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 Riess 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

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

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

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

	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

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

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

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

	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

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

#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

Riess 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%

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

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).

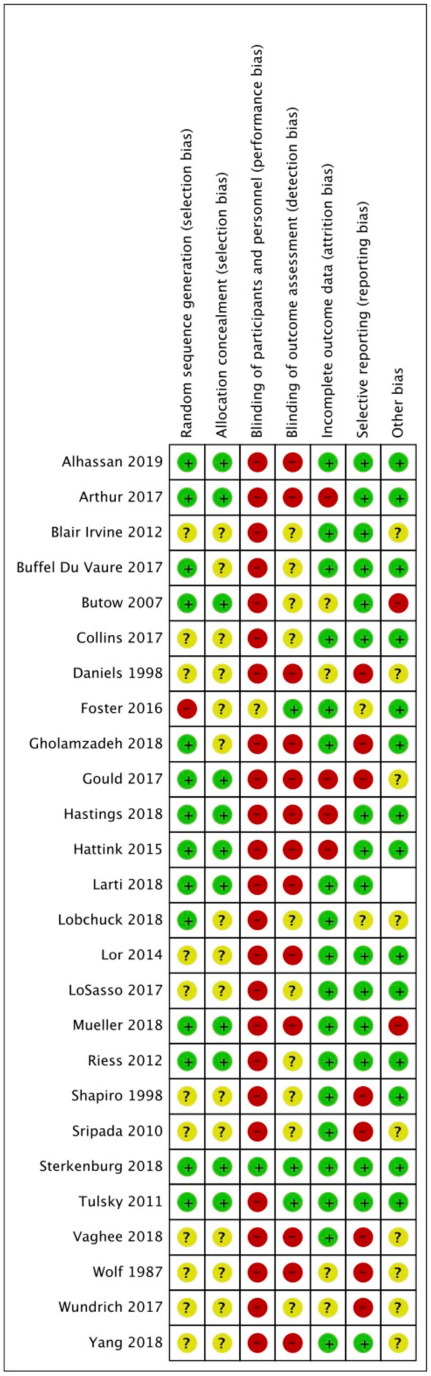
CI: Confidence interval; **SMD:** Standardised mean difference

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

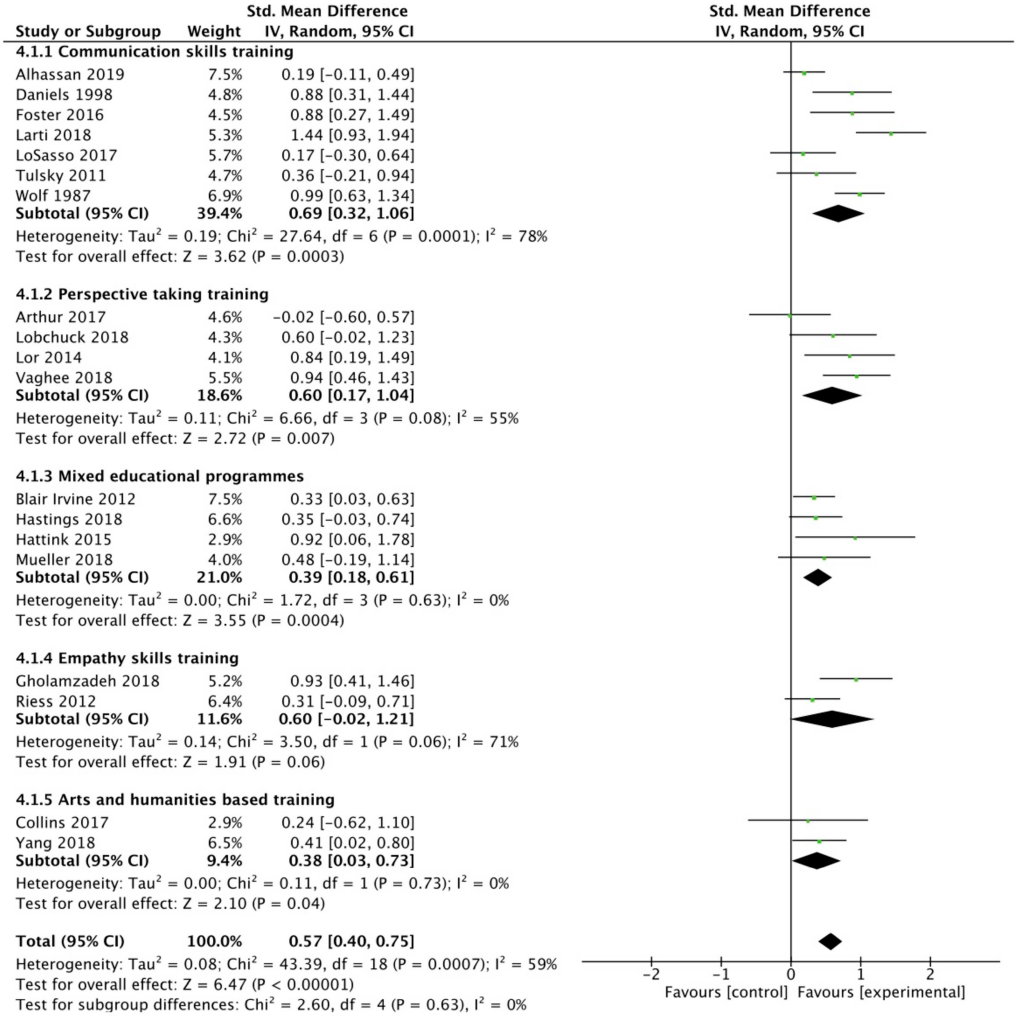
Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

eFigure 1. Risk of bias assessment



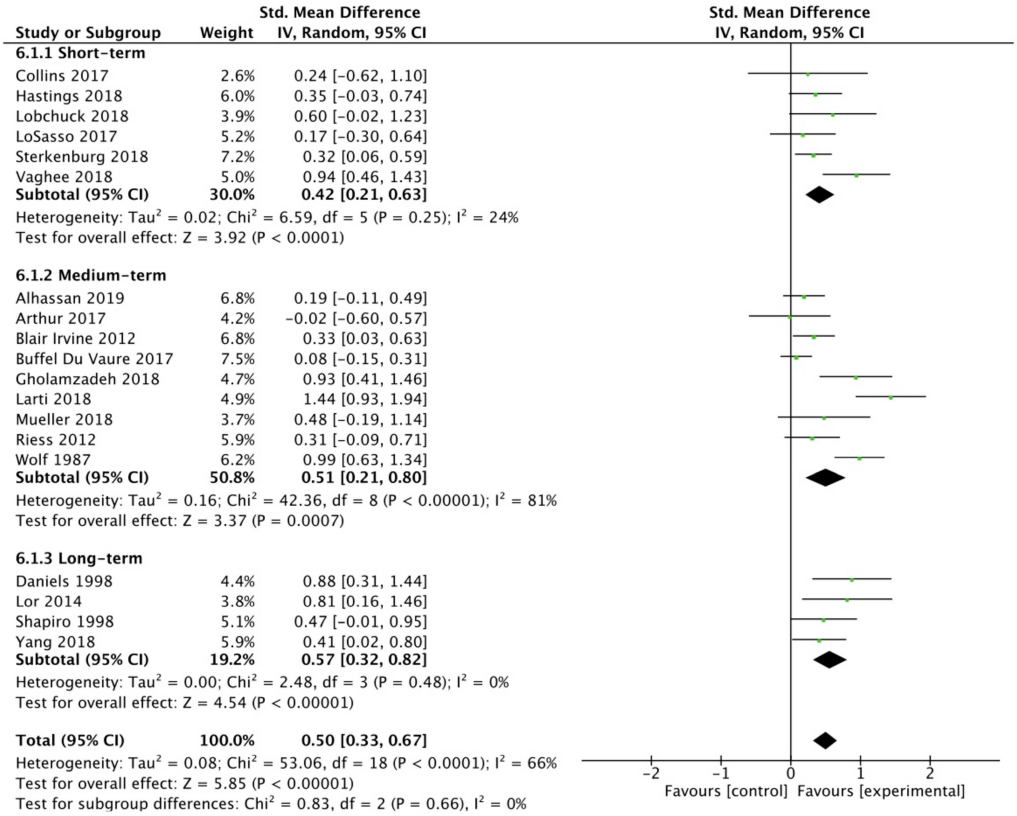
Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

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



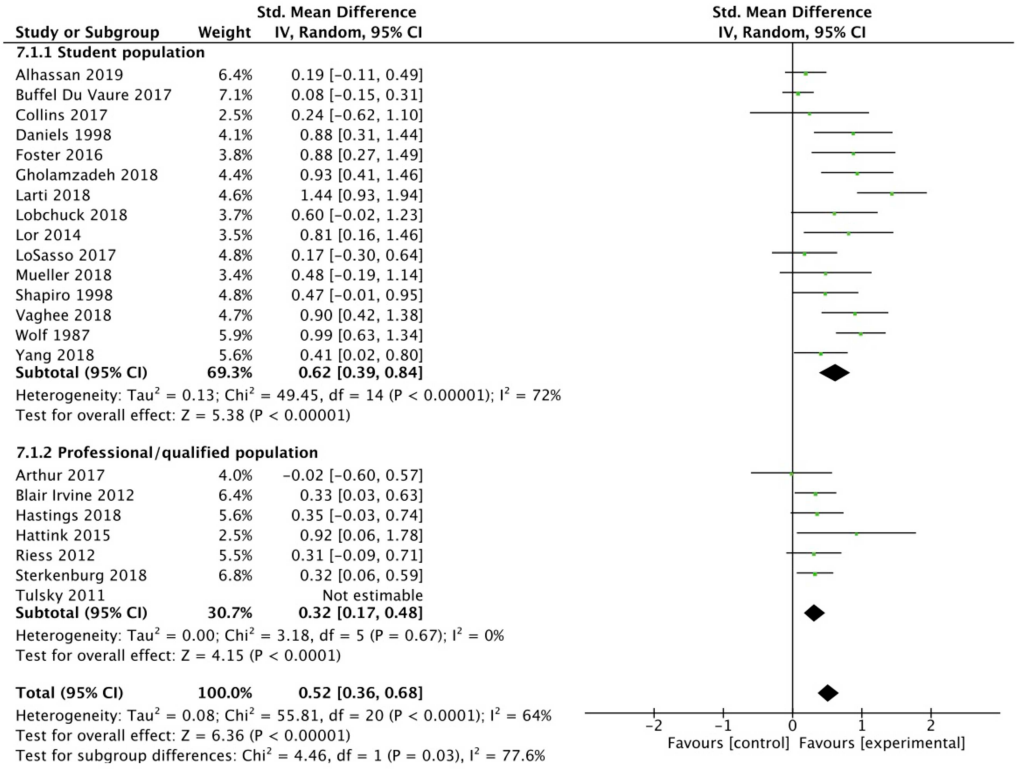
Winter R, Isa E, Roberts N, Norman RI, Howick J 2019

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**