Random effects meta-analyses of the impact of empathy training on participants' empathy levels were performed.

Results: Twenty-six trials were included, with 22 providing adequate data for meta-analysis. An overall moderate effect on participant empathy post-intervention (standardised mean difference 0.52, 95% confidence interval 0.36 to 0.67) was found. Heterogeneity across trial results was substantial (I2=63%). Data on sustainability of effect was provided by 11 trials and found a moderate effect size for improved empathy up until 12 weeks (0.69 95% confidence interval 0.23 to 1.15), and a small but statistically significant effect size for sustainability at 12 weeks and beyond (standardised mean difference 0.34 95% confidence interval 0.11 to 0.57). In total 15 studies were considered to be either unclear or high risk of bias. The quality of evidence of included studies was low.

Conclusions: Findings suggest empathy-enhancing interventions can be effective at cultivating and sustaining empathy with intervention specifics contributing to effectiveness. This review focuses on an important, growing area of medical education, and provides guidance to those looking to develop effective interventions to enhance empathy in the

healthcare setting. Further high quality trials are needed that include patient-led outcome assessments and further evaluate the long-term sustainability of empathy training.

Protocol registration: PROSPERO registration number (CRD42019126843).

Strengths and limitations of this study

- This is an up-to-date review that excludes non-randomised studies, follows a pre-published protocol, and measures the longer term effects of empathy training.
- The quality of the review was limited by the reporting quality of some of the included studies.
- The studies in our review were heterogeneous, which we anticipated.
- We found only four studies that followed-up participants for at least three months,

INTRODUCTION

Rationale

Clinical empathy has multiple benefits for patient care[1-4] and practitioner health.[5, 6] Indeed, person-centred and empathic care are central to all professional healthcare education.[7] Empathy in the clinical setting has been defined in various ways[8] and can be considered as a multidimensional construct incorporating affective, cognitive, behavioural and moral components.[9] A widely accepted definition of clinical empathy involves the ability to understand the patient's situation, perspective and feelings, communicate that understanding to them, and act on it in a helpful and therapeutic way.[10] There is still however, little consensus on the precise nature of clinical empathy, not least reflected in

the variety of tools and scales available to measure it. No guidance exists on how to select measures for assessing clinical empathy and choice of tools is likely to be led by the definition of empathy used or specific domain being measured.[11] A recent systematic review[11] on empathy measurement tools for care professionals identifies certain measures as scoring highest for quality, but concedes even these had low scores in some of the criteria they used.

Although contested by some,[12,13] there is evidence that empathy in medical and healthcare students declines during undergraduate education.[14-16] Researchers agree that empathetic skills can be taught [17-20] and cultivating empathy to protect against a possible decline would seem sensible. No standard empathy-curriculum for healthcare training currently exists and empathy-based training does not appear routinely in healthcare education.[14] Understanding what type of empathy training is most effective in healthcare at both cultivating and sustaining empathy would be a useful start in preparing one.

Four systematic reviews of empathy-promoting interventions have been conducted.[17,20-22] Kelm et al[17] conducted a qualitative synthesis of empathy-cultivating interventions for medical students or physicians. Their findings support the hypothesis that interventions can increase physician and medical student empathy. However, they also identified a lack of rigorous study design in most studies (such as lack of control groups). More recently, Vassilios et al[20] published a systematic review of randomised control trials (RCTs) of empathy-promoting interventions for health professionals. However, only two out of 17 included reported change in empathy as a primary outcome, focusing instead on general communication skills. Hence, the review did not provide robust evidence of empathy-

enhancing interventions. In 2019, Patel et al[21] reviewed educational interventions aimed at enhancing both empathy and/or compassion. They included observational as well as randomised studies and looked only at physicians and physicians-in-training. They were not able to pool their results statistically and did not investigate whether potential benefits of empathy were sustained over time. With the most recent review, Frakgos and Paul[22] conclude that empathy interventions significantly increase empathy, but limit their study population to medical students only. In addition, they do not explore whether any improvement in empathy is sustained over time.

These problems listed above present barriers for medical educators looking to implement empathy training into their curricula. It is unclear how large the effect size of effective empathy training is; whether the effect is sustained over time; or how best to train students and continuing learners from various health backgrounds. It is important to measure the effect of empathy training, both post-intervention and sustainability of effect over time. Arthur et al. [23] found no effect of empathy training immediately after the training, but significant improvement 12 weeks after the end of the training. A delayed improvement in empathy could potentially be accounted for by participants only recognising the benefits of training once they are putting any lessons learnt into action.

In this systematic review and meta-analysis we addressed these gaps, with an up-to-date synthesis of RCTs of interventions aimed at promoting empathy, delivered to both medical and healthcare students and professionals, with results that are generalisable to all healthcare contexts. In addition, we will consider both immediate and longer-term impact of interventions on empathy.

Objectives

The overarching objective of this systematic review and meta-analysis is to combine data from all available RCTs of empathy-enhancing educational interventions in health education and training. This was achieved with two subsidiary objectives:

(1) to assess the effectiveness of empathy-enhancing interventions aiming to enhance empathy in undergraduate and postgraduate health education and training; and

(2) to assess any lasting effect of empathy training.

We also had three secondary aims:

(1) to identify the intervention type (e.g. communication skills training) that is most effective at enhancing empathy;

(2) to identify the duration of training that is most effective; and

(3) to identify the tools used to measure empathy levels in participants to consider differences in self-reported and observer-reported measures.

METHODS

Protocol and registration

In accordance with the Cochrane Handbook for systematic reviews of interventions,[24] we published a protocol for this systematic review,[25] registered with PROSPERO international prospective register of systematic reviews (registration number CRD42019126843). We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[26]

Eligibility criteria

Randomised controlled trials (RCTs) investigating the effect of empathy-enhancing interventions on medical and other healthcare students' and professionals' empathy levels as a primary or secondary outcome were eligible for inclusion. Trials measuring empathy via self- and/or observer-reported measures were included. See eMethods in the supplement for further details.

Information sources and search strategies

The following databases were searched from inception to 6 June 2019: MEDLINE, PsycINFO, EMBASE, CINAHL and Cochrane. Search strategies are detailed within eTable 1 in the Supplement. Electronic searches were supplemented by hand-searching the references of retrieved papers.

Study selection

All studies retrieved through the search strategy were stored using EndNote with duplicates removed. Two authors (RW and EI) reviewed titles and abstracts to identify those meeting inclusion criteria. Full text manuscripts were retrieved for potentially relevant articles. If the decision to include or exclude was unclear, the study was discussed with a third author (JH) to reach a consensus. Seven papers were discussed with the third author. A PRISMA flow chart recorded the screening and selection process.

Data collection

One reviewer (RW) extracted, summarised and recorded data to assess quality and synthesise evidence from included studies. A second author (JH) independently extracted a

random sample (10%) of studies to ensure agreement on the information extracted and summarised. See eMethods for details on information extracted. If data was not reported, study authors were contacted.

Risk of bias in individual studies

Risk of bias was assessed using the Cochrane Collaboration's Tool for assessing the risk of bias in clinical trials (see eMethods in the supplement for further details). Using the criteria provided by Higgins (2011)[24], each item was scored as high, low or unclear risk of bias, and evidence from the study was used to justify each score given. Given that evidence increasingly suggests that sequence generation and allocation concealment are of particular importance in determining the overall risk of bias,[24] a study was classed as being at high risk of bias if it scored as high or unclear risk on either of these domains.

Synthesis of results

We calculated the overall effect size of empathy interventions using the standardised mean difference (SMD) and 95% confidence intervals (CI) based on the data provided in the studies: post-intervention sample size, mean and standard deviation (SD) for experimental versus control group (except where only mean difference and SD between pre- and post-intervention for the experimental and control groups were provided). We used a random effects model (REM) to allow for likely different (though related) intervention effects. If a study had more than one intervention arm, we used the results for the most comprehensive training intervention. If a study provided measures of empathy using different tools, the

primary tool to measure empathy was used. If it was unclear which was the primary measure, we used the first reported measure of empathy.

Heterogeneity was anticipated between studies and assessed using Cochran's Q Statistic (heterogeneity was declared if p-value <0.10) and quantified using the I^2 statistic, with an I^2 value of 50% or more being considered to represent levels of heterogeneity.

Primary analysis included all studies providing the data needed to calculate the mean and SD (or standard error (SE)) of the post-intervention control and intervention groups. Where studies provided more than one point for outcome assessment, the data closest to the end-point of the intervention was used. Studies that provided no numerical data on empathy-related outcomes or data from which it was not possible to calculate mean values and SD were excluded from the meta-analysis.

Additional analyses

We performed a sensitivity analysis excluding studies that were considered to be at high risk of bias (scoring unclear or high risk of bias for either sequence generation or allocation concealment, with evidence suggesting these domains are of particular importance in establishing risk of bias).[24]

We conducted separate meta-analyses to look at: sustainability of the effects of the intervention; the intervention type that is most effective; the duration of intervention that is most effective; the outcome assessment tools (comparing objective and subjective

outcome measures); and participant populations (effectiveness of interventions aimed at student populations compared with those aimed at professional populations). See eMethods in supplement for further details.

Risk of bias across studies

Reporting bias was assessed qualitatively based on inspection of the characteristics of the studies included. A funnel plot was produced to investigate small study effects, which may indicate the presence of publication bias. The GRADE system was used to evaluate the overall quality of evidence for the primary outcome.[27]

Patient and public involvement

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

RESULTS

Study selection

The literature search resulted in 4,904 citations with duplicates removed. Figure 1 provides an overview of the selection process (see eResults in the Supplement for further details). Seventy-two articles were retrieved for full-text review. Forty-six studies were excluded

(eTable 2 in the Supplement). Twenty-six trials were included.[23,28-52] (n=2,900) Table 1 provides a summary of characteristics (eTable 3 in the Supplement gives further details).

Table 1. Summary of characteristics of included studies

Study	Year	County	No. participants	Participant type	Intervention type	Duration of intervention (hours)	Outcome assessor	Outcome measure	Effect of intervention
Alhassan	2019	Ghana	210	Nursing and midwifery students	Communication skills training	12	Self	JSE	No significant effect found
Arthur	2017	UK	112	Health care assistants	Perspective-taking training	12	Self	JSE	No significant effect found
Blair Irvine	2012	USA	172	Health care professionals	Mixed	4	Self	VST	Significant effect found
Buffel Du Vaure	2017	France	352	Medical students	Balint group	10.5	Self Observer	JSE CARE	Mixed. No significant effect for JSE, significant effect for CARE
Butow	2007	Australia	30	Physicians	Communication skills training	15	Observer	CRP	No significant effect found
Collins	2017	USA	25	Student pharmacists	Literature intervention	2	Self	JSE	No significant effect found
Daniels	1998	Canada	53	Nursing students	Communication skills training	18	Self Self	ECRS CIC	Significant effect found
Foster	2016	USA	70	Medical students	Communication skills training	NE	Observer	ECCS	Significant effect found
Gholamzadeh	2018	Iran	63	Nursing students	Empathy skills training	8	Self	JSE	Significant effect found
Gould	2017	UK	249	Nursing staff and healthcare assistants	Mixed	NE	Self	JSE	No significant effect found
Hastings	2018	UK	236	Qualified care staff	Mixed	3	Self	SECBQ	No significant effect found
Hattink	2015	Netherlands and UK	142	Qualified care staff	Mixed	NE	Self	IRI	Significant effect found
Larti	2014	Iran	82	Nursing students	Communication skills training	12	Self	JSE	Significant effect found
Lobchuck	2018	Canada	44	Nursing staff and students	Perspective-taking training	2.66	Observer Self	CARE CARE (modified)	Mixed. No significant effect found for CARE. Significant effect found

									on modified CARE
Lor	2014	USA	40	Student pharmacists	Perspective-taking training	18	Self	JSE	Significant effect found
LoSasso	2017	USA	70	Medical students	Communication skills training	1	Self	JSE	No significant effect found
Mueller	2001	USA	37	Physical therapy students	Mixed	11	Self	JSE	Significant effect found
Reiss	2012	USA	99	Physicians	Empathy skills training	4	Observer Self	CARE JSE BEES EFDT	Mixed No significant effect found for CARE, JSE, BEES. Significant effect for EFDT
Shapiro	1998	USA	78	Medical students	Mindfulness training	17.5	Self	ECRS	Significant effect found
Sripada	2010	USA	12	Physicians	Psychotherapy intervention	NE	Observer	BLRI	Significant effect found
Sterkenburg	2018	Netherlands	224	Qualified care staff	Serious game	0.33	Self	SQ	Significant effect found
Tulsky	2011	USA	48	Physicians	Communication skills training	NE	Observer	ES EO PE	Significant effect found
Vaghee	2018	Iran	127	Nursing students	Perspective-taking training	3	Self	JSE	Significant effect found
Wolf	1987	Canada	134	Medical students	Communication skills training	12	Self	HRI MCI	Significant effect found
Wundrich	2017	Germany	158	Medical students	Empathy skills training	6	Self Observer	JSE OSCE	Mixed. No significant effect found for JSE. Significant effect found on OSCE scores
Yang	2018	China	177	Nursing students	Narrative medicine intervention	42	Self	JSE	Significant effect found

Study characteristics

Study publication dates ranged from 1987 to 2019, with 15 out of 26 trials published in the last five years.[23,28,30,32,34-38,40,42,47,49,51,52] Thirteen were carried out in the USA and Canada,[29,32-34,40-43,45,46,48,50] seven in Europe,[23,30,36-38,47,51] three in

Iran,[35,39,49] and one each in Australia,[31] Ghana[28] and China.[52] Fourteen studies provided a definition of empathy.[30,32,34-37,40,43-47,51,52]

Study design

Sample size ranged from 12 to 352 participants (median of 90.5; interquartile range (IQR) 49.25-154). Twelve studies had 100 or more participants.[23,28-29,36-38,47,49-51] Seven had fewer than 50 participants.[31,32,40,41,43,46,48] Fifteen studies evaluated empathy interventions for student populations,[28,30,32-35,39,41,42,43,45,49-52] including seven which looked at medical students,[30,34,35,42,45,50,51] five with nursing students,[33,39,40,50,52] two with student pharmacists,[32,41] one with physiotherapy students,[43] and one with a mixed nursing and midwifery student population.[28] Ten trials used professional/qualified populations,[23,29,30,36-38,44,46-48] with four of these focusing on physicians,[31,44,46,48] one on nurses,[36] and five with qualified care staff, including healthcare assistants.[23,29,37,38,47] One study used a mixed student and professional population (nursing students and nurse practitioners).[40]

Five trials used multiple sites,[23,30,36,37,40] and five were cluster RCTs.[23,36,37,49,52] Ten studies defined both inclusion and exclusion criteria for the study.[23,28-29,35,37,39,41,49,52] Thirteen defined inclusion criteria only[30-33,36,38,40,42,43,45-47,50] and in three studies inclusion/exclusion criteria were either not given or were not clear.[34,48,51]

Study interventions

While the aims of eligible trials in this review were to enhance empathy through an educational intervention, a range of intervention types were employed. The most commonly used approach was a communication skills-based training intervention, with eight studies [28,31,33,34,39,42,48,50] using this. Four studies used perspective-taking training,[23,40,41,49] two had a psychotherapy focus,[30,46] three used empathy skills-based training sessions,[35,44,51] two used an arts and humanities approach,[32,52] one used mindfulness-based training,[45] and one a serious gaming intervention.[47] Five studies could not be classified and were described as 'mixed' interventions, using various elements of theoretical knowledge teaching and experiential learning sessions.[29,36-38,43] Seventeen were specifically designed to foster empathy[23,32,34-37,39-44,46-48,50,52] and the remainder used interventions not specifically designed to improve empathy but with the hypothesis that they would. For example, Buffel Du Vaure et al[30] explored the impact of a psychotherapy-focused 'Balint Group' intervention on medical student empathy.

The most frequently used mode of delivery was face-to-face, with eighteen interventions using this.[23,28,30,31,33,35-37,40-42,44,45,46,49-52] Six interventions were delivered online,[29,34,38,39,42,47] one employed a self-directed mode of delivery,[32] and one a CD-ROM to deliver the intervention.[48]

Studies ranged in duration of intervention (total time spent participating in the intervention) from 20 minutes to 42 hours. The mean duration was 10.2 hours (SD 8.8). Five studies did not explicitly state duration.[34,36,38,46,48] Training packages in six studies were considered to be 'short duration', lasting three hours or less;[32,37,39,42,47,49] ten were considered 'medium duration', lasting between four and 12 hours;[23,28-30,35,39,43,44,50,51] and five were considered 'long duration', lasting more than 12

hours.[31,33,41,45,52] Timespan of the interventions ranged from one to 120 days, with a mean length of 38.5 days (SD 40.2).

Outcome measures

Studies used either self-report or other-(objective)report measures to assess a change in participants' empathy. Objective measures included those completed by patients or experts (for example faculty staff or trained actors playing simulated patients). Most studies (18) used only self-report measures.[23,28,29,32,33,35-39,41-43,45,47,48-50,52] Four used objective measures[31,34,46,48] (with only Tulsky et al[48] using patients rather than simulated patients or experts as the outcome assessors). Four used a combination of self- and objective-report tools to measure empathy.[30,40,44,51]

The Jefferson Scale of Empathy (JSE)[53] was the most frequently used self-reported outcome measurement tool, with 13 studies employing it.[23,28,29,32,35,36,39,41-43,49,51,53] Other self-report tools used included the Balanced Emotional Empathy Scale (BEES),[54] the Ekman Facial Decoding test,[55] and the Toronto Empathy Questionnaire (TEQ).[56] The Consultation and Relational Empathy Scale (CARE)[57] was the most frequently used objective measure of empathy, with three studies employing it.[30,40,44] Other objective outcome measures included the Carkhuff Empathy Rating Scale.[58] In addition, some studies developed their own measures of empathy, for example Tulksy et al[48] used a Likert scale with ten items to assess perceived oncologist empathy. Butow et al[31] created a manual to code transcripts of videoed patient interactions to assess empathic behaviour, in addition to using the CARE scale.[57] All studies except

three[29,31,48] employed a validated tool to measure empathy.

Outcome assessment strategy

Timeframes for measuring outcomes varied. Fifteen studies did not specify a timeframe for post-intervention measurements or were unclear.[31-33,35,36-38,40,41,43-48,49-52] For example, Hastings et al[37] reported measuring empathy six-weeks post-randomisation but were not clear how long after the intervention had ended that this measurement was taken. For studies that were explicit, post-intervention measures varied between two days and six months, with the majority of measures taken within two weeks of the intervention.[23,28-30,32,41,48] Eleven studies measured the effects at one or more follow-up points (in addition to the post-intervention measurement),[23,28,29,31,33,35,37,39,41,49,52] which varied between four weeks and 18 months.

Risk of bias within studies

In total, 11 studies[23,28,31,36-39,43,45,47,48] were considered to be at low risk of bias overall (low risk of bias for sequence generation and allocation concealment).[24] Thirteen were considered to be low risk for random sequence generation[23,28,31,35-7,43,47,48] and 11 were low risk for allocation concealment.[23,28,31,6-39,43,44,47,48] Blinding was not possible in the majority of studies due to the nature of the interventions (often described to participants as empathy-promoting) and the method of outcome assessment (for example self-report questionnaires, making explicit what is being measured, such as the

JSE). Full details of the risk of bias assessment are reported in the eResults of the Supplement with eFigure 1 illustrating the overall findings.

Results of individual studies

The majority of studies (19/26) found that the tested intervention significantly improved empathy on at least one outcome measure.[29,30,33-35,38-41,43-52] Seven studies did not find any significant increase in empathy.[23,28,31,32,36,37,42] Of the studies that reported a significant improvement in empathy on at least one outcome measure, 11 were aimed at student populations (representing approximately 73% of student population studies)[30,33-35,41,43,45,49-52] and seven were aimed at professionals (representing 70% of professional population studies).[29,38,39,46,47,48,44] Fifteen studies reported a significant improvement in empathy using a self-rated outcome measure (this represents 68% of studies (15/22) using a self-report outcome tool).[29,30,33,35,38-41,43,45-47,49,50,52] Four studies reported an increase in empathy when using an objective measure (representing 50% (4/8) of studies using an objective outcome measure).[34,44,48,51] Seventeen studies employed an educational intervention that had been specifically designed to foster empathy.[23,32,34-37,39-44,46-48,50,52] Of these, 12 (70%) were successful.[34,5,39-44,46-48,51,52] Four out of five studies that were classed as 'long duration' (lasting >12 hours) reported a significant improvement in empathy post intervention;[33,41,45,52] 50% of 'medium duration' studies (between 3 and 12 hours) reported a significant increase in empathy;[29,35,39,50,51] and 33% of 'short duration'

studies (<3 hours) reported a significant improvement.[47,49]

Synthesis of results

Of the 26 studies included in this review, four were excluded from meta-analysis as they did not provide adequate data from which to calculate the SMD and SD.[31,36,46,51] For the studies that were excluded from the primary analysis, Butow et al[31] reported a positive but not statistically significant effect and Gould et al[36] found no significant difference between control and intervention groups. Wundrich[51] reported no significant influence of the intervention as measured by the JSE (student version) but did report a positive and statistically significant effect on the observer-assessed outcome. Sripada et al[46] also reported a statistically significant positive effect. Of the 22 studies that had adequate data for pooling, all but one (Arthur et al[23]) showed a benefit of intervention. The primary analysis identified that the overall effect of empathy interventions in terms of improving participant empathy was statistically significant (SMD 0.52, 95% CI 0.36 to 0.67) (figure 2). The Q value indicated significant heterogeneity, with p equal to 0.0001 and I^2 equal to 63%. A summary of findings is presented in table 2.

Table 2. Summary of effect sizes for studies included in meta-analyses

	Standardised mean difference (95% confidence interval)	Heterogeneity (I^2)	References
Overall effect of empathy interventions	0.52 (0.36-0.67)	63%	23,8-30,32-35,37-45,47-49,52
Effect of intervention with least risk of bias	0.44 (0.19-0.69)	63%	23,28,37-39,43,44,47,48
Sustainability of effect			
- Follow-up measurement before 12 weeks	0.69 (0.23-1.15)	84%	28,29,35,37,39,49
- Follow-up measurement at 12 weeks or later	0.34 (0.11-0.57)	0%	23,37,41,52
Effect by type of intervention			
- Communication skills training	0.69 (0.32-1.06)	78%	28,33,34,39,42,48,50
- Perspective-taking training	0.60 (0.17-1.04)	55%	23,40,41,49
- Mixed educational programmes	0.39 (0.18-0.61)	0%	29,37,38,43
- Empathy skills training	0.60 (-0.02-1.21)	71%	35,44

- Arts/humanities interventions	0.38 (0.03-0.73)	0%	32,52
Effect by duration of intervention			
- Short (3 hours or less)	0.42 (0.21-0.63)	24%	32,37,40,42,47,49
- Medium (4 to 12 hours)	0.51 (0.21-0.80)	82%	23,28,29,30,35,39,43,44,50
- Long (more than 12 hours)	0.57 (0.32-0.82)	0%	33,41,45,52
Effect by participant population			
- Student population	0.62 (0.38-0.85)	74%	28,30,32-35,39-43,45,49,50,52
- Professional/qualified population	0.33 (0.18-0.47)	0%	23,29,37,38,44,47,48
Effect by outcome assessor			
- Self-assessment	0.52 (0.37-0.68)	58%	23,28-30,32,33,35,37-45,47,49,50,52
- Observer-assessment	0.28 (-0.18-0.75)	81%	30,34,40,44,48

Additional analyses

Sensitivity analysis

For the sensitivity analysis of the least biased studies (table 2), 11 were judged to have low risk of bias for random allocation or allocation concealment[23,28,31,36-39,43,44,47,48] and nine of these provided sufficient data to be included in a meta-analysis (figure 3).[23,28,37-39,43,44,47,48]

Sustainability of improved empathy analysis

Eleven studies provided follow-up data assessing sustainability of changes to empathy, in addition to post-intervention measurement.[23,28,29,31,33,35,37,39,41,49,52] Eight were eligible for inclusion in a sub-group analysis [23,29,35,37,39,41,49,52] (see eResults for further details) which found a moderate effect size for sustainability up to 12 weeks and a smaller, but still significant effect size for sustainability of impact of training at 12 weeks or later (figure 4 and table 2).

Type of intervention analysis

A meta-analysis comparing sub-groups of different types of intervention (eFigure 2 in the Supplement and eResults for further details) found the greatest effect was with empathy

training that was communication skills-based (table 2). The smallest effect reported was for interventions that were described as 'mixed educational programmes' and ones based in the arts and humanities (table 2). It is worth noting however that only two studies used arts and humanities interventions (compared to seven in the communications skills group) and this may well impact on the effect size.

Duration of intervention analysis

Interventions of medium and longer duration (eFigure 3 in the Supplement) were most effective. Interventions of short duration had the smallest effect size (table 2).

Participant population analysis

Studies using healthcare student participant populations appeared to have a larger effect size than those directed at professional/qualified populations (eFigure 4 in the Supplement). Studies included in a sub-analysis of interventions for students showed a moderate effect size of training, compared to a smaller but still significant effect size for training directed at professional/qualified populations (table 2).

Outcome assessor analysis

Studies using a self-assessment outcome scale showed a moderate and significant benefit to empathy for the intervention tested (eFigure 5), compared to a small and statistically not significant effect size for observer-assessed outcome studies (table 2).

Risk of bias across studies

A funnel plot used in the primary meta-analysis (22 studies) did not reveal evidence of publication bias (figure 5). An evaluation of evidence using GRADE software found the quality of evidence was low (eTable 4). This was due to a high or uncertain risk of bias based on random sequence generation and/or allocation concealment in a number of studies and a high degree of heterogeneity across studies.

DISCUSSION

Summary of evidence

Training healthcare practitioners and trainees improved their empathy by a modest amount. The effect of training seemed to diminish, but lasts to beyond 12 weeks.

Comparison with other evidence

Our review supports the evidence of previous similar reviews, finding benefits of empathy training[17,20,21,22] and that practitioner empathy training makes a difference to patients.[59] Our study adds to this evidence by providing an estimate of empathy training from higher quality (randomised) trials, and by showing that the effect lasts well beyond the intervention.

Strengths and limitations

This review, to the best of our knowledge, is the first systematic review and meta-analysis limited to RCTs of clinical empathy training for all healthcare students and professionals. This is an up-to-date review that excludes non-randomised studies, follows a pre-published protocol and assesses both the immediate and longer term effects of empathy training. Our

broad study population, with both healthcare students and professionals means findings are generalisable to all areas of healthcare education and training.

We chose to include only the results of the primary measure of empathy reported by each study. Where it was unclear which was the primary measure, we used the measure that was reported first. We recognise that this might have been biased, as authors may have chosen to report the most positive outcomes first. However, we found that this was not necessarily the case. For example, the first measure of empathy reported by Buffel du Vaure et al [30] (who did not specific which measure was primary) had a smaller effect than the second.

We recognise the heterogeneity of the studies in our review and anticipated this. This means that further research is required to identify the most effective empathy training methodology. Also, the strength of findings in this review may be limited by the reporting quality of some of the included studies. A sensitivity analysis of studies of highest quality found a slightly smaller but still significant effect size. Another limitation in reviewing the evidence in this field is the multiple tools used by investigators to measure clinical empathy. With the lack of a definitive definition of clinical empathy and a range of tools measuring different aspects of empathy, the impact of an intervention may vary depending on the measurement tool used. This is demonstrated by Reiss et al [44] who found a statistically significant improvement in empathy when measured using the CARE scale but no significant changes using the JSE. In contrast Buffel du Vaure [30] reported the opposite. Perhaps because of the larger sample size or other factors, our review found a benefit of training independently of how it was measured. A further limitation with this review is that we only identified four studies that followed participants up for at least three months. The trials

identified however found a positive effect. Lastly, we did not measure the qualitative experiences of participants in this review.

Implications for research and practice

Interventions for cultivating student and trainee empathy should be further developed and implemented. Optimizing implementation will require additional qualitative research on the experiences of empathy teachers and learners. Also, the longer term effects (>12 weeks) of empathy training has not been studied adequately and future research should address this. With competition for time and space in both undergraduate and postgraduate healthcare curriculums, future research in this area needs to be robust. Designers of future trials of empathy training in healthcare can use the results of this review as a guide to their intervention development.

CONCLUSION

Teaching students and other learners how to enhance empathy is moderately effective over a sustained period of time and is likely to benefit present and future patients. Future research should focus on empathy-interventions with patient-led outcome assessment and on assessing effectiveness of training over more sustained periods of time. Medical educators and curriculum designers can use this research to think of ways to integrate empathy training into busy curricula.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in BMJ editions and any other BMJPGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

CONTRIBUTORSHIP STATEMENT

RW had a lead role in the planning, conduct and reporting of the work described in this article. EI had an equal role in the conduct of the work described in this article. NR had an equal role in the conduct of the work described. RN had a supporting role in the planning and reporting of the work described in this article. JH had an equal role in the planning, conduct and reporting of the work described in this article. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

COMPETING INTEREST STATEMENT

All authors have completed the Unified Competing Interest form and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

TRANSPARENCY DECLARATION

This manuscript is an honest, accurate and transparent account of the studies being reported. No important aspects of the review have been omitted.

ETHICAL APPROVAL

Ethical approval was not required.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sections.

DATA AVAILABILITY STATEMENT

This is a systematic review of randomised controlled trials. No original datasets generated or analysed for this study.

FIGURES LEGEND

Figure 1. PRISMA flow diagram

Figure 2. Meta-analysis of eligible studies providing adequate data to calculate standardised mean difference with 95% confidence interval

Figure 3. Meta-analysis of eligible studies, excluding those considered to be at high risk of bias

Figure 4. Meta-analysis of studies that provided follow-up observation points to determine long-term effectiveness of intervention

Figure 5. Funnel plot of effect sizes and standard errors.

REFERENCES

1. Kim SS, Kaplowitz S, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. Eval Health Prof. 2004;27:237-251
2. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther. 2001;26(5):331-342
3. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijen J. Influence of context effects on health outcomes: A systematic review. Lancet. 2001;357(9258):757-62
4. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Reiss H. The influences of the patient-clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomised controlled trials. PLoS One. 2014;9(4):e94207
5. Thomas MR, Dyrbye LN, Huntington JL, Lawson KL, Novotny PJ, Sloan JA, Shanafelt TD. How do distress and well-being relate to medical student empathy? A multicentre study. J Gen Intern Med. 2007;22(2):177-183
6. 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
7. The NHS Constitution for England. Health Education England. Department of Health. 2015. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/480482/NHS_Constituti on_WEB.pdf Accessed 5th February 2019
8. Cummins J and Bennett V. Compassion in Practice: Nursing, Midwifery and Care Staff, our Vision and Strategy. Leeds: NHS Commissioning Board, 2012. Available at:

https://www.england.nhs.uk/wp-content/uploads/2012/12/compassion-in-practice.pdf. Accessed 5th February 2019

9. Francis R. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Summary. London: The Stationery Office, 2013. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/279124/0947.pdf. Accessed 5th February 2019
10. Mercer SW, Reynolds W. Empathy and quality of care. Br J Gen Pract. 2002; 52(Suppl):S9-12
11. Hong H, Han A. A systematic review on empathy measurement tools for care professionals. Educ. Gerentol. 2020;46(2):72-83
12. Colliver JA, Conlee MJ, Verhulst SJ, Dorsey JK. Reports of the decline of empathy during medical education are greatly exaggerated: A re-examination of the research. Acad Med. 2010;85(4):588–593
13. Smith KE, Norman GJ, Decety J. The complexity of empathy during medical school training: Evidence for positive changes. Med Educ, 2017;51(11):1146-1159
14. 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 J Assoc Am Med Coll. 2011;86(8):996-1009
15. 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 J Assoc Am Med Coll. 2009;84(9):1182-1191
16. Nunes P, Williams S, Sa B, Stevenson K. A study of empathy decline in students from five health disciplines during their first year of training. Int J Med Educ. 2011:2;12-17

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

18. Chen D, Lew R, Hershman W, Orlander J. A cross-sectional measurement of medical student empathy. J Gen Intern Med. 2007;22(10):1434-8

19. Batt-Rawden SA, Chisholm MD, Anton B, Flickinger TE. Teaching empathy to medical students: An updated, systematic review. Acad Med. 2013;88(8):1171-7

20. Kiosses VN, Karathanos VT, Tatsioni A. Empathy promoting interventions for health professionals: A systematic review of RCTs. Journal of Compassionate Health Care. 2016;3:7

21. Patel S, Pelletier-Bui A, Smith S, Roberts MB, Kilgannon H, Trzeciak S, Roberts BW. Curricula for empathy and compassion training in medical education: A systematic review. PLoS One. 2019;14(8):e0221412

22. Fragkos KC, Crampton PES. The effectiveness of teaching clinical empathy to medical students: A systematic review and meta-analysis of randomized controlled trials. Acad Med. 2020;95(6)947-957

23. Arthur A, Maben J, Wharrad H, Aldus C, Sarre S, Schneider J et al. Can Healthcare Assistant Training (CHAT) improve the relational care of older people? Study protocol for a pilot cluster randomised controlled trial. Trials. 2015;16:559

24. Higgins JJ and Green S. The Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Chichester. The Cochrane Collaboration, 2011

25. Winter R, Roberts N, Howick J. Assessing the effect of empathy-enhancing interventions in health education and training: A systematic review of randomised

controlled trials. PROSPERO: International prospective register of systematic reviews. 2019

26. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses:* The PRISMA Statement. BMJ 2009;339:b2535
27. Schunemann H, Brozek J and Oxman A. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendation. Hamilton: Group TGW, 2008
28. Alhassan M. Effect of a 2-day communication skills training on nursing and midwifery students' empathy: a randomised controlled trial. BMJ Open 2019;9(3):e023666
29. Blair Irvine A, Billow MB, Eberhage MG, Seeley JR, McMahon E, Bourgeois M. Mental illness training for licensed staff in long-term care. Issues Ment Health Nurs 2012;33(3):181-194
30. Buffel du Vaure C, Lemogne C, Bunge L, Catu-Pinault A, Hoertel N, Ghasarossian C et al. Promoting empathy among medical students: A two-site randomized controlled study. J Psychosom Res. 2017;103:102-107
31. Butow P, Cockburn J, Girgis A, Bowman D, Schofield P, D'Este C et al. Increasing oncologists' skills in eliciting and responding to emotional cues: Evaluation of a communication skills training program. Psycho-oncology. 2008;17(3):209-218.
32. Collins KL, Zweber A, Irwin AN et al. Impact of a fictional reading intervention on empathy development in student pharmacists. Curr Pharm Teach Learn 2017;9(3): 498-503
33. Daniels TG, Denny A, Andrews D. Using microcounseling to teach RN nursing students skills of therapeutic communication. Journal of Nursing Education 1988;27(6):246-252

34. Foster A, Chaudhary N, Kim T, Waller JL, Wong J, Borish M et al. Using Virtual Patients to Teach Empathy: A Randomized Controlled Study to Enhance Medical Students' Empathic Communication. Simulation in Healthcare: The Journal of The Society for Medical Simulation 2016;11(3):181-189

35. Gholamzadeh S, Khastavaneh M, Khademian Z, Ghadakpour S. The effects of empathy skills training on nursing students' empathy and attitudes toward elderly people. BMC Med Educ 2018;18(1):198

36. Gould LJ, Griffiths P, Barker HR, Libberton P, Mesa-Eguiagaray I, Pickering R et al. Compassionate care intervention for hospital nursing teams caring for older people: A pilot cluster randomised controlled trial. BMJ Open 2018;8(2):e018563

37. Hastings RP, Gillespie D, Flynn S, McNamara R, Taylor Z, Knight R et al. Who's challenging who training for staff empathy towards adults with challenging behaviour: cluster randomised controlled trial. JIDR 2018;62(9):798-813

38. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J et al. Web-Based STAR E-Learning Course Increases Empathy and Understanding in Dementia Caregivers: Results from a Randomized Controlled Trial in the Netherlands and the United Kingdom. J Med Internet Res 2015;17(10):e241

39. Larti N, Ashouri E, Aarabi A. The effect of an empathy role-play program for operating room nursing students. J Educ Eval Health Prof 2018;15:29

40. Lobchuk M, Hoplock L, Halas G, West C, Dika C, Schroeder W et al. Heart health whispering: A randomized, controlled pilot study to promote nursing student perspective-taking on carers' health risk behaviors. BMC Nurs 2018;17:21

41. Lor KB, Truong JTT, Ip EJ, Barnett MJ. A randomized prospective study on outcomes of an empathy intervention among second-year student pharmacists. Am J Pharm Educ 2015;79(2):18s

42. LoSasso AA, Lamberton CE, Sammon M, Berg K, Caruso JW, Cass J 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

43. Mueller K, Prins R, de Heer H. An Online Intervention Increases Empathy, Resilience, and Work Engagement Among Physical Therapy Students. Journal of Allied Health 2018; 47(3):196-203

44. Reiss H, Kelley JM, Bailey RW, Dunn EJ, Phillips M. Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum. J Gen Intern Med 2012;27(10):1280-1286

45. Shapiro SL, Schwartz GE, Bonner G. Effects of mindfulness-based stress reduction on medical and premedical students. Journal of Behavioral Medicine 1998;21(6):581-599

46. Sripada BN, Henry DB, Jobe TH, Winer JA, Schoeny ME, Gibbons RD. A randomized controlled trial of a feedback method for improving empathic accuracy in psychotherapy. Psychol Psychother 2011;84(2):113-127

47. Sterkenburg PS, Vacaru VS. The effectiveness of a serious game to enhance empathy for care workers for people with disabilities: A parallel randomized controlled trial. Disability and Health J 2018;11(4):576-582

48. Tulsky JA, Arnold RM, Alexander SC, Olsen MK, Jeffreys AS, Rodriguez KL. Enhancing communication between oncologists and patients with a computer-based training program: A randomized trial. Annals of Internal Medicine 2011;155(9):593-601

49. Vaghee S, Loffabadi MK, Salarhaji A, Vaghei N, Hashemi BM. Comparing the Effects of Contact-Based Education and Acceptance and Commitment-Based Training on Empathy toward Mental Illnesses among Nursing Students. Iranian Journal of Psychiatry 2018;13(2):119-127
50. Wolf FM, Woolliscroft JO, Calhourn JG, Boxer GJ. A controlled experiment in teaching students to respond to patients' emotional concerns. J Med Educ 1987;62(1):25-3
51. 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
52. Yang N, Xiao H, Cao Y, Li S, Yan H, Wang Y. Does narrative medicine education improve nursing students' empathic abilities and academic achievement? A randomised controlled trial. J Int Med Res 2018;46(8):3306-3317
53. Hojat M, Gonnella JS, Nasca TJ, *et al*. Physician empathy: Definition, components, measurement, and relationship to gender and specialty. Am J Psychiatry 2002;159:1563–9
54. Mehrabian A, Epstein N. A measure of emotional empathy. J Pers. 1972;40:525–43.
55. Ekman P. Emotions revealed: Recognizing faces and feelings to improve communication and emotional life. New York: Holt; 2004
56. 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;91:62–71
57. Mercer SW, Maxwell M, Heaney D, Watt GCM. 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
58. Carkhuff RR. Helping and human relations: A primer for lay and professional helpers. Montreal-Toronto: Holt, Rinehart and Winston Inc
59. Howick J, Moscrop A, Mebius A, Fanshawe T, Lewith G, Bishop FL et al. Effects of empathic and positive communication in healthcare consultations: a systematic review and meta-analysis. J R Soc Med. 2018;111(70:240-252

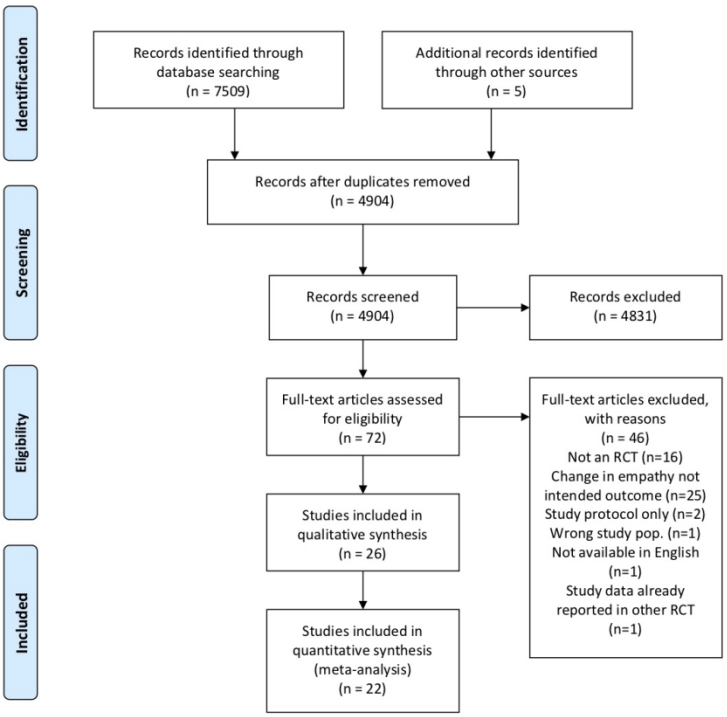


Figure 1. PRISMA flow diagram

209x296mm (150 x 150 DPI)

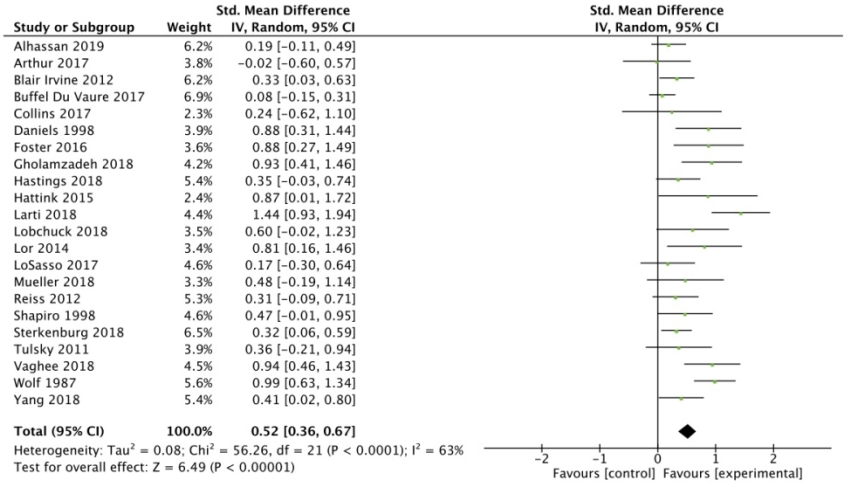


Figure 2. Meta-analysis of eligible studies providing adequate data to calculate standardised mean difference with 95% confidence interval

215x279mm (150 x 150 DPI)

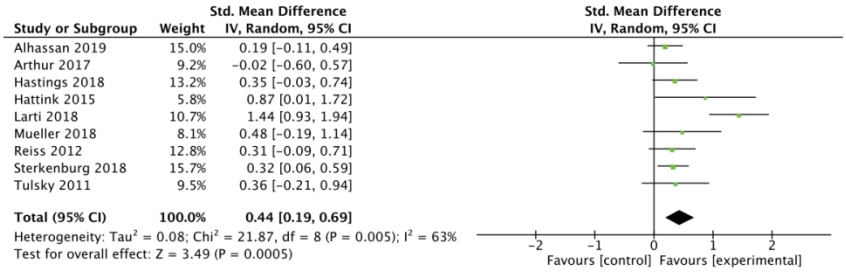


Figure 3. Meta-analysis of eligible studies, excluding those considered to be at high risk of bias

215x279mm (150 x 150 DPI)

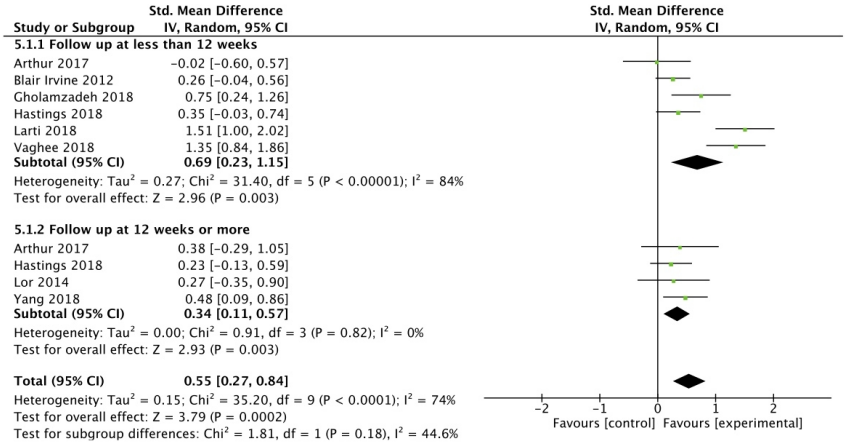


Figure 4. Meta-analysis of studies that provided follow-up observation points to determine long-term effectiveness of intervention

215x279mm (150 x 150 DPI)

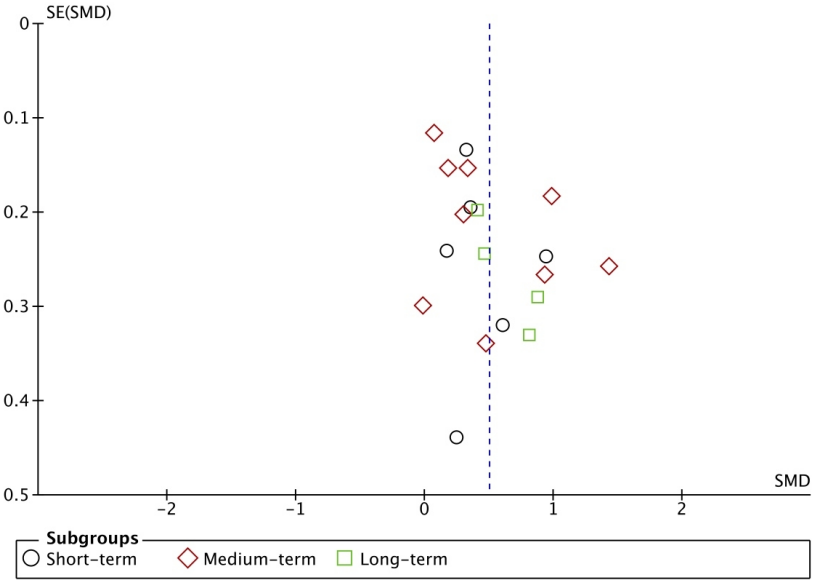


Figure 5. Funnel plot of effect sizes and standard errors

215x279mm (150 x 150 DPI)

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

SUPPLEMENT

Additional methods (eMethods)

Eligibility criteria

Randomised controlled trials (RCTs), including cluster RCTs, which investigated the effect of empathy-enhancing interventions on medical and other healthcare students and professionals' empathy levels as a primary or secondary outcome were eligible for inclusion. We included studies with students and trainees at any level and qualified practitioners from any health profession (including medicine, dentistry, nursing, pharmacy, midwifery and allied healthcare professions). Studies measuring any aspect of 'clinical empathy' were eligible for inclusion. In addition, terminology and measures used in each study were assessed to ensure that outcomes reported under different terms but using the same definitions (for example, reporting on compassion taken to mean empathy) would be captured. Trials measuring empathy via self- and/or observer-reported measures were included.

Risk of bias in individual studies

Risk of bias was assessed using the Cochrane Collaboration's Tool for assessing the risk of bias in clinical trials. This recommends the explicit reporting of each individual element of an RCT: random sequence generation and allocation concealment (selection bias); blinding of participants and blinding of outcome assessment (detection bias); incomplete outcome data (attrition bias); and selective reporting (reporting bias). Using the criteria provided by Higgins (2011)[24], each item was scored as high, low or unclear risk of bias, and evidence

from the study was used to justify each score given. For cluster RCTs, an additional domain was assessed: selective recruitment of cluster participants.

Additional analyses

To assess for sustainability, studies that provided follow-up measurements of the impact of an empathy intervention were grouped into measurements taken before 12 weeks, and at 12 weeks or later. To evaluate the type of intervention most effective at cultivating empathy, we divided interventions into communication skills-based training interventions, perspective-taking interventions, empathy skills-based training, psychotherapy-focused training, arts and humanities-focused interventions, stress management-focused training, serious gaming, and mixed educational programmes. Interventions were categorised based on the descriptions given of the training programmes in each individual study. Where an intervention could not be put into one or other category, it was allocated to the 'mixed educational programme' category. To assess impact of duration on cultivating empathy, interventions were divided on the basis of the length of time participants spent engaging with the intervention.

Data collection

Data was extracted about: general demographics of the study (first author, date published, country of origin, whether empathy is defined); study design (participants and recruitment, inclusion/exclusion criteria, study duration, control conditions); description of the

intervention (setting, duration and frequency); outcome measures (type of measure, whether measure is validated); results (sample size, completeness of outcome data, data that can be used to calculate an effect size); risk of bias and funding source.

Additional results (eResults)

Study selection

The literature search resulted in 7,509 citations. EMBASE included 2,754, PsychINFO 1767, CINAHL 381, MEDLINE 2441 and Cochrane 346. An additional five records were identified through other sources. After duplications were removed 4904 citations remained. 4831 citations were excluded after screening abstracts. Seventy-two articles were retrieved for full-text review. Forty-six studies were excluded (eTable 2). The total number of eligible papers included in this review was 26[23,28-52] (n=2,900). See eTable3 for descriptive characteristics.

Risk of bias within studies

Allocation

Thirteen studies were considered to be low risk for random sequence generation,[23,28,30,35-40,43,47,48] of which seven employed some form of computer randomisation,[28,36,37,38,40,44,47] one used the minimisation method,[48] one used a random numbers table[31] and three used a low-tech method[27,39,43] (for example a shuffled pack of cards). Thirteen trials were considered to have an unclear risk[29,30,32-34,41,42,45,46,49-52] with 12 of these stating that participants were randomly assigned but not describing the method.[29,32-34,41,42,45,46,49-52] One trial used participants from

two different sites, using computer randomisation at one site but not describing the method of randomisation at the other.[30] The risk of bias for allocation concealment was considered low for 11 studies[23,27,31,36-39,43,44,47,48] and was well described in each of these. Fifteen studies did not describe or clearly describe allocation concealment and so were considered unclear in terms of risk.[29,30,32-35,40,41,42,45,46,49-52]

Blinding

Whilst blinding of participants was not possible in the majority of the trials, due to the nature of the interventions, one study did blind participants.[47] This was achieved by using an online package to deliver either a 'serious game'(experimental) intervention or a 'digital reading' (control) intervention. Participants were unaware of which was the control and which was the experimental intervention so were unaware which they were participating in once they had been randomly allocated to one or the other. In two trials it was unclear whether participants had been adequately blinded.[29,34] Similarly, blinding of outcome assessors was not always possible due to the self-reported nature of outcome assessments used by many studies. However three studies reported blinding of outcome assessors [34,47,48] three were unclear if blinding had occurred[29,31,46] and 15 were rated as high risk as no blinding of outcome assessment had occurred.[27,28,32,33,35-39,41,43,45,49,50,52] Five studies reported a 'mixed' picture with blinding of the outcome assessment reported for some outcome measures and not for others.[30,40,42,44,51] For example Reiss et al [44] used the observer rated CARE scale, blinding the assessors to physician randomisation and three non-blinded self-rated scales to measure empathy.

Incomplete outcome data

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Incomplete outcome data was considered to be 'low risk' in 19 studies,[23,29-32,34,35,39-49,52] with attrition rates ranging from 0-16%. The risk was unclear in three studies[32,50,51]and considered high in four.[28,36,37,38]

Selective reporting

Eighteen trials described all pre-specified outcomes as stated in the methodology.[27-32,34,37-42,47,48] One trial presented an 'unclear risk' (Daniels et al[33] described dropping all males from the analysis) and seven studies were high risk for selective reporting.[35,36,45,49-51] Gould et al[36] for example did not report the data associated with the JSE questionnaire which was one of the specified outcomes.

Other potential sources of bias

Five trials were cluster RCTs,[28,36,37,49,52] of which three were considered low risk for recruitment bias[28,36,37] and two were identified as either unclear or high risk.[49,52] Eight studies were identified to be at either a high risk or unclear risk from 'other potential sources of bias.[29,31,33,36,40,46,50,51] For example Butow et al[31] reported differences between the study groups in baseline characteristics and six other studies did not report baseline demographics and/or empathy measurements at baseline.

Sustainability of improved empathy analysis

Eleven studies provided follow-up data assessing sustainability of changes to empathy, in addition to post intervention measurement.[27-29,31,33,35,37,39,41,49,52] Eight were

eligible for inclusion in a sub-group analysis.[28,29,35,37,39,41,49,52] One was excluded from all meta-analyses due to lack of data,[31] one was excluded from this meta-analysis as the empathy-intervention was delivered to the control group prior to the follow-up measures being taken,[23] and one was excluded as the follow-up data was not reported.[33] Studies were divided into two groups; those reporting follow up measures at less than 12 weeks and those reporting follow up at 12 weeks or later (figure 4). Arthur et al[23] and Hastings et al[37] provided multiple follow up data at time points that could be included in both groups (at 8 weeks and 12 weeks, and at 6 weeks and 20 weeks respectively). Meta-analysis found a moderate effect size for improved empathy until 12 weeks (effect size 0.69 95% CI 0.23-1.15) and a small but statistically significant effect size for sustainability at 12 weeks and later (effect size 0.34 95% CI 0.11 to 0.57).

Type of intervention analysis

A meta-analysis comparing sub-groups of different types of intervention (eFigure 2) found the greatest effect was with empathy training that was communication skills-based (effect size 0.69 [95% confidence interval 0.32 to 1.06]). The smallest effect reported was for interventions that were described as 'mixed educational programmes' and ones based in the arts and humanities (effect size 0.39 [95% confidence interval 0.18 to 0.61] and 0.38 [95% confidence interval 0.03 to 0.73] respectively). Interventions labelled as 'empathy skills-based training' had a positive but not statistically significant overall effect (0.60, 95% confidence interval -0.02 to 1.21).

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

eTable 1. Search strategies

MEDLINE		
# ▲	Searches	Results
1	exp Students/	116946
2	student?.ti,ab.	254787
3	(physician? or doctor? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff*).ti.	295930
4	exp Health Personnel/	481003
5	1 or 2 or 3 or 4	906748
6	exp Education/	767285
7	ed.fs.	264737
8	((intervention? or program*) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	137613
9	(train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula).ti.	369134
10	(intervention or program*).ti.	260613
11	6 or 7 or 8 or 9 or 10	1249776
12	5 and 11	335534
13	((physician? or doctor? or surgeon? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff* or practitioner? or student?) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	137434
14	12 or 13	393662
15	Empathy/	17455
16	(empath* or compassion*).ti,ab.	21716
17	15 or 16	31561
18	randomized controlled trial.pt.	481154
19	controlled clinical trial.pt.	93050
20	randomized.ab.	441413
21	placebo.ab.	197236
22	drug therapy.fs.	2104120
23	randomly.ab.	309893

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

24	trial.ab.	461528
25	groups.ab.	1906393
26	multicenter study.pt.	249476
27	pragmatic clinical trial.pt.	1037
28	(multicenter or multi center or multicentre or multi centre).ti.	47574
29	(intervention? or effect? or impact? or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab.	8937416
30	18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29	11030368
31	14 and 17 and 30	2441

EMBASE		
# ▲	Searches	Results
1	*student/ or exp *health student/	68463
2	student?.ti,ab.	326421
3	exp *health care personnel/	479224
4	(physician? or doctor? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff*).ti.	301997
5	1 or 2 or 3 or 4	941482
6	education/ or continuing education/ or curriculum/ or education program/ or in service training/ or lifelong learning/ or exp medical education/ or exp paramedical education/ or postgraduate education/	736812
7	((intervention? or program*) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	184005
8	(train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula).ti.	399259
9	(intervention or program*).ti.	318923
10	6 or 7 or 8 or 9	1266300
11	5 and 10	281380
12	((physician? or doctor? or surgeon? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff* or practitioner? or student?) adj5 (train* or educat* or course? or workshop? or staff development or professional	179470

	development or curriculum or curricula)).ti,ab.	
13	11 or 12	369015
14	Empathy/	23785
15	(empath* or compassion*).ti,ab.	28390
16	14 or 15	39458
17	13 and 16	4903
18	randomized controlled trial/	545326
19	single blind procedure/ or double blind procedure/	192596
20	crossover procedure/	58851
21	random*.tw.	1400168
22	(((singl* or doubl*) adj (blind* or mask*)) or crossover or cross over or factorial* or latin square or assign* or allocat* or volunteer*).ti,ab.	983905
23	pragmatic trial/ or multicenter study/	213866
24	intervention study/	40085
25	(multicenter or multi center or multicentre or multi centre).ti.	74011
26	(intervention? or effect? or impact? or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab.	11312699
27	18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26	12032330
28	(exp animals/ or nonhuman/) not human/	6212385
29	27 not 28	9294426
30	17 and 29	2574

PsychINFO		
# ▲	Searches	Results
1	students/ or medical students/	35317
2	student?.ti,ab.	481295
3	exp health personnel/	128154
4	(physician? or doctor? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff*).ti.	47232
5	1 or 2 or 3 or 4	616902
6	education/ or exp curriculum/ or distance education/ or nursing education/ or paraprofessional education/ or exp personnel training/ or exp medical education/	186066

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

7	((intervention? or program*) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	100952
8	(train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula).ti.	207043
9	(intervention or program*).ti.	121597
10	6 or 7 or 8 or 9	455304
11	5 and 10	166574
12	((physician? or doctor? or surgeon? or intern? or internship or resident? or residency or nurse? or health* professional? or health* worker? or health* staff* or practitioner? or student?) adj5 (train* or educat* or course? or workshop? or staff development or professional development or curriculum or curricula)).ti,ab.	98357
13	11 or 12	209818
14	Empathy/	12489
15	(empath* or compassion*).ti,ab.	37254
16	14 or 15	38291
17	13 and 16	3043
18	random*.ti,ab,hw,id.	187448
19	trial*.ti,ab,hw,id.	172104
20	controlled stud*.ti,ab,hw,id.	11726
21	placebo*.ti,ab,hw,id.	38934
22	((singl* or doubl* or trebl* or tripl*) and (blind* or mask*)).ti,ab,hw,id.	27892
23	(cross over or crossover or factorial* or latin square).ti,ab,hw,id.	28819
24	(assign* or allocat* or volunteer*).ti,ab,hw,id.	156473
25	treatment effectiveness evaluation/	22860
26	mental health program evaluation/	2062
27	exp experimental design/	54976
28	(clinical trial or treatment outcome).md.	41809
29	intervention/	58790
30	(multicenter or multi center or multicentre or multi centre).ti.	2788
31	(intervention? or effect? or impact? or controlled or control group? or (before adj5 after) or (pre adj5 post) or ((pretest or pre test) and (posttest or post test)) or quasiexperiment* or quasi experiment* or evaluat* or time series or time point? or repeated measur*).ti,ab.	1834258

32	18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31	2026090
33	17 and 32	1767

CINAHL		
#	Query	Results
S17	S13 AND S16	381
S16	S14 NOT S15	556,315
S15	(MH animals+ OR MH (animal studies) OR TI (animal model*)) NOT MH (human)	154,114
S14	MH randomized controlled trials OR MH double-blind studies OR MH single-blind studies OR MH random assignment OR MH pretest-posttest design OR MH cluster sample OR TI (randomised OR randomized) OR AB (random*) OR TI (trial) OR (MH (sample size) AND AB (assigned OR allocated OR control)) OR MH (placebos) OR PT (randomized controlled trial) OR AB (control W5 group) OR MH (crossover design) OR MH (comparative studies) OR AB (cluster W3 RCT)	579,579
S13	S9 AND S12	2,335
S12	S10 OR S11	17,823
S11	TI (empath* or compassion*) OR AB (empath* or compassion*)	13,814
S10	(MH "Empathy")	8,360
S9	S7 OR S8	188,626
S8	TI ((physician? or doctor? or intern? or internship or resident? or residency or nurse? or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*") N5 (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula)) OR AB ((physician? or doctor? or intern? or internship or resident? or residency or nurse? or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*") N5 (train* or	55,142

	educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula))	
S7	(S3 AND S6)	158,577
S6	S4 OR S5	550,634
S5	TI (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula) OR AB (((intervention? or program*) N5 (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula))) OR TI(intervention? or program*)	349,186
S4	(MH "Curriculum+") OR (MH "Education, Clinical+") OR (MH "Education, Health Sciences+") OR (MH "Staff Development") OR (MH "Education")	294,559
S3	S1 OR S2	663,254
S2	TI student? OR AB student? OR TI (physician? or doctor? or intern? or internship or resident? or residency or nurse? or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*")	226,699
S1	(MH "Students, Health Occupations+") OR (MH "Health Personnel+")	529,459

COCHRANE	
ID	Search
#1	MeSH descriptor: [Students] explode all trees
#2	(student*):ti,ab,kw
#3	MeSH descriptor: [Health Personnel] explode all trees
#4	(physician* or doctor* or intern or interns or internship or resident* or residency or nurse* or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*"):ti
#5	#1 or #2 or #3 or #4
#6	MeSH descriptor: [Education] explode all trees
#7	(train* or educat* or course* or workshop* or "staff development" or "professional development" or curriculum or curricula):ti OR (intervention* or program*):ti OR (((intervention8 or program*) N5 (train* or educat* or course* or workshop* or "staff development" or "professional development" or curriculum or curricula))):ti,ab,kw

#8	#6 or #7
#9	#5 and #8
#10	((physician* or doctor* or intern or interns or internship or resident* or residency or nurse* or "health professional*" or "health worker*" or "health staff*" or "healthcare professional*" or "healthcare worker*" or "healthcare staff*" or "health care professional*" or "health care worker*" or "health care professional*") NEAR/5 (train* or educat* or course? or workshop? or "staff development" or "professional development" or curriculum or curricula)):ti,ab,kw
#11	#9 or #10
#12	MeSH descriptor: [Empathy] explode all trees
#13	(empath* or compassion*):ti,ab,kw
#14	#11 and #13

eTable 2. Characteristics of excluded studies

Study	Reason for exclusions
Arthur 2015	Study protocol.
Bonvicini 2008	Observational data taken from an RCT. Intervention not specifically designed with outcome of change in empathy. Secondary analysis of data to see if there is an impact on empathy.
Bosse 2012	Change in empathy not a specified outcome of study
Bruera 2007	Change in empathy not measured or intended outcome.
Chen 2016	Not an RCT. Quasi-experimental design, not randomised.
Chunharas 2013	Not an RCT
Daeppen 2012	Change in empathy is not an intended outcome
Danucalov 2017	Empathy is not an intended outcome of the study. Participants not healthcare students or professionals.
Delvaux 2005	Change in empathy not an intended outcome and not measured
Downar 2016	Change in empathy not an intended outcome
Downar 2017	Change in empathy is not an intended outcome of the study.
Dundas 2017	Participants are not healthcare students/professionals.
Fallowfield 2002	Empathy is not directly measured
Fine 1977	Not an RCT
Gibon 2013	Change in empathy not an intended outcome
Gorniewicz 2016	Change in empathy not an intended outcome and is not measured
Hojat 2013	Not an RCT. Experimental control groups without randomisation.
Jaury 2018	Analysis of data already reported in RCT
Johnson 2013	Not an RCT. Controls selected from a waitlist group and intervention participants from a group who were due to undergo training in a set time-period.
Kahriman 2016	Change in empathy is not intended outcome
Klein 1999	Change in empathy is not measured
Liao 2016	Not an RCT. Quasi-experimental design
Lienard 2010	Change in empathy not an intended outcome
Lim 2011	Change in empathy not an intended outcome
Little 2015	Change in empathy not intended outcome of study and not specifically measured
Misra-Herbert 2012	Not an RCT
Nasr Esfahani 2014	No control arm, comparison between wo groups receiving same training, one as distant learning, one as attendants on course.
Nixon 2018	Not an RCT. Quasi-experimental design "partial randomisation was conducted" with participants designated to their preference group
Oz 2001	Not an RCT.
Perula de Torres 2019	Study protocol only

Potash 2014	No control arm "mixed-methods quantitative-qualitative study"
Rask 2009	Empathy not measured as an outcome
Razavi 2002	Change in empathy is not an intended outcome
Razavi 2003	Empathy not explicitly measured as an outcome
Rosenzweig 2016	Not an RCT
Roter 1995	Unclear whether intervention is looking to cultivate empathy and whether change in empathy is an intended outcome
Schroeder 2018	Change in empathy is not an intended outcome of the study
Shapiro 2004	Not an RCT
Shapiro 2009	Not an RCT
Shapiro 2011	Change in empathy is not an intended outcome
Smith 1995	Change in empathy is not intended outcome
Tamuma 2017	Only available in Japanese
Van Dijk 2017	Change in empathy is not an intended aim of the study
Van Vilet 2017	Not an RCT. Exploratory, controlled, quasi-experimental study using students not on a specific course as control group
Weatherdale 2018	Correspondence and not research study
West 2014	Change in empathy is not an intended outcome.

eTable 3. Characteristics of included studies

Alhassan 2019

Methods	Randomised controlled trial
Participants	The country of origin was Ghana. 104 students were randomised to the intervention group and 106 to the control group. The inclusion criteria were nursing and midwifery students in their second year of training, above age 18 and available for follow-up data collection after 6 months. The exclusion criteria included students not studying at Tamale Nursing and Midwifery College
Interventions	Communication Skills Training (CST) developed by author (MA) using 'Four Habits Model' and 'PCNF' (person-centred nursing framework). The mode of delivery were small group discussions, brainstorming, personal experience from participants, group reports, roleplaying, questions and answers, videos and summaries. The duration was 2 days and frequency was one off.
Outcomes	The outcome was empathy measured with JSE HPS version Outcome assessment 2 days post intervention and 6 months post intervention
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"NMS were separated before random assignment to ensure that both professions were approximately equally represented in the groups" "The researcher (MA) and research assistants conducted this by allowing participants to pick numbers written on papers, which had been randomly shuffled in a box."
Allocation concealment (selection bias)	Low risk	"There was allocation concealment to the researcher, research assistants and the participants. The researcher (MA) and research assistants conducted this by allowing participants to pick numbers written on papers, which had been randomly shuffled in a box."
Blinding of participants and personnel (performance bias)	High risk	"The participants were made aware of empathy being an outcome of this study and since JSE is self-reported, it may have impacted their self-report."

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Blinding of outcome assessment (detection bias)	High risk	"The participants were made aware of empathy being an outcome of this study and since JES is self-reported, it may have impacted their self-report." "The data was analysed by the author (MA) without blinding."
Incomplete outcome data (attrition bias)	Low risk	11 participants in intervention group and 26 in control were excluded from analysis due to incomplete data or outcome measures not returned.
Selective reporting (reporting bias)	Low risk	Outcomes reported as pre-determined
Other bias	Low risk	No other bias detected

Arthur 2017

Methods	Pilot cluster randomised controlled trial
Participants	The country of origin was UK. Clusters were wards within three acute hospital trusts in England. General medical, stroke or care of the elderly/older people wards were eligible. Specialist dementia wards and medical admissions units were excluded. Health Care Assistants (HCAs) employed full or part time within enrolled wards were eligible to enter trial. Bank staff and not part of the named staff on ward roster were ineligible. In total 59 Health Care Assistants were randomised to the intervention group and 53 to the control group.
Interventions	'Older People's Shoes' training intervention that focuses on relational care of older people. The mode of delivery was small group teaching led by nurses who had received full training in content and delivery of the intervention from a member of the research team. The setting was the hospital, the duration of the intervention was 2 weeks and frequency was 1 half day session for 2 consecutive days followed by a weeks break and then repeated.
Outcomes	HCA outcomes were empathy, as measured by The Toronto Empathy Questionnaire (TEQ) at baseline and post intervention at 8 and 12 weeks post randomisation.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Stratified by NHS hospital trust, wards were randomly allocated by the Norwich Clinical Trials Unit. Each ward had an equal chance of receiving either Older People's Shoes training for HCAs or TAU. Random allocation was generated via computer-written code using block sizes of four"
Allocation concealment (selection bias)	Low risk	"To conceal allocation from those responsible for recruitment, randomisation took place immediately after baseline measures were completed and 4 weeks ahead of the start of the intervention (set-up period) to allow appropriate arrangements, including HCA staffing cover to be arranged."
Blinding of participants and personnel (performance bias)	High risk	"At a number of ward-based meetings during the 4-week baseline period, HCAs were given information about the study"
Blinding of outcome assessment (detection bias)	High risk	Not described. Outcome measure is self-reported
Incomplete outcome data (attrition bias)	High risk	"For HCAs, completion of questionnaires was 72 out of 112 (64.2%) at baseline, 52 out of 112 (46.4%) at the first follow-up and 40 out of 112 (35.7%) at the second follow-up."
Selective reporting (reporting bias)	Low risk	Outcomes are reported as per methodology
Other bias	Low risk	Recruitment bias considered to be low risk: "Each ward had an equal chance of receiving either Older People's Shoes training for HCAs or TAU".

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Blair Irvine 2012

Methods	Randomised controlled trial
Participants	The country of origin was the USA. 84 healthcare professionals were randomised to the intervention group and 88 to the control group. Eligibility criteria included: identification of professional license from a pre-determined list, working in nursing home and assisted living settings Exclusion criteria included: Working as Certified Nursing Assistant, Nursing Assistant, and Home Health Aide, working in a psychiatric/Alzheimer's care units and hospitals, working less than 20 hours per week, a 'moderate' or 'a lot' of self-reported level of mental illness, 'extremely confident' self-reported confidence to deal with resident behaviours associated with mental illness
Interventions	Online training designed to develop skills and confidence to deal with symptoms of whatever mental illness was causing a particular behaviour. The mental illness training approach included video modelling vignettes, right-way and wrong-way exemplars, testimonials and narration supplemented by short on-screen text designed to create empathy for residents with mental illness. A minimum 'viewing time' for all online courses was 4 hours with two online 'visits' one week apart.
Outcomes	Video situational testing (VST) was used to assess participant reactions to short video vignettes of resident behaviour. Four items in VST were used to assess participant empathy towards a resident.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No detail given on how randomisation occurred
Allocation concealment (selection bias)	Unclear risk	No detail given on allocation of participant
Blinding of participants and personnel (performance bias)	High risk	“After submitting the baseline assessment, treatment participants were e-mailed login information to the Internet training program for Visit 1. One week after logging on to the Visit 1 courses, each participant was sent a second e-mail with log-in information for Visit 2.”
Blinding of outcome assessment (detection bias)	Unclear risk	No detail given on how/who assessed video situational vignettes and whether outcome assessors were blinded
Incomplete outcome data (attrition bias)	Low risk	"Of the 172 study participants 91% completed all three assessment surveys, 6% completed two surveys, and 3% completed one survey Participants who completed all three surveys were compared to those who completed one or two surveys on study condition, demographic characteristics, and all baseline outcome measures. Attrition was not significantly related to any of the measures, which suggests that dropping out of the study did not bias results."
Selective reporting (reporting bias)	Low risk	Outcomes reported as stated in methodology
Other bias	Unclear risk	“our measures of empathy and stigma did not provide an in-depth assessment of these constructs, nor is it clear what elements of the training were influential”

Buffel Du Vaure 2017

Methods	Two site parallel group randomised controlled trial

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Participants	The country of origin was France 176 fourth year medical students were randomised to the intervention group and 176 to the control group from two medical schools. No exclusion criteria were stated.
Interventions	Balint group training was the intervention with control conditions as 'teaching as usual'. The intervention was delivered in small group discussions held at the university. The duration of the intervention was 10.5 hours delivered in 1.5-hour weekly sessions over 7 weeks.
Outcomes	Empathy was assessed using the observer-rated CARE scale post intervention and JSPE student version self-rated scale pre and post intervention.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Students from Paris Diderot were randomized with a simple randomization using computer generated random numbers" "students from Paris Descartes, we took advantage of the randomization routinely performed each year by university staff to allocate each student to one of three groups, each corresponding to a particular order of the three mandatory 3-month programs of the fourth-year curriculum"
Allocation concealment (selection bias)	Unclear risk	"students from Paris Descartes, we took advantage of the randomization routinely performed each year by university staff to allocate each student to one of three groups, each corresponding to a particular order of the three mandatory 3-month programs of the fourth-year curriculum"
Blinding of participants and personnel (performance bias)	High risk	"Participants in the intervention group received a training of 7 sessions of 1.5 hour Balint groups, over 3 months"
Blinding of outcome assessment (detection bias)	Unclear risk	Outcome assessed both by observer and self. "Whereas students and facilitators were aware of the allocated group, standardized patients, OSCE's observers and data analysts were kept blinded to the allocation". Self-assessment for JSPE so unable to blind outcome assessors (students themselves)
Incomplete outcome data (attrition bias)	Low risk	52 lost to follow up but study over recruited to ensure significance level of 5% and power of 80%. 14.7% attrition (21 intervention and 32 controls)
Selective reporting (reporting bias)	Low risk	Primary and secondary outcomes reported as stated in the methods
Other bias	Low risk	No other bias detected

Butow 2007

Methods	Randomised controlled trial
Participants	The country of origin was Australia. 16 medical and radiation oncologists were randomised to the intervention group and 14 to the control group. All medical and radiation oncologists from six tertiary care hospitals in six Australian cities which incorporated oncology outpatient clinics were invited to participate in the study No exclusion criteria stated
Interventions	Communication skills training was an intensive face-to-face workshop incorporating presentation of principles, a DVD modelling ideal behaviour and role-play practice, followed by four 1.5 hour monthly video-conferences incorporating role-play of doctor-generated scenarios.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Outcomes	The outcome was a change in doctor behaviour in eliciting and responding to emotional cues in patients and was measured via coding of a transcript from a filmed role-play at baseline, after completing the training and at 12 months post intervention.
Notes	No funding source stated

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"oncologists individually randomised immediately after giving consent and baseline data collection, to receive the training or not. Oncologists were stratified by hospital to ensure approximately equal numbers in the control and intervention arms within each institution, and then randomised within permuted blocks of size 6 constructed by the central research team using a random number table"
Allocation concealment (selection bias)	Low risk	"oncologists individually randomised immediately after giving consent and baseline data collection, to receive the training or not."
Blinding of participants and personnel (performance bias)	High risk	"Control group doctors were offered training at the completion of the study." "It is possible that intervention doctors shared some study materials with control doctors although they were strictly instructed not to do so" "all doctors were aware that they were being assessed, which likely motivated them to be on 'their best behaviour"
Blinding of outcome assessment (detection bias)	Unclear risk	Does not state whether assessors were blinded
Incomplete outcome data (attrition bias)	Unclear risk	Two controls and two intervention participants lost to follow-up. 11.4% overall attrition
Selective reporting (reporting bias)	Low risk	Outcomes reported as stated in methodology
Other bias	High risk	Baseline imbalance: "EE and DP scores were significantly higher in the intervention group compared to the control group at baseline".

Collins 2017

Methods	Randomised controlled trial
Participants	The country of origin was USA 13 student pharmacists were randomised to the intervention group and 12 to the control group. First through to third year pharmacist students invited to participate. No exclusion criteria stated
Interventions	Students randomized to the literature intervention group were then sent a weekly email that included the reading assignment. Reading assignments were divided into three segments (approximately three to five minutes apiece), and students were requested to complete the readings in three separate sittings throughout the week. The intervention duration was 8 weeks with weekly sessions.
Outcomes	A change in empathy was measured using the JSE-HPS two weeks post end of the intervention.
Notes	-

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Participants were randomized into either an intervention or control group." No detail of how randomisation occurred
Allocation concealment (selection bias)	Unclear risk	No details given
Blinding of participants and personnel (performance bias)	High risk	"The announcement was then followed by an email further explaining the study and inviting students to participate."
Blinding of outcome assessment (detection bias)	Unclear risk	No details given. However, outcome assessment is self-assessed by participants and participants not blinded.
Incomplete outcome data (attrition bias)	Low risk	Overall attrition rate 16%. (15.4% for intervention group, 16.7% for control group dropout rate)
Selective reporting (reporting bias)	Low risk	Outcomes reported as stated in results
Other bias	Low risk	No other bias detected

Daniels 1998

Methods	Randomised controlled trial
Participants	The country of origin was Canada 53 full-time second year nursing students were randomly allocated to either the intervention or control group. Full-time second year female students in a two-year, eight-month registered nurse (RN) diploma program. Males not excluded from study randomisation but were excluded from analysis.
Interventions	Micro-counselling training divided into six segments with one micro-skill taught per segment including attending behaviour, questioning, minimal encouragers, paraphrasing, reflection of feeling and summarizing. The intervention was delivered face-to-face and training was divided into 6 segments of 3-5 hours with a minimum of 18 hours training.
Outcomes	The Empathy Construct Rating Scale and The Carkhuff Index of Communication (Empathy) self-rated scales were administered to assess changes in empathy post intervention.
Notes	No details on funding source given.

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Subjects were randomly assigned to either an experimental group or a non-attention control group." No details of how random sequence generated
Allocation concealment (selection bias)	Unclear risk	"Subjects were randomly assigned to either an experimental group or a non-attention control group." No details on allocation of students to experimental/control
Blinding of participants and personnel (performance bias)	High risk	"During the period of micro-counselling training of the experimental subjects, the control subjects were non-attended. Essentially, the control subjects spent this period of time entirely on their own and received no supervision or structured training experience of any kind."

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Blinding of outcome assessment (detection bias)	High risk	No details given of blinding outcome assessors however outcome assessment is self-assessment
Incomplete outcome data (attrition bias)	Unclear risk	"The sample consists of all full-time second year female students (n=60). In all, there are 56 females and 4 males. The males were dropped from the analysis and there was a further attrition of three subjects."
Selective reporting (reporting bias)	High risk	The males were dropped from the analysis and there was a further attrition of three subjects
Other bias	Unclear risk	No results tables/figures published for the 9-month follow-up data ("At the nine-month follow-up period, the experimental group performed better on all the dependent measures than the control group. However, these differences failed to reach statistical significance")

Foster 2016

Methods	Randomised controlled trial
Participants	The country of origin was USA. 35 and 18 medical students were allocated to 2 intervention arms and 17 to a control arm.
Interventions	Student engagement with a virtual patient (VP). Students interacted with VP online test-based interface. They conducted interviews as they would with live patients, but typed what they wanted to say rather than speaking. The three arms to the study consisted of: -The empathy-feedback VP: Human-assisted empathy feedback is a technique where human 'assessors' anonymously follow online the trainee's interaction with the VP in real time. The assessors' feedback about opportunities to express empathy was available to students for review at the end of the VP interaction -The Backstory VP: Combines embodied conversational agents and narrative video vignettes. When specific questions are asked of the VP, noninteractive video vignettes are presented which show scenes of the VP illustrating their condition. -Control VP: Provides typed interaction with VP without empathy feedback or patient backstory.
Outcomes	The primary outcome was to assess students' verbal responses to all the opportunities to show empathy presented to them by the simulated patients. The Empathic Communication Encoding System (ECCS) (developed to code empathic opportunities, defined as an explicit, clear and direct statement of emotion, progress or challenge by the patient) was used to assess empathy.
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	High risk	"Students were randomized into one of three groups." No detail on random sequence generation given.
Allocation concealment (selection bias)	Unclear risk	No detail on allocation given.
Blinding of participants and personnel (performance bias)	Unclear risk	"The (VP) assessors were not aware of the students' identity or study group assignment and could not see the students, and the students were not aware of the assessors' presence"
Blinding of outcome assessment (detection bias)	Low risk	"The (VP) assessors were not aware of the students' identity or study group assignment and could not see the students, and the students were not aware of the assessors' presence." "Measures were taken to label the transcripts (of SP interactions) in each study group such that the source of the transcript was not identifiable to the assessors"

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

		"The SPs (standardised patients) were blinded to students' study group assignment."
Incomplete outcome data (attrition bias)	Low risk	No attrition reported. N=70 randomised and n=70 analysed
Selective reporting (reporting bias)	Unclear risk	Study outcomes reported as stated in methodology
Other bias	Low risk	No other bias detected

Gholamzadeh 2018

Methods	Quasi-experimental randomised controlled design
Participants	The country of origin was Iran 63 third and fourth year medical students were allocated to either the control or intervention group. The inclusion criteria of the study were willingness to participate, being a third- or fourth-year nursing student, and not having taken any empathy courses in the past 6 months. In case the students were unwilling to continue participation in the study or were participating in another educational program at the same time, they were excluded.
Interventions	Workshop on empathy skills including self-awareness, and definition and examples of empathy towards patients. The intervention consisted of an 8-hour workshop on empathy skills that was held at the college for 2 days. The content of the workshop was designed by the researchers and reviewed and revised by some of the college professors. The workshop was mainly based on constructivist learning theory.
Outcomes	The JSE-HP self-rating scale was used to examine the effects of empathy skills training immediately and 2 months after the intervention.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"the 70 students were randomly divided into a control and an intervention group through block randomization."
Allocation concealment (selection bias)	Unclear risk	"the 70 students were randomly divided into a control and an intervention group through block randomization." No details of allocation to groups post randomisation.
Blinding of participants and personnel (performance bias)	High risk	"All students in the intervention group participated in the same workshop. The students were informed about the date of the workshop in advance."
Blinding of outcome assessment (detection bias)	High risk	Self-rated questionnaire (outcome assessor is participant)
Incomplete outcome data (attrition bias)	Low risk	All participants randomised completed the study
Selective reporting (reporting bias)	High risk	Outcomes not specifically stated in methodology.
Other bias	Low risk	No other bias detected

Gould 2017

Methods	Multi-site pilot randomised controlled trial (as part of a wider feasibility study)
Participants	Six ward teams were randomised to either intervention or control groups with a total of 168 nursing staff randomised to the intervention group and 81 to the control group. Medical and surgical wards with high proportion of older patients were eligible.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Interventions	The Creating Learning Environments for Compassionate Care (CLECC): educational programme focused on developing manager and team practices at a group level that create an expansive learning environment, theorised to enhance team capacity to provide compassionate care
Outcomes	Nurses' self-reported empathy was measured using the Jefferson Scale of Empathy (JSE) (Physician/HP version).
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomisation of clusters was undertaken using the ralloc command in Stata (Release 12, StataCorp) by the team statistician (IM-E) blinded to hospital and ward information other than ward specialty."
Allocation concealment (selection bias)	Low risk	"Procedures for allocation concealment and blinding proceeded as planned, with the exception of two researcher observers at follow-up reporting that they learnt of ward allocation from ward staff."
Blinding of participants and personnel (performance bias)	High risk	"It was not possible to conceal allocation from ward team nursing staff. Patients were not informed of allocation."
Blinding of outcome assessment (detection bias)	High risk	Empathy measurement is self-rated questionnaire so unable to blind outcome assessor Researchers gathering questionnaire data were aware of ward allocation.
Incomplete outcome data (attrition bias)	High risk	No attrition of wards during the study
Selective reporting (reporting bias)	High risk	No data reported on JSE other than: "There was no significant difference between groups (P=0.800)"
Other bias	Unclear risk	Baseline demographic and baseline measurement difference not fully reported for JSE. Recruitment bias low risk: Six wards in two NHS hospital Trusts in England were enrolled and allocated to intervention (n=4) or control (n=2). The number of clusters was determined by funding availability and the plan to run the study in at least two hospital organisations, and at least two ward specialties. Randomisation of clusters was undertaken using the ralloc command in Stata (Release 12, StataCorp) by the team statistician (IM-E) blinded to hospital and ward information other than ward specialty.

Hastings 2018

Methods	Cluster randomised controlled trail
Participants	118 residential care settings for people with intellectual disability (with a total of 236 staff) were randomised to either the intervention or control group. Residential settings were eligible for inclusion if: they were based in a community setting, provided services via publicly funded contracts, supported between one and 10 people with ID, employed staff who provided at least some 24-h support, provided care for at least one person with ID who displayed aggressive CB, could identify one manager/lead staff member and one other support staff member who could attend WCW training together. Staff were eligible for inclusion if: they were either a manager (or lead staff member as defined by the service provider organisation) or a direct support worker whose roles were no more than 50% administrative/management. Staff who worked less than 70% of full-time equivalent were also ineligible.
Interventions	WCW (Who's challenging who) training course for support staff in ID context covering communication, frustrations of people with CB (challenging behaviours), experience of

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

	being physically restrained, medication, feeling excluded and unhelpful attitudes and behaviour or support staff). The intervention was delivered in small group facilitated learning sessions by trained trainers. It was delivered in a one off half day session.
Outcomes	The Staff Empathy for people with Challenging Behaviour Questionnaire (SECBQ) was used to measure staff self-reported empathy at baseline and at 6 weeks and 20 weeks post randomisation.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomisation occurred at one point in time for each phase, was carried out by a study-independent statistician from the Centre for Trials Research and used a dynamic balancing algorithm specifically designed for cluster randomised trials"
Allocation concealment (selection bias)	Low risk	The trial statistician remained blind to allocation up until the point of data analysis.
Blinding of participants and personnel (performance bias)	High risk	"Settings, and staff members within them, could not be masked to the intervention but were recruited prior to randomisation."
Blinding of outcome assessment (detection bias)	High risk	Self-reported outcomes to measure empathy
Incomplete outcome data (attrition bias)	High risk	Intervention group: 77% received intervention 6 week follow up 44.1% 20 week follow up 48.3%
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	Recruitment bias low: Randomisation occurred at one point in time for each phase, was carried out by a study-independent statistician from the Centre for Trials Research and used a dynamic balancing algorithm specifically designed for cluster randomised trials No evidence that further residential settings were added to the trial following randomisation.

Hattink 2015

Methods	Randomised controlled trial
Participants	The countries of origin were UK and the Netherlands. 142 care givers (informal or professional) were randomised to the intervention or control group. 24 were professional care givers. Participants who fulfilled the following criteria were recruited for the evaluation study: (1) were sufficiently computer literate to utilize the STAR website and (2) were currently an informal caregiver for someone with dementia living in the community, or a volunteer working with people with dementia with direct contact with community-dwelling people with dementia, or a professional caregiver for people with dementia with direct contact with community-dwelling people with dementia.
Interventions	STAR training portal, a Web-based portal consisting of 8 modules, 2 of which had a basic level and 6 additional modules at intermediate and advanced levels about dementia care. In addition, users had access to online peer and expert communities for support and information exchange. Up to 4 months to complete on-line training modules at participants own pace.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Outcomes	The Interpersonal Reactivity Index (IRI) was used to measure empathy pre and post intervention (empathy was measured as a secondary outcome) with changes to knowledge about dementia and attitudes to it being primary outcomes.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomization software was used to classify participants into either the experimental or control group."
Allocation concealment (selection bias)	Low risk	"Randomization software was used to classify participants into either the experimental or control group"
Blinding of participants and personnel (performance bias)	High risk	"Participants in the experimental group received a link to the STAR registration" "People in the control group were informed that they were assigned to the group that could follow the course free of charge after post-test measurements 4 months later."
Blinding of outcome assessment (detection bias)	High risk	Self-rated instrument used to measure empathy
Incomplete outcome data (attrition bias)	High risk	"During the pilot, 59 participants dropped out. The total response at post-test was 61%. Reasons for dropouts in the Netherlands (n=29) were no time (n=4) or unknown (n=25; no response to repeated emails of researchers to remind them of filling in the questionnaires). Reasons for dropouts in the United Kingdom (n=30) were no time (n=1), no computer at home (n=1), or unknown (n=28; no response to repeated requests by researchers to fill in the questionnaires)."
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Larti 2018

Methods	Comparative study with random allocation to control and intervention groups.
Participants	The country of origin was Iran 82 operating room nursing students were randomised to either the intervention or control group. Inclusion criteria: second-semester or higher students who had entered the stage of clinical practice, had experience with communicating with patients, had not been diagnosed with any psychological conditions, and had no history of participation in communication or patient empathy workshops The exclusion criteria included incomplete responses to questionnaires, absence at any of the training sessions, and withdrawal from continuation of the study.
Interventions	Training programme for empathetic communication with patients in the operating room, mainly during the perioperative phase, using role-playing technique. The training was delivered face-to-face by the researchers with assistance from psychologists specialising in running empathy workshops. The duration of training was 12 hours delivered in 3 x 4 hour sessions with weekly sessions over 3 weeks.
Outcomes	The purpose of this study was to investigate the effects of a role-playing training program for empathetic communication with patients on the empathy scores of operating room nursing students. The JSE-HPS was used to measure self-rated empathy pre and one month post intervention.
Notes	-

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"A number was then randomly assigned to each of the students, and the numbers were poured into a bowl. The first paper drawn out of the bowl was for the experimental group, the second paper was for the control group, and this procedure was continued to select students from all years of study"
Allocation concealment (selection bias)	Low risk	"A number was then randomly assigned to each of the students, and the numbers were poured into a bowl. The first paper drawn out of the bowl was for the experimental group, the second paper was for the control group, and this procedure was continued to select students from all years of study"
Blinding of participants and personnel (performance bias)	High risk	"The objectives of the training program were then explained"
Blinding of outcome assessment (detection bias)	High risk	Self-assessment so no blinding of outcome assessor
Incomplete outcome data (attrition bias)	Low risk	Low attrition rate (6%)
Selective reporting (reporting bias)	Low risk	No other bias detected
Other bias	Unclear risk	

Lobchuck 2018

Methods	Two centre randomised controlled pilot study
Participants	The country of origin was Canada 25 nursing students were allocated to the intervention group and 19 to the control group. Students at: (a) the end of the second year or in the third year of a three-year accelerated baccalaureate program at the college or (b) the end of the second year or in the third or fourth year of a four-year baccalaureate program at the university were included. No exclusion criteria listed.
Interventions	Heart Health Whispering intervention was delivered as a novel person-cantered approach for counselling and health promotion. The training programme on perspective taking involved 4 phases. Phase 1 – individual teaching on perspective taking followed by 2 week period and instructions to practice skills. Phase 2 10 minute videoed conversation with actor. Phase 3, researcher and actor watch video and 'video-tag' thoughts and feelings actor remembered having experienced, shared, displayed etc. Phase 4 exit interviews
Outcomes	Empathy post intervention was assessed using the CARE scale completed by observer An adapted version of the CARE scale was also completed by the participant to capture their inference of the actors response to his or her clinical empathy.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The Research Assistant (RA) conducted a computerized randomization process to assign students to Group I (n=24) or Group PI (n=18)"
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias)	High risk	"Due to practical reasons, students, the interventionist (JL), and interviewers (ML and LH) were not blinded"

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Blinding of outcome assessment (detection bias)	Unclear risk	Mixed High – self reported measure of empathy (JSE) Low – observer reported - actor was blinded to group assignment.
Incomplete outcome data (attrition bias)	Low risk	Low attrition rate 5%
Selective reporting (reporting bias)	Unclear risk	Outcomes reported as per methodology
Other bias	Unclear risk	Baseline demographic differences not reported

Lor 2014

Methods	Randomised controlled trial
Participants	The country of origin was USA 40 student pharmacists were randomised to either the intervention or the control group. Students with pre-existing medical conditions were asked not to participate, and students with any self-reported medical conditions were automatically excluded.
Interventions	A 3 day simulation with each day including a designed activity with loss of the dominant hand usage, vision and speech. Simulations were followed by small group discussions regarding the daily activity, which covered its purpose, their feelings about the activity, items they learned, key take-away points, and how the items would affect their practice as future health care providers. This was followed by a large group discussion
Outcomes	The purpose of this study was to determine the immediate and sustained impact of a single, 3-day empathy intervention on empathy levels among students. The JSE-HPS was used to measure self-reported empathy at baseline, 7 days post-intervention and 90 days post-intervention.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Forty student pharmacists who volunteered and provided informed consent were then randomly assigned to either the intervention or control group" No information provided on random sequence generation
Allocation concealment (selection bias)	Unclear risk	"Subjects were randomized to an intervention group (n520) or control group (n520) and completed the JSE-HPS at baseline, 7 days postintervention, and 90 days postintervention." No information provided on allocation of students
Blinding of participants and personnel (performance bias)	High risk	"The purpose of this study was to determine the immediate and sustained impact of a single, 3-day empathy intervention on empathy levels among students and to address the lack of a control group by using a randomized, non-blinded, quasi-controlled design"
Blinding of outcome assessment (detection bias)	High risk	"The Jefferson Scale of Empathy-Health Profession Students version (JSE-HPS) was administered to the intervention and control groups at baseline, 7 days following the intervention (as post-test 1), and 90 days following the intervention (as post-test 2)."
Incomplete outcome data (attrition bias)	Low risk	No attrition from randomisation to reporting
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

LoSasso 2017

Methods	Randomised controlled trial
Participants	The country of origin was USA. 70 medical students were randomised to either the intervention or control groups. Third-year students were eligible to participate in the study while on their regularly scheduled six-week paediatric clerkship if their outpatient assignment was at a site using the Epic EMR system
Interventions	Training session on EMR (electronic medical records) specific communication skills, including discussion of EMR use, the SALTED (set-up, ask, listen, type, exceptions, documentation) mnemonic and technique and role-play.
Outcomes	Empathy was measured pre and post intervention using the self-rated JSE questionnaire. In addition an observer rating of empathy was taken using the JSPPPE (Jefferson Scale of Patient Perception of Physician Empathy).
Notes	No funding source reported.

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Participants in each six-week clerkship block were randomly assigned to the intervention group (n = 38) or to the control group (n = 32)." Not stated how randomisation occurred
Allocation concealment (selection bias)	Unclear risk	Details on allocation process not given
Blinding of participants and personnel (performance bias)	High risk	"In consenting for the study, students in both groups were made aware that the study examined how the training may improve empathy, which could have led to some bias."
Blinding of outcome assessment (detection bias)	Unclear risk	The SP and faculty raters' were blinded to whether students were in the intervention or control group – and completed the observer-rated scale JSPPPE (low risk) Self-reported scale JSE outcome assessors not blinded (high risk)
Incomplete outcome data (attrition bias)	Low risk	No attrition from randomisation to analysis
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Mueller 2018

Methods	Randomised controlled trial.
Participants	The country of origin was USA. 19 physical therapy students were randomised to the intervention group and 18 to the control group (which was a 'delayed' intervention group). All students entering the third year were approached. No exclusion criteria listed.
Interventions	On-line Called to Care curriculum used to improve patient outcomes through the development of optimal physical therapist behaviours. (employs film clips, quidded questions, research articles and other readings to promote the clinical application of educational concepts. Participants post and respond via a discussion board for each of the 11 modules.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Outcomes	The JSE-HP was used to measure a change in empathy pre and post intervention.
Notes	

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Participants were randomly assigned (via a blinded shuffle of cards) to an immediate intervention group or a delayed intervention group."
Allocation concealment (selection bias)	Low risk	"Participants were randomly assigned (via a blinded shuffle of cards) to an immediate intervention group or a delayed intervention group. The deck included only the numbered cards (to ensure an event 50/50 split) and group assignment based on events or odds)."
Blinding of participants and personnel (performance bias)	High risk	An orientation to the Called to Care curriculum was provided to all participants at the end of the spring 2015 semester. The participants were informed of their designation into the immediate or delayed intervention group.
Blinding of outcome assessment (detection bias)	High risk	Self- reported scale
Incomplete outcome data (attrition bias)	Low risk	Of the 37 participants 1 withdraw due to pregnancy-related delay in her internship (2.7%)
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	High risk	No other bias detected

Reiss 2012

Methods	Randomised controlled trial
Participants	The country of origin was USA. 54 residents and fellows were randomised to the intervention group and 45 to the control group. Residents and fellows were eligible if they (1) were currently in training, (2) were available to attend all three training modules, and (3) had clinical interactions with adult outpatients or inpatients able to complete physician rating surveys. Trainees on clinical rotations outside MEEI or MGH were excluded. Trainees on night float, paediatrics, ICU or research rotations were excluded unless they had a clinic with adult patients.
Interventions	Empathy and relational skills training protocol developed by first author and previously tested in a pilot study. Aims of training (1) scientific foundation of empathy, (2) increase awareness of physiology of emotions, (3) improve skills in decoding facial expressions of emotion, (4) teach empathic responses. Training was delivered any a trained physician in both the inpatient and outpatient setting. The duration of intervention was 4 hours and was delivered in 60 minute modules spaced over 4 weeks.
Outcomes	Change in empathy was assessed by patients using the CARE measure as the primary outcome. As secondary outcomes the following was measured: Physician skill at decoding facial expression (The Ekman Facial Decoding Test). Self-rated physician attitude about empathy (JSPSE, validated scale). Self-rated general empathic responsiveness in personal life (The Balanced Emotional Empathy Scale, BEES)
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Random sequence generation (selection bias)	Low risk	“Group assignment was determined by a computer-generated random number sequence”
Allocation concealment (selection bias)	Low risk	“Participating physicians were randomly assigned in a 1:1 allocation ratio to either the training intervention or to standard residency or fellowship training”
Blinding of participants and personnel (performance bias)	High risk	"Participating physicians were randomly assigned in a 1:1 allocation ratio to either the training intervention or to standard residency or fellowship training." "The training was comprised of three 60-minute modules spaced over 4 weeks"
Blinding of outcome assessment (detection bias)	Unclear risk	“Patients were blind to physician randomization, and physicians were blinded to which patients completed the surveys” “The primary outcome measure was change in empathetic and relational skills as assessed by patients blinded to physician randomization” Secondary outcomes – self rated scales of empathy so unable to blind outcome assessor
Incomplete outcome data (attrition bias)	Low risk	Overall attrition rate 7.5% (4 participants lost in control group, 1 participant lost in intervention group).
Selective reporting (reporting bias)	Low risk	Primary and secondary outcomes reported as stated in methods.
Other bias	Low risk	No other bias detected

Shapiro 1998

Methods	Matched randomised experiment with wait-list controls.
Participants	78 premedical and medical students were randomised to either the intervention or control groups. Inclusion criteria: first- and second-year medical students, the premedical honours society, and the Fostering and Achieving Cultural Equity and Sensitivity (FACES) premedical student group. Only those students willing to be randomly assigned to either the intervention or control group were included in the study.
Interventions	Elective module in Stress Reduction and Relaxation. The core of the program focused on training the students in mindfulness. Participants received training in: “Sitting Meditation”, “Body Scan” and “Hatha Yoga”. Emphasis on mindful breathing, “lovingkindness” and “forgiveness”. In addition, students participated in experiential exercises designed to cultivate mindful listening skills and empathy. The training was delivered via a mixture of didactic teaching and small group sessions. The duration was approximately 18 hours delivered in 2.5 hour weekly sessions over 8 weeks.
Outcomes	Empathy was measured using an adapted version (half of the original version of 84 items) of The Empathy Construct Rating Scale (ECRS).
Notes	No funding source reported

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	“The design was a matched randomized experiment in which participants were assigned to a 7-week mindfulness-based intervention or a wait-list control group.” Random sequence generation not reported

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Allocation concealment (selection bias)	Unclear risk	Details of allocation concealment not stated
Blinding of participants and personnel (performance bias)	High risk	"The design was a matched randomized experiment in which participants were assigned to a 7-week mindfulness-based intervention or a wait-list control group"
Blinding of outcome assessment (detection bias)	Unclear risk	"all assessment measures were self-report psychological questionnaires which are intrinsically limited and open to response bias."
Incomplete outcome data (attrition bias)	Low risk	"One student did not complete the intervention due to severe medical problems for which she was hospitalized. Four of the participants in the control group did not complete the post-measures. The final count of participants was 73, consisting of 32 males and 41 females, 35 premedical students and 38 medical students."
Selective reporting (reporting bias)	High risk	"Outcomes reported as a cohort in general."
Other bias	Low risk	No other bias detected

Sripada 2010

Methods	Pilot randomised controlled trial
Participants	The country of origin was USA. 12 psychiatry residents were randomised to either the intervention or control group. All second- through fourth-year psychiatry residents treating out-patients at the University of Illinois College of Medicine during the academic years 2002–2005 were eligible to participate in this study. Patients were eligible if they were between the ages of 18 and 65, were in treatment for an Axis I psychiatric disorder, had no intellectual disability, and were not suicidal or psychotic.
Interventions	A feedback intervention designed to increase therapist empathic understanding and improve patient outcomes in psychotherapy was delivered. The feedback intervention condition involved completing the empathy measure along with other measures, and engaging in the feedback intervention which involved: At the end of each therapy session, patients and therapists recorded their views of the patient's GAF and predicted the GAF ratings of the other. In the intervention condition, at the beginning of the next session, therapists and patients exchanged ratings from the preceding session, providing an opportunity to discuss their respective views. The average number of sessions completed by each therapist–patient pair was 14.1 The average duration of patient participation in the study was 13.75 (±7.0) sessions or 183.87 (± 111.1) days. The average duration of therapist participation was 195.8 (± 117.4) days.
Outcomes	The Barrett-Lennard Relationship Inventory - 6-item scale designed to assess patients' ratings of therapist empathy as well as therapists' self-ratings of empathy.
Notes	-

Risk of bias table

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	"Patient-therapist pairs were randomly assigned by the first author to the intervention or control group by flipping a coin." However how therapists were assigned to intervention or control not reported.
Allocation concealment (selection bias)	Unclear risk	Allocation to intervention/control not descried
Blinding of participants and personnel (performance bias)	High risk	"Patients were blind to intervention condition, but therapists were not, as they administered the intervention".

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Blinding of outcome assessment (detection bias)	Unclear risk	"A more methodological limitation of this study is the potential for contamination that existed because a single therapist treated five patients, three of whom were assigned to control, and two of whom were assigned to intervention."
Incomplete outcome data (attrition bias)	Low risk	Methodology states: "Additionally, at the end of the 1st, 5th, 10th, 15th, and 20th sessions, patient and therapist subjects in both groups completed their respective forms of the BLRI (Barrett-Lennard, 1976). Only patient scores reported in results"
Selective reporting (reporting bias)	High risk	Data not explicitly reported for each group
Other bias	Unclear risk	difference in baseline demographics of therapists and patients not reported

Sterkenburg 2018

Methods	Parallel randomised controlled trial
Participants	The country of origin was the Netherlands. 111 care workers were randomised to the intervention group and 113 to the control group. Inclusion: Care workers working with people with disabilities
Interventions	Playing a computer-based serious game "The World of EMPA", aimed at enhancing empathy towards people with disabilities. The game illustrates characters with several types of disability, with six levels in which players have to respond to multiple-choice questions. The intervention was delivered online and took 20 minutes to complete. It was a one-off intervention.
Outcomes	The Empathy Quotient (EQ) short version self-rating questionnaire was administered to assess changes in empathy at baseline and immediately following the intervention.
Notes	Funding source not stated.

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Upon completion of the pre-test phase, participants were automatically randomized via a computerized random assignment to one of the two conditions, based on the Mersenne Twister pseudorandom number generator (PRNG)"
Allocation concealment (selection bias)	Low risk	"The automatic computer-based randomization was implemented in the programming script of the experiment, resulting in the concealed allocation of the participants into one of the two intervention arms"
Blinding of participants and personnel (performance bias)	Low risk	"The participants were also unaware whether the condition they were allocated to was the experimental or control condition"
Blinding of outcome assessment (detection bias)	Low risk	"The researcher was blind to condition once participants started the computer program".
Incomplete outcome data (attrition bias)	Low risk	a total of 224 care workers working with people with disabilities were recruited, and 223 completed the study
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Tulsky 2011

Methods	Parallel randomised controlled trial
Participants	The country of origin was USA.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

	24 medical, gynaecological and radiation oncologists were randomised to the intervention group and 24 to the control group. Inclusion and exclusion criteria were not stated.
Interventions	A communication lecture (1 hour) was delivered to all intervention and control students. An interactive CD-ROM about responding to patients' negative emotions was then given to intervention participants. The CD-ROM included tailored feedback on the oncologists own recorded conversations. Participants had up to one month to view the CD-ROM.
Outcomes	Empathic statements - Post-intervention audio recordings were used to identify the number of empathic statements and responses to patients' expressions of negative emotion. Perceived empathy - 10 Likert scale items was used to assess perceived oncologist empathy (as assessed by patient)
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The oncologists were then randomly assigned by using the minimization method"
Allocation concealment (selection bias)	Low risk	"The oncologists were stratified by balanced randomization in a 1:1 ratio by site (Durham or Pittsburgh), sex (men or women), and specialty (medical oncology, solid and liquid tumours; medical oncology, solid tumours only; malignant haematology, liquid tumours only; gynaecologic oncology; or radiation oncology)."
Blinding of participants and personnel (performance bias)	High risk	"All of the oncologists viewed a 1-hour lecture on communication skills delivered by one of the investigators. In addition, oncologists in the intervention group received a CD-ROM training program on communication skills that was tailored with exemplars from their own audio-recorded clinic visits."
Blinding of outcome assessment (detection bias)	Low risk	"Two independent, blinded coders were trained over 6 weeks"
Incomplete outcome data (attrition bias)	Low risk	No attrition from randomisation to analysis
Selective reporting (reporting bias)	Low risk	Outcomes reported as per methodology
Other bias	Low risk	No other bias detected

Vaghee 2018

Methods	Cluster randomised controlled trial
Participants	The country of origin was Iran. Nursing faculties training mental health clerkship in Ibne-Sina psychiatric hospital were invited to attend in the study, and accordingly, 12 faculties accepted the invitation, and 4 faculties were randomly selected. 127 nursing students were randomised to one of three groups: two intervention groups or a control group. Inclusion criteria were no work experience in psychiatric wards, no psychological disorders, and no mental illness in their first and second degree relatives. Exclusion criteria were reluctance to continue the study, absence of the post-test, and being absent or lack of participation in 1 or more intervention sessions.
Interventions	The two intervention groups were: Contact based education: In contact-based education, 3 patients with improved disorders who were working daily for 4 hours as a connector between different wards of the hospital were selected. They had schizophrenia, bipolar type I, and major depression. The patients

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

	were asked to talk about their experiences and personal life with students Acceptance and commitment education: According to Steven Hayse protocol (1986), ACT with the content of mental illnesses stigma was held as a workshop by one master of clinical psychology and 2 masters of psychiatric nursing,
Outcomes	The study aimed at comparing the effects of contact-based education and commitment and acceptance-based training on empathy toward mental illnesses among nursing students. The JSE was used as a self-rating measure of empathy pre and post intervention.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"Two groups of male and female students were randomly selected (according to clerkship division group) from each university by quota sampling based on gender distribution. Finally, each group was separately divided into 3 groups of contact-based education, ACT, and control." No details on random sequence generation
Allocation concealment (selection bias)	Unclear risk	No details on allocation concealment reported
Blinding of participants and personnel (performance bias)	High risk	"The patients were asked to talk about their experiences and personal life with students"
Blinding of outcome assessment (detection bias)	High risk	Self-reported outcome measures
Incomplete outcome data (attrition bias)	Low risk	Low attrition rate (12.5%)
Selective reporting (reporting bias)	High risk	Outcomes are not clearly stated in methodology
Other bias	Unclear risk	Recruitment bias: Random cluster and quota sampling methods were used. Nursing faculties training mental health clerkship in Ibne-Sina psychiatric hospital were invited to attend in the study, and accordingly, 12 faculties accepted the invitation, and 4 faculties were randomly selected.

Wolf 1987

Methods	Randomised controlled trial
Participants	The county of origin was Canada 65 medical students were randomised to the intervention group and 69 to the control group. Part of course was conducted in community nursing homes, so not all students could be scheduled to participate in it at the same time. Therefore, some of the students participated in the main part of the study. The remaining (excluded) students participated in the course after the study was completed.
Interventions	Programme in medical interviewing and history taking that integrates humanistic principles and medical content. The course is designed to use community resources and maximise efficient use of faculty members' time. Consists of set of large group lectures and then small group teaching sessions which included discussing strategies for responding empathically to patients. The teaching was delivered in small group sessions by social workers and educational psychologists. It consisted of 3 x 4 hour sessions and was delivered weekly.
Outcomes	The Medical Communication Index (MCI) served as the dependent variable to measure the students' responses to patients' emotional concerns The Helping relationship Inventory (HRI) served to measure the dependent variable to measure the students' preferences for responses that expressed empathy or understanding.

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Notes	No funding source stated

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"All students in both the intervention and control groups attended these large group lectures. Following this instruction, the students were randomly assigned to an intervention or control group" Details of random sequence generation not reported
Allocation concealment (selection bias)	Unclear risk	"Part of course conducted in community nursing homes, not all students could be scheduled to participate in it at the same time. Therefore, only 134 of these students participated in the main part of the study. The remaining (excluded) students participated in the course after the study was completed." Allocation concealment not reported
Blinding of participants and personnel (performance bias)	High risk	"The 69 students in the control group received no other instruction in communication skills during the study. The 65 students in the intervention group were divided into four smaller groups. Each group met for four weekly, three-hour sessions."
Blinding of outcome assessment (detection bias)	High risk	Self-rated outcome assessment
Incomplete outcome data (attrition bias)	Unclear risk	24 lost to follow up (not clearly stated) on analysis of MCI). Not explicitly stated on what number of students' basis analysis carried out, how many lost to follow up or reasons
Selective reporting (reporting bias)	High risk	Outcomes not clearly stated in methodology.
Other bias	Unclear risk	no baseline demographics reported so cannot comment on baseline differences

Wundrich 2017

Methods	Randomised controlled trial.
Participants	The country of origin was Germany. 158 third year medical students were randomised to either an intervention or control group. No inclusion or exclusion criteria were stated.
Interventions	A three week training course with focus on empathy: The empathy skills training consisted of an introduction course on empathy and empathy skills training with simulated patients. The duration of the intervention was 6 hours delivered over 3 weeks.
Outcomes	The self-rated JSPE (student version) was used to measure empathy in addition to an empathy-related communications skills questionnaire completed by an observer.
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"A total of 158 3rd year medical students at the University of Freiburg Medical Centre were assigned into an intervention group receiving an empathy training and a control group" Details of random sequence generation not reported
Allocation concealment (selection bias)	Unclear risk	Allocation concealment not stated
Blinding of participants and personnel (performance bias)	High risk	"The intervention group participated in an empathy skills training with simulated patients (SPs). The control group participated in a history course."

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

Blinding of outcome assessment (detection bias)	Unclear risk	Experts and SPs were blinded to the students' group membership - low risk for observer rated outcome. Self-rated outcome high risk
Incomplete outcome data (attrition bias)	Unclear risk	Number analysed not reported. Missing data not reported
Selective reporting (reporting bias)	High risk	Number analysed not reported. Missing data not reported
Other bias	Unclear risk	no baseline demographics reported so cannot comment on baseline differences

Yang 2018

Methods	Cluster randomised controlled trial
Participants	The country of origin was China. 59 'grade 3' nursing students each were randomised to two intervention arms and 59 to a control arm of the study. Exclusion criteria: students who were taking doctor–patient communication- related courses and students who were planning to take those courses during the study.
Interventions	The intervention was a narrative medicine programme. Two intervention groups: One group received the theoretical education part of the programme and one intervention group received both theoretical teaching and clinical experience. The theoretical component was delivered by a teacher 'well trained in narrative medicine'. The clinical component was delivered by teaching nurses who had been trained in narrative medicine.
Outcomes	The JSE (Chinese version) was administered to students at baseline and then at various follow up points post intervention: T1: January 2015 (pre-intervention), T2: July 2015 (post-step 1 intervention) T3: January 2016 (post-step 2 intervention), T4: July 2016 (0.5 years after the intervention), T5: January 2017 (1 year after the intervention), and T6: July 2017 (1.5 years after the intervention).
Notes	-

Risk of bias table

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Random sequence generation not stated. "the six classes were randomly divided into three groups"
Allocation concealment (selection bias)	Unclear risk	"Of the sixteen classes, six (30 students per class) were randomly selected to participate in this study." "Taking each class as a unit, the six classes were randomly divided into three groups: one observation group (Group 1) and two experimental groups (Groups 2 and 3)." Method of allocation not stated.
Blinding of participants and personnel (performance bias)	High risk	No blinding of participants or personnel
Blinding of outcome assessment (detection bias)	High risk	Outcome assessors were not blinded.
Incomplete outcome data (attrition bias)	Low risk	5 participants from intervention groups and 7 controls lost to follow up. Attrition 6.6%

Selective reporting (reporting bias)	Low risk	Outcomes reported as stated in methods.
Other bias	Unclear risk	Recruitment bias: Method of randomisation not described "six [classes] were randomly selected" According to methodology, no participants were recruited after the clusters had been randomised.

eTable 4 Empathy effect summary of findings

Summary of findings:

Empathy training compared to Control for Healthcare students and professionals

Patient or population: Healthcare students and professionals
Setting: University, primary care settings, secondary care settings
Intervention: Empathy training
Comparison: Control

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Control	**Risk with Empathy training**				
empathy	-	SMD **0.52 SD more** (0.36 more to 0.67 more)	-	2024 (22 RCTs)	⨁⨁◯◯ LOW [a,b]	Empathy training may increase empathy.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

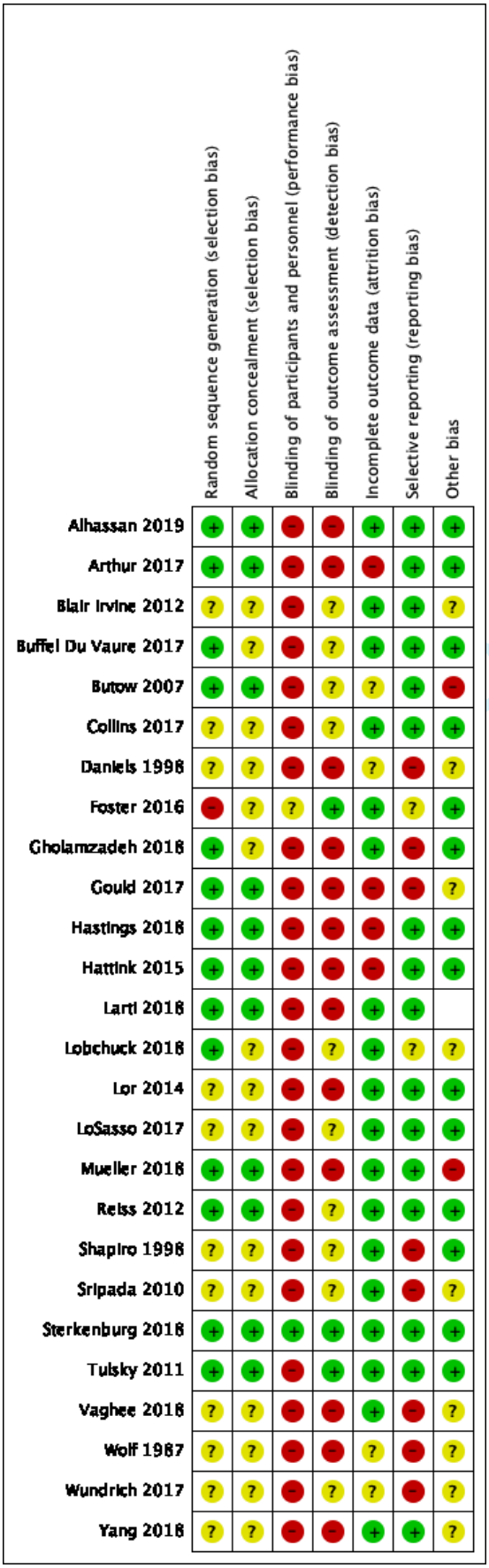
CI: Confidence interval; **SMD:** Standardised mean difference

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Footnotes
a High risk of bias suspected in 11 studies (with a high or unclear risk of bias for sequence generation and allocation concealment)
b There was variation across all studies with type of intervention and population studied

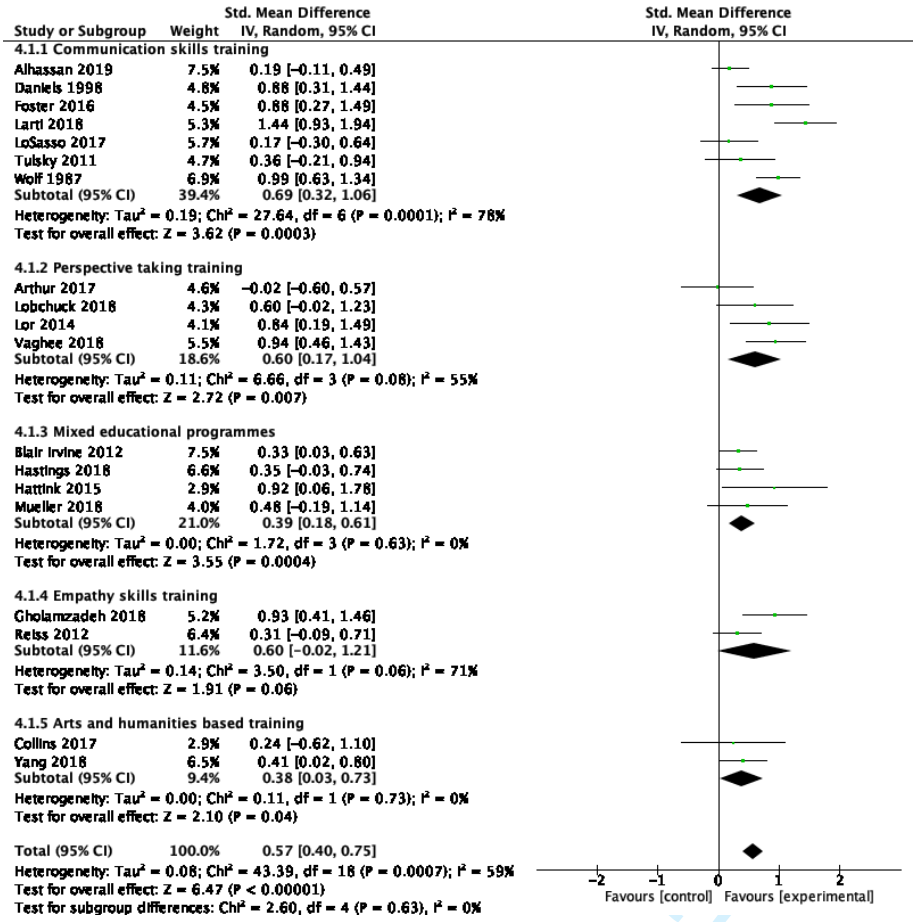
Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

eFigure 1. Risk of bias assessment



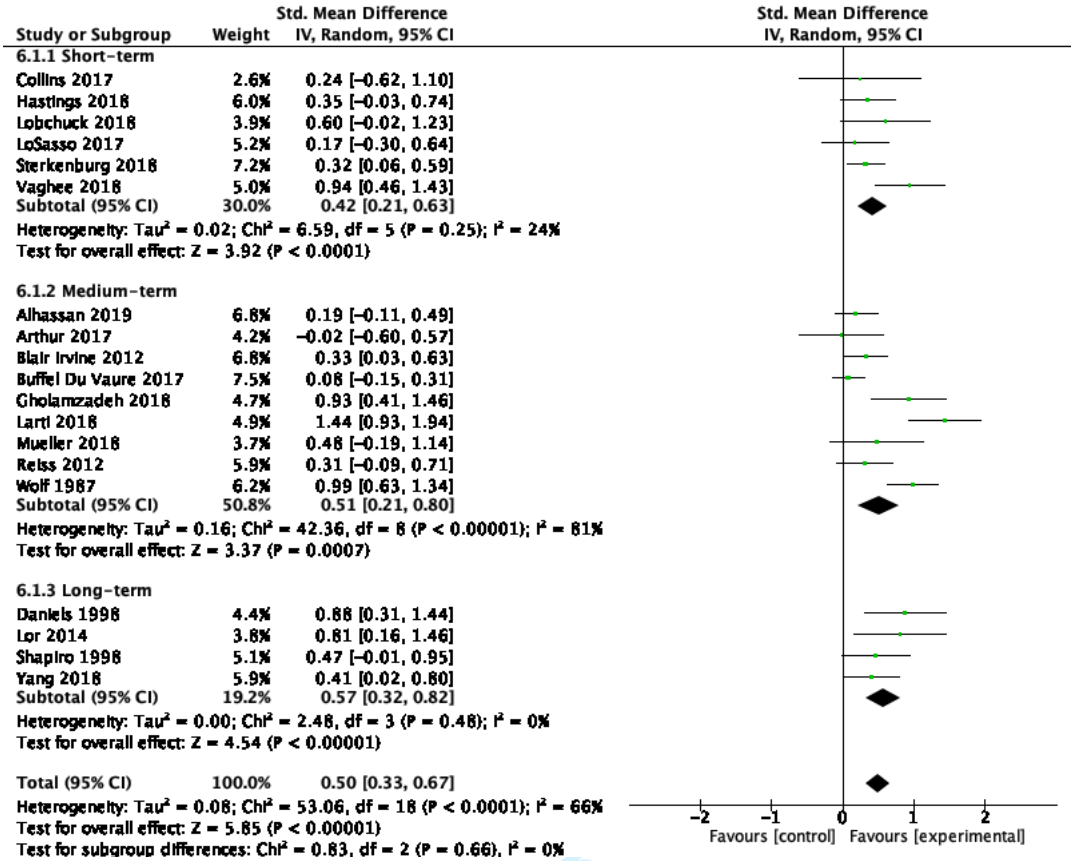
Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

eFigure 2. **Meta-analysis of sub-groups according to type of intervention**

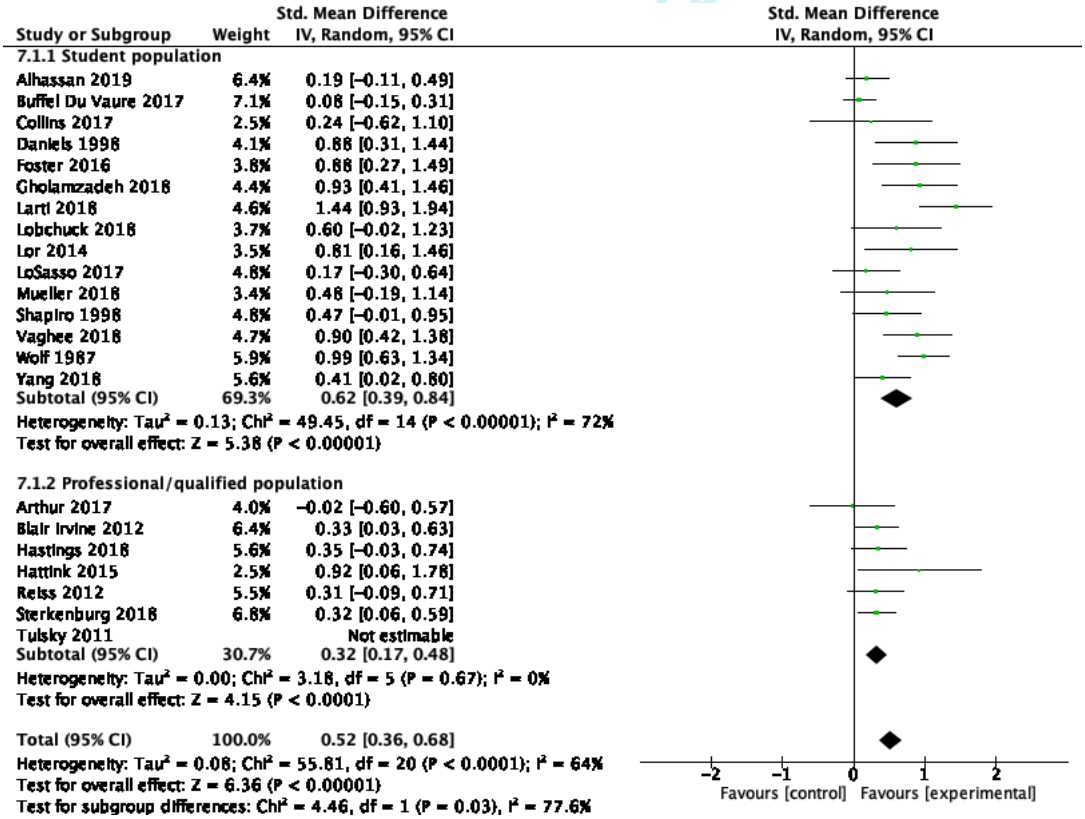


eFigure 3. **Meta-analysis of subgroups according to duration of intervention**

Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

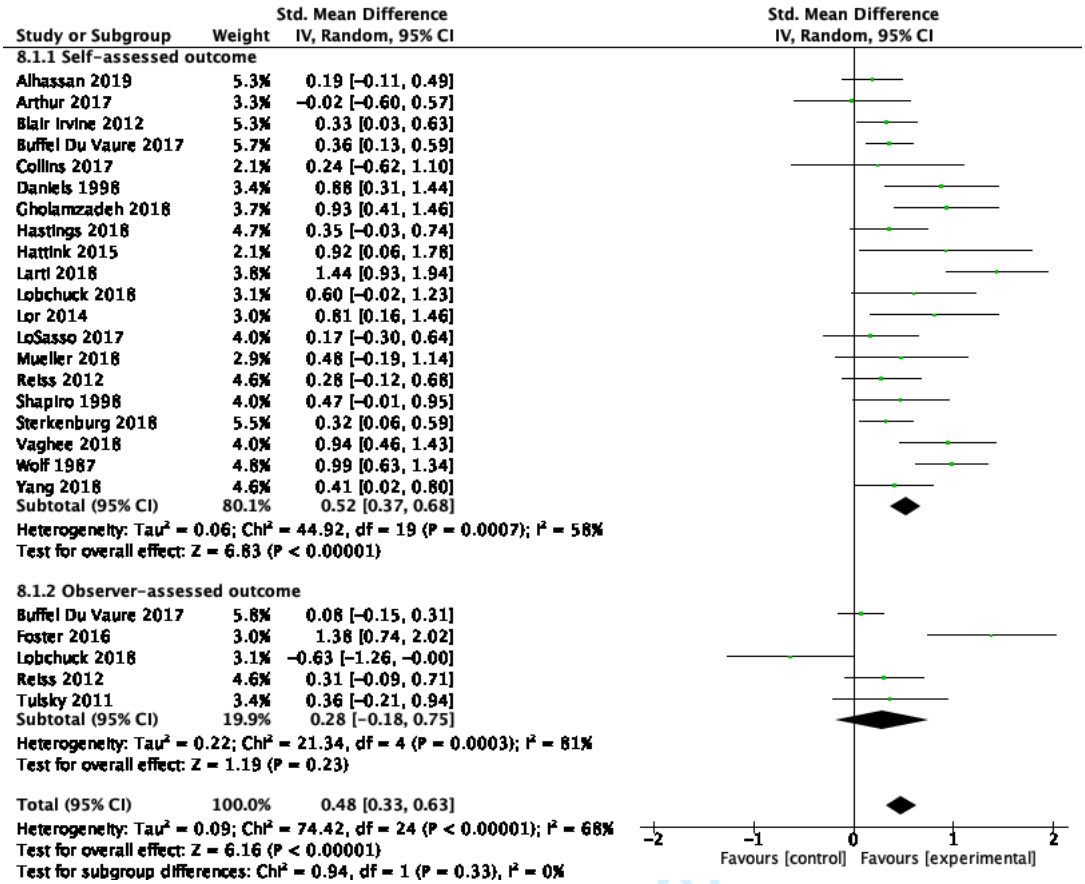


eFigure 4 **Meta-analysis of subgroups according to participant population**



Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

eFigure 5 **Meta-analysis of subgroups according to outcome assessor**



PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	7
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	8
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e) grouping and comparing tests, f) handling of different reference standards	8

PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	9
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	11
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	12
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	16
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	16
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	17
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	18
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	21
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	21
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	22
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

For more information, visit: www.prisma-statement.org.