Cochrane Library

Cochrane Database of Systematic Reviews

Interventions for improving medical students' interpersonal communication in medical consultations (Review)

Gilligan C, Powell M, Lynagh MC, Ward BM, Lonsdale C, Harvey P, James EL, Rich D, Dewi SP, Nepal S, Croft HA, Silverman J

Gilligan C, Powell M, Lynagh MC, Ward BM, Lonsdale C, Harvey P, James EL, Rich D, Dewi SP, Nepal S, Croft HA, Silverman J. Interventions for improving medical students' interpersonal communication in medical consultations. *Cochrane Database of Systematic Reviews* 2021, Issue 2. Art. No.: CD012418.
DOI: 10.1002/14651858.CD012418.pub2.

www.cochranelibrary.com

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

WILEY

TABLE OF CONTENTS

ABSTRACT 1
PLAIN LANGUAGE SUMMARY 2
SUMMARY OF FINDINGS 5
BACKGROUND 14
OBJECTIVES 17
METHODS 17
Figure 1. 19
RESULTS 22
Figure 2. 26
Figure 3. 27
Figure 4. 32
Figure 5. 33
Figure 6. 34
Figure 7. 35
Figure 8. 35
Figure 9. 37
Figure 10. 38
Figure 11. 38
Figure 12. 40
Figure 13. 41
DISCUSSION 44
AUTHORS' CONCLUSIONS 47
ACKNOWLEDGEMENTS 48
REFERENCES 49
CHARACTERISTICS OF STUDIES 63
DATA AND ANALYSES 224
Analysis 1.1. Comparison 1: Communication intervention versus control or usual care, Outcome 1: Overall communication skills 226
Analysis 1.2. Comparison 1: Communication intervention versus control or usual care, Outcome 2: Subgroup_Overall communication skills_control only 227
Analysis 1.3. Comparison 1: Communication intervention versus control or usual care, Outcome 3: Sensitivity_RoB_Overall communication skills 228
Analysis 1.4. Comparison 1: Communication intervention versus control or usual care, Outcome 4: Sensitivity_cluster 5%_Overall communication skills 229
Analysis 1.5. Comparison 1: Communication intervention versus control or usual care, Outcome 5: Sensitivity_cluster 10%_Overall communication skills 230
Analysis 1.6. Comparison 1: Communication intervention versus control or usual care, Outcome 6: Sensitivity_cluster 20%_Overall communication skills 231
Analysis 1.7. Comparison 1: Communication intervention versus control or usual care, Outcome 7: Empathy 231
Analysis 1.8. Comparison 1: Communication intervention versus control or usual care, Outcome 8: Relationship building/Rapport 232
Analysis 1.9. Comparison 1: Communication intervention versus control or usual care, Outcome 9: Sensitivity_cluster 5%_Relationship building/Rapport 232
Analysis 1.10. Comparison 1: Communication intervention versus control or usual care, Outcome 10: Sensitivity_cluster 10%_Relationship building/Rapport 233
Analysis 1.11. Comparison 1: Communication intervention versus control or usual care, Outcome 11: Sensitivity_cluster 20%_Relationship building/Rapport 233
Analysis 1.12. Comparison 1: Communication intervention versus control or usual care, Outcome 12: Information gathering about patient perspectives/concerns 234
Analysis 2.1. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 1: Overall communication skills 235
Analysis 2.2. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 2: Sensitivity_cluster 5%_Overall communication skills 236

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.3. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 3: Sensitivity_cluster 10%_Overall communication skills 236
Analysis 2.4. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 4: Sensitivity_cluster 20%_Overall communication skills 237
Analysis 2.5. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 5: Empathy 237
Analysis 2.6. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 6: Relationship building/rapport 238
Analysis 3.1. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 1: Overall communication skills 239
Analysis 3.2. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 2: Sensitivity_RoB_Overall communication skills 240
Analysis 3.3. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 3: Sensitivity_cluster 5%_Overall communication skills 240
Analysis 3.4. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 4: Sensitivity_Cluster 10%_Overall communication skills 241
Analysis 3.5. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 5: Sensitivity_Cluster 20%_Overall communication skills 241
Analysis 4.1. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 1: Overall communication skills 242
Analysis 4.2. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 2: Subgroup_Overall communication skills_SP only 242
Analysis 4.3. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 3: Sensitivity_RoB_Overall communication skills 243
Analysis 4.4. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 4: Sensitivity_Cluster 5%_Overall communication skills 243
Analysis 4.5. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 5: Sensitivity_Cluster 10%_Overall communication skills 243
Analysis 4.6. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 6: Sensitivity_Cluster 20%_Overall communication skills 243
ADDITIONAL TABLES 244
APPENDICES 249
WHAT'S NEW 257
HISTORY 257
CONTRIBUTIONS OF AUTHORS 257
DECLARATIONS OF INTEREST 257
SOURCES OF SUPPORT 257
DIFFERENCES BETWEEN PROTOCOL AND REVIEW 258
NOTES 258
INDEX TERMS 258

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

[Intervention Review]

Interventions for improving medical students' interpersonal communication in medical consultations

Conor Gilligan[1], Martine Powell[2], Marita C Lynagh[1], Bernadette M Ward[3], Chris Lonsdale[4], Pam Harvey[3], Erica L James[1], Dominique Rich[5], Sari P Dewi[5], Smriti Nepal[6], Hayley A Croft[7], Jonathan Silverman[8]

[1]School of Medicine and Public Health, University of Newcastle, Hunter Medical Research Institute, Callaghan, Australia. [2]Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, Brisbane, Australia. [3]School of Rural Health, Monash University, Bendigo, Australia. [4]Institute for Positive Psychology and Education, Australian Catholic University, Strathfield, Australia. [5]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia. [6]The Matilda Centre for Research in Mental Health and Substance Use, University of Sydney, Darlington, Australia. [7]School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan, Australia. [8]School of Clinical Medicine, University of Cambridge, Cambridge, UK

Contact: Conor Gilligan, conor.gilligan@newcastle.edu.au.

Editorial group: Cochrane Consumers and Communication Group.
Publication status and date: Edited (no change to conclusions), published in Issue 2, 2021.

Citation: Gilligan C, Powell M, Lynagh MC, Ward BM, Lonsdale C, Harvey P, James EL, Rich D, Dewi SP, Nepal S, Croft HA, Silverman J. Interventions for improving medical students' interpersonal communication in medical consultations. *Cochrane Database of Systematic Reviews* 2021, Issue 2. Art. No.: CD012418. DOI: 10.1002/14651858.CD012418.pub2.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Communication is a common element in all medical consultations, affecting a range of outcomes for doctors and patients. The increasing demand for medical students to be trained to communicate effectively has seen the emergence of interpersonal communication skills as core graduate competencies in medical training around the world. Medical schools have adopted a range of approaches to develop and evaluate these competencies.

Objectives

To assess the effects of interventions for medical students that aim to improve interpersonal communication in medical consultations.

Search methods

We searched five electronic databases: Cochrane Central Register of Controlled Trials, MEDLINE, Embase, PsycINFO, and ERIC (Educational Resource Information Centre) in September 2020, with no language, date, or publication status restrictions. We also screened reference lists of relevant articles and contacted authors of included studies.

Selection criteria

We included randomised controlled trials (RCTs), cluster-RCTs (C-RCTs), and non-randomised controlled trials (quasi-RCTs) evaluating the effectiveness of interventions delivered to students in undergraduate or graduate-entry medical programmes. We included studies of interventions aiming to improve medical students' interpersonal communication during medical consultations. Included interventions targeted communication skills associated with empathy, relationship building, gathering information, and explanation and planning, as well as specific communication tasks such as listening, appropriate structure, and question style.

Data collection and analysis

We used standard methodological procedures expected by Cochrane. Two review authors independently reviewed all search results, extracted data, assessed the risk of bias of included studies, and rated the quality of evidence using GRADE.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Main results

We found 91 publications relating to 76 separate studies (involving 10,124 students): 55 RCTs, 9 quasi-RCTs, 7 C-RCTs, and 5 quasi-C-RCTs. We performed meta-analysis according to comparison and outcome. Among both effectiveness and comparative effectiveness analyses, we separated outcomes reporting on overall communication skills, empathy, rapport or relationship building, patient perceptions/ satisfaction, information gathering, and explanation and planning. Overall communication skills and empathy were further divided as examiner- or simulated patient-assessed. The overall quality of evidence ranged from moderate to very low, and there was high, unexplained heterogeneity.
Overall, interventions had positive effects on most outcomes, but generally small effect sizes and evidence quality limit the conclusions that can be drawn. Communication skills interventions in comparison to usual curricula or control may improve both overall communication skills (standardised mean difference (SMD) 0.92, 95% confidence interval (CI) 0.53 to 1.31; 18 studies, 1356 participants; I^2 = 90%; low-quality evidence) and empathy (SMD 0.64, 95% CI 0.23 to 1.05; 6 studies, 831 participants; I^2 = 86%; low-quality evidence) when assessed by experts, but not by simulated patients. Students' skills in information gathering probably also improve with educational intervention (SMD 1.07, 95% CI 0.61 to 1.54; 5 studies, 405 participants; I^2 = 78%; moderate-quality evidence), but there may be little to no effect on students' rapport (SMD 0.18, 95% CI -0.15 to 0.51; 9 studies, 834 participants; I^2 = 81%; low-quality evidence), and effects on information giving skills are uncertain (very low-quality evidence).

We are uncertain whether experiential interventions improve overall communication skills in comparison to didactic approaches (SMD 0.08, 95% CI -0.02 to 0.19; 4 studies, 1578 participants; I^2 = 4%; very low-quality evidence). Electronic learning approaches may have little to no effect on students' empathy scores (SMD -0.13, 95% CI -0.68 to 0.43; 3 studies, 421 participants; I^2 = 82%; low-quality evidence) or on rapport (SMD 0.02, 95% CI -0.33 to 0.38; 3 studies, 176 participants; I^2 = 19%; moderate-quality evidence) compared to face-to-face approaches. There may be small negative effects of electronic interventions on information giving skills (low-quality evidence), and effects on information gathering skills are uncertain (very low-quality evidence).

Personalised/specific feedback probably improves overall communication skills to a small degree in comparison to generic or no feedback (SMD 0.58, 95% CI 0.29 to 0.87; 6 studies, 502 participants; I^2 = 56%; moderate-quality evidence). There may be small positive effects of personalised feedback on empathy and information gathering skills (low quality), but effects on rapport are uncertain (very low quality), and we found no evidence on information giving skills.

We are uncertain whether role-play with simulated patients outperforms peer role-play in improving students' overall communication skills (SMD 0.17, 95% CI -0.33 to 0.67; 4 studies, 637 participants; I^2 = 87%; very low-quality evidence). There may be little to no difference between effects of simulated patient and peer role-play on students' empathy (low-quality evidence) with no evidence on other outcomes for this comparison.

Descriptive syntheses of results that could not be included in meta-analyses across outcomes and comparisons were mixed, as were effects of different interventions and comparisons on specific communication skills assessed by the included trials. Quality of evidence was downgraded due to methodological limitations across several risk of bias domains, high unexplained heterogeneity, and imprecision of results.
In general, results remain consistent in sensitivity analysis based on risk of bias and adjustment for clustering. No adverse effects were reported.

Authors' conclusions

This review represents a substantial body of evidence from which to draw, but further research is needed to strengthen the quality of the evidence base, to consider the long-term effects of interventions on students' behaviour as they progress through training and into practice, and to assess effects of interventions on patient outcomes. Efforts to standardise assessment and evaluation of interpersonal skills will strengthen future research efforts.

PLAIN LANGUAGE SUMMARY

Communication skills interventions for medical students

What are the best ways to teach medical students how to talk to patients?

Teaching interpersonal skills

We use interpersonal skills every day when we talk and interact with other people. Interpersonal skills include communication skills such as listening, speaking, and asking questions. Good communication between doctor and patient is vital for all medical consultations. Doctors need to build relationships, show empathy, gather information, explain concepts, and plan treatment with their patients.

Medical students need to prove they are competent in interpersonal and communication skills before they graduate. Approaches to teaching medical students these skills include face-to-face teaching, online courses, programmes that give students personalised or tailored feedback, and use of role-play with peers or with actors playing the role of the patient.

Why we did this Cochrane Review

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Medical schools and universities around the world use different approaches to teach interpersonal and communication skills. We wanted to find out which types of educational programmes work best.

What did we do?

We searched for studies that assessed educational programmes to teach interpersonal and communication skills to medical students.

Search date: we included evidence published up to September 2020.

What we found

We found 90 relevant studies involving 10,124 students, conducted in the USA and in countries in Europe, the Middle East, and Asia Pacific. These studies assessed educational programmes to teach interpersonal and communication skills, including individual role-play and feedback, and large group demonstrations. Programmes were delivered by face-to-face teaching, by video, or online. In most studies, results were assessed immediately after the educational programme concluded, and up to 12 months later.

We compared the results of studies conducted to find out how the different programmes affected:

· overall communication skills;

· understanding of what another person is feeling (empathy);

· relationship building or understanding one another's feelings or ideas (rapport);

· gathering of information, including patients' satisfaction, understanding, or views; and

· explaining and planning (giving of information).

What are the main results of our review?

Comparing interpersonal programmes against the usual educational programmes or being on a waiting list for an interpersonal programme showed that these programmes:

· may slightly improve students' overall communication skills (evidence from 18 studies, with 1356 students) and empathy (6 studies; 831 students);

· probably improve skills in gathering information about the patient's views (5 studies, 405 students); but

· may have little to no effect on rapport skills (9 studies, 834 students).

We are uncertain about how the interpersonal programmes affected skills in giving information (5 studies, 659 students).

We are uncertain how online or self-taught programmes affected students' overall communication skills (4 studies, 1578 students) or skills in gathering information (1 study, 164 students) compared with face-to-face teaching.

Online or self-directed programmes compared with face-to-face teaching:

· may have little to no effect on students' empathy skills (3 studies, 421 students);

· probably have little to no effect on rapport skills (3 studies, 176 students); but

· may slightly reduce skills in giving information (1 study, 122 students).

Programmes that give students personalised or tailored feedback compared with those that involve general or no feedback:

· probably slightly improved overall communication skills (6 studies, 502 students); and

· may slightly improve empathy skills (1 study, 66 students) and skills in gathering information (1 study, 48 students).

We are uncertain how programmes with personalised or tailored feedback affected rapport skills (1 study, 190 students). No studies reported results for effects on skills in giving information.

We are uncertain whether programmes involving role-play with people acting as patients improved students' overall communication skills (4 studies, 637 students) compared with programmes involving role-play with peers. Role-play with people acting as patients may slightly improve empathy skills (2 studies, 213 students). No studies reported results for how role-play with people acting as patients affected skills in rapport or gathering and giving of information.

No studies reported any unwanted effects for any of the education programmes assessed.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

How confident are we in our results?

We are moderately confident that interpersonal education programmes and programmes involving personalised, specific feedback probably improve overall communication skills.

We are not confident in our results for the other types of programmes. Some studies had limitations in the ways they were designed, such as how students taking part were assigned to different groups. Other studies had widely varying results, and some had short follow-up times. Further research is likely to change our results.

Key messages

Interpersonal education programmes for medical students had positive effects on most of the interpersonal skills we looked at, although these effects were small and our confidence in some of our results is low.

Programmes that include personalised feedback probably improve medical students' overall communication skills more than programmes that involve general or no feedback.

Online or self-directed programmes may make little to no difference in improving skills in empathy or rapport compared with face-to-face teaching.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings

Comparison 1. Communication skills intervention compared with control or usual curriculum for improving medical students' interpersonal communication skills

Population: pre-registration medical students

Settings: university campus, clinical setting, classroom

Intervention: educational intervention for interpersonal communication

Comparison: usual curriculum or waitlist control

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control or usual curriculum	Communication skills intervention				
Overall communication skills Different scales used by different studies: total scores vary across communication items or average score on a Likert scale Based on assessments occurring immediately and up to 12 months after intervention delivery		Mean communication score in the intervention groups was 5.75 higher (3.31 to 8.19 higher)[a]		1356 (18)	⊕⊕⊝⊝ **low** [b]	Scores estimated using a standardised mean difference of 0.92 (95% CI 0.53 to 1.31) Generally small differences across a range of scales (e.g. 1/10, up to 5/100) Four further studies could not be pooled, reporting mixed results (2 favouring the intervention, 1 reporting no difference between groups, and 1 providing no data)
Empathy Different scales used by different studies: average empathy score across several items or average score on a 5- to 9-point Likert scale		Mean empathy score in the intervention groups was 0.53 higher (0.19 to 0.87 higher)[c]		831 (6)	⊕⊕⊝⊝ **low** [d]	Scores estimated using a standardised mean difference of 0.64 (95% CI 0.23 to 1.05) Generally small differences across a range of scales (e.g. < 1/5)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Based on assessments occurring immediately and up to 12 months after intervention delivery					Two further studies reported data that could not be pooled, 1 reporting results in favour of the intervention, the other not reporting a direct between-group comparison
Relationship building/rapport Different scales used by different studies: average empathy score across several items or average score on a Likert scale Based on assessments occurring immediately and up to 12 months after intervention delivery		Mean rapport score in the intervention groups was 0.12 higher (0.10 lower to 0.34 higher)[e]		834 (9)	⊕⊕⊝⊝ **low** [f]	Scores estimated using a standardised mean difference of 0.18 (95% CI -0.15 to 0.51) Generally small differences across a range of scales (e.g. < 1/12, 1/8)
Information gathering about patient perspectives/concerns Different scales used by different studies, some dichotomous outcomes Based on assessments occurring immediately and up to 12 months after intervention delivery		Mean score for information gathering about patient perspectives in the intervention groups was 1.12 higher (0.64 to 1.62 higher)[g]		405 (5)	⊕⊕⊕⊝ **moderate** [h]	Scores estimated using a standardised mean difference of 1.07 (95% CI 0.61 to 1.54) Effect sizes and differences between groups varied across scales from 4/100 to 2.6/4
Explanation and planning (information giving) Different scales used by different studies, some dichotomous outcomes Based on assessments occurring immediately and up to 12 months after intervention delivery	These studies reported some favourable outcomes of intervention groups relative to control groups, but results were mixed both within and between studies, and effect sizes were generally small			695 (5)	⊕⊝⊝⊝ **very low** [i]	Small differences between groups (e.g. 0.07/4, 7%) and large standard deviations (e.g. 0.84, > 16%)
Specific communication techniques Different scales used by different studies, some dichotomous outcomes	Both positive and no effects were reported for the intervention compared with control or usual care, on a range of specific communication skills. Findings were mixed within and across studies, with most reporting some positive effects of the intervention on some, but not all, specific skills assessed			371 (5)	⊕⊕⊝⊝ **low** [j]	Small differences between groups (e.g. 0.17/4, < 0.1/5) and large standard deviations (e.g. 0.72, 0.07)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Based on assessments occurring immediately and up to 12 months after intervention delivery		
Adverse events	No studies reported on this outcome	

*Assumed risks are not provided due to the wide variation in scales used. The **corresponding risk** (and its 95% confidence interval) is based on a representative study, as outlined in the footnotes.
CI: confidence interval.

GRADE Working Group grades of evidence.
High quality: further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: we are very uncertain about the estimate.

[a]We have used results (mean scores and standard deviation 6.23) from Bosse 2012 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a common outcome measure, based on the Calgary-Cambridge observation guides, and low risk of bias. Outcomes in this study were rated on visual analogue scales that range from 100 = completely agree to 1 = strongly disagree.
[b]Downgraded one level due to high or unclear risk of bias ratings in several domains; downgraded one level due to substantial heterogeneity (high I^2 value, wide variation in effect estimates and some CIs with no overlap, not explained by subgroup analysis).
[c]We have used results (mean scores and standard deviation 0.83) from Evans 1989 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a common outcome measure, aligned with the Calgary-Cambridge observation guides, and low risk of bias. The possible score on each item was within a range from 0 to 8.
[d]Downgraded one level due to serious heterogeneity (high I^2 value, wide variation in effect estimates, and some CIs with no overlap); downgraded one level due to suspected publication bias (asymmetrical funnel plot). Heterogeneity remained moderate in subgroup analysis of outcomes assessed by experts and by SPs.
[e]We have used results (mean scores and standard deviation 0.67) from Legg 2005 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a common outcome measure, based on a modified version of the Calgary-Cambridge observation guides, and low risk of bias.
[f]Downgraded one level due to serious heterogeneity (high I^2 value, wide variation in effect estimates, and some CIs with no overlap); downgraded one level due to imprecision of results (CI crosses zero; therefore the true effect could be either a benefit or a harm). Heterogeneity was reduced in subgroup analysis of outcomes assessed by experts but remained high in those assessed by SPs.
[g]We have used results (mean scores and standard deviation 1.05) from Evans 1989 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a common outcome measure, aligned with the Calgary-Cambridge observation guides, and low risk of bias. The possible score on each item was within a range from 0 to 8.
[h]Downgraded one level due to considerable heterogeneity (high I^2 value).
[i]Downgraded one level due to high risk of bias across several domains in several studies; downgraded one level due to heterogeneity of results and methods; downgraded one level due to imprecision of results (small differences and large standard deviations).
[j]Downgraded one level due to heterogeneity of results and methods; downgraded one level due to imprecision of results (small differences and large standard deviations).

Summary of findings 2. Summary of findings

Comparison 2. Communication skills intervention compared with communication skills intervention using an alternate mode for improving medical students' interpersonal communication skills

Population: pre-registration medical students

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Settings: university campus, clinical setting, classroom

Intervention: educational intervention for interpersonal communication using alternate mode (e.g. online, experiential)

Comparison: educational intervention for interpersonal communication using traditional mode (face-to-face or didactic)

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Traditional communication intervention (face-to-face or didactic)	Alternate communication intervention (experiential, e-learning, or video-based)				
Overall communication skills Different scales used by different studies: total scores vary across communication items or average score on a Likert scale Based on assessments occurring immediately and up to 12 months after intervention delivery		Mean overall communication score in the experiential learning groups was 0.32 higher (0.08 lower to 0.75 higher) than in the didactic group[a]		1578 (4)	⊕⊝⊝⊝ **very low** [b]	Scores estimated using a standardised mean difference of 0.08 (95% CI -0.02 to 0.19) Generally small differences between groups (e.g. 0.44/20, 1.5/36) and large standard deviations
Empathy Different scales used by different studies: average empathy score across several items or average score on a 5- to 9-point Likert scale Based on assessments occurring immediately and up to 12 months after intervention delivery		Mean empathy score in the e-learning/video groups was 0.13 lower (0.69 lower to 0.44 higher) than in the face-to-face group[c]		421 (3)	⊕⊕⊝⊝ **low** [d]	Scores estimated using a standardised mean difference of -0.13 (95% CI -0.68 to 0.43) Varied scales and effect sizes, generally small with large standard deviations
Relationship building/rapport		Mean rapport score in the e-learning/online groups was 0.01 higher (0.18 low-		176 (3)	⊕⊕⊕⊝ **moderate** [f]	Scores estimated using a standardised mean difference of 0.02 (95% CI -0.33 to 0.38)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Different scales used by different studies: average empathy score across several items or average score on a Likert scale Based on assessments occurring immediately and up to 12 months after intervention delivery	er to 0.21 higher) than in the face-to-face group[e]			Small differences between groups (e.g. 0.1/4)
Information gathering about patient perspectives/concerns	Only 1 study in this comparison reported on information gathering. Use of video modules was associated with slightly higher scores on some, but not all, areas of student performance	164 (1)	⊕⊖⊖⊖ **very low** [g]	Differences between groups were small, ranging from 1% to 8%, and standard deviations ranged from 9% to 18%
Explanation and planning (information giving)	Only 1 study in this comparison reported on explanation and planning. Students exposed to a personal digital assistant intervention may perform less well than those exposed to paper-based guidance plus personal digital assistant only	122 (1)	⊕⊕⊖⊖ **low** [h]	Differences between groups were small: 7% at 1 month post intervention and 2% at the end of the year with standard deviations over 16%
Specific communication skills	Results varied across outcomes both between and within studies, with overall little evidence for any appreciable difference between online and in-person learning on specific communication skills Inclusion of patients in teaching may have some impact on communication skills, but the effect of varied tutor training or tutor qualification level is unclear	427 (6)	⊕⊖⊖⊖ **very low** [i]	Differences between groups were generally small, ranging from 0.2% to 7%, and standard deviations were large, at 5.5% and 18.6% for these examples
Adverse events	No studies reported on this outcome.			

*Assumed risks are not provided due to the wide variation in scales used. The **corresponding risk** (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.

GRADE Working Group grades of evidence.
High quality: further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: we are very uncertain about the estimate.

[a]We have used results (mean scores and standard deviation 3.96) from Roche 1997 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a common outcome measure, aligned with the Calgary-Cambridge observation guides. 18 interactional skills items were rated, leading to a maximum possible score of 36.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

[b]Downgraded two levels due to methodological limitations (with high risk of bias across several domains in 3 of 4 studies); downgraded one level due to imprecision of results (CI crosses zero; therefore the true effect could be either a benefit or a harm).

[c]We have used results (mean scores and standard deviation 1.02) from Kaltman 2018 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a well-defined outcome measure and low risk of bias. Measures were based on coding (counts) of specified communication behaviours including empathy.

[d]Downgraded one level for serious heterogeneity (high I^2 value, wide variation in effect estimates, not explained by subgroup analysis); downgraded one level due to imprecision of results (CI crosses zero; therefore the true effect could be either a benefit or a harm).

[e]We have used results (mean scores and standard deviation 0.56) from Liu 2016 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a well-defined outcome measure and low risk of bias. The Student Patient Observed Communication Assessment (SOCA) form was adapted from the Calgary Cambridge Guides. Maximum scores are not clear.

[f]Downgraded one level due to imprecision of results (CI crosses zero; therefore the true effect could be either a benefit or a harm) and small sample size.

[g]Downgraded two levels due to methodological limitations (quasi-RCT); downgraded one level due to imprecision (single study, small sample size).

[h]Downgraded two levels due to imprecision (single study, small sample size).

[i]Downgraded one level due to methodological limitations (significant risk of bias across several domains); downgraded two levels due to substantial heterogeneity within and between studies.

Summary of findings 3. Summary of findings

Comparison 3. Intervention involved personalised or tailored feedback compared with intervention involving generic or no feedback for improving medical students' interpersonal communication skills

Population: pre-registration medical students

Settings: university campus, clinical setting, classroom

Intervention: educational intervention for interpersonal communication involving tailored or personalised feedback

Comparison: educational intervention for interpersonal communication involving generic or no feedback

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Generic or no feedback	Tailored feedback				
Overall communication skills Different scales used across studies: total scores vary across communication items or average score on a Likert scale		Mean communication score in tailored feedback groups was 1.19 higher (0.60 higher to 1.79 higher)[a]		502 (6)	⊕⊕⊕⊖ **moderate**[b]	Scores estimated using a standardised mean difference of 0.58 (95% CI 0.29 to 0.87) Small effect sizes and differences between groups (e.g. 0.18/4, 0.48/8, 0.96/15) One further study reported no overall differences between groups

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Based on assessments occurring immediately and up to 12 months after intervention delivery				
Empathy	One study reported that the group receiving specific feedback had slightly higher scores on handling patients' emotions than the general feedback group. With small effect size and a single study, certainty is low	66 (1)	⊕⊕⊝⊝ **low**[c]	Difference between groups was 0.54 on a 4-point Likert scale (3.1 vs 2.56), and standard deviations were larger than this difference at 0.89 and 1.32
Relationship building/Rapport	One study reported small positive effects of self-feedback and peer feedback on some, but not all measures, in comparison to teacher feedback	190 (1)	⊕⊝⊝⊝ **very low**[d]	Difference between groups was 0.36 on a 4-point Likert scale (3.82 vs 3.46), and standard deviations were larger than this difference at 0.53 and 0.8
Information gathering about patient perspectives/concerns	One study reported consistent positive effects of increasingly tailored feedback including video feedback on information gathering, compared with control	48 (1)	⊕⊕⊝⊝ **low**[e]	Differences between intervention groups were 0.9 and 4.1 out of 44, and between the most intensive intervention and control, the difference was 12.6/44. Standard deviations ranged from 12.6 to 7.2
Explanation and planning (information giving)	No studies reported on this outcome			
Specific communication skills	One study reported consistent positive effects of video-assisted feedback in comparison to verbal feedback across a range of specific skills	100 (1)	⊕⊕⊝⊝ **low**[f]	The difference between groups of 0.72 is equivalent to the standard deviation for both groups
Adverse events	No studies reported on this outcome			

*Assumed risks are not provided due to the wide variation in scales used. The **corresponding risk** (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.

GRADE Working Group grades of evidence.
High quality: further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: we are very uncertain about the estimate.

[a]We have used results (mean scores and standard deviation 2.06) from Ruesseler 2017 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a well-defined outcome measure and low risk of bias. Verbal and non-verbal communication was assessed using a checklist with a maximum of 15 points.
[b]Downgraded one level due to high or unclear risk of bias across multiple domains.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

[c]Downgraded two levels due to imprecision (single, small study).
[d]Downgraded one level due to methodological limitations (significant risk of bias across several domains), downgraded two levels due to imprecision (single, small study).
[e]Downgraded two levels due to imprecision (single, small study).
[f]Downgraded two levels due to imprecision (single, small study).

Summary of findings 4. Summary of findings

Comparison 4. Intervention involving role-play with simulated patients compared with intervention involving role-play with peers for improving medical students' interpersonal communication skills

Population: pre-registration medical students

Settings: university campus, clinical setting, classroom

Intervention: educational intervention for interpersonal communication using simulated patient role-play

Comparison: educational intervention for interpersonal communication using peer role-play

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Peer role-play	Simulated patient role-play				
Overall communication skills Different scales used by different studies: total scores vary across communication items or average score on a Likert scale Based on assessments occurring immediately and up to 12 months after intervention delivery		Mean communication score in simulated patient role-play groups was 0.56 higher (1.10 lower to 2.22 higher)[a]		637 (4)	⊕⊖⊖⊖ **very low** [b]	Scores estimated using a standardised mean difference of 0.17 (95% CI -0.33 to 0.67)[b] Small effect sizes and differences between groups (e.g. 1.45/15, 0.2/45)
Empathy	Two studies reported no differences between simulated patient and peer role-play			213 (2)	⊕⊕⊖⊖ **low** [c]	Small differences between groups (0.2/5 and 0.02/6) and large standard deviations (1.14, 0.95)
Relationship building/Rapport	No studies reported on this outcome within this comparison					
Information gathering about patient perspectives/concerns	No studies reported on this outcome within this comparison					

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Explanation and planning (information giving)	No studies reported on this outcome within this comparison			
Specific communication skills	Only 1 reported on this outcome within this comparison. Very small differences in open questions but not in facilitative behaviours were reported for those receiving simulated patient practice and feedback, compared with faculty or peer practice/feedback	120 (1)	⊕⊕⊝⊝ **low** [d]	Small differences between groups (0.97/5; 3.194 vs 3.097)
Adverse events	No studies reported on this outcome			

*Assumed risks are not provided due to the wide variation in scales used. The **corresponding risk** (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.

GRADE Working Group grades of evidence.
High quality: further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: we are very uncertain about the estimate.

[a]We have used results (mean scores and standard deviation (3.32) from Bosse 2012 to illustrate effect sizes in terms of measures used in that study. This study was chosen for its use of a common outcome measure, based on the Calgary-Cambridge observation guides, and low risk of bias. Outcomes in this study were rated on visual analogue scales that range from 100 = completely.
[b]Downgraded two levels for methodological quality due to significant proportion of quasi-RCTs and high risk of bias across several domains; downgraded one level for serious heterogeneity (high I^2 value, wide variation in effect estimates, not explained by subgroup analysis); downgraded one level due to imprecision of results (CI crosses zero; therefore the true effect could be either a benefit or a harm).
[c]Downgraded one level due to heterogeneity within and between studies; downgraded one level due to imprecision (small number of participants from only two studies).
[d]Downgraded two levels due to imprecision (single, small study).

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

BACKGROUND

Description of the condition

The importance of effective communication between clinicians and patients was highlighted in the Institute of Medicine's *Crossing the Quality Chasm* report (Richardson 2001). This report proposed free and open sharing of knowledge between patients and clinicians as one of ten principles for redesigning the healthcare system to ensure delivery of optimal patient-centred care. Effective doctor-patient communication has been shown to positively affect patient satisfaction (Pollak 2011; Street 2009), decision-making (NHS 2010), and treatment adherence (Street 2009; Zolnierek 2009), and a range of patient outcomes such as blood pressure and emotional health (Stewart 1995b), as well as doctors' job satisfaction (Maguire 2002). Information gathered as part of the medical history is critical to the establishment of an accurate diagnosis, with most information used to form a diagnosis gathered whilst the medical history is collected (Peterson 1992). Subsequent steps in the medical consultation facilitate patient education and shared decision-making, with associations established between patient-centred communication, therapeutic alliance, and adherence to treatment (Pinto 2012; Thompson 2016). It has been estimated that a medical doctor will conduct approximately 200,000 medical consultations during his or her working career (Silverman 2013). Thus, it is incumbent upon educators to appropriately train medical students to ensure that, upon graduating, consultations are conducted effectively via appropriate patient-centred communication (Simpson 1991).

Medical consultations should be a joint, collaborative effort between doctor and patient, using patient-centred communication to acknowledge and understand the patient's desire for information, shared decision-making, and discussions of care (Stewart 2001). Although global agreement on definitions of the terms is elusive, there is broad agreement about the concepts and importance of patient-centred care and patient-centred consultation. These concepts are integral to the display of respect for patients, and thus, along with patient involvement in care systems, should be regarded as an ethical and democratic right (Gregory 2007). Definitions generally include elements of a biopsychosocial perspective of illness; consideration of each individual's personal meaning of illness; sensitivity to patients' preferences for information and shared decision-making; and development of a therapeutic relationship between doctor and patient (Mead 2002).

Patient-centred medical consultations call upon a set of skills that are considered both teachable and learnable by medical professionals at any point along their career trajectory (Aspegren 1999). We will use the term 'interpersonal communication' to refer to these skills that facilitate patient-centred communication and care. Interpersonal communication can be defined as communication that occurs from one individual to another (dyadic or small group), is non-mediated (face-to-face), and is shaped by individual characteristics, social roles, and relationships between the people involved (Hartley 1999). Interpersonal communication is the process by which we establish a communicative relationship and exchange messages to establish shared goals and understandings (Burleson 2010; Hargie 2011). We will refer to communication interventions as those aiming to improve the skills associated with such communication.

Significant progress has been made in the development and evaluation of formal curricula for interpersonal communication in medicine (Aspegren 1999; Smith 2007). Skills in interpersonal communication in clinical contexts are recognised as different from everyday communication skills and should be developed through careful teaching and experiential learning (Benbassat 2009; Silverman 2013). Descriptive data suggest that students find acquiring skills in interpersonal communication challenging for reasons including difficulty relearning or reconditioning engrained communication styles (Lumma-Sellenthin 2009; Macdonald 2002; Royston 1997), difficulty attending to medical and psychosocial needs simultaneously (Aper 2015), lack of exposure to models of patient-centred care (Thistlethwaite 1999), and wide variability between clinical and non-clinical role models (Rees 2002). With absence of up-to-date, high-quality systematic reviews, evidence on effects of communication curricula for improving medical students' skills in conducting effective patient-centred consultations and for improving clinical practice is unclear.

Increasing demand for medical students to be specially trained to communicate effectively and efficiently has seen the emergence of skills associated with interpersonal communication as core graduate competencies in medical training around the world (Australian Medical Council 2012; General Medical Council 2015; Health Professions Council of South Africa 2014; Laidlaw 2009). New doctors are required to possess a range of skills for communicating in different formats such as face-to-face, online, and by telephone, and in different medical contexts, such as doctor-patient consultations, communicating about the patient (e.g. with other medical professionals), and communicating about medicine and science in general (e.g. lectures, conferences). In this review, we focused on the medical consultation, referring to the verbal and non-verbal interaction between physician and patient that occurs during face-to-face encounters. The medical consultation has been described via a range of models, all of which summarise the process as including elements of relationship building, information gathering, information giving, and treatment planning, which can occur in both initial and follow-up encounters (Keller 1994; Kurtz 1998; Makoul 1998; Novack 1992; Stewart 1995a). The overall goal of such encounters is shared understanding of issues and plans, and the specific goals of any individual consultation can vary from diagnosis to understanding the cause of illness, risk, prognosis, benefits and risks of various treatment options, health behaviour change, screening, and any number of other therapeutic and health-promoting activities.

Description of the intervention

Models of the medical consultation and communication training

Teaching and assessing interpersonal communication have been guided by frameworks and models evolving over several decades (reviewed in Boon 1998). The core elements of relationship establishment, information gathering, and patient education have been expanded over time (Lipkin 1995). More recent additions include the need for the doctor to gather information about the patient's understanding of his or her own health (Makoul 2001a), recognition of the influence of clinicians' personal experiences on their interpersonal communications (Windover 2014), and the need to incorporate the electronic health record (Duke 2013). Although the most recent studies have incorporated the use of technology, the fundamental structure and content of these models have

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

not changed since the Kalamazoo Consensus Statement was established in 2001 (Makoul 2001b).

The Kalamazoo Consensus drew upon five contemporary models for doctor-patient communication, namely, the Bayer Institute for Health Care Communication E4 Model (Keller 1994), the Three Function Model/Brown Interview Checklist (Cole 2013; Novack 1992), the Calgary–Cambridge Observation Guide (Kurtz 1998), the patient-centred clinical method (Stewart 1995b), and the SEGUE Framework for teaching and assessing communication skills (Makoul 1998). Hence, the Kalamazoo Consensus brought together a comprehensive set of skill competencies (Makoul 2001b). Specific communication tasks as well as knowledge, skills, and attitudes associated with the following essential elements are listed in the consensus statement: build the relationship, open the discussion, gather information, understand the patient's perspective, share information, reach agreement on problems and plans, and provide closure.

Subsequent to the synthesis of communication tasks in the Kalamazoo Consensus Statement, Kurtz and colleagues expanded their Calgary-Cambridge Guide to more explicitly connect with the process of doctor-patient consultations (Kurtz 2003). This guidance clarified and expanded upon the specific skills used at each step of the medical consultation process. For example, when gathering information, necessary skills include using open and closed questions appropriately, structuring, clarifying and summarising information, picking up verbal and non-verbal cues from the patient, facilitating patients' responses verbally and non-verbally, and listening attentively. The marriage of process and content proposed by Kurtz and colleagues represents a true amalgamation of communication skills (performance of specific tasks and behaviours) with interpersonal skills (relational, process-oriented skills such as respect, empathy, and considering the patient's perspective) required to establish a therapeutic relationship (Kurtz 2003; Makoul 2001b).

In this review, we included all interventions for medical students that specifically target the skills associated with what we have defined as interpersonal communication. These skills are likely to include appropriate use of open and closed questions, active listening, picking up on verbal and non-verbal cues, facilitating patients' responses, eliciting patients' concerns, considering patients' ideas, discerning concerns and expectations (gathering and understanding patients' perspectives), working in partnership with patients to explain and plan, and making shared decisions, maintaining structure of the consultation, and clarifying and summarising information.

Training methods

Interpersonal communication training for medical students takes a range of forms depending upon available resources, the current training level of students (e.g. undergraduate degree, postgraduate training programme), and the context of learning (on campus, in clinical placement, online). Interventions can be categorised as delivered face-to-face or in a self-directed format. Face-to-face interventions typically are delivered as lectures for large groups of students or workshops for small groups. The latter can provide opportunities for participants to practise communication with real or simulated patients or their peers. Face-to-face interventions also enable feedback provided by peers, facilitators, and/or patients (whether real or simulated). These interventions may be based around live role-plays or feedback on videotaped consultations (Deveugele 2005; Maguire 1986). Self-directed interventions are those by which the learner receives individual training based on written or audiovisual materials in hard copy or by online or e-learning format such as online video demonstrations (Cook 2010).

Assessment of interpersonal communication

The Kalamazoo Consensus outlined three methods of assessing interpersonal communication: checklists (observer ratings); patient surveys; and examinations (of knowledge and perceptions through traditional written questions or questions linked to a stimulus such as a video vignette). Both checklists and patient surveys can be used in assessment of interactions with real or simulated patients, can be completed live or based on recorded interactions, and can be used in formative or summative assessments such as objective standardised clinical examinations (OSCEs) (Duffy 2004). Given the variation in validity and reliability among techniques, the assessment method utilised affects the capacity to compare different intervention studies. In this review, we categorised each study based on the method of assessment (observer ratings or survey/examination scores) and the nature of the observed consultation (real patient, simulated patient, live, recorded). Given our focus on behaviour change, we did not include data obtained through student examinations or surveys. When studies used more than one of the included methods of assessment, we extracted data for each outcome separately and compared this information with other outcome data as appropriate.

How the intervention might work

Interventions to improve interpersonal communication skills of medical students aim to produce doctors capable of delivering effective, safe, patient-centred health care when they enter the workforce. Education-based interventions work by bringing about change in learners' attitudes, increasing their knowledge, and increasing their competence in performing particular skills. In the case of medical consultation skills, educational interventions are likely to improve learners' skills and knowledge through educator modelling and feedback and experiential learning, with opportunities to practise, reflect, and receive constructive feedback, draw upon knowledge and previous experience, and learn in a self-directed fashion (Kaufman 2003). Although the highest level of evaluation of learning is application of skills in clinical practice leading to improved patient outcomes, the only immediately measurable outcomes for medical students are improvements in skill, knowledge, attitudes, and confidence (Kirkpatrick 1996; Naugle 2000; Smidt 2009). Student learning outcomes can be conceptualised hierarchically (Alliger 1989). At the very least, participation in a communication intervention should increase knowledge of patient-centred approaches to communication. In addition, these interventions should increase confidence in undertaking effective doctor-patient consultations. However, the ultimate goal of communication interventions should always be to improve actual behavioural skills for application in evidence-based doctor-patient consultations. These behavioural outcomes can be assessed by methods outlined in the Kalamazoo Consensus Statement and form the basis of the outcome measures assessed in this review.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Why it is important to do this review

Given community and professional concerns regarding the physical, emotional, and financial impact of poor communication by medical practitioners, there is a critical need to evaluate communication training programmes in medical education. To date, the effectiveness of interventions for improving medical students' interpersonal communication has not been demonstrated unequivocally. Moreover, significant variability has been noted in communication curricula across medical schools (Hargie 2010; Hoffman 2004), and new methods have become increasingly popular since publication of previous reviews (Lanken 2015). Given the rapid evolution of innovative teaching and learning approaches, it is timely to review the effectiveness of approaches that have been utilised. Thus, there is a need to determine:

- the evidence base for communication interventions for medical students;
- teaching and learning approaches associated with improvements in medical students' interpersonal communications;
- the most effective approaches to teaching medical communication in the context of pragmatic limitations of medical curricula; and
- gaps in knowledge about communication interventions for medical students to guide future teaching and research endeavours.

This review provides necessary guidance to medical educators and medical education-accrediting bodies regarding the most effective communication programmes in medical curricula and identifies the resources needed for successful presentation of these programmes.

Determining the evidence base for communication interventions for medical students

A number of completed and ongoing reviews have sought to examine the effectiveness of communication training programmes in medicine (Aspegren 1999; MacDonald-Wicks 2012; Smith 2007; Van Nuland 2005). Aspegren 1999 reviewed 83 randomised, quasi-randomised, and non-randomised trials and descriptive studies of communication training for medical students and concluded that teaching interpersonal communication to medical students can improve students' ability to undertake doctor-patient consultations. However, the methodological quality of included studies was not adequately assessed, and inclusion of non-randomised trials and descriptive studies limited the extent to which improvements in interpersonal communication could be attributed to the interventions described. Smith 2007 identified 24 randomised controlled trials (RCTs) that were available from 1977 to 2005 and conducted meta-analyses on 15 trials that met study inclusion criteria. Smith 2007 included only RCT interventions; however, given the settings in which these interventions are expected to be delivered (e.g. university classrooms, hospital clinics), it is possible that other study designs (such as cluster-RCTs and quasi-RCTs) may also be relevant. In addition, a preliminary search of research undertaken since 2005 suggests that at least 30 additional RCTs of communication training for medical students have been published since the Smith 2007 review.

The authors of a planned, but subsequently withdrawn, Cochrane Review had intended to assess the effects of communication training programmes specifically for general practice (GP) trainees, while excluding studies involving medical students (Van Nuland 2005). GP trainees are completing their training for specialisation and, as such, have advanced in their training beyond the basic medical degree. The authors of another yet to be completed review, MacDonald-Wicks 2012, plan to review the effectiveness of assessment tools and methods of teaching interpersonal communication to students in the health professions. This review includes students from undergraduate and postgraduate medical, nursing, and allied health programmes (e.g. nutrition and dietetics, occupational therapy, physiotherapy). Given the heterogeneous nature of professional roles and scope of practice, undergraduate training programmes, and student cohorts, it is important to examine medical education separately.

Other reviews have investigated specific topics or skills in medical consultation or have compared specific intervention types. Alelwani 2014 reviewed seven studies involving both students and physicians that focused on skills involved in breaking bad news. Lane 2007 included a range of health professions as well as registered practitioners and students in this review of 23 studies comparing simulated patient and peer role-play. Both Batt-Rawden 2013 and Kelm 2014 reviewed studies that focused on teaching empathy to medical students, residents, fellows, and physicians. Artemiou 2014 reviewed 10 studies evaluating web-based learning for medical students. Similarly, Kyaw 2019 reviewed 12 studies comparing digital learning with traditional approaches. Keifenheim 2015 reviewed 23 studies focused on teaching history taking to medical students and included a range of study designs such as single-group non-randomised studies. This review, in line with others, reported heterogeneous measures and the low quality of studies, and did not differentiate between interview skills, interpersonal skills, and communication skills.

The Smith 2007 review demonstrated that providing structured feedback on participants' performance and engaging in small group discussions were associated with greater improvement in skills compared to other methods (e.g. lectures, clerkship experience, assigned readings). An overview of systematic reviews of strategies for teaching communication skills to qualified doctors reported little evidence for interventions based on lectures or those based on modelling appropriate interpersonal communication for participants (Berkhof 2011). Like the review by Smith and colleagues, stronger evidence was reported for interventions based on role-plays and feedback from educators, particularly when used in combination with self-directed didactic techniques (e.g. written information, review of videos).

Our review differs from these in a number of ways. First, we have ensured methodological rigour by appropriately reviewing identified studies with careful consideration of research design and additional features of methodological quality. Second, we have focused on students completing an undergraduate or graduate-entry medical degree, which will help to ensure that we know what works for these students as compared with more refined samples completing a medical specialisation (e.g. Van Nuland 2005), or students in other allied health programmes (e.g. MacDonald-Wicks 2012). Third, identifying the intervention characteristics associated with improvements in interpersonal communication and separating interpersonal communication skills from other outcomes associated with the process and content of information gathering will help to focus the development of future curricula.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

OBJECTIVES

To assess the effects of interventions for medical students that aim to improve interpersonal communication in medical consultations.

METHODS

Criteria for considering studies for this review

Types of studies

We included:

- randomised controlled trials (RCTs);
- cluster-RCTs (C-RCTs); and
- non-randomised controlled trials including quasi-RCTs (when randomisation was attempted but was inadequate, such as allocation by cohort or rotation).

Types of participants

We included interventions for medical students, defined as people enrolled in an undergraduate or graduate-entry medical degree programme. When studies included participants from multi-disciplinary courses or medical residents or trainees, we included only studies in which subgroup analyses allowed data from medical students to be identified and extracted separately.

We excluded programmes delivered as continuing medical education or postgraduate programmes delivered to registered professionals. Students from other allied healthcare disciplines, such as nursing, physiotherapy, pharmacy, and psychology, were excluded. We did not exclude studies on the basis of age of students, country, setting (classroom, hospital), frequency of intervention contact, duration/dose of intervention, timing, skills targeted, or medical system in which the intervention was delivered.

Types of interventions

We included studies of interventions that aim to improve medical students' interpersonal communication when undertaking medical consultations. Specifically, included interventions were those targeting the communication tasks and skills associated with relationship building, information gathering, and planning and explaining, as well as specific tasks of communication such as listening, using appropriate non-verbal communication, and providing closure (Makoul 2001b). Included interventions were those focused on interpersonal communication in consultations with adult patients. Although our inclusion criteria allowed for dyadic or triadic communication scenarios, no studies were identified that related to triadic communication.

We included the following comparisons.

- Communication intervention for medical students versus no intervention.
- Communication intervention for medical students versus usual training.
- Communication intervention A versus communication intervention B (both for medical students).

Types of outcome measures

We included outcomes assessed using:

- observer ratings of interpersonal communication during student-patient consultations with real patients;
- observer ratings of interpersonal communication during student-patient consultations with simulated patients;
- observer ratings of interpersonal communication during videotaped or audiotaped student-patient consultations with simulated patients; or
- scores obtained on surveys or questionnaires completed by real or simulated patients in relation to learners' interpersonal communications.

From previous reviews, we anticipated that most assessments would be obtained through OSCEs. Although such assessments are less than ideal in terms of the capacity to extrapolate from simulated situations to actual behaviours used in real patient consultations, they are the most widely recognised and adopted proxy throughout medical education internationally. Given the inherent challenges of assessment in education interventions, these outcomes are regarded as best practice. Indeed, all included studies reported outcomes obtained through some form of observed interaction. These were not all labelled as OSCEs by study authors but could be categorised as such due to their objective, structured nature.

Primary outcomes

The primary outcome was interpersonal communication at any point in a medical consultation, including actual behaviour or patient satisfaction with said behaviour. Elements of interpersonal communication were categorised as relating to relationship building, information gathering, planning and explaining, or patient appraisal, with an additional category of specific communication tasks to capture elements that cannot be placed into a single one of these categories.

Trials that measured any of the following outcomes were included, irrespective of whether they were regarded as primary or secondary outcomes in the trial itself. For all outcomes, we assessed terminology and measures used to ensure that any outcomes reported under different labels but pertaining to the measures below were captured (e.g. compassion taken to mean empathy, specific context-related questions taken to be part of information gathering). Therefore, primary outcomes for the review included:

- overall communication skills (e.g. total scores from OSCE rating checklist relating to communication components, a global or overall rating of communication on a single Likert scale);
- empathy (e.g. making an empathetic statement, responding to patients' feelings);
- relationship building/rapport (e.g. rapport or relationship building, communicating without judgement);
- information-gathering skills (e.g. process skills such as appropriately using open-ended and closed questions, eliciting medical content, eliciting patients' concerns or feelings);
- explanation and planning skills (e.g. giving appropriate information about a diagnosis and/or management plan, shared decision-making, acknowledging patient preferences);
- specific communication tasks (e.g. opening the consultation, appropriately using open and closed questions, actively listening, using non-verbal communication including eye contact, performing facilitative behaviours, structuring the consultation, providing closure);

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

- simulated or real patient appraisals of the consultation (e.g. measures of satisfaction, therapeutic alliance, perceived support); and
- in line with Cochrane Methodological Expectations, adverse events (student complaints or referrals to student welfare).

Secondary outcomes

We did not record secondary outcomes.

Search methods for identification of studies

Electronic searches

We searched the following electronic databases without restriction of language or publication status in May 2019 and updated the searches in September 2020.

- Cochrane Central Register of Controlled Trials (CENTRAL; latest issue), in the Cochrane Library.
- MEDLINE OvidSP.
- Embase OvidSP.
- PsycINFO OvidSP.
- Cumulative Index to Nursing and Allied Health Literature (CINAHL) complete EBSCOhost.
- Educational Resource Information Centre (ERIC) OvidSP.

Subject strategies for databases were modelled on the search strategy designed for MEDLINE. When appropriate, these were combined with subject strategy adaptations of the Cochrane highly sensitive search strategy for identifying RCTs and controlled clinical trials (as described in the *Cochrane Handbook for Systematic Reviews of Interventions*, Version 5.1.0, Box 6.4.b; Higgins 2011). The search strategy for MEDLINE is provided in Appendix 1.

Searching other resources

In addition to searching these databases, in late 2019 we also:

- handsearched relevant journals in the field published since 2007 (e.g. *Medical Education; BMC Medical Education, Medical Teacher*);
- contacted study authors to clarify reported information and to seek unpublished data;
- reviewed reference lists of relevant publications and systematic reviews;
- contacted experts in the field to seek advice relating to other relevant studies;
- searched trial registries:
  - World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) (www.who.int/ictrp/en); and
  - ClinicalTrials.gov (clinicaltrials.gov);
- searched relevant grey literature (e.g. dissertation/thesis portals); and
- searched Google Scholar (i.e. review sample of 200 citations).

Data collection and analysis

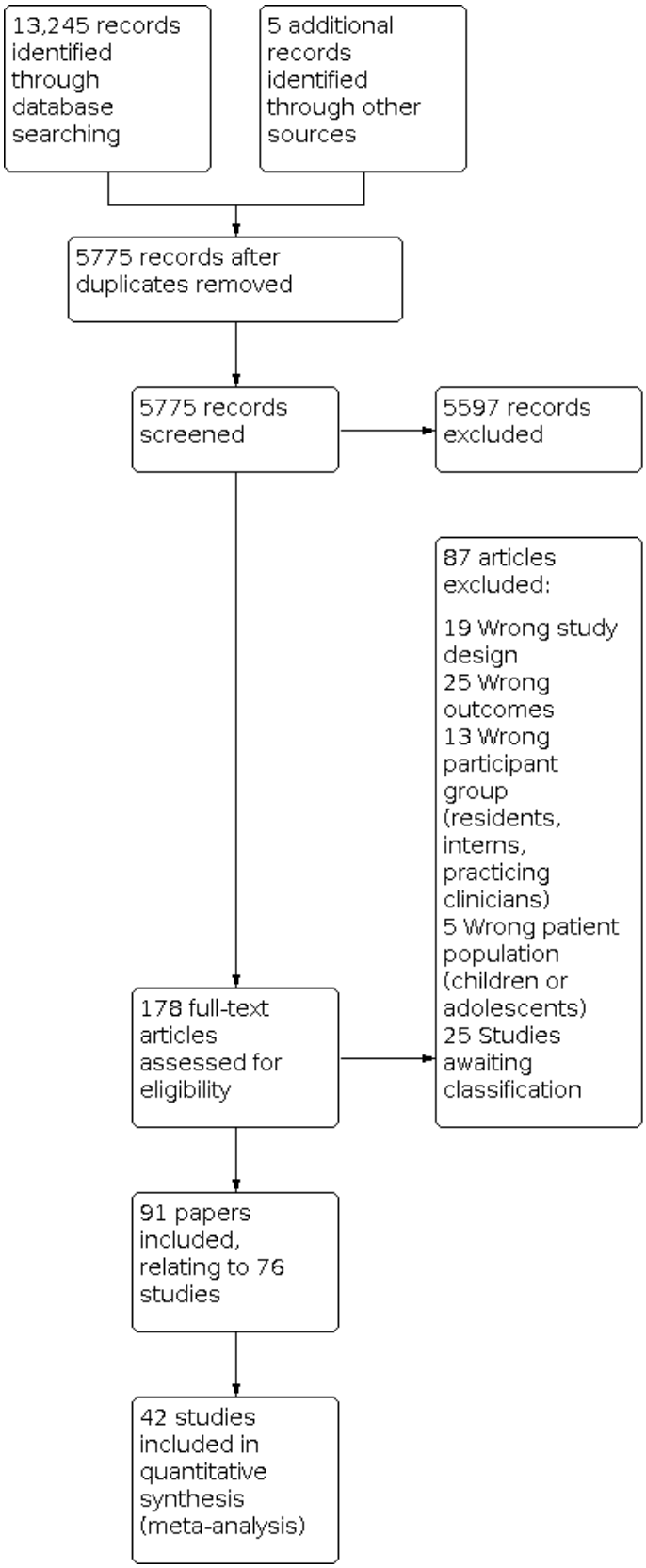
All processes were conducted in keeping with the published protocol (Gilligan 2016). When differences exist between the protocol and the review, they are described in Differences between protocol and review.

Selection of studies

We combined search results in an Endnote library and removed duplicate records. Initial screening of titles and abstracts was conducted independently by two review authors to determine which records met the inclusion criteria. We excluded studies that clearly did not meet the inclusion criteria. We then retrieved in full text any papers identified as potentially relevant by at least one person. Two review authors independently screened full-text articles for inclusion or exclusion. Two other review authors acted as independent arbiters to resolve disagreements regarding study inclusion. Studies that were excluded at this stage are detailed in the Characteristics of excluded studies tables with reasons for exclusion. Three review authors verified the final list of included studies. Disagreements regarding inclusion in the final list of studies were resolved by these three review authors. We report the screening and selection process in an adapted PRISMA flow chart Figure 1.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 1. Study flow diagram.



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Data extraction and management

Pairs of review authors (of CG, PH, ML, SD, SN, MP, HC, and EJ) independently extracted data from all included studies using an a priori defined data extraction form based on the Cochrane Consumers and Communication Data Extraction Template (available at cccrg.cochrane.org/author-resources). We resolved differences in opinion arising during data extraction through discussion and involvement of a third review author for resolution, when required. We extracted the following information: general information about each publication and country of origin, study design and methods, types of experimental and control interventions including deliverers (facilitators) of interventions and characteristics of outcome assessments and assessors, numbers and characteristics of participants, intervention settings, length of follow-up, types of outcomes, outcome data (sample sizes, means standard deviations, odds ratios, confidence intervals as available), and methodological characteristics associated with assessment of risk of bias (randomisation procedures, blinding, data collection procedures, attrition, outcome reporting, and analysis characteristics associated with clustered studies). One review author (CG) entered data into Review Manager 5 software (RevMan 2014), and a second review author (DR) checked these data against extraction sheets.

Assessment of risk of bias in included studies

For each study included in the review, two review authors independently assessed the risk of bias (of CG, BW, CL, and DR) in accordance with the *Cochrane Handbook for Systematic Reviews of Interventions* and the Cochrane Consumer and Communication Guidelines (Higgins 2011; Ryan 2013), which recommend explicit reporting of the following individual elements for RCTs: random sequence generation; allocation sequence concealment; blinding of participants/personnel; blinding of outcome assessment (assessed for each outcome measure); completeness of outcome data (attrition bias; less than 20% loss of participants with no differential attrition between experiment groups was regarded as low risk), selective outcome reporting; and other sources of bias (we considered potential contamination and baseline imbalances between groups, as well as the relationship between intervention deliverers and student participants). We assessed and reported quasi-RCTs as having high risk of bias for the random sequence generation item of the 'Risk of bias' tool. For cluster-RCTs (C-RCTs), we also assessed and reported the risk of recruitment bias, selective recruitment of cluster participants, loss of clusters, incorrect analysis, and compatibility with individually randomised trials (herd effect).

As most outcomes reported were generated from a single assessment (e.g. separate components of an OSCE assessment checklist), blinding was assessed at the individual-study level. We judged each item as being at high, low, or unclear risk of bias, as set out in the criteria provided by Higgins 2011, and we provided a quote from the study report along with a justification for our judgement for each item in the 'Risk of bias' table.

Measures of treatment effect

We calculated treatment effects using RevMan 2014 when possible. Primary outcome measures falling into each of the outcome categories were the basis of measurement of treatment effect. In a post-hoc process, two review authors independently assigned the outcomes reported in each included study to one of the outcome categories. When outcomes were reported for different time periods, we used outcomes closest to the end of intervention delivery as the most homogenous measure across studies.

Dichotomous outcome data

Given that most outcomes were reported using continuous data, dichotomous outcomes reporting on the same measures as those using continuous data were converted to continuous outcomes by calculating the odds ratio (OR) and the confidence interval (CI) for each study, converting these to standardised mean difference (SMD) and standard error (SE), and entering these into a comparison using Generic Inverse Variance (GIV) in RevMan.

Continuous outcome data

We analysed continuous data based on the mean, the standard deviation (SD), and the number of people assessed for both intervention and comparison groups to calculate SMDs (given the use of different measurement scales across studies), each with 95% CIs. When several studies measured the same outcome using different tools, we calculated the SMD and the 95% CI using the inverse variance method in RevMan 2014.

Unit of analysis issues

Trials that we classified as cluster-RCTs were not treated as such by study authors; therefore, no adjustment for clustering was performed. As such, no intracluster correlation (ICC) could be used as a reference and for application to other studies, and unit of analysis issues likely exist in each of these studies. As a result, whenever cluster trials were included in the meta-analyses, sensitivity analyses were performed based on an inflation of the standard error based on a design effect calculated using an ICC of 0.05, 0.1, and 0.2 to represent 5%, 10%, and 20% inflation. The design effect was calculated using these ICC estimates and accounting for average cluster size (1+[average cluster size-1]ICC).

Dealing with missing data

When important summary data or study level characteristics were missing, we attempted to contact the authors of those included studies. When standard deviations were missing from continuous data, we scanned studies for any other statistics (CIs, standard errors, T values, P values, F values) that enabled their calculation. In most cases, given the absence of intention-to-treat analysis, data were analysed as reported. As most studies did not use intention-to-treat analysis, analysed numbers were used in meta-analysis. Levels of missing data were assessed as a source of potential bias.

Assessment of heterogeneity

Assessment of heterogeneity involved inspecting each included study for variability in study populations (baseline characteristics), interventions (structure, focus, mode of delivery), and outcome measures (methods of assessment, scales, and outcome definitions). We considered methodological heterogeneity by inspecting variability in study design and risk of bias. When acceptable homogeneity was found within subgroups (based on participants, interventions, assessment approaches), we conducted meta-analysis for subgroups of studies. We assessed statistical heterogeneity using the Chi^2 test and its P value, by visually inspecting the forest plots and by using the I^2 statistic. A P value of the test lower than 0.10 or an I^2 statistic of at least 50% indicated substantial statistical heterogeneity.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Assessment of reporting biases

We used funnel plots (plots of the effect estimate from each study against the sample size or the effect standard error) to detect possible publication bias. We used tests for funnel plot asymmetry only when a minimum of 10 studies were included in the meta-analysis, as fewer than 10 studies would render the power of the tests too low to distinguish chance from real asymmetry.

Data synthesis

We calculated pooled SMDs (to account for heterogeneity of outcome measures) for each comparison using a random-effects model with a generic inverse variance weighting method to accommodate a mixture of continuous and dichotomous reporting of outcomes (RevMan 2014). When all studies reporting a particular outcome presented continuous data, the continuous model was used for meta-analysis. We selected post-intervention values over changes from baseline data for inclusion in the meta-analysis to reduce the risk of selective reporting and to maximise the number of studies that could be pooled. Similarly, as few studies presented adjusted results, we used unadjusted data for pooling across all studies.

We synthesised studies that provided suitable data for pooling in meta-analysis grouped by outcome. Given the large number of studies and variation in approaches to reporting the a priori defined outcome categories, for the purpose of meta-analysis, outcomes were grouped into categories of empathy, relationship building, overall communication skills (global rating or total score of items from multiple categories relating to communication skills), information gathering, explanation and planning, and simulated patient (SP) perception. Studies could contribute to multiple meta-analyses if they reported eligible outcomes in more than one category. From studies that reported multiple follow-up points, we extracted data from the follow-up period closest to the end of the intervention for inclusion in meta-analyses (based on homogeneity of this time point). Results for the few studies reporting on longer-term outcomes are presented narratively.

In all instances where data could not be pooled in a meta-analysis, we have provided a narrative summary of trial findings according to the review objectives.

Subgroup analysis and investigation of heterogeneity

Given the large numbers of studies and outcomes, when possible we reported outcomes separately when assessed by experts and by simulated (or real) patients. This negated the need for subgroup analysis for this difference. We investigated the extent of heterogeneity through visual examination of forest plots and through use of the Chi^2 statistic, the P value, and the I^2 statistic. When there was evidence of heterogeneity (I^2 statistic > 50%), we investigated the potential source of heterogeneity through subgroup analyses. When possible, we conducted subgroup analyses based on intervention mode of delivery (face-to-face, online), but due to heterogeneity across studies and outcomes, we presented subgroup analyses for length of follow-up and mode of delivery primarily narratively. We did not conduct subgroup analyses based on student year level, as the nature of outcome measurement meant that assessments would be calibrated for expectations relating to each year level, and the use of post-intervention comparison between groups rather than change from baseline renders the student year level irrelevant to the demonstration of improvement in communication skills.

Sensitivity analysis

We performed sensitivity analysis of the main review outcomes after removing trials judged to be at high risk of bias (graded as high risk on three or more 'Risk of bias' measures). For C-RCTs, two or more ratings of high risk on any of the five cluster-specific risk of bias domains contributed one high risk rating to the overall assessment.

'Summary of findings' and assessment of the certainty of evidence

We prepared Summary of findings 1 for the main comparison to present the results of meta-analysis, based on the methods described in Chapter 11 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2011). Summary of findings 2 Summary of findings 3 and Summary of findings 4 present comparisons between alternate intervention approaches. We presented the results of meta-analysis and for narratively described results for primary outcomes including potential harms as outlined in Types of outcome measures. We provided a source and a rationale for each assumed risk cited in the tables, and we used the GRADE system to rank the quality of evidence. GRADE ratings were determined by two review authors working independently.

Ensuring relevance to decisions in health care

The review author team consists of a number of international experts who have contributed to the review including input regarding the role of interpersonal communication in advancement of patient-centred health care and evidence for education and training in communication and interviewing. Small discussion groups have been held with medical students and simulated patients, as well as with subcommittees of the International Association for Communication in Healthcare (EACH). Initially, members reviewed the protocol to ensure that the objectives, inclusion/exclusion criteria, outcome measures, and proposed data extraction items were appropriate and adequately reflected their experiences and priorities. Discussions throughout the review process and at completion of the review have reinforced the findings and limitations of this review, and confirm that the findings reflect the importance and challenges of providing interpersonal communication training to medical students. The group shares views about the need to better align approaches and efforts to teach and evaluate teaching, and to investigate ways to ensure long-term retention of skill improvement.

Summary of findings and assessment of the certainty of the evidence

We prepared Summary of findings 1 of the main comparison to present the results of meta-analysis, based on the methods described in Chapter 11 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2011). Summary of findings 2; Summary of findings 3; and Summary of findings 4 present comparisons between alternate intervention approaches. We present the results of meta-analysis, and for narratively described results, for the primary outcomes including potential harms, as outlined in Types of outcome measures. We provide a source and rationale for each assumed risk cited in the tables, and use the

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

GRADE system to rank the quality of the evidence. GRADE ratings were performed by two reviewers working independently.

Ensuring relevance to decisions in health care

The author team consists of a number of international experts who have contributed to the review including input regarding the role of interpersonal communication in the advancement of patient-centred health care, and the evidence for education and training in communication and interviewing. Small discussion groups have been held with medical students and simulated patients, as well as with sub-committees of the International Association for Communication in Healthcare (EACH). Initially, members reviewed the protocol to ensure that the objectives, inclusion/exclusion criteria, outcome measures, and proposed data extraction items were appropriate and adequately reflected their experiences and priorities. Discussions throughout the review process and at completion of the review have reinforced the findings and limitations of this review, and confirm that the findings reflect the importance and challenges of providing interpersonal communication training to medical students. The group shares views about the need to better align approaches and efforts to teach and evaluate teaching, and to investigate ways to ensure long-term retention of skill improvement.

RESULTS

Description of studies

See Characteristics of included studies, Characteristics of excluded studies, Characteristics of studies awaiting classification, Characteristics of ongoing studies, and features of included studies in Additional Table 1.

Attempts were made to contact the authors of 10 studies to request further information to enable inclusion of the study in the review as a whole, or in the meta-analysis. Responses including additional unpublished data were received from authors of six studies (Engerer 2019; Gartmeir 2015; Ishikawa 2010; Lupi 2012; Schmitz 2018; Solomon 2004).

Results of the search

See the PRISMA flow diagram (Figure 1).

The search strategy resulted in a total of 13,245 records, and we identified a further five studies by checking the reference lists of included studies and relevant systematic reviews. After removal of duplicate records, 5775 records remained. Screening of titles and abstracts revealed 178 studies for full-text review and formal inclusion or exclusion. Of these, 76 papers met the inclusion criteria as primary studies (Allen 1990; Alroy 1984; Bearman 2001; Berney 2017; Betchart 1984; Blatt 2010; Bosse 2012; Bowyer 2006; Bowyer 2010; Buffel du Vaure 2017; Cave 2007; Chibhall 2005; Colletti 2001; Daeppen 2012; Edwardsen 2006; Eells 2002; Engerer 2019; Evans 1989; Evans 1996; Feddock 2009; Filipetto 2006; Foster 2016; Gartmeir 2015; Gerber 1985; Ho 2008; Hobgood 2009; Ishikawa 2010; Kahan 2003; Kaltman 2018; Klein 2000; Knowles 2001; Lee 2015; Legg 2005; Levenkron 1990; Lim 2011; Liu 2016; Lorin 2006; LoSasso 2017; Lupi 2012; Maguire 1977; Maguire 1978; Mason 1988; Meirovich 2016; Moreland 1973; Morrow 2009; Mounsey 2006; Nomura 2017; Ockene 2016; Ozcakar 2009; Palmer 2018; Papadakis 1997; Perera 2010; Pirdehghan 2018; Poole 1980; Roche 1996; Roche 1997; Ruesseler 2017; Schmitz 2018; Schwartz 2010; Seim 1995; Shaddheau 2015; Shapiro 2009; Siassakos 2010; Snow 2016; Solomon 2004; Spollen 2010; Steifel 2013; Stolz 2012; Strayer 2010; Taylor 2019; Vanatta 1996; Wagner 2011; Walsh 1999; Weihs 1986; Windish 2005; Wundrich 2017), and a further 14 as companion papers.

Included studies

As detailed in Table 1, 55 of the included studies were RCTs, nine were quasi-RCTs, seven were C-RCTs, and five were quasi-C-RCTs. RCTs generally randomised at the level of the student (n = 42), or by student followed by group or rotation (n = 13). In quasi-RCTs, randomisation procedures often were not detailed or were absent but were based on allocation of students to groups that were then assigned to study conditions (Alroy 1984; Bowyer 2006; Colletti 2001; Lorin 2006; Walsh 1999). One study reported using alphabetical assignment of students to study conditions (Klein 2000). In another study, students were allocated to their placement site by lottery and the sites were alternated monthly between intervention and control (Schwartz 2010).

We classified 12 studies as cluster-randomised trials (C-RCTs) because in those trials, groups of students (rather than individuals themselves) were randomised to different arms of the trial. In five studies, the unit of allocation was the tutorial groups (or discussion groups) (Cave 2007 Engerer 2019 Lee 2015 Papadakis 1997 Shaddheau 2015). One study was randomised by tutor (Meirovich 2016), and one by university (Ockene 2016). Five of these studies were defined as quasi-C-RCTs because they did not adequately randomise and allocate groups of students to experimental conditions. In four studies, the unit of allocation was the cohort year (Filipetto 2006 Levenkron 1990 Lim 2011 Wagner 2011), and in one study, the unit of allocation was rotation blocks within an academic year (Solomon 2004).

Most of the included studies compared two arms or experimental conditions (n = 59), and 17 studies compared more than two arms. Comparisons made across study arms are detailed below.

Country

Thirty-seven studies were undertaken in the United States (Allen 1990; Betchart 1984; Blatt 2010; Bowyer 2006; Bowyer 2010; Chibhall 2005; Colletti 2001; Edwardsen 2006; Eells 2002; Feddock 2009; Filipetto 2006; Foster 2016; Gerber 1985; Hobgood 2009; Kaltman 2018; Lee 2015; Levenkron 1990; Lorin 2006; LoSasso 2017; Lupi 2012; Mason 1988; Moreland 1973; Morrow 2009; Mounsey 2006; Ockene 2016; Palmer 2018; Papadakis 1997; Schwartz 2010; Seim 1995; Shaddheau 2015; Solomon 2004; Spollen 2010; Strayer 2010; Vanatta 1996; Wagner 2011; Weihs 1986; Windish 2005), eight in Australia (Bearman 2001; Evans 1989; Evans 1996; Liu 2016; Roche 1996; Roche 1997; Taylor 2019; Walsh 1999), and seven in the United Kingdom (Cave 2007; Klein 2000; Knowles 2001; Maguire 1977; Maguire 1978; Siassakos 2010; Snow 2016). Five studies were conducted in Germany (Bosse 2012; Engerer 2019; Gartmeir 2015; Ruesseler 2017; Wundrich 2017), five in Switzerland (Berney 2017; Daeppen 2012; Schmitz 2018; Steifel 2013; Stolz 2012), and two in each of Canada (Kahan 2003; Shapiro 2009), Israel (Alroy 1984; Meirovich 2016), and Japan (Ishikawa 2010; Nomura 2017). One study was conducted in each of France (Buffel du Vaure 2017), Iran (Pirdehghan 2018), Malaysia (Perera 2010), New Zealand (Lim 2011), South Africa (Legg 2005), Taiwan, (Ho 2008), and Turkey (Ozcakar 2009). The country in which the study was undertaken was unclear for Poole 1980.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Participants

A total of 10,124 participants were reported on across the included studies. The characteristics of participants are poorly described across the included studies. Only 28 studies reported the age of participants (Berney 2017; Bosse 2012; Buffel du Vaure 2017; Daeppen 2012; Eells 2002; Engerer 2019; Foster 2016; Ho 2008; Hobgood 2009; Kaltman 2018; Lee 2015; Legg 2005; Meirovich 2016; Mounsey 2006; Ockene 2016; Ozcakar 2009; Palmer 2018; Pirdehghan 2018; Roche 1996; Ruesseler 2017; Schmitz 2018; Schwartz 2010; Shaddheau 2015; Shapiro 2009; Spollen 2010; Strayer 2010; Walsh 1999; Windish 2005), with an average age of 24.8 years across these studies. Forty studies reported the gender make-up of participants (Berney 2017; Blatt 2010; Bosse 2012; Chibhall 2005; Daeppen 2012; Eells 2002; Engerer 2019; Feddock 2009; Foster 2016; Gartmeir 2015; Ho 2008; Hobgood 2009; Ishikawa 2010; Kahan 2003; Kaltman 2018; Klein 2000; Lee 2015; Legg 2005; Lim 2011; Liu 2016; Lorin 2006; Lupi 2012; Meirovich 2016; Moreland 1973; Mounsey 2006; Nomura 2017; Ockene 2016; Ozcakar 2009; Palmer 2018; Pirdehghan 2018; Roche 1996; Schmitz 2018; Schwartz 2010; Shaddheau 2015; Shapiro 2009; Spollen 2010; Stolz 2012; Strayer 2010; Walsh 1999; Windish 2005). With the exception of one study that included only male participants (Moreland 1973), the average proportion of males across these studies was 49.4%. Only 15 studies reported the ethnicity of participants (Blatt 2010; Foster 2016; Ho 2008; Hobgood 2009; Kaltman 2018; Lee 2015; Legg 2005; Mounsey 2006; Palmer 2018; Schwartz 2010; Shaddheau 2015; Shapiro 2009; Spollen 2010; Walsh 1999; Windish 2005). Most of these were US studies reporting that a majority of their participants were White/Caucasian, with some representation from Black, Hispanic, Asian, Pacific Islander, and other groups. One exception is Legg 2005, in which 43% of participants were Black. In Ho 2008, 87% of participants were Taiwanese; in Shapiro 2009, 67% were Canadian born; and in Walsh 1999, 76.4% were Australian born. Other sociodemographic factors were reported in three studies, with Shaddheau 2015 reporting on religiosity, Hobgood 2009 reporting on academic achievement, and Gartmeir 2015 commenting on differences in socioeconomic background between study groups.

Recruitment and eligibility

Recruitment was predominantly based on cohorts of students at particular year levels (n = 55), with some cohorts limited by additional criteria such as performance in earlier assessments (Betchart 1984), participation in previous training (Poole 1980), or lack of participation in previous training (Ozcakar 2009; Pirdehghan 2018). Recruitment was based on the site or timing of a clinical rotation or was elective in 11 studies (Bosse 2012; Buffel du Vaure 2017; Colletti 2001; Ho 2008; LoSasso 2017; Maguire 1978; Roche 1997; Ruesseler 2017; Seim 1995; Shaddheau 2015; Siassakos 2010), was based on class or group in two studies (Alroy 1984; Maguire 1977), and was based on programme or medical school in three studies (Berney 2017; Gartmeir 2015; Ockene 2016).

In 13 studies, students were in the first year of their medical programme (Betchart 1984; Edwardsen 2006; Filipetto 2006; Foster 2016; Kaltman 2018; Morrow 2009; Palmer 2018; Papadakis 1997; Perera 2010; Shapiro 2009; Vanatta 1996; Wagner 2011; Weihs 1986), with a further two studies specifying that students were 'preclinical' or were in the first three years of their programme (Ishikawa 2010; Meirovich 2016, respectively). In 11 studies, students were recruited during their second year (Allen 1990; Daeppen 2012; Eells 2002; Gerber 1985; Levenkron 1990; Liu 2016; Moreland 1973; Ozcakar 2009; Spollen 2010; Taylor 2019; Windish 2005); 18 studies recruited third-year students (Blatt 2010; Bowyer 2006; Bowyer 2010; Cave 2007; Chibhall 2005; Engerer 2019; Feddock 2009; Kahan 2003; Klein 2000; Lee 2015; LoSasso 2017; Lupi 2012; Mounsey 2006; Shaddheau 2015; Solomon 2004; Stolz 2012; Strayer 2010; Wundrich 2017); 17 recruited fourth-year students (Alroy 1984; Berney 2017; Buffel du Vaure 2017; Evans 1989; Evans 1996; Hobgood 2009; Knowles 2001; Lorin 2006; Mason 1988; Maguire 1977; Nomura 2017; Ruesseler 2017; Schmitz 2018; Schwartz 2010; Seim 1995; Siassakos 2010; Steifel 2013; Maguire 1978 did not specify the year level but recruited students during a psychiatry clerkship; and students were recruited from their fifth year or later in 10 studies (Bosse 2012; Ho 2008; Legg 2005; Lim 2011; Pirdehghan 2018; Poole 1980; Roche 1996; Roche 1997; Snow 2016; Walsh 1999).

Setting and mode of delivery

The setting was unclear in many studies, but most interventions were delivered in a classroom or simulation laboratory at either a university campus or a teaching hospital. The few studies that involved interviewing real patients were generally conducted in teaching hospitals. Most studies (n = 57; see additional table for list of studies) delivered face-to-face interventions, eight delivered online interventions (Bearman 2001; Foster 2016; Kaltman 2018; Lee 2015; Palmer 2018; Schmitz 2018; Snow 2016; Wagner 2011), one used videos (Mason 1988), and six used a combination of face-to-face and online interventions (Gartmeir 2015; Liu 2016; Ockene 2016; Solomon 2004; Stolz 2012; Strayer 2010). Twenty-four studies that delivered face-to-face interventions included the use of videos as a resource providing key information or examples of interactions (Allen 1990; Alroy 1984; Bowyer 2010; Daeppen 2012; Ho 2008; Kahan 2003; Klein 2000; Maguire 1977; Maguire 1978; Mason 1988; Schmitz 2018; Seim 1995; Wagner 2011; Windish 2005), video recordings of student role-plays for use in feedback and individual or class discussion (Berney 2017; Betchart 1984; Cave 2007; Kaltman 2018; Knowles 2001; Ozcakar 2009; Roche 1996; Roche 1997; Ruesseler 2017), or a combination of the two (Moreland 1973). Some studies compared different modes of delivery for different interventions.

Interventions and comparisons

Across the studies, 106 separate interventions were provided. A majority of interventions used small group workshops (n = 69 interventions, 48 studies), and fewer (n = 36 interventions, 27 studies) included more didactic lectures or seminars, in some cases in addition to smaller group workshops. The remainder used online or video delivery.

Most interventions involved role-play (n = 97 interventions, 72 studies) with simulated patients (n = 50 interventions, 33 studies), peers (n = 20 interventions, 16 studies), or real patients (8 studies; see below). In some cases, it is not clear whether role-plays were conducted with peers, simulated patients, or faculty (Bowyer 2006; Daeppen 2012; Ho 2008; Kahan 2003; Legg 2005; LoSasso 2017; Morrow 2009; Ockene 2016). Virtual patients were used in five interventions across three studies, with text-based interaction in Bearman 2001 and Foster 2016 or a combination of text and video clips of simulated patient responses in Kaltman 2018. One further study used simulated patients who interacted with students in an online environment (Liu 2016). Overall, real patients were involved in 15 interventions across 12 studies. In some cases, real patients were involved in presenting to students in live face-

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

to-face seminars, as in Shaddheau 2015, or in videos as part of online modules (Legg 2005; Snow 2016). In most studies using real patients, however, patients were interviewed by students in the classroom or at the bedside as part of a clinical rotation (Klein 2000; Maguire 1978; Meirovich 2016; Moreland 1973; Schwartz 2010; Shapiro 2009; Stolz 2012; Weihs 1986).

Feedback was an important feature of 64 interventions, with different interventions using varied approaches. A combination of self-reflection and peer, faculty (tutor or facilitator), and simulated patient feedback was used in 11 interventions (Berney 2017; Bosse 2012; Gerber 1985; Levenkron 1990; Moreland 1973; Perera 2010; Ruesseler 2017; Vanatta 1996; Walsh 1999; Windish 2005; Wundrich 2017); feedback was provided by faculty alone in 16 interventions (Berney 2017; Betchart 1984; Cave 2007; Engerer 2019; Gartmeir 2015; Hobgood 2009; Kaltman 2018; Klein 2000; Knowles 2001; Maguire 1978; Nomura 2017; Ozcakar 2009; Roche 1996; Roche 1997; Seim 1995; Steifel 2013), by peers alone in three interventions (Lupi 2012; Mounsey 2006; Roche 1996), by simulated patients alone in four interventions (Colletti 2001; Mounsey 2006; Papadakis 1997; Siassakos 2010), and by self-reflection alone in three (Legg 2005; Mason 1988; Papadakis 1997). One study used Teaching Associate Simulated Patients who provided feedback (Gerber 1985). Some studies compared effects of feedback delivered by different people or via particular styles (e.g. Engerer 2019; Maguire 1978; Roche 1996; Ozcakar 2009). In nine studies, feedback was based on review of video-recorded interactions (Cave 2007; Gartmeir 2015; Kaltman 2018; Klein 2000; Maguire 1977; Maguire 1978; Roche 1996; Ozcakar 2009; Walsh 1999). Feedback was provided online via pre-set responses or human assessors in two studies (Foster 2016; Liu 2016).

For the purposes of synthesis and meta-analysis, comparisons were categorised as those comparing a communication skills intervention with a usual curriculum or control; those comparing two interventions with different modes of delivery (online versus face-to-face or experiential versus didactic); those comparing two interventions with different feedback approaches; and those comparing interventions using simulated or real patients with those relying on peer role-play.

Thirty-three studies presented a comparison between an intervention group and a control group. Ten of these studies compared the intervention to usual care (Bosse 2012; Bowyer 2006; Bowyer 2010; Eells 2002; Evans 1989; Filipetto 2006; Knowles 2001; Lee 2015; Legg 2005; Seim 1995), and 15 compared the intervention group with a control group receiving no specific communication training (Allen 1990; Alroy 1984; Buffel du Vaure 2017; Colletti 2001; Daeppen 2012; Ho 2008; Hobgood 2009; Lorin 2006; Mason 1988; Poole 1980; Roche 1996; Schwartz 2010; Shaddheau 2015; Solomon 2004; Wundrich 2017). Although it was not explicitly stated in the text, Pirdehghan 2018 is also assumed to report a no intervention comparison. Seven studies used a wait list control (Evans 1989; Evans 1996; Gartmeir 2015; Lupi 2012; Shapiro 2009; Spollen 2010; Windish 2005).

Sixteen studies compared modes of delivery - Bowyer 2010; Eells 2002; Gartmeir 2015; Kaltman 2018; Liu 2016; Ockene 2016; Palmer 2018; Roche 1997; Stolz 2012; Strayer 2010; Wagner 2011; Walsh 1999 - or alternate tutors - Gerber 1985; Meirovich 2016; Nomura 2017; Snow 2016. Nine studies compared styles of feedback (Cave 2007; Engerer 2019; Hobgood 2009; Maguire 1978; Moreland 1973; Ozcakar 2009; Perera 2010; Ruesseler 2017; Vanatta 1996). Seven studies compared role-play with simulated patients and peers (Bosse 2012; Levenkron 1990; Mounsey 2006; Papadakis 1997; Stolz 2012; Taylor 2019; Vanatta 1996), and 25 studies compared other alternate structures such as the style of a virtual patient or the style of examples given in online modules (Bearman 2001; Berney 2017; Betchart 1984; Blatt 2010; Cave 2007; Chibhall 2005; Edwardsen 2006; Feddock 2009; Foster 2016; Ho 2008; Ishikawa 2010; Kahan 2003; Klein 2000; Levenkron 1990; Lim 2011; LoSasso 2017; Maguire 1977; Mason 1988; Moreland 1973; Morrow 2009; Nomura 2017; Schmitz 2018; Siassakos 2010; Steifel 2013; Weihs 1986).

Outcomes

Most studies reported on multiple outcomes within our established outcome categories, with a total of 510 outcomes recorded across all included studies. Although few studies specified a theoretical framework, several studies based their assessment criteria on the Calgary-Cambridge guides to medical interview (e.g. Bosse 2012; Windish 2005), and even for those that did not do so explicitly, outcomes generally fit well into components of the Calgary-Cambridge model. Fifty-one outcomes were categorised as information gathering, 207 as specific communication tasks, 44 as explanation and planning, and 120 as relationship building. Overall communication skills, which generally comprised a combination of several of our pre-defined outcome categories, or total score of students' communication performance was reported in 74 outcomes. Simulated or real patient satisfaction with consultations was used in four studies (seven outcomes in total; Blatt 2010; Legg 2005; Lupi 2012; Palmer 2018), student self-assessment or satisfaction in two studies (three outcomes; Steifel 2013; Stolz 2012), and student knowledge in three studies (three outcomes in total; Stolz 2012; Strayer 2010; Wagner 2011). No studies reported any adverse events.

For the purposes of synthesis and meta-analysis, we grouped the most homogenous outcomes within a priori defined categories, as well as with an overall rating category, which included studies reporting a total score on an assessment checklist that included components from each outcome category and those reporting a global score on a communication OSCE station or for the communication components of a station. Within relationship building, we grouped outcomes as relating to empathy or relationship development (rapport building). Outcomes were further divided as judged by an academic/clinical assessor (usually a tutor, clinician, or other faculty member) or by the simulated patients involved.

Twenty-seven studies reported on overall communication skills as judged by academic assessors (Bosse 2012; Eells 2002; Engerer 2019; Evans 1996; Filipetto 2006; Gartmeir 2015; Knowles 2001; Lee 2015; Levenkron 1990; Liu 2016; Lorin 2006; Lupi 2012; Maguire 1977; Maguire 1978; Ockene 2016; Ozcakar 2009; Perera 2010; Pirdehghan 2018; Roche 1996; Roche 1997; Ruesseler 2017; Solomon 2004; Spollen 2010; Steifel 2013; Stolz 2012; Taylor 2019; Walsh 1999). Thirteen studies reported on overall communication skills outcomes as judged by simulated patients (Blatt 2010; Bowyer 2010; Colletti 2001; Ho 2008; Hobgood 2009; Legg 2005; LoSasso 2017; Maguire 1977; Shaddheau 2015; Shapiro 2009; Siassakos 2010; Snow 2016; Wagner 2011).

Seventeen studies reported on empathy outcomes as rated by assessors (Alroy 1984; Bowyer 2010; Daeppen 2012; Engerer 2019;

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Evans 1989; Foster 2016; Kaltman 2018; Klein 2000; Moreland 1973; Mounsey 2006; Palmer 2018; Poole 1980; Seim 1995; Spollen 2010; Vanatta 1996; Weihs 1986; Wundrich 2017). Seven studies reported on empathy outcomes as judged by simulated patients (Bowyer 2010; Buffel du Vaure 2017; Chibhall 2005; Foster 2016; Klein 2000; LoSasso 2017; Wundrich 2017). Empathy was measured in some cases as a single item, for example, via an open-ended question about coping (Allen 1990); time spent using supportive verbal communication behaviours including empathy as in Alroy 1984, provision of comfort as in Bowyer 2010, or empathetic statements as in Edwardsen 2006 and Kaltman 2018 judgements about students' eliciting and handling emotions (Engerer 2019; Moreland 1973; Seim 1995; Weihs 1986); or responses to opportunities to express empathy (Foster 2016). Some empathy outcomes were simply judgements of student empathy (Mounsey 2006; Palmer 2018); others used scales such as the CARE (Consultation And Relational Empathy) Scale (a 10-item patient-rated questionnaire of physician empathy; Buffel du Vaure 2017), the Jefferson Scale of patient perceptions of physician empathy (LoSasso 2017), the empathy subscale of the Rotor Interactional Analysis System (RIAS; Vanatta 1996), or a series of items relating to empathy as assessed by assessors in Wundrich 2017 or by simulated patients in Chibhall 2005 and Wundrich 2017.

Relationship or rapport building as judged by academic assessors was reported in 14 studies (Bosse 2012; Evans 1989; Foster 2016; Gartmeir 2015; Hobgood 2009; Legg 2005; Levenkron 1990; Liu 2016; Lupi 2012; Mason 1988; Meirovich 2016; Morrow 2009; Perera 2010; Windish 2005). Simulated patient perception or satisfaction (including patient impression of relationship building) was recorded in seven studies (Blatt 2010; Hobgood 2009; Legg 2005; Palmer 2018; Shaddheau 2015; Siassakos 2010; Windish 2005).

Outcomes used in meta-analysis were assessed immediately, or shortly after intervention delivery, with the longest follow-up measurements occurring at the end of the academic year in which the intervention occurred. Longer-term follow-up of outcomes was reported in eight studies (Cave 2007; Colletti 2001; Filipetto 2006; Gerber 1985; Klein 2000; Maguire 1978; Meirovich 2016; Strayer 2010).

Outcome validity and reliability

Most studies reported some form of training for examiners and simulated patients involved in assessments. Thirty-three studies reported some form of calculation of interrater reliability, often for a proportion of assessments, and reported results of these calculations as adequate to strong (Alroy 1984; Bearman 2001; Berney 2017; Betchart 1984; Bosse 2012; Daeppen 2012; Edwardsen 2006; Engerer 2019; Evans 1996; Foster 2016; Gartmeir 2015; Ho 2008; Hobgood 2009; Ishikawa 2010; Kaltman 2018; Klein 2000; Knowles 2001; Legg 2005; Lorin 2006; Maguire 1978; Mason 1988; Moreland 1973; Mounsey 2006; Ockene 2016; Ozcakar 2009; Poole 1980; Roche 1996; Roche 1997; Schmitz 2018; Seim 1995; Vanatta 1996; Wagner 2011; Weihs 1986). Two further studies reported conducting interrater reliability assessments but did not report the results (LoSasso 2017; Walsh 1999), and in one further study, interrater reliability was assessed for examiner training only (Strayer 2010). In three studies, a single examiner performed all assessments (Schwartz 2010; Siassakos 2010; Steifel 2013). Three studies reported some form of internal consistency calculation but not interrater reliability (Buffel du Vaure 2017; Feddock 2009; Pirdehghan 2018). Allen 1990 did not assess interrater reliability but conducted ANOVA on scores of different assessors at different sites. Cave 2007 did not formally assess interrater reliability but made efforts to reduce variability in training and scale design. Maguire 1977 referred to interrater reliability established in a previous study. Twenty-six studies did not report any effort to assess interrater reliability.

Excluded studies

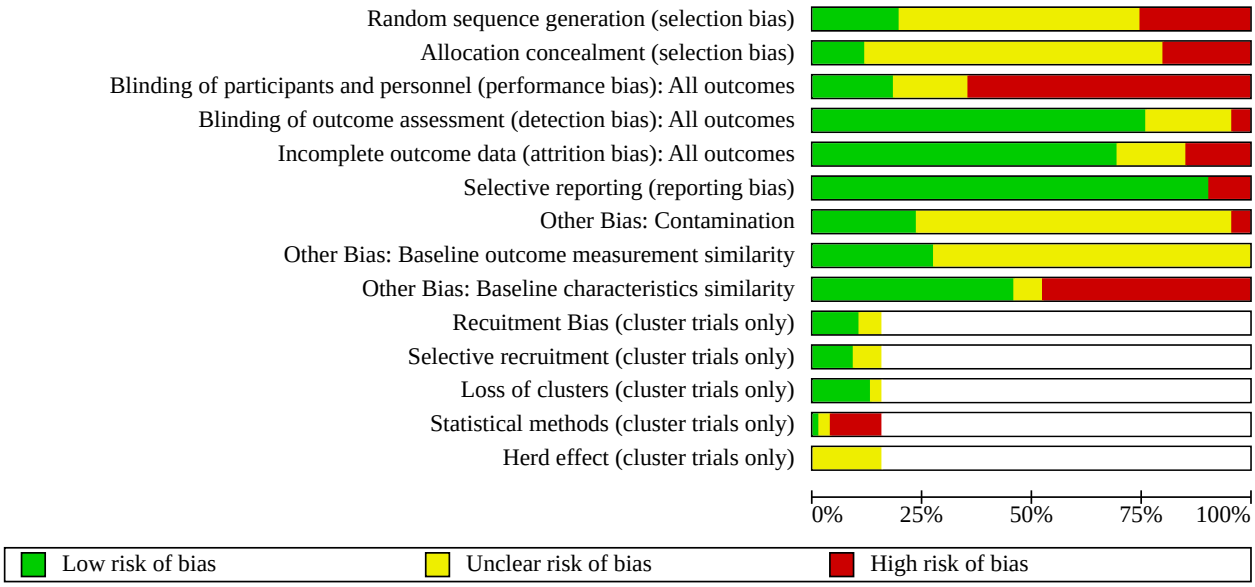
A total of 178 records remained after title and abstract screening, of which 164 full-text articles were obtained for further review. We considered 62 articles to be ineligible after assessment of the full text (reasons for exclusion were study design (n = 19), participants (n = 13), outcomes (n = 25), or patient population (n = 5)). Twenty-five studies were regarded as awaiting classification. See Characteristics of excluded studies for further details.

Risk of bias in included studies

We assessed and reported on the methodological risk of bias of included studies in accordance with the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011), as well as the guidelines of the Cochrane Consumers and Communication Review Group (Ryan 2013). Assessment results of risk of bias for all included studies are presented in Figure 2 and Figure 3.

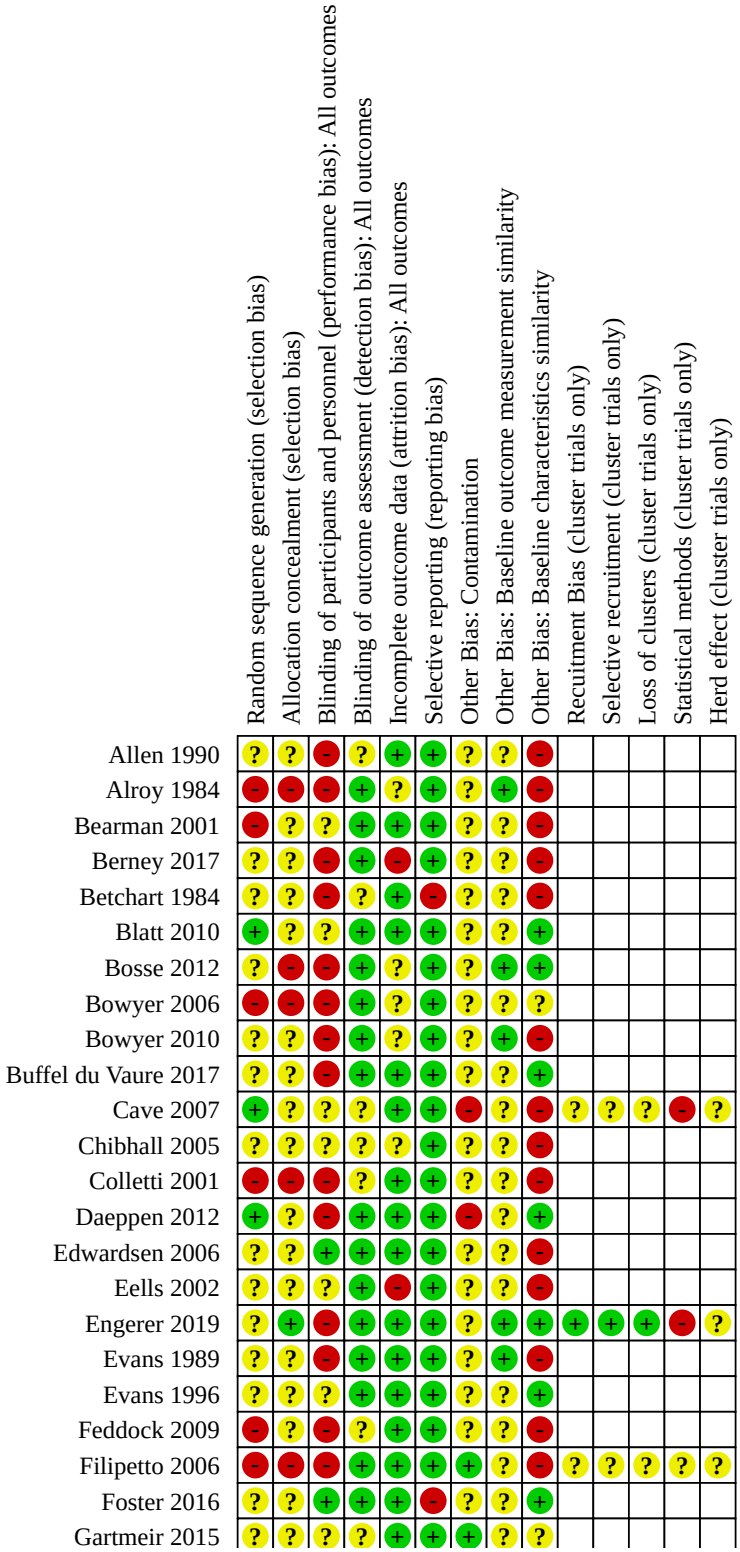
Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.



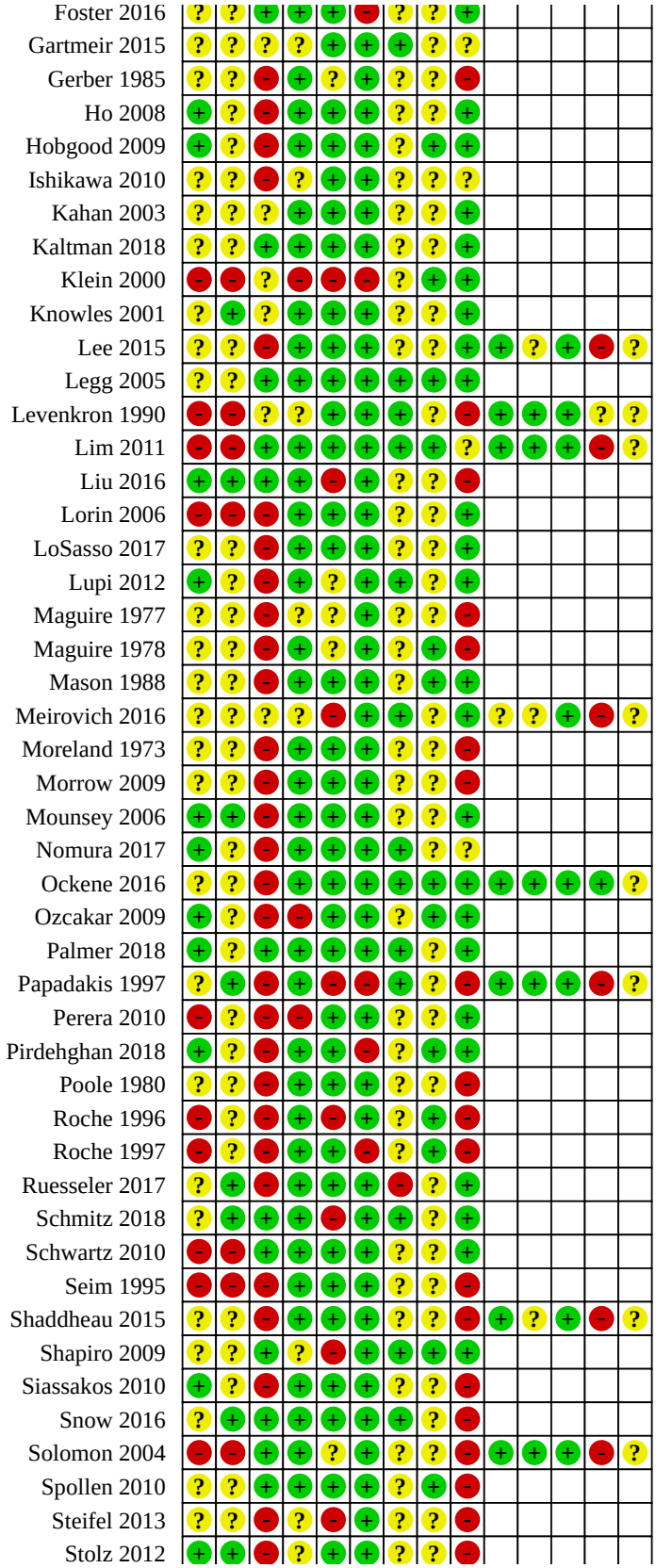
Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.



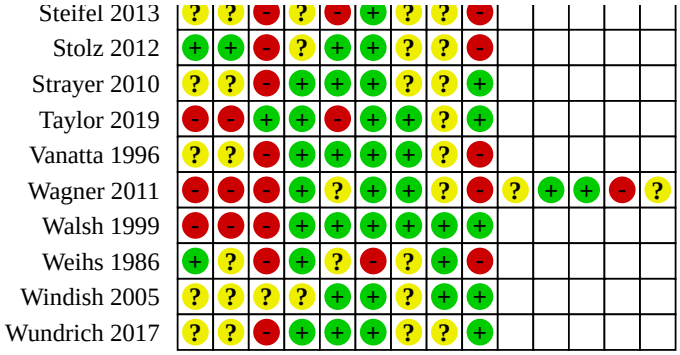
Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 3. (Continued)



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 3. (Continued)



Overall, 21 studies were regarded as high risk (with three or more 'high' ratings) for the purpose of sensitivity analysis. Most studies had low risk of detection bias, reporting bias, and attrition bias. Many studies had unclear risk of bias judgements for performance bias, selection bias, and other bias including baseline characteristics and baseline outcome measures. This lack of information is likely to indicate high risk of bias across studies (Higgins 2011).

Allocation

We judged 15 studies as having low risk of bias for random sequence generation. Of these, 14 studies used a computer-generated random number sequence (Bearman 2001; Daeppen 2012; Ho 2008; Hobgood 2009; Liu 2016; Lupi 2012; Mounsey 2006; Nomura 2017; Ozcakar 2009; Palmer 2018; Pirdehghan 2018; Siassakos 2010; Stolz 2012; Weihs 1986), and one study performed randomisation by drawing cards out of a box (Cave 2007). Five studies were rated as having high risk of bias for this domain, as students were not randomly allocated to groups, rather they were allocated to groups by rotation term or by academic scores (Bearman 2001; Feddock 2009; Perera 2010; Roche 1996; Roche 1997). An additional 14 quasi-RCTs were rated as high risk, in accordance with the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011), as well as the Cochrane Consumers and Communication guidelines (Ryan 2013), as their method of random sequence generation was not truly random (Alroy 1984; Bowyer 2006; Colletti 2001; Filipetto 2006; Klein 2000; Levenkron 1990; Lim 2011; Lorin 2006; Schwartz 2010; Seim 1995; Solomon 2004; Taylor 2019; Wagner 2011; Walsh 1999. The remaining 42 studies stated that participants were randomised but did not adequately describe the method used to generate the random sequence and therefore were rated as having unclear risk of bias.

Of the 76 included studies, only nine provided sufficient detail to establish that participant allocation to experimental groups was concealed from those conducting the research. We rated these as having low risk of selection bias for this domain (Engerer 2019; Knowles 2001; Liu 2016; Mounsey 2006; Papadakis 1997; Ruesseler 2017; Schmitz 2018; Snow 2016; Stolz 2012). Fifteen studies were rated as having high risk of bias due to the nature of the study design and inability to conceal the allocation sequence. Of these 15 studies, 14 were quasi-RCTs. We were unable to make a judgement on the remaining 52 studies due to insufficient detail provided; therefore, we judged these studies as having unclear risk of bias with regard to allocation concealment.

Blinding

For most studies (n = 49), blinding of participants and personnel (performance bias) was not achievable due to the educational nature of the interventions; therefore these studies were rated as having high risk of performance bias. Although attempts were made to blind participants from their intervention status (e.g. by using control tutorials, alternative lectures), it is likely that either participants (students) or personnel (tutors) would have been aware that they were receiving an intervention that differed from the standard medical curriculum, and this knowledge may have introduced performance bias. For 14 studies, we judged that lack of blinding was unlikely to impact performance bias; thus we rated these studies as having low risk of bias for this domain (Edwardsen 2006; Foster 2016; Kaltman 2018; Legg 2005; Lim 2011; Liu 2016; Palmer 2018; Schmitz 2018; Schwartz 2010; Shapiro 2009; Snow 2016; Solomon 2004; Spollen 2010; Taylor 2019). For example, for studies involving online computer-based interventions, blinding of personnel was not relevant and lack of blinding of participants was not likely to lead to bias. Risk of bias for the remaining 13 studies was unclear due to insufficient information provided (Bearman 2001; Blatt 2010; Cave 2007; Chibhall 2005; Eells 2002; Evans 1996; Gartmeir 2015; Kahan 2003; Klein 2000; Knowles 2001; Levenkron 1990; Meirovich 2016; Windish 2005). For example, studies were judged as having unclear risk of bias if they reported that participants were blind to group allocation but provided no information regarding blinding of personnel delivering the intervention.

Blinding of outcome assessment was achieved in most studies, with 58 of the 76 included studies reporting that assessors were blind to group assignment. Therefore, those studies were rated as having low risk of detection bias. Three studies reported that assessors were not blind to the student's group allocation and as such were judged as having high risk of detection bias (Klein 2000; Ozcakar 2009; Perera 2010). For the remaining 15 studies, blinding of outcome assessment was unclear due to insufficient information provided (Allen 1990; Betchart 1984; Cave 2007; Chibhall 2005; Colletti 2001; Feddock 2009; Gartmeir 2015; Ishikawa 2010; Levenkron 1990; Maguire 1977; Meirovich 2016; Shapiro 2009; Steifel 2013; Stolz 2012; Windish 2005).

Incomplete outcome data

We rated 53 studies as having low risk of attrition bias, as they reported less than 20% loss of participants and showed no differential attrition between experimental groups. Eleven studies

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

were rated as having high risk of attrition bias due to high attrition rates (> 20%) and/or unequal attrition between groups (Berney 2017; Eells 2002; Klein 2000; Liu 2016; Meirovich 2016; Papadakis 1997; Roche 1996; Schmitz 2018; Shapiro 2009; Steifel 2013; Taylor 2019). We rated the remaining 12 studies as having unclear risk of attrition bias, as the details provided were insufficient to permit a judgement on the completeness of outcome data (Allen 1990; Bosse 2012; Bowyer 2006; Bowyer 2010; Chibhall 2005; Gerber 1985; Lupi 2012; Maguire 1977; Maguire 1978; Solomon 2004; Wagner 2011; Weihs 1986).

Selective reporting

For most included studies, no protocol was available. Studies without an available protocol were deemed to have low risk if all outcomes described in the methods were reported as planned. We judged 69 studies as having low risk of reporting bias for this reason. We rated seven studies as having high risk of reporting bias because detailed outcome data were not reported, or only statistically different items were reported (Betchart 1984; Foster 2016; Klein 2000; Papadakis 1997; Pirdehghan 2018; Roche 1997; Weihs 1986).

Other potential sources of bias

In the domain of "Other risks", we assessed the risk of contamination and baseline imbalances between intervention and control groups as other potential sources of bias. Most included studies used participant groups within the same institution, and therefore sharing of knowledge could have occurred. However, this factor did not automatically result in a high risk rating due to lack of information about the curricula and rotational structure at the institution for most studies. Therefore it is likely that our judgements for risk of contamination bias reflect underestimation of the bias present. We judged three studies as having high risk of contamination bias due to the potential for contamination to have occurred between groups or rotations (Cave 2007; Daeppen 2012; Ruesseler 2017). We judged 18 studies as having low risk of contamination bias for one of two reasons: (a) control group data were obtained before intervention group data were obtained, thus reducing the risk of contamination bias from students in the intervention group sharing information with students in the control group, or (b) the difference between study groups was the mode of delivery or feedback, rather than any difference in teaching content, thus reducing the risk of contamination bias (Filipetto 2006; Gartmeir 2015; Legg 2005; Levenkron 1990; Lim 2011; Lupi 2012; Meirovich 2016; Nomura 2017; Ockene 2016; Palmer 2018; Papadakis 1997; Schmitz 2018; Shapiro 2009; Snow 2016; Taylor 2019; Vanatta 1996; Wagner 2011; Walsh 1999). For the remaining 55 studies, the risk of contamination bias was judged as unclear due to insufficient information provided.

We assessed two types of baseline imbalance as a potential source of bias - baseline outcome measurement and baseline characteristics. With regard to baseline outcome measurement, 21 studies measured performance outcomes before the intervention and reported no significant baseline differences between study groups, and one study performed appropriately adjusted analyses to account for baseline imbalances (Mason 1988). We rated these studies as having low risk of bias (Alroy 1984; Bosse 2012; Bowyer 2010; Engerer 2019; Evans 1989; Hobgood 2009; Klein 2000; Legg 2005; Lim 2011; Maguire 1978; Mason 1988; Ockene 2016; Ozcakar 2009; Pirdehghan 2018; Roche 1996; Roche 1997; Shapiro 2009; Spollen 2010; Walsh 1999; Weihs 1986; Windish 2005). We rated the remaining 55 studies as having unclear risk because no baseline measure of outcome was reported.

With regard to baseline characteristics, 35 studies reported similar baseline characteristics between intervention and control groups and thus were rated as having low risk of bias for this domain (Blatt 2010; Bosse 2012; Buffel du Vaure 2017; Daeppen 2012; Engerer 2019; Evans 1996; Foster 2016; Ho 2008; Hobgood 2009; Kahan 2003; Kaltman 2018; Klein 2000; Knowles 2001; Lee 2015; Legg 2005; Lorin 2006; LoSasso 2017; Lupi 2012; Mason 1988; Meirovich 2016; Mounsey 2006; Ockene 2016; Ozcakar 2009; Palmer 2018; Perera 2010; Pirdehghan 2018; Ruesseler 2017; Schmitz 2018; Schwartz 2010; Shapiro 2009; Strayer 2010; Taylor 2019; Walsh 1999; Windish 2005; Wundrich 2017). We rated 36 studies as having high risk of bias; of these, two studies reported significant differences in baseline characteristics between groups (Feddock 2009; Shaddheau 2015), and 34 studies did not report baseline characteristics (Allen 1990; Alroy 1984; Bearman 2001; Berney 2017; Betchart 1984; Bowyer 2010; Cave 2007; Chibhall 2005; Colletti 2001; Edwardsen 2006; Eells 2002; Evans 1989; Filipetto 2006; Gerber 1985; Levenkron 1990; Liu 2016; Maguire 1977; Maguire 1978; Moreland 1973; Morrow 2009; Papadakis 1997; Poole 1980; Roche 1996; Roche 1997; Seim 1995; Siassakos 2010; Snow 2016; Solomon 2004; Spollen 2010; Steifel 2013; Stolz 2012; Vanatta 1996; Wagner 2011; Weihs 1986). Those not reporting baseline characteristics were regarded as having high risk of bias in this domain in keeping with the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). Risk of bias was unclear for the remaining five studies, as they did not provide sufficient information to permit a judgement.

For the 12 cluster-RCTs (n = 7) and quasi-C-RCTs (n = 5), we assessed risk of recruitment bias, selective recruitment, loss of clusters, incorrect analysis, and compatibility with individually randomised trials (herd effect). We considered eight studies to have low risk of recruitment bias based on appropriate recruitment techniques applied before allocation to clusters (Engerer 2019; Lee 2015; Levenkron 1990; Lim 2011; Ockene 2016; Papadakis 1997; Shaddheau 2015; Solomon 2004). Risk of recruitment bias was unclear for the remaining C-RCTs. For selective recruitment, we considered seven studies to be at low risk of bias based on similar characteristics of groups at baseline (Engerer 2019; Levenkron 1990; Lim 2011; Ockene 2016; Papadakis 1997; Solomon 2004; Wagner 2011), and we rated five studies as having unclear risk due to insufficient information to permit judgement. Nine studies did not account for effects of clustering in the analysis, so we rated these studies as having high risk for incorrect statistical analysis (Cave 2007; Engerer 2019; Lee 2015; Lim 2011; Meirovich 2016; Papadakis 1997; Shaddheau 2015; Solomon 2004; Wagner 2011). We judged one study as having low risk, as it did adjust for effects of clustering in the analysis (Ockene 2016), and we rated the remaining two studies as having unclear risk (Filipetto 2006; Levenkron 1990). Information was insufficient to permit judgement of the herd effect for all 12 C-RCTs and quasi-C-RCTs.

Effects of interventions

See: **Summary of findings 1** Summary of findings; **Summary of findings 2** Summary of findings; **Summary of findings 3** Summary of findings; **Summary of findings 4** Summary of findings

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

See Summary of findings 1 Summary of findings 2 Summary of findings 3 and Summary of findings 4.

Examination of interventions, settings, and student groups suggest that the included studies were sufficiently homogenous, and we conducted meta-analysis by pooling data from studies in which comparisons and outcome measures were considered similar. Otherwise, we provide a narrative synthesis of trial findings.

For the purpose of meta-analysis, we pooled outcome measures according to our a priori defined outcome categories, with subgroup analyses according to whether assessments were made by experts or by simulated patients. Given the similarity of outcomes, we pooled relationship building/rapport as assessed by experts with similar outcomes (judgements of rapport and connection) as assessed by simulated patients. Comparisons were categorised as those comparing a communication skills intervention with a usual curriculum or control; those comparing two interventions with different modes of delivery (online versus face-to-face or experiential versus didactic); those comparing two interventions with different feedback approaches; and those comparing interventions using simulated or real patients with those relying on peer role-play. When sufficient studies within each comparison reported on a single outcome, results were pooled in meta-analysis.

Comparison 1. Communication skills intervention versus usual curriculum or control

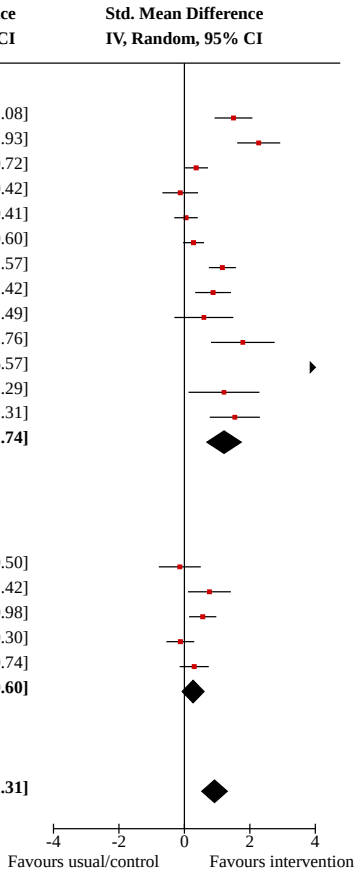
Overall communication skills

Communication skills interventions may improve students' overall scores on objective standardised clinical examinations (OSCEs) to a greater extent than the usual curricula or control (standardised mean difference (SMD) 0.92, 95% confidence interval (CI) 0.53 to 1.31; 18 studies, 1356 participants; I^2 = 90%; Analysis 1.1; Figure 4; low-quality evidence, downgraded due to risk of bias across several domains and unexplained heterogeneity, with no publication bias detected; Figure 5). Subgroup analyses indicate that results were similar when assessed by experts (faculty or trained assessors; SMD 1.21, 95% CI 0.69 to 1.74; 13 studies, 959 participants; I^2 = 92%) but were not consistently superior when assessed by simulated patients (SMD 0.27, 95% CI -0.07 to 0.60; 5 studies, 397 participants; I^2 = 55%). Subgroup differences were statistically significant (P = 0.003). In subgroup analysis limited to studies comparing an intervention group only to a control group (excluding any usual curriculum that included communication skills training), results remained consistent (SMD 1.06, 95% CI 0.54 to 1.58; 13 studies, 1057 participants; I^2 = 92%; with significant subgroup differences P = 0.0008; Analysis 1.2). Results were similar in re-analysis with those at highest risk of bias removed (SMD 1.14, 95% CI 0.62 to 1.66; 14 studies, 1035 participants; I^2 = 92%; Analysis 1.3). Results also remained consistent in sensitivity analysis conducted to account for clustering in cluster-RCTs and cluster-quasi-RCTs. Although the impact of interventions on overall communication skills as assessed by simulated patients moves farther away from the null hypothesis with increased adjustment of the standard error, the confidence interval of the pooled effect widens, suggesting decreased confidence in the result (see Analysis 1.4; Analysis 1.5; Analysis 1.6).

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

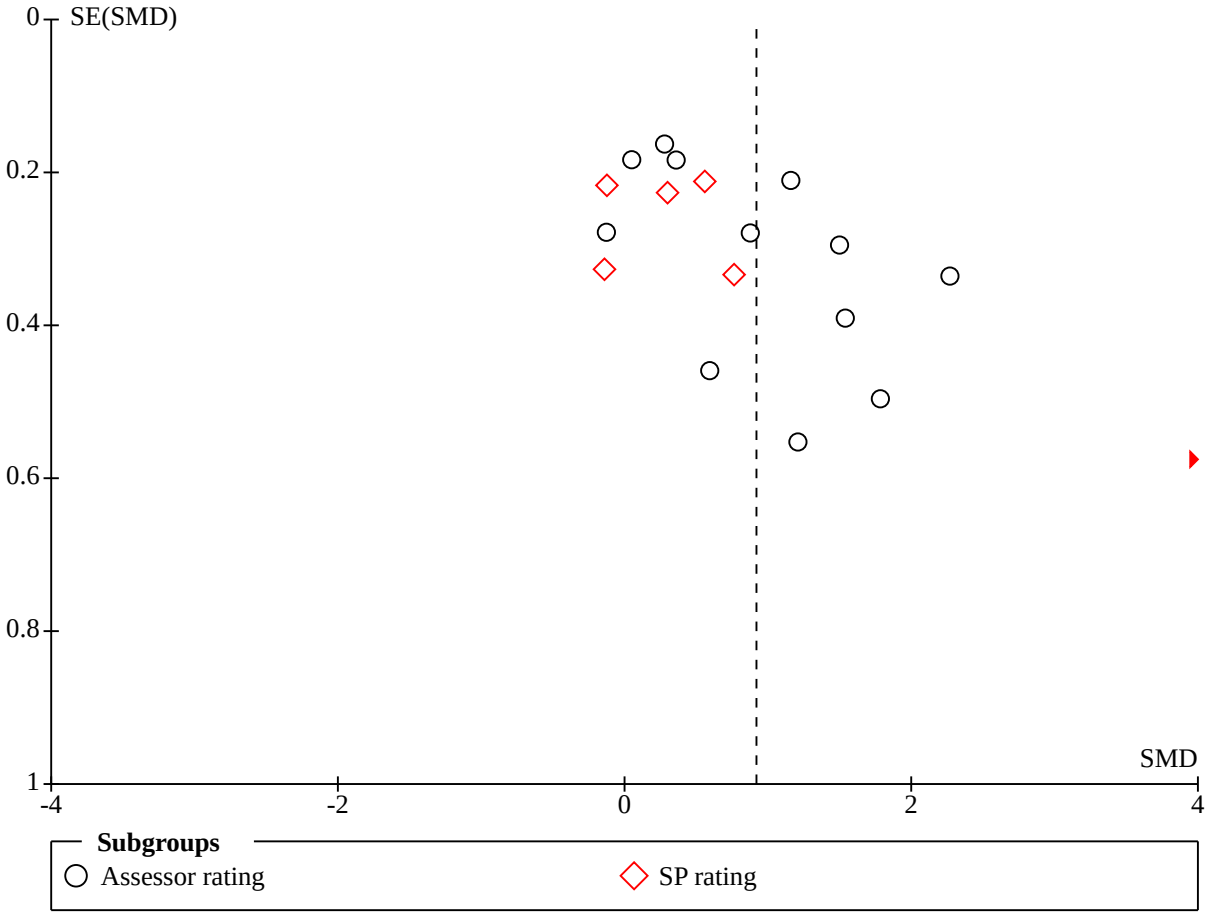
Figure 4. Forest plot of comparison: 1 Communication intervention versus control or usual care, outcome: 1.1 Overall communication skills.

Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.1.1 Assessor rating						
Bosse 2012	1.4999	0.295	28	32	5.7%	1.50 [0.92 , 2.08]
Evans 1996	2.27	0.3356	30	30	5.5%	2.27 [1.61 , 2.93]
Filipetto 2006	0.36	0.1838	68	54	6.2%	0.36 [-0.00 , 0.72]
Gartmeir 2015	-0.1267	0.2783	54	17	5.8%	-0.13 [-0.67 , 0.42]
Lee 2015	0.05	0.1834	60	59	6.2%	0.05 [-0.31 , 0.41]
Liu 2016	0.2792	0.1629	59	107	6.3%	0.28 [-0.04 , 0.60]
Lorin 2006	1.16	0.2105	53	53	6.1%	1.16 [0.75 , 1.57]
Lupi 2012	0.8773	0.2792	50	45	5.8%	0.88 [0.33 , 1.42]
Maguire 1977	0.5942	0.4593	10	10	4.8%	0.59 [-0.31 , 1.49]
Maguire 1978	1.7848	0.4961	12	12	4.6%	1.78 [0.81 , 2.76]
Pirdehghan 2018	5.444	0.5753	30	30	4.2%	5.44 [4.32 , 6.57]
Solomon 2004	1.2095	0.5526	5	16	4.3%	1.21 [0.13 , 2.29]
Spollen 2010	1.54	0.3906	17	18	5.2%	1.54 [0.77 , 2.31]
Subtotal (95% CI)			**476**	**483**	**70.8%**	**1.21 [0.69 , 1.74]**
Heterogeneity: Tau² = 0.80; Chi² = 144.06, df = 12 (P < 0.00001); I² = 92%						
Test for overall effect: Z = 4.54 (P < 0.00001)						
1.1.2 SP rating						
Colletti 2001	-0.14	0.3267	21	17	5.6%	-0.14 [-0.78 , 0.50]
Ho 2008	0.7643	0.3337	15	27	5.5%	0.76 [0.11 , 1.42]
Hobgood 2009	0.56	0.2118	35	68	6.1%	0.56 [0.14 , 0.98]
Shaddheau 2015	-0.123	0.2169	75	60	6.1%	-0.12 [-0.55 , 0.30]
Shapiro 2009	0.3	0.2265	38	41	6.0%	0.30 [-0.14 , 0.74]
Subtotal (95% CI)			**184**	**213**	**29.2%**	**0.27 [-0.07 , 0.60]**
Heterogeneity: Tau² = 0.08; Chi² = 8.94, df = 4 (P = 0.06); I² = 55%						
Test for overall effect: Z = 1.56 (P = 0.12)						
Total (95% CI)			**660**	**696**	**100.0%**	**0.92 [0.53 , 1.31]**
Heterogeneity: Tau² = 0.60; Chi² = 165.36, df = 17 (P < 0.00001); I² = 90%						
Test for overall effect: Z = 4.64 (P < 0.00001)						
Test for subgroup differences: Chi² = 8.92, df = 1 (P = 0.003), I² = 88.8%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 5. Funnel plot of comparison: 1 Communication intervention versus control or usual care, outcome: 1.1 Overall communication skills.



Additional studies that were unable to be pooled also reported on this measure. Knowles 2001 reported that scores on OSCEs conducted one week after the intervention were higher in the intervention group that received traditional teaching as well as role-play compared to the control group that received traditional teaching only (Wilcoxon T statistically significant at P < 0.001). OSCE scores reflected both overall performance and interviewing skills. Roche 1996 compared audio, peer, and video feedback with control and reported that all intervention groups performed significantly better (at P < 0.05 level) than the control group in overall OSCE scores. Eells 2002 compared experiential tutorials including peer role-play with self-study and theoretical information only and reported that there were no significant differences in any supervisor Interview Evaluation Questionnaire variables between groups (data not provided). Steifel 2013 compared a workshop including role-play with a wait list control but did not report outcomes by group.

Taken together, these quantitative and narrative results indicate there may be a positive effect of communication skills interventions compared to control on students' overall communication skills. Although positive, with some exceptions, the scale of the effect is in general relatively small. In general, a difference of 1 point on a 10-point scale or up to 5 points on a 100-point scale was the norm across studies, with most also reporting large standard deviations of the mean, indicative of significant variance in results.

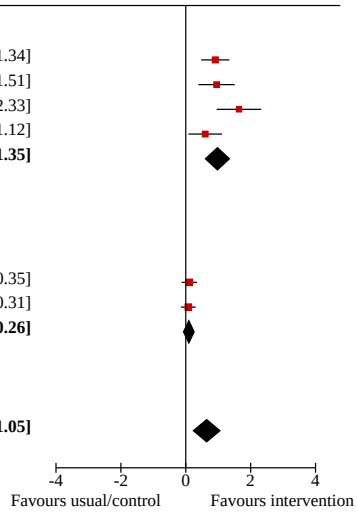
Empathy

Communication skills interventions may have a greater impact on students' empathy scores in comparison to usual curricula or control (SMD 0.64, 95% CI 0.23 to 1.05; 6 studies, 831 participants; I^2 = 86%; Analysis 1.7; Figure 6; low-quality evidence, downgraded due to only partially explained heterogeneity and suspected publication bias). Despite the small number of studies, asymmetry was apparent upon inspection of the funnel plot (not shown); therefore, publication bias is suspected. Subgroup analyses indicate that results were similar when assessed by experts (SMD 0.98, 95% CI 0.61 to 1.35; 4 studies, 251 participants; I^2 = 47%), but the two studies reporting empathy assessed by simulated patients showed no effect (SMD 0.09, 95% CI -0.07 to 0.26; 2 studies, 580 participants; I^2 = 0%). Subgroup differences were statistically significant (P < 0.0001). All studies assessed by experts were rated as having low risk of bias.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 6. Forest plot of comparison: 1 Communication intervention versus control or usual care, outcome: 1.7 Empathy.

Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.7.1 Assesor rating						
Daeppen 2012	0.91	0.2213	42	49	16.9%	0.91 [0.48 , 1.34]
Evans 1989	0.95	0.2856	28	27	15.0%	0.95 [0.39 , 1.51]
Poole 1980	1.64	0.3504	25	20	13.2%	1.64 [0.95 , 2.33]
Seim 1995	0.6	0.2643	30	30	15.7%	0.60 [0.08 , 1.12]
Subtotal (95% CI)			**125**	**126**	**60.8%**	**0.98 [0.61 , 1.35]**
Heterogeneity: Tau² = 0.07; Chi² = 5.66, df = 3 (P = 0.13); I² = 47%						
Test for overall effect: Z = 5.16 (P < 0.00001)						
1.7.2 SP rating						
Bowyer 2010	0.1109	0.121	118	163	19.5%	0.11 [-0.13 , 0.35]
Buffel du Vaure 2017	0.0789	0.1158	155	144	19.7%	0.08 [-0.15 , 0.31]
Subtotal (95% CI)			**273**	**307**	**39.2%**	**0.09 [-0.07 , 0.26]**
Heterogeneity: Tau² = 0.00; Chi² = 0.04, df = 1 (P = 0.85); I² = 0%						
Test for overall effect: Z = 1.13 (P = 0.26)						
Total (95% CI)			**398**	**433**	**100.0%**	**0.64 [0.23 , 1.05]**
Heterogeneity: Tau² = 0.21; Chi² = 34.58, df = 5 (P < 0.00001); I² = 86%						
Test for overall effect: Z = 3.09 (P = 0.002)						
Test for subgroup differences: Chi² = 18.21, df = 1 (P < 0.0001), I² = 94.5%						



Additional studies that were unable to be pooled also reported on this measure. Alroy 1984 compared tutorials based on discussion of trigger films with no specific interpersonal skills training (control) and reported on total time spent using supportive verbal and non-verbal communication behaviours. Although direct differences between groups were not provided, study authors reported that supporting behaviours of the intervention group increased significantly, whereas those of the control group decreased significantly (P < 0.025 and P < 0.01, respectively) for both verbal and non-verbal behaviour. Wundrich 2017 compared an empathy training course with a control seminar on an arbitrary psychiatry topic and reported that both simulated patients and assessors scored intervention group students as significantly more empathic than control group students (P < 0.0001).

Taken together, these studies indicate that communication skills interventions may increase students' empathy towards patients when assessed by expert examiners but not by simulated patients. Again, effect sizes were generally small, at less than 1 point on a 5-point scale.

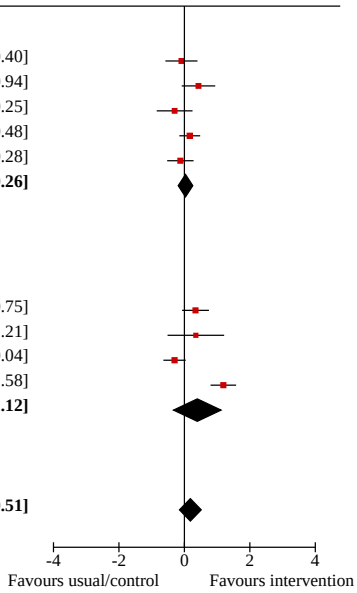
Relationship building/rapport

Communication skills interventions may have little to no impact on relationship building scores, in addition to the usual curriculum or control (SMD 0.18, 95% CI -0.15 to 0.51; 9 studies, 834 participants; I^2 = 81%; Analysis 1.8; Figure 7; low-quality evidence, downgraded due to only partially explained heterogeneity and imprecision of results, with a confidence interval crossing zero indicating that the true effect could be either a benefit or a harm). Subgroup analyses indicate that results were similar when assessed by experts (SMD 0.03, 95% CI -0.19 to 0.26; 5 studies, 456 participants; I^2 = 23%) and simulated patients (SMD 0.39, 95% CI -0.33 to 1.12; 4 studies, 378 participants; I^2 = 91%). Subgroup differences were not statistically significant (P = 0.35). No studies were deemed at high risk of bias. Results remained consistent when clustering was adjusted for by inflating the standard error from 5% to 20% (see Analysis 1.9; Analysis 1.10; Analysis 1.11). Differences between groups were consistently small across the varied scales used in different studies (e.g. less than 1 point on a 12-point scale (Hobgood 2009), 1 point on an 8-point scale (Evans 1989)).

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 7. Forest plot of comparison: 1 Communication intervention versus control or usual care, outcome: 1.8 Relationship building/Rapport.

Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.8.1 Assessor rating						
Bosse 2012	-0.09	0.2501	32	32	10.9%	-0.09 [-0.58 , 0.40]
Evans 1989	0.43	0.2614	30	30	10.7%	0.43 [-0.08 , 0.94]
Gartmeir 2015	-0.2989	0.2793	54	17	10.3%	-0.30 [-0.85 , 0.25]
Liu 2016	0.1641	0.1624	59	107	12.7%	0.16 [-0.15 , 0.48]
Lupi 2012	-0.123	0.2057	50	45	11.8%	-0.12 [-0.53 , 0.28]
Subtotal (95% CI)			**225**	**231**	**56.5%**	**0.03 [-0.19 , 0.26]**
Heterogeneity: Tau² = 0.01; Chi² = 5.18, df = 4 (P = 0.27); I² = 23%						
Test for overall effect: Z = 0.30 (P = 0.77)						
1.8.2 SP perception/satisfaction/rapport						
Hobgood 2009	0.3392	0.2094	35	68	11.8%	0.34 [-0.07 , 0.75]
Legg 2005	0.35	0.441	11	10	7.3%	0.35 [-0.51 , 1.21]
Shaddheau 2015	-0.3003	0.1742	75	60	12.5%	-0.30 [-0.64 , 0.04]
Windish 2005	1.19	0.1994	59	60	12.0%	1.19 [0.80 , 1.58]
Subtotal (95% CI)			**180**	**198**	**43.5%**	**0.39 [-0.33 , 1.12]**
Heterogeneity: Tau² = 0.48; Chi² = 31.68, df = 3 (P < 0.00001); I² = 91%						
Test for overall effect: Z = 1.06 (P = 0.29)						
Total (95% CI)			**405**	**429**	**100.0%**	**0.18 [-0.15 , 0.51]**
Heterogeneity: Tau² = 0.20; Chi² = 41.18, df = 8 (P < 0.00001); I² = 81%						
Test for overall effect: Z = 1.07 (P = 0.28)						
Test for subgroup differences: Chi² = 0.86, df = 1 (P = 0.35), I² = 0%						

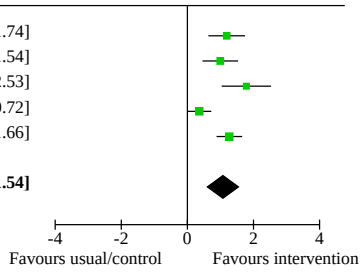


Information gathering

Communication skills interventions probably improve scores on information gathering about patient perspectives in OSCEs in comparison to the usual curriculum or control (SMD 1.07, 95% CI 0.61 to 1.54; 5 studies, 405 participants; I^2 = 78%; Analysis 1.12; Figure 8; moderate-quality evidence, downgraded due to considerable heterogeneity). All studies that could be pooled were rated by experts, and none were deemed as having high risk of bias overall. The size of differences varied, ranging from 4 out of 100 possible points in Bosse 2012 to 2.6 out of 4 possible points in Ho 2008.

Figure 8. Forest plot of comparison: 1 Communication intervention versus control or usual care, outcome: 1.12 Information gathering about patient perspectives/concerns.

Study or Subgroup	Intervention Mean	SD	Total	Usual curriculum/control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	85.72	3.51	28	79.64	6.07	32	19.3%	1.19 [0.64 , 1.74]
Evans 1989	4.5	1.82	30	3	1.05	30	19.6%	1.00 [0.46 , 1.54]
Ho 2008	3.27	0.8	15	1.85	0.77	27	15.8%	1.78 [1.04 , 2.53]
Lee 2015	2.48	0.98	60	2.1	1.11	59	22.8%	0.36 [-0.00 , 0.72]
Schwartz 2010	0.9	0.1614	65	0.62	0.2686	59	22.4%	1.27 [0.88 , 1.66]
Total (95% CI)			**198**			**207**	**100.0%**	**1.07 [0.61 , 1.54]**
Heterogeneity: Tau² = 0.22; Chi² = 18.33, df = 4 (P = 0.001); I² = 78%								
Test for overall effect: Z = 4.49 (P < 0.00001)								
Test for subgroup differences: Not applicable								



Explanation and planning

Outcomes relating to explanation and planning were insufficiently homogenous for pool in a meta-analysis. Reported outcomes included students' comfort in delivering test results (Shaddheau 2015), breaking news of a death or an event and explaining the situation (Bowyer 2010; Hobgood 2009), providing information (Seim 1995), and delivering advice and assistance for smoking cessation (Strayer 2010). In general, these studies reported some favourable outcomes in their intervention groups relative to their control groups, but results were mixed both within and between studies, and effect sizes were generally small. We are uncertain about effects on these outcomes overall.

Bowyer 2010 reported that students who received any of their training models (15-minute training video on SPIKES model, 45-minute lecture, lecture and video) were rated by the simulated patient wife as superior to those who were given no pre-training (control group) in several skills including those associated with explanation and planning: students appeared to have a plan, students assessed how much I knew, students were aware of my ability to discuss the news, and students prepared me to

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

receive the news. Only video and combined groups were rated as superior in the following: students explaining events of the death and students checking my understanding. Only lecture and combined groups were superior in the following: students reinforced and clarified the information and students comforted me. Similarly, Hobgood 2009 compared the death notification skills of students after pre-exposure to a simulated survivor with or without written feedback, or after no pre-exposure. Both pre-exposed groups improved their death notification competence from before to after the intervention, and the feedback group showed significantly superior improvement. Differences between the three groups in competence and confidence did not reach significance (P = 0.61 and P = 0.144, respectively). It is important to note that communication skills scores in both groups declined after training, with no significant differences between groups.

Seim 1995 compared presentation of a 2.5-hour multimedia workshop with distribution of an algorithm only (usual curriculum) on the ability of students to establish a smoking cessation plan for a simulated patient. The intervention group obtained higher scores in providing information (P < 0.05) and in eliciting/responding to feelings (see Analysis 1.7) but not in eliciting information. Intervention group students also obtained significantly higher scores in three out of six content areas, including negotiating a plan (P < 0.01). Strayer 2010 investigated students' skills in smoking cessation counselling, comparing those who were provided with a personal digital assistant-based tool after a smoking cessation workshop with those who received the workshop and a paper-based tool only. Few significant differences were noted between groups in smoking cessation counselling behaviour, with the following exceptions in which the control group performed better than the intervention group: asking about confidence in changing smoking behaviour (P = 0.02), listening reflectively (P = 0.01), and using summary statements (P = 0.03). All other communication and explanation and planning skills were non-significantly different between groups but were generally higher in the control group. The intervention group performed 62% of the key smoking cessation counselling activities and the control group performed 69%.

Shaddheau 2015 compared students' performance in a pregnancy options counselling OSCE after receiving the standard curriculum only or the standard curriculum plus a novel patient interaction with a patient who discussed her personal experiences on the topic. Overall scores did not differ significantly between groups (P = 0.08), but fewer intervention group students appeared uncomfortable when delivering results (P = 0.006).

Quality of evidence for this outcome was graded as very low, downgraded by one level due to risk of bias, downgraded by one level due to heterogeneity of results and methods, and downgraded by one level due to imprecision of results (small differences and large standard deviations).

Specific communication tasks

Five studies comparing an intervention group with a control or usual care group reported on a range of specific communication skills, with mixed findings across and within studies. These studies could not be pooled due to high heterogeneity. In general, students exposed to interventions show improvement in some but not all of the specific skills measured, and the extent of differences between groups is small (e.g. less than 1 point on a 4-point (Allen 1990) or a 5-point scale (Bowyer 2006)). When differences were greater (e.g. over 2 points on an 8-point scale (Evans 1989)), standard deviations were also large. Quality of evidence for those outcomes was graded as low; downgraded by one level due to heterogeneity of results and methods, and downgraded by one level due to imprecision of results (small differences and large standard deviations).

Allen 1990 used ANOVA to explore differences in assessor ratings of students between control and workshop groups, reporting significantly better performance in the workshop group on the use of open questions (P < 0.001 across four types of questions) and on the overall process (structure of the consultation; P < 0.005 on both of two items) but not in giving information. Bowyer 2006 used simulated patients to rate students' performance and reported that students who received any of three interventions (compared to usual curriculum) performed better on four checklist items (having a plan for the encounter, assessing knowledge, being aware of patients' ability to discuss the news, and preparing for bad news; P < 0.03 across all comparisons) but found no differences in 12 other items including providing comfort, choosing the language used, detecting emotions, responding appropriately, and inspiring trust. Evans 1989 compared a series of lectures and experiential workshops with standard patient clerking and reported that students' communication skills improved after the lectures (opening and closing the interview, detecting leads, providing warmth, choosing appropriate question style, and ensuring clarity) and further after the workshops. After workshop exposure, students' performance was significantly superior to performance of the control group in opening the interview (P < 0.0001), seating appropriately (P < 0.001), maintaining body posture (P < 0.001), facilitating conversation (P < 0.001), addressing psychosocial issues (P < 0.05), providing empathy (P < 0.0001), using silence (P < 0.0001), addressing personal issues (P < 0.0001), detecting leads (P < 0.05), choosing appropriate question style (P < 0.0001), ensuring clarity (P < 0.0001), and closing the interview (P < 0.0001) but not in making eye contact, avoiding interruptions, ensuring relevance, or providing warmth. Gartmeir 2015 compared the effectiveness of e-learning with use of video and role-play practice versus a wait list control in terms of students' shared decision-making skills. This group found that both combined (P < 0.001) and individual interventions (P = 0.003 and P = 0.023, respectively) were associated with higher scores in structuring the conversation compared to the control group. Ishikawa 2010 was not pooled, as this was the only study to directly report on non-verbal behaviour. This study compared usual training versus an intervention focusing on non-verbal communication and found no significant differences in non-verbal communication between intervention and control groups (no data provided). Legg 2005 compared interactive training in taking a history regarding aphasia with theoretical information only and found that experimental group students were superior to control group students in establishing initial rapport (P < 0.05), exploring problems (P < 0.01), providing structure to the consultation (P < 0.001), developing rapport (P < 0.001), acknowledging patient competence (P < 0.05), ensuring patient understanding (P < 0.01), and verifying information (P < 0.05) as assessed by experts but were not superior in any patient ratings nor in identifying reasons for the consultation, understanding the patient's perspective, involving the patient, or ensuring that the patient has a means of responding.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

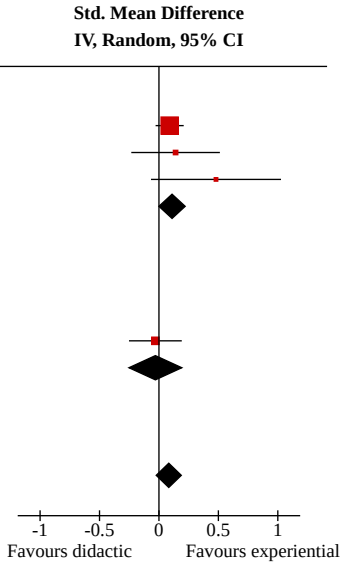
Comparison 2. Communication intervention 1 versus communication intervention 2 – mode of delivery comparisons

Overall communication skills

Sufficient studies comparing experiential and didactic intervention approaches and reporting on overall communication outcomes were available for pooling. We are uncertain whether experiential communication interventions are superior to didactic approaches in achieving high overall communication scores on OSCEs assessed by experts or by simulated patients (SMD 0.08, 95% CI -0.02 to 0.19; 4 studies, 1578 participants; I^2 = 4%; Analysis 2.1; Figure 9; very low-quality evidence, downgraded by two points due to methodological limitations, with high risk of bias across several domains in three of four studies, and imprecision of results, with a confidence interval crossing zero, meaning that the true effect could be either a benefit or a harm). Subgroup analysis including studies with outcomes assessed by experts only remained consistent (SMD 0.11, 95% CI 0.00 to 0.22; 3 studies, 1266 participants; I^2 = 0%). Only one study that could be included in the meta-analysis reported outcomes assessed by simulated patients, and subgroup differences were not significant (P = 0.27). All but one of these studies - Ockene 2016 - were deemed at high risk of bias. Results remained consistent in sensitivity analyses with adjustment for clustering by inflating the standard error from 5% to 20% (see Analysis 2.2; Analysis 2.3; Analysis 2.4). Overall differences between groups were small and standard deviations were large.

Figure 9. Forest plot of comparison: 2 Communication intervention 1 versus communication intervention 2 mode of delivery, outcome: 2.1 Overall communication skills.

Study or Subgroup	SMD	SE	Experiential/video Total	Didactic Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 Assessor rating						
Ockene 2016	0.09	0.0603	571	531	67.1%	0.09 [-0.03 , 0.21]
Roche 1997	0.14	0.1901	56	55	7.8%	0.14 [-0.23 , 0.51]
Walsh 1999	0.48	0.279	27	26	3.7%	0.48 [-0.07 , 1.03]
Subtotal (95% CI)			**654**	**612**	**78.6%**	**0.11 [-0.00 , 0.22]**
Heterogeneity: Tau² = 0.00; Chi² = 1.89, df = 2 (P = 0.39); I² = 0%						
Test for overall effect: Z = 1.96 (P = 0.05)						
2.1.2 SP rating						
Wagner 2011	-0.03	0.1133	151	161	21.4%	-0.03 [-0.25 , 0.19]
Subtotal (95% CI)			**151**	**161**	**21.4%**	**-0.03 [-0.25 , 0.19]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.26 (P = 0.79)						
Total (95% CI)			**805**	**773**	**100.0%**	**0.08 [-0.02 , 0.19]**
Heterogeneity: Tau² = 0.00; Chi² = 3.12, df = 3 (P = 0.37); I² = 4%						
Test for overall effect: Z = 1.54 (P = 0.12)						
Test for subgroup differences: Chi² = 1.23, df = 1 (P = 0.27), I² = 18.6%						



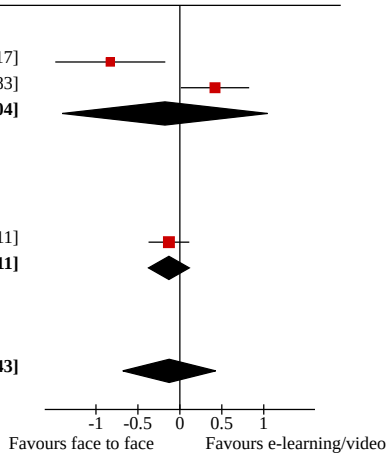
Empathy

Sufficient studies comparing face-to-face and e-learning or video interventions and reporting on empathy outcomes were available for pooling. There may be little to no difference between interventions in achieving high empathy scores (SMD -0.13, 95% CI -0.68 to 0.43; 3 studies, 421 participants; I^2 = 82%; Analysis 2.5; Figure 10; low-quality evidence, downgraded due to unexplained heterogeneity and imprecision of results, with a confidence interval crossing zero meaning that the true effect could be either a benefit or a harm). Results were similar in subgroup analysis including outcomes assessed by experts only (SMD -0.18, 95% CI -1.40 to 1.04; 2 studies, 149 participants; I^2 = 90%; non-significant subgroup differences P = 0.94), and no studies were deemed at high risk of bias. Scales used differed across studies in the proportion of students meeting the standard (33% versus 69% in Palmer 2018), in the number of empathetic statements made (mean 1.48, standard deviation (SD) 1.83 versus mean 0.82, SD 1.02, in Kaltman 2018), and in score on a 5-point scale (mean 4.14, SD 0.9 versus mean 4.03, SD 0.8, in Bowyer 2010).

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 10. Forest plot of comparison: 2 Communication intervention 1 versus communication intervention 2 mode of delivery, outcome: 2.5 Empathy.

Study or Subgroup	SMD	SE	E-learning/video Total	Face to face Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.5.1 Assessor rating						
Palmer 2018	-0.8292	0.3345	24	26	26.5%	-0.83 [-1.48 , -0.17]
Kaltman 2018	0.4189	0.2079	60	39	34.4%	0.42 [0.01 , 0.83]
Subtotal (95% CI)			**84**	**65**	**60.9%**	**-0.18 [-1.40 , 1.04]**
Heterogeneity: Tau² = 0.70; Chi² = 10.04, df = 1 (P = 0.002); I² = 90%						
Test for overall effect: Z = 0.28 (P = 0.78)						
2.5.2 SP report						
Bowyer 2010	-0.1304	0.1239	163	109	39.1%	-0.13 [-0.37 , 0.11]
Subtotal (95% CI)			**163**	**109**	**39.1%**	**-0.13 [-0.37 , 0.11]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.05 (P = 0.29)						
Total (95% CI)			**247**	**174**	**100.0%**	**-0.13 [-0.68 , 0.43]**
Heterogeneity: Tau² = 0.19; Chi² = 10.92, df = 2 (P = 0.004); I² = 82%						
Test for overall effect: Z = 0.45 (P = 0.65)						
Test for subgroup differences: Chi² = 0.01, df = 1 (P = 0.94), I² = 0%						

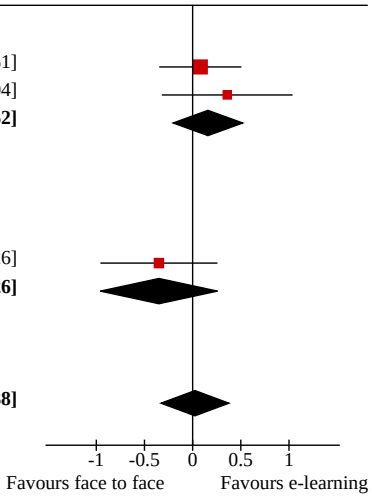


Relationship building/rapport

Sufficient studies comparing face-to-face and e-learning or video interventions and reporting on relationship building/rapport outcomes were available for pooling. There is probably little to no difference between interventions in achieving high rapport scores (SMD 0.02, 95% CI -0.33 to 0.38; 3 studies, 176 participants; I^2 = 19%; Analysis 2.6; Figure 11; moderate-quality evidence, downgraded due to imprecision of results, with a confidence interval that crosses zero, indicating that the true effect could be either a benefit or a harm). Results were similar in subgroup analysis including outcomes assessed by experts only (SMD 0.16, 95% CI -0.20 to 0.52; 2 studies, 126 participants; I^2 = 0%; no significant subgroup differences P = 0.16), and no studies were deemed to be at high risk of bias. Differences between groups were again small, ranging from 50% to 65% of students meeting the standard (Palmer 2018), a score of 3.34 (SD 0.6) to 3.24 (SD 0.56) on a 4-point scale (Liu 2016), or a score of 1.8 (standard error (SE) 0.07) to 1.92 (SE 0.08) on an inverse performance scale (Gartmeir 2015).

Figure 11. Forest plot of comparison: 2 Communication intervention 1 versus communication intervention 2 mode of delivery, outcome: 2.6 Relationship building/rapport.

Study or Subgroup	SMD	SE	E-learning Total	Face to face Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.6.1 Assessor rating						
Liu 2016	0.0797	0.2175	33	59	48.7%	0.08 [-0.35 , 0.51]
Gartmeir 2015	0.3587	0.3461	17	17	23.2%	0.36 [-0.32 , 1.04]
Subtotal (95% CI)			**50**	**76**	**72.0%**	**0.16 [-0.20 , 0.52]**
Heterogeneity: Tau² = 0.00; Chi² = 0.47, df = 1 (P = 0.49); I² = 0%						
Test for overall effect: Z = 0.86 (P = 0.39)						
2.6.2 SP perception/satisfaction/rapport						
Palmer 2018	-0.35	0.31	24	26	28.0%	-0.35 [-0.96 , 0.26]
Subtotal (95% CI)			**24**	**26**	**28.0%**	**-0.35 [-0.96 , 0.26]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.13 (P = 0.26)						
Total (95% CI)			**74**	**102**	**100.0%**	**0.02 [-0.33 , 0.38]**
Heterogeneity: Tau² = 0.02; Chi² = 2.46, df = 2 (P = 0.29); I² = 19%						
Test for overall effect: Z = 0.13 (P = 0.89)						
Test for subgroup differences: Chi² = 1.99, df = 1 (P = 0.16), I² = 49.8%						



Information gathering

Only one study comparing intervention modes reported on information gathering outcomes. Wagner 2011 compared rating of video interviews with a didactic lecture only and reported on students' ability to cover smoking cessation-specific content and found differences in simulated patient-reported checklist scores across different simulated patient encounters (OSCE stations). Use

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

of the modules was associated with slightly higher scores in some but not all areas of student performance. Quality of evidence was rated as very low and was downgraded by two levels due to methodological limitations (quasi-RCT) and by one level due to imprecision (single study, small sample size).

Explanation and Planning

One study comparing intervention modes reported on explanation and planning outcomes. Strayer 2010 reported that based on an intention-to-treat analysis, students exposed to a paper-based smoking cessation counselling tool performed markedly better on several measures of motivational interviewing (including explanation and planning) than those exposed to the paper-based tool plus a personal digital assistant-based tool (P = 0.022), but the difference did not persist when those not exposed to the intervention were excluded. Evidence for this outcome was rated as low quality and was downgraded by two levels due to imprecision (single study, small sample size).

Specific communication tasks

Due to the heterogeneity of these outcomes, we did not pool studies reporting on specific communication tasks. Results varied across outcomes both between and within studies, with overall little evidence for any appreciable differences between online and in-person learning for specific communication skills. Gartmeir 2015 found that little difference between e-learning and role-play in outcomes for medical students. Palmer 2018 found no significant differences between skills in introducing (P = 0.62) and closing the consultation (P = 0.62) between students who were exposed to in-class or online learning.

Four studies compared effects of different tutors or tutor training on student performance. It seems that inclusion of patients in teaching may have some impact on communication skills, but effects of varied tutor training or tutor qualification level are unclear, with a small number of studies suggesting this may involve little difference.

Gerber 1985 measured overall interpersonal skills used during a physical examination and reported that students trained and assessed by Teaching Assistant Simulated Patients (TASPs) performed significantly better than students in the control group on six of the eight interpersonal skills measured (e.g. explanation given, questions invited), and these improvements were maintained two years later for three of the eight items (mean differences significant at least at the 0.05 level). Meirovich 2016 reported no significant differences (P values not provided) in students' use of open- or closed-ended questions between experimental (students taught by tutors coached in learner-centred pedagogy) and control groups (taught by tutors not trained in learner-centred pedagogy). Nomura 2017 compared effects of a role-play workshop led by fifth year student tutors versus one run by physicians and did not find an appreciable difference between the two. Snow 2016 compared two e-learning modules: for control, a video of a clinician describing a colposcopy, and for intervention, video clips of patients describing their colposcopy. Students performance on an OSCE was rated using the Doctors' Interpersonal Skills Questionnaire, and those in the experimental group performed better (P = 0.016).

Overall, the quality of evidence for this outcome was rated as very low; downgraded by one level due to methodological limitations (significant risk of bias across several domains) and downgraded by two levels due to substantial heterogeneity within and between studies.

Comparison 3. Communication intervention 1 versus communication intervention 2 – feedback approach comparisons

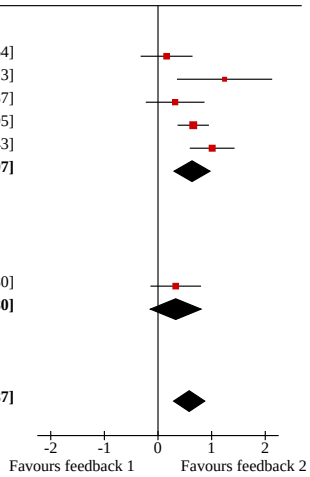
Overall communication skills

Interventions that include personalised or structured feedback probably improve overall OSCE scores to a greater extent than those using more general feedback (SMD 0.58, 95% CI 0.29 to 0.87; 6 studies, 502 participants; I^2 = 56%; Analysis 3.1; Figure 12 moderate-quality evidence, downgraded due to high or unclear risk of bias across several domains). Subgroup analyses indicate that results were similar when limited to those assessed by experts only (SMD 0.63, 95% CI 0.3 to 0.97; 5 studies, 432 participants; I^2 = 60%), but only one study that could be included in meta-analysis reported outcomes assessed by simulated patients (no significant subgroup differences P = 0.30). Sensitivity analysis excluding studies at highest risk of bias remained consistent (SMD 0.56, 95% CI 0.17 to 0.96; 5 studies, 312 participants; I^2 = 64%; Analysis 3.2). Results also remained consistent in sensitivity analyses with adjustment for clustering through standard error inflation of 5% to 20% (see Analysis 3.3; Analysis 3.4; Analysis 3.5).

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 12. Forest plot of comparison: 3 Communication intervention 1 versus communication intervention 2 feedback approach, outcome: 3.1 Overall communication skills.

Study or Subgroup	SMD	SE	Intervention 1_feedback 1 Total	Intervention 2_feedback 2 Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.1.1 Assessor rating						
Engerer 2019	0.16	0.2467	34	32	16.8%	0.16 [-0.32 , 0.64]
Maguire 1978	1.2414	0.4529	12	12	8.0%	1.24 [0.35 , 2.13]
Ozcakar 2009	0.32	0.2795	27	25	14.9%	0.32 [-0.23 , 0.87]
Perera 2010	0.6572	0.1491	97	93	24.0%	0.66 [0.36 , 0.95]
Ruesseler 2017	1.01	0.2129	49	51	19.1%	1.01 [0.59 , 1.43]
Subtotal (95% CI)			**219**	**213**	**82.8%**	**0.63 [0.30 , 0.97]**
Heterogeneity: Tau² = 0.08; Chi² = 9.89, df = 4 (P = 0.04); I² = 60%						
Test for overall effect: Z = 3.71 (P = 0.0002)						
3.1.2 SP rating						
Hobgood 2009	0.3306	0.2408	35	35	17.2%	0.33 [-0.14 , 0.80]
Subtotal (95% CI)			**35**	**35**	**17.2%**	**0.33 [-0.14 , 0.80]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.37 (P = 0.17)						
Total (95% CI)			**254**	**248**	**100.0%**	**0.58 [0.29 , 0.87]**
Heterogeneity: Tau² = 0.07; Chi² = 11.29, df = 5 (P = 0.05); I² = 56%						
Test for overall effect: Z = 3.93 (P < 0.0001)						
Test for subgroup differences: Chi² = 1.06, df = 1 (P = 0.30), I² = 5.4%						



Roche 1996 compared audio, peer, and video feedback and reported that video feedback group scores on assessor ratings of video-recorded OSCEs were significantly higher than those for both audio and peer feedback groups at post test (P < 0.05), but overall differences between intervention groups were small and non-significant. All three feedback groups had significantly higher scores than the control group. These data were unable to be pooled, as no measure of error was available.

These results suggest that tailored feedback is probably more effective than general feedback, but it should be noted that effect sizes are predominantly small.

Empathy

Engerer 2019 compared specific structured and behaviour-oriented feedback on communication skills with general experience-oriented feedback. Unpublished data reveal that the group receiving specific feedback had non-significantly higher scores on handling patient emotions than the general feedback group. Evidence was rated as low quality, downgraded by two levels due to imprecision (single, small study).

Relationship building/rapport

Perera 2010 compared teacher feedback with objectively structured self-assessment and peer feedback and reported that the group engaging in self and peer feedback performed significantly better than the comparison group in building rapport (3.82 versus 3.46 out of 4; P = 0.0001) and in listening (3.71 versus 3.43 out of 4; P = 0.006) and interview style (0.90 versus 0.60 out of 1; P = 0.0001) but not in language (3.94 versus 3.89 out of 4; P = 0.285) or interview structure (0.93 versus 0.86 out of 1; P = 0.132). The quality of evidence was rated as very low, downgraded by one level due to methodological limitations (significant risk of bias across several domains) and by two levels due to imprecision (single, small study).

Information gathering

Maguire 1978 compared feedback based on video or audio recordings of interviews with paper-based ratings of practice interviews, as well as with a control. This group reported that all three intervention groups performed better than the control group in terms of the amount of information elicited, as well as in significant between groups differences (P < 0.05), with the video feedback group outperforming both of the other groups. Evidence was rated as low quality, downgraded by two levels due to imprecision (single, small study).

Explanation and planning

No studies reported results for this outcome.

Specific communication tasks

Ruesseler 2017 compared video-assisted feedback on role-plays with verbal feedback alone and reported higher scores in the video-assisted feedback group on all five specific skills measured: introduction (P = 0.029), intelligibility (P < 0.001), questioning (P < 0.001), impression and impact (P = 0.001), and interaction (P = 0.009). Evidence was rated as low quality, downgraded by two levels due to imprecision (single, small study).

Comparison 4. Communication intervention 1 versus communication intervention 2 – simulated or real patient versus peer role-play comparisons

Overall communication skills

We are uncertain whether communication skills interventions using simulated or real patients were superior to those using peer role-play in achieving high overall scores on OSCEs (SMD 0.17, 95% CI -0.33 to 0.67; 4 studies, 637 participants; I^2 = 87%; Analysis 4.1; Figure 13; very low-quality evidence, downgraded due to methodological limitations with a substantial proportion of evidence from quasi-RCTs, and high risk of bias across several domains, with high unexplained heterogeneity and imprecision of results with the confidence interval crossing zero). Subgroup

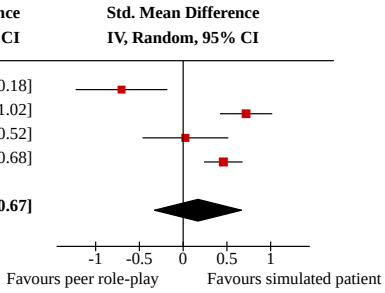
Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

analyses indicate that results were similar when one study using real patients (as opposed to simulated patients) was excluded (SMD 0.20, 95% CI -0.41 to 0.82; 3 studies, 573 participants; I^2 = 91%; Analysis 4.2), as were results of sensitivity analysis when studies at highest risk of bias were excluded (SMD -0.33, 95% CI -1.05 to 0.38; 2 studies, 124 participants; I^2 = 75%; Analysis 4.3). Results remained consistent in sensitivity analyses with adjustment for clustering by inflation of the standard error 5% to 20% (see Analysis 4.4; Analysis 4.5; Analysis 4.6). Differences between groups across these studies were generally small.

Figure 13. Forest plot of comparison: 4 Communication intervention 1 versus communication intervention 2 SP vs peer, outcome: 4.1 Overall communication skills.

Study or Subgroup	SMD	SE	Simulated (real) patient Total	Peer role-play Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	-0.7043	0.2672	32	28	22.3%	-0.70 [-1.23 , -0.18]
Levenkron 1990	0.7188	0.1526	91	92	26.7%	0.72 [0.42 , 1.02]
Stolz 2012	0.0256	0.2501	33	31	23.0%	0.03 [-0.46 , 0.52]
Taylor 2019	0.4591	0.1121	148	182	28.0%	0.46 [0.24 , 0.68]
Total (95% CI)			**304**	**333**	**100.0%**	**0.17 [-0.33 , 0.67]**

Heterogeneity: Tau² = 0.22; Chi² = 23.90, df = 3 (P < 0.0001); I² = 87%
Test for overall effect: Z = 0.67 (P = 0.50)
Test for subgroup differences: Not applicable



Empathy

Vanatta 1996 reported that students who practised with and received feedback from simulated patients performed better than those in the faculty feedback group but not better than the peer role-play group on empathy scores from the Arizona Clinical Interview Rating Scale (scores of 3.36, 3.09, and 3.32, respectively; F = 4.27, P = 0.02). Mounsey 2006 also reported on empathy scores after students engaged with either a simulated patient or peer role-play and found no significant differences between groups (4.04 versus 4.02; P = 0.92). Evidence was rated as low quality, downgraded by one level due to heterogeneity within and between studies and by one level due to imprecision (small sample size).

Relationship building/rapport

No studies reported results for this outcome.

Information gathering

No studies reported results for this outcome.

Explanation and planning

No studies reported results for this outcome.

Specific communication tasks

Vanatta 1996 reported that students who practised with and received feedback from simulated patients (SPs) achieved superior results on open questions (F = 3.58, P = 0.03 in comparison between SP practice and feedback (3.69), SP practice and faculty feedback (3.39), and peer practice and feedback (3.51)), but no significant differences were observed in facilitative behaviours (3.19, 3.81, 3.09 across groups). Evidence was rated as low quality, downgraded by two levels due to imprecision (single, small study).

Comparison 5. Communication intervention 1 versus communication intervention 2 – alternate intervention structure

Twenty-six interventions compared alternate structures of the same fundamental approach. Due to the heterogeneity of interventions and outcomes, these were not pooled in the meta-analysis. Individual studies are described briefly below, but overall, these results show that although educational interventions tend to have an effect in comparison to no intervention or a control with a different focus, more subtle differences in intervention structure do not produce consistent results.

Berney 2017 compared their 'usual' approach involving one student per group interacting with an SP as stimulus for group discussion with a more intensive intervention in which each student was involved in a role-play and subsequent individual supervision. Students in the intervention group achieved higher scores compared to students in the comparison group on the overall evaluation of the interview (P < .001) and on process skills (P < 0.001), as judged using the Calgary-Cambridge tool. Similarly, Levenkron 1990 compared a lecture and resource package followed either by a group or by an individual simulated patient experience to improve students' behavioural counselling skills. This group reported that the individually trained group performed better as demonstrated by the overall score (P < 0.001), as well as by the subscales of relationship building (P < 0.001), positive focus (P < 0.001), investigating change (P < 0.05), and general technique (P < 0.001). Similarly, Steifel 2013 compared individual training with group training for breaking bad news. Data are not provided, but the study authors report that individual training is supported by differences between groups in results on the Roter Interactional Assessment Scale. Together, these studies consistently support individual training as superior to group training in improving students' skills, although the extent of effects may vary.

Nine studies compared interventions with and without additional components. Betchart 1984 compared their standard interviewing course versus the addition of four tutorial sessions for students with identified weaknesses and reported that following the intervention, the five experimental students scored higher than the other seven initially weak students on both process (P < 0.01) and content (P < 0.05) elements, with overall differences of 8 to 12 out of 60. Further comparison indicated that the scores of all 12 of these weak students remained below the rest of the cohort at the end of the course. Chibhall 2005 compared an experiential tutorial with and without religious awareness training and reported that the training had an effect for women only on affective (difference of 0.6 out of 4; P < 0.05) but not on communication behaviours (data

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

not provided). Ho 2008 compared the basic communication skills of students who had attended a cultural competence workshop with and without an additional session for practice with role-play. The extensive intervention group scored higher than the basic intervention group and the control group in eliciting patient perspectives (67% of extensive intervention compared with 11% and 13% of control and basic intervention, respectively; P < 0.001), as well as in history taking (scores of 8.9 compared with 7.3 and 7.6; P = 0.03) and in differential diagnosis (score of 2.9 compared with 2.1 and 1.5; P < 0.001) but not in basic communication skills (score of 4.8 compared with 4.2 and 4.1). Ishikawa 2010 compared a training programme involving simulated patient role-play and discussion, with one involving specific feedback regarding non-verbal communication skills. No differences were found between groups in students' non-verbal communication skills, but no data were presented. Lim 2011 used analysis of variance (ANOVA) to compare workshops involving peer role-play with and without the addition of 'how to act in role' training, with the intervention group scoring higher for empathy than the comparison group (P < 0.001). Mason 1988 compared an intervention based on an instructional videotape with one using the instructional videotape as well as a critique of a pre-intervention role-play video, as well as additional groups exposed to the review of pre-intervention video only, and a control group receiving no instruction. Both groups that reviewed the video performed better in interviews than the control group, but reviewing their own pre-intervention video did not add benefit over viewing the instructional video (between-groups data not available). Moreland 1973 compared a standard training course involving each student interviewing volunteer patients with one in which students focused on a particular micro skill in each session and were provided with a description and demonstration of the skill, and one student practised an interview. Both groups improved across a range of counselling skills, but the micro-counselling group showed greater improvement in attending behaviour (F = 7..41, df = 1/16; P < 0.025) and reflection of feeling (F = 12.09, df = 1/16; P < 0.025), representing two out of ten items across the rating scale and coding system used. Morrow 2009 compared a standard course teaching students to integrate electronic health records (EHRs) into physician-patient communication with a course with EHR-specific communication skills added through brief didactics and practice role-plays. Intervention group students performed better than control group students in 6 of 10 EHR communication skills (P < 0.05 for all; for example, introducing self before training to computer, adjusting the chair to be at eye level, adjusting the screen to allow the patient to see it easily; differences range from no students in the control group and all students in the intervention group performing a task to three control (out of eight) and eight intervention (out of nine) students performing some tasks). However, for 10 of 11 general communication skills, there were no differences between groups (generally most students in both groups performed the basic communication tasks). Siassakos 2010 compared students' performance after attending an obstetric simulation on shoulder dystocia with and without an additional hybrid simulation session with a simulated patient. Students attending the hybrid simulation had higher total patient satisfaction (median 11 versus 9; P = 0.0239) and median communication (4 versus 3; P = 0.0128) scores but no differences in respect (median 3 versus 3) and safety (median 4 versus 3; not statistically significant) scores.

Overall, results for these studies are mixed, with additional intervention components tending to have some effects on the skills directly associated with additional training but not on more general skills.

Two studies compared different approaches to preparing students for OSCE assessment. Blatt 2010 compared students' performance in clinical examination after either a standard briefing about the assessment or a briefing with additional instruction regarding consideration of patients' perspective. Students briefed on patient perspective showed higher patient satisfaction scores (mean patient satisfaction 4.04 (5-point scale) compared with a mean score of 3.94 for the control group; P = 0.01). Similarly, Cave 2007 added to their usual communication skills sessions with a greater focus on preparation for assessment and compared directing students to the assessment criteria, providing the assessment criteria, and running a session as a mini-OSCE. Study authors reported no differences between the three groups in end of year OSCE performance (mean scores of 61.8, 62.9, and 62.3 for the standard arm and two intervention arms, respectively). These mixed results make it difficult to draw any conclusions about the value of specific focus on preparation for assessment.

Three studies compared more with less structured interventions, some using a mnemonic to guide student learning and recall. LoSasso 2017 added to their regular training on electronic medical records - the use of a mnemonic to assist with EMR-specific communication skills (SALTED: set-up, ask, listen, type, exceptions, documentation). Study authors reported higher empathy and communication scores in the intervention group as judged by faculty (mean 4.4 versus 3.9; P = 0.02, and mean 4.6 versus 4.3; P = 0.01) but only empathy scores differed when judged by simulated patients (mean 3.5 versus 3.1; P = 0.07). Faculty-judged and simulated patient-judged history taking scores were also higher for the intervention group compared to the comparison group (mean 4.4 versus 3.8; P = 0.0001, and 3.6 versus 3.2; P = 0.05, respectively). Edwardsen 2006 compared a lecture and experiential training with and without guidance on a structured mnemonic for discussing intimate partner violence. The intervention group performed better on making an empathic statement (82% versus 45%; P = 0.027) and on asking about prior abuse (55% versus 19%; P = 0.027), but there were no differences between groups on measures focused on the content of the consultation. Weihs 1986 compared a structured, progressive approach to interview training with a more traditional experiential model, reporting better scores on responding to feeling with the intervention (scores of 2.56 (experimental group) and 1.31 (control group) on the 5-point Carkuff Developmental Helping Model Scale; P = 0.001).

Overall, these studies suggest that structured training including memory aids can enhance students' communication skills, but results are mixed both within and between studies, and effects are small.

Seven studies compared interventions that were fundamentally the same but had a subtly different focus. Bearman 2001 compared a narrative design to a problem-solving design in the case presented by their virtual patient and reported improvements in the narrative group relative to the problem-solving group in overall communication skills assessed one week after the tutorials (mean 38.8 versus 35.7; P = 0.034). Differences between these intervention groups were also observed on the individual items; using open questions and appropriate language (two out of ten individual items), but the differences between groups were

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

no longer present three weeks after the intervention. Foster 2016 reported that students who interacted with an empathy-feedback virtual patient showed slightly higher empathy than those interacting with a virtual patient with a back-story only and those in the control group (mean (SD) scores coded on a 0 to 6 scale were 2.91 (0.16) versus 2.20 (0.22) and 2.27 (0.21), respectively). Kahan 2003 assessed students' management of problem drinking and alcohol dependence after a three-hour workshop on either alcohol dependence (experimental) or depression (control). The experimental group performed better on all six items associated with appropriate assessment and advice (difference in means on a 0 or 1 scale range from 0.08 to 0.23; P = 0.000 or 0.002). Klein 2000 compared students who were taught a third-year interview methods course with patients with cancer (experimental) or other diagnoses (control). When they reached their fifth year (two years later), the experimental group had better ratings in terms of responding empathically (50% versus 21%; P < 0.05), showing regard and concern for the patient (95% versus 46%; P < 0.001), and assessing the impact of the symptoms on the patient's life (89% versus 54%; P < 0.001). Students in the experimental group were also more likely to shake the patient's hand (P < 0.05). However, these were the only 4 out of 33 individual rating items that were different between groups. Feddock 2009 compared students' skills after participating in a four-hour adolescent health medicine workshop (experimental) and an alternative workshop (control). Experimental group students performed better on both history taking (mean difference 14.9, 95% CI 11.5 to 18.3) and counselling (mean difference 19.6, 95% CI 14.9 to 24.3) subscales of the adolescent-medicine examination rating used. Maguire 1977 compared communication skills teaching using didactic versus 'discovery' approaches, as well as traditional seminars. Students who had completed the discovery or didactic courses obtained more relevant information (discovery mean 51.1, didactic 48.7, traditional 34.1; P < 0.01) and used many more of the desired techniques (discovery mean 45.7, didactic 42.2, traditional 32.1; P < 0.01) than those who had completed the traditional course. Schmitz 2018 compared an online module that delivered a worked example in text-based, video, or text-enriched video form. Students randomised to the enriched video arm broke bad news to simulated patients more appropriately than students in either of the other groups based on SPIKES (setting up the interview, assessing patient perception, obtaining an invitation, providing knowledge, addressing emotion, deciding strategy and summarising) (P < 0.01) and global Breaking Bad News Assessment Scale (gIBAS) scores (P < 0.05), but no further differences between groups were noted. Group-level scores were not specified.

Overall, mixed findings within and between these studies indicate wide variation in the impact of different structured interventions on students' communication skills. It seems that teaching approaches that use a patient-centred, narrative approach can improve skills associated with patient-centred communication. It is important to note that those studies comparing a topic-specific workshop to an alternative one tend to find that students' topic-specific skills improve but other, more generic skills may not improve relative to a comparison group.

Nomura 2017 conducted a randomised controlled non-inferiority trial to investigate the effectiveness of medical interview training led by final year student tutors compared with that led by physicians. Mean OSCE scores for the groups were 91.4 and 91.2, respectively (mean difference 0.2, 95% CI -1.8 to 2.2), which was within the pre-determined non-inferiority margin of 3.0. Study authors concluded that cross-year peer tutoring is effective for teaching communication skills to medical students.

Snow 2016 compared e-learning modules in which patients described their experience of colposcopy (experimental) with those in which a clinician described the procedure (comparison). The experimental group performed better in the OSCE in an overall interpersonal skills index, compared to the comparison group (OR 2.7, p = 0.016).

Longer-term outcomes

Four studies assessed outcomes at the end of the year in which the intervention was delivered, with three of the four reporting results in favour of the intervention group. Cave 2007 delivered communication skills teaching sessions to students four times in the year and assessed skills at the end of the year. In the three study arms, study authors compared standard teaching, presentation of assessment criteria, and running of sessions as mini-OSCEs. There were no differences in OSCE scores between the three arms. Colletti 2001 assessed students' skills in breaking bad news at the end of their third year in which they had training with a simulated patient instructor on a rectal cancer or pregnancy loss case, or neither experience. Students who previously had the experience of breaking the bad news of a pregnancy loss or a diagnosis of rectal cancer performed better on the clinical performance examination than students who had not had this experience (mean scores 1.8 (pregnancy loss), 2.0 (colon cancer), 2.4 (control) on a reverse scored scale from 1 (skilfully done) to 4 (not done); P < 0.05). The training experience accounted for 28% of the total variance in performance on the examination. Study authors report that some of the communication skills learned through these exercises appear to transfer across cases into different clinical situations. Filipetto 2006 examined use of a first-year preceptorship involving five hours per month of shadowing a preceptor for 10 consecutive months. The control group (previous cohort) did not participate in a preceptorship. The intervention group performed better on interpersonal/communication skills in an end of year OSCE than control students (mean 69.9 versus 65.2; P = 0.05). Strayer 2010 assessed smoking cessation counselling skills in students trained using a paper-based reminder or a personal digital assistant (PDA)-based tool and found no differences in skills between groups one month after or at the end of the year in which training occurred.

Four studies assessed outcomes a year or more beyond delivery of the intervention, and all reported at least some maintenance of advantage over control groups. Gerber 1985 assessed students immediately and two years after their simulated patient training programme. This study compared the skills of students trained using peers and faculty with those of students using teaching-associated simulated patients. Teaching associates worked in pairs to enable students to practise intimate physical examination on them and to provide guidance and feedback. Control group students practised physical examination on part-trainers. At immediate follow-up, students in the intervention group were superior in six of the eight interpersonal skills items, and two years later remained superior in three areas (overall P < 0.001; more specific explanations of the procedure about to be performed P < 0.01; inviting questions from the patient P < 0.01), retaining 96.6%, 64.7%, and 66.7% of the differences between groups in each area. As reported above, Klein 2000 compared students who were taught a third-year interview methods course using patients with

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

cancer versus students receiving the same training using patients with other diagnoses. At two-year follow-up, the intervention group was rated more highly on 4 of 33 individual rating scales. Maguire 1978 compared feedback provided to students using video replays, audio replays, or written ratings versus no feedback to students on their practice interviews. In a related study, 36 of the original participants (18 who had received feedback training and 18 in the control group) were followed up four to six years later, when they were qualified doctors. Each participating doctor interviewed two simulated patients and one real patient, and video-recorded interviews were rated. The experimental group maintained its advantage over the control group overall (mean score 29.61 or 67% versus 23.15 or 53%; P < 0.001) in 10 out of 11 specific skills (e.g. using open questions, enquiring about psychosocial symptoms, using verbal and visual encouragement). Meirovich 2016 ran an intervention with students in their first three clinical years and assessed their performance at the end of their third year. Students who were tutored by those who were being trained in learner-centred pedagogy were regarded as more patient-centred (mean 44.6 versus 40.6; P < 0.05) and received higher scores in 'building a relationship' (22.0 versus 19.5; P < 0.05) but lower scores in 'gathering data' (15.1 versus 17.4; P < 0.05) compared to control group students, who were trained by tutors who did not receive the training. No differences were found in the use of open or closed questions, doctor-centred-ness, or specific measures associated with motivational interviewing.

DISCUSSION

Summary of main results

This systematic review identified a large number of studies investigating effects of diverse communication skills interventions on medical students' interpersonal skills in medical consultations. The review found evidence of low certainty that compared to usual curricula or control, educational interventions can improve medical students' overall communication skills as assessed by experts but may have little to no effect on skills as assessed by simulated patients. In general, interventions may also improve students' empathy and information gathering scores as assessed by experts but may have little to no effect on relationship building or rapport as assessed by experts or simulated patients and on empathy as assessed by simulated patients; effects on explanation and planning skills are uncertain..

Evidence of very low certainty suggests that interventions using experiential approaches to communication skills training have uncertain effects on students' overall communication skills compared to didactic interventions when assessed by experts or simulated patients. Evidence of low certainty suggests that interventions using e-learning or video approaches may have little to no difference in impact on empathy or rapport scores compared to face-to-face interventions.

Evidence of moderate certainty indicates that interventions using tailored or intensive feedback probably improve overall communication skills scores more than those using general or no feedback when assessed by experts or simulated patients. Due to the very low quality of evidence, we are uncertain whether interventions involving role-play with simulated patients improve students' overall communication skills in comparison to peer role-play.

The most promising result appears to be that tailored or specific feedback approaches may improve students' communication skills more than generic or no feedback in practice opportunities. Common elements of effective feedback appear to be use of video-assisted feedback, which is supported by studies included in this meta-analysis (e.g. Maguire 1978; Ruesseler 2017), as well as studies outside of this meta-analysis (Roche 1996), along with review as in Hermann-Werner 2019 and specific behaviour-oriented feedback as described by Engerer 2019. This is an unsurprising finding given general support for the value of feedback in medical education, as well as studies indicating that students place great value on feedback received (Schopper 2016; Uhm 2015). Other studies indicating that feedback is rarely received and does not always align with the communication skills curriculum highlight that this is a worthy target for educational and faculty development interventions (Al-Mously 2014; Rosenbaum 2013).

A perhaps more surprising finding is that this review has not identified comparative improvement in simulated patient over peer role-play. Studies assessing this comparison provided only very low-quality evidence; thus considerable uncertainty remains about the comparative effects of these approaches on students' communication skills. Simulated patient role-play has long been regarded as the gold standard of communication skills training by medical educators, but the evidence presented here does not clearly support this experiential learning approach over other less costly options. Few studies have directly compared the effectiveness of these approaches, but available evidence suggests that both of these approaches may be effective for teaching certain communication skills (Bosse 2012; Taylor 2019), and that peer role-play offers the advantage of cost-effectiveness (Bosse 2015). Absence of comparative effectiveness between these and other approaches such as electronic versus face-to-face delivery may be reassuring for educators given the logistical and resource challenges often associated with facilitating in-person, small-group learning.

Although meta-analyses comparing didactic and experiential interventions failed to show a clear differential benefit, the narrative synthesis does suggest that there may be some promise in individual practice opportunities compared to group-level experiential learning in which only one or a few students actively participate in role-play (Berney 2017; Ho 2008; Levenkron 1990).

Primary outcomes assessed differed across studies, but common themes were apparent, with objective standardised clinical examination (OSCE) assessments generally aligning with the overall process of the medical consultation expressed in terms paralleling the Kalamazoo Consensus Statement and the Calgary-Cambridge guides (Kurtz 2003; Makoul 2001a). Across all outcomes, intervention effects were generally small, and the quality of evidence was low. Comparative effectiveness studies comparing effects of different interventions illustrate that in general, educational efforts impact student learning, with all interventions tending to have some positive effect on student communication outcomes, but in most cases, it is not possible to conclude that a particular style of intervention is superior to another.

The broader educational literature supports this notion, with a large body of medical and health professional education literature generally demonstrating positive effects of most teaching efforts. The comparative effectiveness of different approaches, the transferability of learning into real-world practice, and the

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

longevity of impacts are, however, less clear. A majority of included studies reported short-term effectiveness of their interventions but did not attempt to investigate the application of skills in real patient interactions or in practice beyond medical education. Literature from studies using similar methods in other fields of information gathering demonstrates that gains made through intervention with skills such as use of open questions are usually short-lived, with practice returning to baseline levels within 12 weeks of training (Lamb 2002; Smith 2009). There is also evidence that appropriately spaced, repeated practice can enhance retention of learning and ongoing development of skills (Benson 2015; Donovan 1999; Lamb 2002b). Further, when performance is evaluated in both simulated and actual interviews, it is clear that transfer of learning to the workplace is rarely made (Benson 2015). Only four studies in this review assessed outcomes beyond the year in which the intervention occurred, and only one followed learners once they were qualified doctors (Maguire 1978). Although this study did demonstrate some retention of learning in comparison to the control group, evidence is insufficient to permit judgement about retention of communication skills. Evaluation of long-term outcomes is made difficult by the inability to control or measure other learning and experience gained outside of the intervention in question.

Some differences in outcomes were identified based on trained examiner versus simulated patient assessment. It is expected that examiner ratings should be based on evidence of best practice communication and should therefore align with patient ratings. The differences identified, however, are consistent with previous literature, suggesting that patient perceptions do not entirely align with formal definitions of 'good' communication or patient-centred care (Mead 2002). Similarly, research suggests that perceptions of rapport are not entirely aligned between interviewer and interviewee (Vanderhaellen 2011). One reason why interviewers and interviewees might hold different views about the level of rapport achieved is that they may have different goals for the conversation (Abbe 2013). As such, instead of focusing on subjective perceptions, successful rapport is often measured by the extent to which conversation fosters a mutual working relationship, wherein the interviewee is doing most of the talking and is providing the interviewer with case-relevant information (Collins 2019).

Pooling of outcomes in this review assumes commonality in the definitions and approaches taken in each study. As is demonstrated in the examples of patient-centred care and rapport, communication outcomes can be somewhat subjective and are not necessarily consistent across studies. Although the review refers to the people trained to observe and rate students' communication skills as 'expert' assessors, there is potential for substantial variation even among these experts. Externally validated tools such as the Jefferson Scale of Patient Perceptions of Physician Empathy (LoSasso 2017), as well as the Arizona Clinical Interview Rating Scale (Vanatta 1996; Weihs 1986), were used in only a small number of studies, with most studies using adaptations or their own versions of rating scales based on models of the medical consultation. This represents a potentially substantial source of variation in interpretation of elements of communication and in approaches to assessment. The medical education field does not appear to have reached clear consensus on the definition or measurement of specific communication tasks such as open questions, listening, rapport, or empathy.

The studies included in this review represent 76 discrete trials. The body of literature reviewed here does not tell a clear story of the development of knowledge in this field, and little evidence shows studies building upon the findings of those published before them. There is an opportunity for this field to strengthen the evidence base by building upon the summary presented here, to fill gaps in knowledge and build upon the certainty of findings.

Overall completeness and applicability of evidence

This review used a comprehensive search strategy that followed Cochrane search methods and was not restricted by language or publication status. We included both randomised and quasi-randomised trials, given the challenging nature of randomisation in an educational context. This approach means that the quality of findings may be decreased due to poor randomisation (high risk of bias), but we are less likely to have missed important studies that were unable to randomise due to curriculum or programme structure and/or ethical limitations.

A total of 90 publications, with data from 10,124 participants, met the inclusion criteria for this review. A large proportion of these studies were undertaken in the USA (n = 37), with others reported from Australia (n = 8), the UK (n = 7), and a range of European, Asian, and Middle Eastern countries. This variation means that findings have good external validity and potential generalisability to international medical education, although representation and therefore generalisability of the results to low- and middle-income countries are limited. Generalisability to these lower-resource settings may be improved by further studies in such countries.

This review is limited to medical students, which enhances the heterogeneity of approach and outcome measures but limits the generalisability of findings across health professions. Although many communication skills are universally important for healthcare professionals, the consultation structure upon which the outcome measures assessed here were based is specific to medical consultations. Further, the structure of medical education differs from that of other professions, highlighting the value of reviewing studies pertaining to this context alone. Future research should investigate the applicability of this evidence in other contexts.
The interventions reported in this review adopted a range of strategies from one-on-one role-play and feedback to large-group learning and use of virtual patients. Interventions were targeted at students from the first to the final year of training. The variability in target group and approach is reassuring, indicating that the review has indeed captured the breadth of approaches used to teach communication skills in medical education. Such breadth is realistic in the context of interventions that need to be tailored for different cultural and curriculum contexts. Medical schools have unique needs and resourcing and curriculum design challenges, which prevent the application of a universal approach.

One important element that was absent from the included studies was the involvement of consumers (patients or community members) in studies or in study designs. No studies described consultation with consumers in design or development, nor in interventions or evaluation approaches, and very few studies involved actual patients in intervention delivery or assessment. 'Patient' involvement was largely limited to simulated patients. It could be argued that trained simulated patients are appropriate representatives of the patient population, but future research could

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

more comprehensively consider the perspectives of patients with actual conditions.

Quality of the evidence

Overall, the quality of evidence presented is low. Although only 21 studies were rated at high risk of bias overall for the purpose of sensitivity analysis (determined by three or more high risk ratings across categories), it is important to note that many studies had multiple 'unclear' ratings due to lack of information about their methods. An unclear risk of bias is suggestive of significant risk (Higgins 2011). Although sensitivity analyses did not lead to discernible differences in study results, it is important to note the common sources of bias across included studies, with most high risk ratings relating to high risk of bias for blinding of participants and personnel due to the nature of the interventions and the study design. In most cases, this is unlikely to impact the results. Most studies provided insufficient detail to enable assessment of allocation (selection) bias. A strength across these studies was low risk of detection, attrition, and reporting biases, with most assessors blinded to group allocation, and studies reporting low attrition rates. Study authors generally analysed available data (not accounting for those lost to follow-up), but with most studies assessing outcome measures at post-test only, this is likely to have little impact on findings. Although no studies had a published protocol, most reported all measures described in their methods.

Low quality ratings were generally the result of overall risk of bias, small sample sizes, high heterogeneity, and imprecision of results. It is expected that to improve the quality of evidence, future studies would need to adopt more consistent measures to reduce statistical heterogeneity. In the included studies, the scales used varied widely, with some studies failing to specify what the maximum possible score was, or how the reported numbers were derived. It is accepted that small sample sizes and variable interventions are inherent in medical education research, in which research is often presented as an extracurricular component and interventions must be context-specific.

It may be difficult to overcome the imprecision of results, with most studies across the review reporting relatively small differences between intervention groups and large measures of error. Even when differences reach statistical significance, the impact of differences on student outcomes and implications for future practice and patient safety and other outcomes are questionable. Small differences can differentiate between a 'pass' and a 'fail' grade for students, but usually this would occur only at low score levels. Across the studies included in this review, most students improved their communication skills or demonstrated a reasonable level of skill after any educational intervention. As such, these small differences at higher levels of performance are less likely to have major implications for student and patient outcomes.

Potential biases in the review process

We pooled the results of studies reporting on our primary outcomes, despite high levels of statistical heterogeneity. This may have introduced some bias because the actual effect size might be substantially larger or smaller than that of the pooled (mean) result. We conducted post-hoc formal subgroup analyses to investigate one potential source of this heterogeneity (assessment by experts versus simulated patients). We also conducted sensitivity analysis to account for risk of bias and to explore the potential impact of unit of analysis errors in cluster trials. Heterogeneity was largely unexplained and remained high in all cases. Further subgroup analyses could be conducted to explore differences in effects based on student year level or intervention features such as role-play, but the large numbers of outcomes and comparisons, the already low quality of evidence, and substantial heterogeneity in outcome measures led us to limit the subgroup analyses performed. Future updates of this review could include additional exploratory subgroup analyses.

Another possible source of bias is our selection of comparable measures of each primary outcome. Data used for the purposes of analysis were often selected from more than one communication outcome measurement per included trial. We cannot rule out the possibility that other researchers might choose alternative measurements from the same studies, and so might reach different conclusions about effects of the intervention on this outcome. This is relatively unlikely, as there was reasonable consistency in items rated across studies, with many based on common models such as the Calgary Cambridge Guides to the medical interview, or the SPIKES model for breaking bad news (e.g. Bosse 2012; Schmitz 2018).

Although we conducted sensitivity analyses to account for potential unit of analysis errors in studies defined as cluster trials or quasi-cluster-randomised controlled trials, there is a chance that other studies may be impacted by similar errors. Studies were classified as cluster trials when they randomised at group rather than individual student level; however, in at least 51 included studies, interventions were delivered to small groups of students, meaning that an effect could be associated with the group nature of delivery. Although intervention protocols were identical for each small group, natural differences can occur in the experience of each group due to group dynamics, the relationship of the tutor with the group, and variation in tutor style.
All other methods and literature searching for this update did not deviate from the processes outlined in the original review and should be relatively free from bias. We sought further information for seven of the included studies by contacting the lead or corresponding author; we received information from two of these study authors.

Agreements and disagreements with other studies or reviews

Our review specifically focused on teaching interpersonal communication to medical students completing an undergraduate or graduate-entry medical degree through an analysis of randomised controlled trials (RCTs), cluster-RCTs, and non-randomised controlled trials including quasi-RCTs. We specified that outcome measures from interventions were limited to observer ratings, as opposed to student knowledge or self-reported confidence or skills. As such, this review differs from previous reviews of communication skills training in medicine (Alelwani 2014; Artemiou 2014; Aspegren 1999; Batt-Rawden 2013; Berkhof 2011; Keifenheim 2015; Kelm 2014; Kyaw 2019; Lane 2007; Smith 2007). Two review proposals were also identified in our initial scoping review, neither of which have been subsequently published (MacDonald-Wicks 2012; Van Nuland 2005).

Although Aspegren 1999 focused on communication within the discipline of medicine only, this review captured pre-service

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

students as well as trainees and practising medical specialists and did not differentiate between groups. The 83 studies identified included randomised, quasi-randomised, and non-randomised trials and descriptive studies, including some of the studies included in our review, but these were not adequately assessed for quality. The conclusion of Aspegren 1999 that "there is overwhelming proof that communication skills in the patient-doctor relationship can be taught and learnt" (p 566) is at odds with the quality of evidence as assessed in our review. Similarly, Alelwani 2014 focused on skills in breaking bad news and reported unclear findings, with only one out of four randomised studies showing outcomes that favoured the intervention group. This is in keeping with the small number of studies included in the present review that focus on breaking bad news and the inconsistent findings. Also, these findings are consistent with the limited evaluation of breaking bad news education reported in a narrative review (Rosenbaum 2004).

Our findings regarding simulated patient versus peer role-play were consistent with the Lane 2007 review, which reported equivalent behavioural outcomes of simulated patient and peer role-play, although simulated patient experience was rated more positively by students. Both Batt-Rawden 2013 and Kelm 2014 reported promising results of empathy education but noted methodological flaws and inconsistent measures across studies, with questionable translation into behaviour change and limited evaluation of longer-term impact. These weaknesses echo those identified in the present review. Keifenheim 2015 reported heterogeneous measures and low quality of studies. This review reported promising results overall but was not able to differentiate between different teaching methods.

In keeping with Comparison 2 in the present review, Artemiou 2014 reported mixed impact of web-based learning across different communication skills measures but overall no clear advantage of this approach. Similarly, Kyaw 2019 reported that despite the low quality of evidence, digital learning was as effective as traditional learning in achieving improvements in medical students' communication skills.

The current review includes studies reviewed by Smith 2007, with the exception of those relying on student knowledge or self-reported confidence or skills as primary outcomes. Although our findings support, to some extent, those proffered by Smith 2007 indicating that communication skills can be taught effectively, the addition of contemporary studies beyond RCTs and more extensive assessment of the quality of evidence bring the strength of these conclusions into question.

In keeping with previous reviews, the individual included studies and the meta-analysis conducted in our review demonstrate the learning impact of most interventions. Smith 2007 reported that structured feedback and small-group discussions were associated with larger improvements in skills than were attained with more didactic or self-directed methods. Keifenheim 2015 also reported promising results of interventions involving video feedback. We found that tailored feedback on performance was more effective than general or no feedback, but due to the quality of evidence, we are uncertain whether experiential learning through role-play is more effective than more didactic learning approaches. As well as the low quality of evidence, actual effectiveness of the interventions is limited by the short-term nature of most outcome measurements. Further, although most studies reported good interrater reliability of their measures, and some referred to outcome validity, the absence of consistent measures places into question the interpretation of results. At face value, outcomes were sufficiently homogenous to warrant pooling; however, the subjective nature of measures such as empathy and rapport building means that variation across studies is inevitable. This was highlighted in a review of patient-centred consultations and outcomes in primary care (Mead 2002), which reported inconsistent associations between doctor behaviours defined as patient-centred and a variety of patient health outcomes. Effects on patient satisfaction were split between significant and non-significant results. This is likely to result from lack of a clear theoretical framework linking patient-centred care with specific outcomes. Similarly, although the Calgary-Cambridge guides and other models of clinical consultation provide a common framework, the operationalisation of assessing factors such as empathy and rapport, and definitions of open questions, remains open to interpretation in different studies. Differences in effects found in this review when outcomes were assessed by trained examiners versus simulated patients further reinforce this issue and highlight the need to more fully examine outcomes and expectations.

AUTHORS' CONCLUSIONS

Implications for practice

Educational efforts to improve students' interpersonal skills are undoubtedly worthwhile, but with the exception of providing tailored feedback, the differential effectiveness of particular approaches remains unclear. Further, long-term effects on students' behaviour and on their behaviour as practising clinicians and effects on patient outcomes remain unclear. These measures are perhaps even more important than short-term learning outcomes, with the need for medical education to facilitate the development of foundational skills that can be further enhanced through practice, and to foster a culture of lifelong learning. Although some of these measures were beyond the scope of the current review, the value of this evidence for medical education and practice is limited without such investigation. Differences in findings between examiner-assessed and simulated patient-assessed outcomes warrant further investigation. Expert assessors should be assessing based on evidence of best practice, which should in turn be based on evidence about patient outcomes associated with specific communication skills. Lack of alignment identified here raises questions for medical educators about development of assessments and training of assessors.

Implications for research

It is accepted that in general, educational interventions will lead to student learning, at least in the short term. Effects on longer-term behaviour change and application in clinical practice, however, are less clear and are more challenging to measure. A beneficial contribution to knowledge could be made if future research were designed to rigorously evaluate the medium- and long-term effects of particular types of educational approaches on student behaviour and lifelong learning as they progress into increasingly autonomous clinical practice. Further, efforts to standardise assessment and evaluation of interpersonal skills would strengthen research efforts and allow greater value to be derived from the synthesis of results across separate studies.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ACKNOWLEDGEMENTS

We thank the editors and staff of Cochrane Consumers and Communication for their input into this review. We are grateful for the input of students and simulated patients making up our advisory group. Finally, thanks to Hamida Zekiroski, from the Centre for Investigative Interviewing, for assistance with screening and data extraction.

We would also like to thank contributors from Cochrane Crowd who assisted in a first assessments of titles and abstracts retrieved by the search. The following people screened 200 or more records: Bernardo Costa, Hariklia Nguyen, Ivan Murrieta-Alvarez, Kerstin Rolfe, Eirik B. Eikenaes, Esther Pozzani, Elizabeth Stovold, Eva Wenig, Karen Ma, Mansoureh Karimollahi, Maura Scott, Mersiha Mahmic-Kaknjo, Riccardo Guarise, Silvana Urru, Tomislav Mestrovic, Lauren Westafer.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

REFERENCES

References to studies included in this review

Allen 1990 *{published data only}*

Allen SS, Bland CJ, Dawson SJ. A mini-workshop to train medical students to use a patient-centered approach to smoking cessation. *American Journal of Preventive Medicine* 1990;**6**(1):28-33.

Alroy 1984 *{published data only}*

Alroy G, Ber R, Kramer D. An evaluation of the short-term effects of an interpersonal skills course. *Medical Education* 1984;**18**(2):85-9.

Bearman 2001 *{published data only}*

Bearman M, Cesnik B, Liddell M. Random comparison of 'virtual patient' models in the context of teaching clinical communication skills. *Medical Education* 2001;**35**(9):824-32.

Berney 2017 *{published data only}*

Berney A, Carrard V, Schmid Mast M, Bonvin R, Stiefel F, Bourquin C. Individual training at the undergraduate level to promote competence in breaking bad news in oncology. *Psycho-oncology* 2017;**26**(12):2232-7.

Betchart 1984 *{published data only}*

Betchart NS, Anderson DG, Thompson TL, Mumford E. A tutorial approach to improving medical students' interviewing skills. *Journal of Medical Education* 1984;**59**(5):431-3.

Blatt 2010 *{published data only}*

Blatt B, LeLacheur SF, Galinsky AD, Simmens JS, Greenberg L. Does perspective-taking increase patient satisfaction in medical encounters? *Academic Medicine* 2010;**89**(5):1445-52.

Bosse 2012 *{published data only}*

Bosse HM, Schultz JH, Nickel M, Lutz T, Moltner A, Junger J, et al. The effect of using standardized patients or peer role play on ratings of undergraduate communication training: a randomized controlled trial. *Patient Education and Counseling* 2012;**87**(3):300-6.

Bowyer 2006 *{published data only}*

Bowyer MW, Rawn L, Hanson J, Pimentel, EA, Flanagan A, Ritter EM, et al. Combining high-fidelity human patient simulators with a standardized family member: a novel approach to teaching breaking bad news. In: Westwood, JD, et al, editors(s). Studies in Health Technology and Informatics; Medicine Meets Virtual Reality 14 - Accelerating Change in Healthcare. Long Beach, CA: IOS Press, 2006:67-72.

Bowyer 2010 *{published data only}*

Bowyer MW, Hanson JL, Pimentel EA, Flanagan AK, Rawn LM, Rizzo AG, et al. Teaching breaking bad news using mixed reality simulation. *Journal of Surgical Research* 2010;**159**(1):462-7.

Buffel du Vaure 2017 *{published data only}*

Buffel du Vaure C, Lemogne C, Bunge L, Catu-Pinault A, Hoertel N, Ghasarossian C, et al. Promoting empathy among medical students: a two-site randomized controlled study. *Journal of Psychosomatic Research* 2017;**103**:102-7.

Cave 2007 *{published data only}*

Cave J, Washer P, Sampson P, Griffin M, Noble L. Explicitly linking teaching and assessment of communication skills. *Medical Teacher* 2007;**29**(4):317-22.

Chibhall 2005 *{published data only}*

Chibnall JT, Cook MA, Miller DK. Religious awareness training for medical students: effect on clinical interpersonal behavior. *Southern Medical Journal* 2005;**98**(12):1255.

Colletti 2001 *{published data only}*

Colletti L, Gruppen L, Barclay M, Stern D. Teaching students to break bad news. *American Journal of Surgery* 2001;**182**(1):20-3.

Daeppen 2012 *{published data only}*

Daeppen JB, Fortini C, Bertholet N, Bonvin R, Berney A, Michaud PA, et al. Training medical students to conduct motivational interviewing: a randomized controlled trial. *Journal of General Internal Medicine* 2011;**26**:S49.

Daeppen J-B, Fortini C, Bertholet N, Bonvin R, Berney A, Michaud P-A, et al. Training medical students to conduct motivational interviewing: a randomized controlled trial. *Patient Education and Counseling* 2012;**87**(3):313-8.

Edwardsen 2006 *{published data only}*

Evans BJ, Stanley RO, Burrows Graham D. Measuring medical students' empathy skills. *British Journal of Medical Psychology* 1993;**66**(2):121-33.

Eells 2002 *{published data only}*

Eells T, Strauss Gordon D, Teller D, Feldman T. Problem-oriented instruction as a predictor of success in early psychiatric interviewing. *Academic Psychiatry* 2002;**26**(2):82-9.

Engerer 2019 *{published and unpublished data}*

* Engerer C, Berberat PO, Dinkel A, Rudolph B, Sattel H, Wuensch A. Specific feedback makes medical students better communicators. *BMC Medical Education* 2019;**19**(1):51.

Evans 1989 *{published data only}*

Evans BJ, Stanley RO, Burrows GD, Sweet B. Lectures and skills workshops as teaching formats in a history-taking skills course for medical students. *Medical Education* 1989;**23**(4):364-70.

Evans BJ, Stanley RO, Burrows GD. Measuring medical students' empathy skills. *British Journal of Medical Psychology* 1993;**66**(2):121-33.

Evans BJ, Stanley RO, Mestrovic R, Rose L. Effects of communication skills training on students' diagnostic efficiency. *Medical Education* 1991;**25**(6):517-26.

Evans 1996 *{published data only}*

Evans BJ, Coman GJ, Goss B. Consulting skills training and medical students' interviewing efficiency. *Medical Education* 1996;**30**(2):121-8.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Evans BJ, Stanley RO, Coman G, Sinnott V. Measuring medical students' communication skills: development and evaluation of an interview rating scale. *Psychology and Health* 1992;**6**(3):213-25.

Evans BJ, Sweet B. Consulting-skills training to improve medical students' diagnostic efficiency. *Academic Medicine* 1993;**68**(2):170-1.

Feddock 2009 *{published data only}*

Feddock C, Hoellein A, Griffith C, Wilson J, Lineberry M, Haist S. Enhancing knowledge and clinical skills through an adolescent medicine workshop. *Archives of Paediatric and Adolescent Medicine* 2009;**163**(3):256-60.

Filipetto 2006 *{published data only}*

Filipetto FA, Weiss LB, Switala CA, Bertagnolli JF. The effectiveness of a first-year clinical preceptorship on the data collection and communication skills of second-year medical students. *Teaching and Learning in Medicine* 2006;**18**(2):137-41.

Foster 2016 *{published data only}*

Foster A, Chaudhary N, Kim T, Waller Jennifer L, Wong J, Borish M, et al. Using virtual patients to teach empathy: a randomized controlled study to enhance medical students' empathic communication. *Simulation in Healthcare* 2016;**11**(3):181-9.

Gartmeir 2015 *{published and unpublished data}*

Gartmeier M, Bauer J, Fischer Martin R, Hoppe-Seyler T, Karsten G, Kiessling C, et al. Fostering professional communication skills of future physicians and teachers: effects of e-learning with video cases and role-play. *Instructional Science* 2015;**43**(4):443-62.

Gerber 1985 *{published data only}*

Gerber W, Matthes S, Albanese M. A comparison of instructional techniques: internal -external vs. external only. In: Eighteenth Annual Conference on Research in Medical Education. 1979:59-64.

Gerber WL, Albanese M, Brown D, Matthes S. Teaching with simulated patients. Evaluation of the long-term effectiveness of instruction. *Evaluation and the Health Professions* 1985;**8**(1):69-92.

Ho 2008 *{published data only}*

Ho M-J, Yao G, Lee K-L, Beach MC, Green AR. Cross-cultural medical education: can patient-centered cultural competency training be effective in non-western countries? *Medical Teacher* 2008;**30**(7):719-21.

Ho M-J, Yao G, Lee K-L, Hwang T-J, Beach MC. Long-term effectiveness of patient-centered training in cultural competence: what is retained? what is lost? *Academic Medicine* 2010;**85**(4):660-4.

Hobgood 2009 *{published data only}*

Hobgood CD, Tamayo-Sarver JH, Hollar DW Jr, Sawning S. GrievIng: death notification skills and applications for fourth-year medical student. *Teaching and Learning in Medicine* 2009;**21**(3):207-19.

Ishikawa 2010 *{published and unpublished data}*

Ishikawa H, Hashimoto H, Kinoshita M, Yano E. Can nonverbal communication skills be taught? *Medical Teacher* 2010;**32**(10):860-3.

Kahan 2003 *{published data only}*

Kahan M, Wilson L, Midmer D, Borsoi D, Martin D. Randomized controlled trial on the effects of a skills-based workshop on medical students' management of problem drinking and alcohol dependence. *Substance Abuse* 2003;**24**(1):15-6.

Kaltman 2018 *{published data only}*

Kaltman S, Talisman N, Pennestri N, Syverson E, Arthur P, Vovides Y. Using technology to enhance teaching of patient-centered Interviewing for early medical students. *Simulation in Healthcare* 2018;**13**(3):188-94.

Klein 2000 *{published data only}*

Klein S, Tracy D, Kitchener H, Walker L. The effects of the participation of patients with cancer in teaching communication skills to medical undergraduates: a randomised study with follow-up after 2 years. *European Journal of Cancer* 2000;**36**(2):273-81.

Klein S. Does cancer patient participation in communication in skills training enhance the performance of medical undergraduates? *Psycho-oncology* 1995;**4**(2):90.

Knowles 2001 *{published data only}*

Knowles C, Kinchington F, Erwin J, Peters B. A randomised controlled trial of the effectiveness of combining video role play with traditional methods of delivering undergraduate medical education. *Sexually Transmitted Infections* 2001;**77**(5):376-80.

Lee 2015 *{published data only}*

Lee AL, Mader EM, Morley CP. Teaching cross-cultural communication skills online: a multi-method evaluation. *Family Medicine* 2015;**47**(4):302-8.

Legg 2005 *{published data only}*

Legg C, Young L, Bryer A. Training sixth-year medical students in obtaining case history information from adults with aphasia. *Aphasiology* 2005;**19**(6):559-75.

Levenkron 1990 *{published data only}*

Levenkron JC, Greenland P, Bowley N. Teaching risk-factor counseling skills: a comparison of two instructional methods. *American Journal of Preventive Medicine* 1990;**6**(2 Suppl):29-34.

Lim 2011 *{published data only}*

Lim Bee T, Moriarty H, Huthwaite M. "Being-in-role": a teaching innovation to enhance empathic communication skills in medical students. *Medical Teacher* 2011;**33**(12):e663-9.

Liu 2016 *{published data only}*

Liu C, Lim Renee L, McCabe Kathryn L, Taylor S, Calvo Rafael A. A web-based telehealth training platform incorporating automated nonverbal behavior feedback for teaching communication skills to medical students: a randomized crossover study. *Journal of Medical Internet Research* 2016;**18**(9):e246.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lorin 2006 *{published data only}*

Lorin S, Rho L, Wisnivesky JP, Nierman DM. Improving medical student intensive care unit communication skills: a novel educational initiative using standardized family members. *Critical Care Medicine* 2006;**34**(9):2386-91.

LoSasso 2017 *{published data only}*

LoSasso AA, Lamberton CE, Sammon M, Berg KT, Caruso JW, Cass J, et al. Enhancing student empathetic engagement, history-taking, and communication skills during electronic medical record use in patient care. *Academic Medicine* 2017;**92**(7):1022-7.

Lupi 2012 *{published and unpublished data}*

Lupi C, Estes C, Broome M, Schreiber M. Conscientious refusal in reproductive medicine: an educational intervention. *American Journal of Obstetrics and Gynaecology* 2009;**201**(502):e1-7.

Lupi Carla S, Runyan A, Schreiber N, Steinauer J, Turk Jema K. An educational workshop and student competency in pregnancy options counseling: a randomized controlled trial. *American Journal of Obstetrics and Gynecology* 2012;**207**(5):414.e1-7.

Maguire 1977 *{published data only}*

Maguire GP, Clarke D, Jolley B. An experimental comparison of three courses in history-taking skills for medical students. *Medical Education* 1977;**11**(3):175-82.

Maguire 1978 *{published data only}*

Maguire P, Fairbairn S, Fletcher C. Consultation skills of young doctors: I - Benefits of feedback training in interviewing as students persist. *British Medical Journal (Clinical research edition)* 1986;**292**(6535):1573-6.

Maguire P, Roe P, Goldberg D, Jones S, Hyde C, O'Dowd T. The value of feedback in teaching interviewing skills to medical students. *Psychological Medicine* 1978;**8**(4):695-704.

Mason 1988 *{published data only}*

Mason JL, Barkley SE, Kappelman MM, Carter DE, Beachy WV. Evaluation of a self-instructional method for improving doctor-patient communication. *Journal of Medical Education* 1988;**63**(8):629-35.

Meirovich 2016 *{published data only}*

Meirovich A, Ber R, Moore M, Rotschild A. Student-centered tutoring as a model for patient-centeredness and empathy. *Advances in Medical Education and Practice* 2016;**7**:423-8.

Moreland 1973 *{published data only}*

Moreland JR, Ivey AE, Phillips JS. An evaluation of microcounseling as an interviewer training tool. *Journal of Consulting and Clinical Psychology* 1973;**41**(2):294-300.

Moreland JR. Video Programmed Instruction in Elementary Psychotherapeutic and Related Clinical Skills. Final Report (dissertation submitted as partial fulfullment of the requirements for Doctor of Philosphy; Clinical Psychology). Denver University Colorado, Department. of Psychology; Massachusetts University Amherst. School of Education; Oklahoma University, Oklahoma City Medical Center 1971.

Morrow 2009 *{published data only}*

Morrow JB, Dobbie AE, Jenkins C, Long R, Mihalic A, Wagner J. First-year medical students can demonstrate EHR-specific communication skills: a control-group study. *Family Medicine* 2009;**41**(1):28-33.

Mounsey 2006 *{published data only}*

Mounsey AL, Bovbjerg V, White L, Gazewood J. Do students develop better motivational interviewing skills through role-play with standardised patients or with student colleagues? *Medical Education* 2006;**40**(8):775-80.

Nomura 2017 *{published data only}*

Nomura O, Onishi H, Kato H. Medical students can teach communication skills - a mixed methods study of cross-year peer tutoring. *BMC Medical Education* 2017;**17**(1):103.

Ockene 2016 *{published data only}*

Ockene JK, Hayes RB, Churchill LC, Crawford SL, Jolicoeur DG, Murray DM, et al. Teaching medical students to help patients quit smoking: outcomes of a 10-school randomized controlled trial. *Journal of General Internal Medicine* 2016;**31**(2):172-81.

Ozcakar 2009 *{published data only}*

Ozcakar N, Mevsim V, Guldal D, Gunvar T, Yildirim E, Sisli Z, et al. Is the use of videotape recording superior to verbal feedback alone in the teaching of clinical skills? *BMC Public Health* 2009;**9**:474.

Palmer 2018 *{published data only}***http://dx.doi.org/10.1007/s40596-018-0949-8**

Palmer BA, Lee JH, Somers KJ, Swintak CC, Rullo J, Bright RP, et al. Effect of in-class vs online education on sexual health communication skills in first-year medical students: a pilot study. *Academic Psychiatry* 2018;**43**:175-9.

Papadakis 1997 *{published data only}*

Papadakis MA, Croughan-Minihane M, Fromm LJ, Wilkie HA, Ernster VL. A comparison of two methods to teach smoking-cessation techniques to medical students. *Academic Medicine* 1997;**72**(8):725-7.

Perera 2010 *{published data only}*

Perera J, Mohamadou G, Kaur S. The use of objective structured self-assessment and peer-feedback (OSSP) for learning communication skills: evaluation using a controlled trial. *Advances in Health Sciences Education* 2010;**15**(2):185-93.

Pirdehghan 2018 *{published data only}*

Pirdehghan A, Vakili M, Mirzaei M, Karimi M. The effect of training in communication skills on medical students: a randomized controlled trial. *International Journal of Pharmaceutical Research* 2018;**10**(3):339-43.

Poole 1980 *{published data only}*

Poole AD, Sanson-Fisher RW. Long-term effects of empathy training on the interview skills of medical students. *Patient Counselling and Health Education* 1980;**2**(3):125-7.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Poole AD, Sanson-Fisher RW. Understanding the patient: a neglected aspect of medical eduction. *Social Science and Medicine* 1979;**13A**:37-43.

Roche 1996 *{published data only}*

Roche AM, Eccleston P, Sanson-Fisher R. Teaching smoking cessation skills to senior medical students: a block-randomized controlled trial of four different approaches. *Preventive Medicine* 1996;**23**(5):251-8.

Roche 1997 *{published data only}*

Roche AM, Stubbs JM, Sanson-Fisher RW, Saunders JB. A controlled trial of educational strategies to teach medical students brief intervention skills for alcohol problems. *Preventive Medicine* 1997;**26**(1):78-85.

Ruesseler 2017 *{published data only}*

Ruesseler M, Sterz J, Bender B, Hoefer S, Walcher F. The effect of video-assisted oral feedback versus oral feedback on surgical communicative competences in undergraduate training. *European Journal of Trauma and Emergency Surgery* 2017;**43**(4):461-6.

Schmitz 2018 *{published and unpublished data}*

Schmitz FM, Schnabel KP, Bauer D, Bachmann C, Woermann U, Guttormsen S. The learning effects of different presentations of worked examples on medical students' breaking-bad-news skills: a randomized and blinded field trial. *Patient Education and Counseling* 2018;**101**(8):1439-51.

Schmitz FM, Schnabel KP, Bauer D, Woermann U, Guttormsen S. Learning how to break bad news from worked examples: does the presentation format matter when hints are embedded? Results from randomised and blinded field trials. *Patient Education and Counseling* 2020;**103**(9):1850-85. [DOI: https://dx.doi.org/10.1016/j.pec.2020.03.022]

Schwartz 2010 *{published data only}*

Schwartz A, Weiner SJ, Harris IB, Binns-Calvey A. An educational intervention for contextualizing patient care and medical students' abilities to probe for contextual issues in simulated patients. *JAMA* 2010;**304**(11):1191-7.

Seim 1995 *{published data only}*

Seim HC, Verhoye JR. Comparison of training techniques using a patient-centered approach to smoking cessation. *Medical Education* 1995;**29**(2):139-43.

Shaddheau 2015 *{published data only}*

Shaddeau A, Nimz A, Sheeder J, Tocce K. Effect of novel patient interaction on students' performance of pregnancy options counseling. *Medical Education Online* 2015;**20**(Article 29401):1-6.

Shapiro 2009 *{published data only}*

Shapiro SM, Lancee WJ, Richards-Bentley CME. Evaluation of a communication skills program for first-year medical students at the University of Toronto. *BMC Medical Education* 2009;**9**(11):1-8.

Siassakos 2010 *{published data only}*

Siassakos D, Draycott T, O'Brien K, Kenyon C, Bartlett C, Fox R. Exploratory randomized controlled trial of hybrid obstetric simulation training for undergraduate students. *Simulation in Healthcare* 2010;**5**(4):193-8.

Snow 2016 *{published data only}*

Snow R, Crocker J, Talbot K, Moore J, Salisbury H. Does hearing the patient perspective improve consultation skills in examinations? An exploratory randomized controlled trial in medical undergraduate education. *Medical Teacher* 2016;**38**(12):1229-35.

Solomon 2004 *{published data only (unpublished sought but not used)}*

Keefe CW, Thompson ME, Noel, MM. Medical students, clinical preventive services, and shared decision making. *Academic Medicine; Assessing Clinical Skills supplement* 2002;**77**(11):1160-1.

Solomon DJ, Laird-Fick HS, Keefe CW, Thompson ME, Noel MM. Using a formative simulated patient exercise for curriculum evaluation. *BMC Medical Education* 2004;**4**(8):1-6.

Spollen 2010 *{published data only}*

Spollen J, Thrush C, Mui D, Woods M, Tariq S, Hicks E. A randomized controlled trial of behavior change counseling education for medical students. *Medical Teacher* 2010;**32**(4):e170-7.

Steifel 2013 *{published data only}*

Stiefel F, Bourquin C, Layat C, Vadot S, Bonvin R, Berney A. Medical students' skills and needs for training in breaking bad news. *Journal of Cancer Education* 2013;**28**(1):187-91.

Stolz 2012 *{published data only}*

Stolz D, Langewitz W, Meyer A, Pierer K, Tschudi P, S'Ng Ching T, et al. Enhanced didactic methods of smoking cessation training for medical students - a randomized study. *Nicotine & Tobacco Research* 2012;**14**(2):224-8.

Strayer 2010 *{published data only}*

Strayer S, Pelletier S, Martindale JR, Rais S, Powell J, Schorling J. A PDA-based counseling tool for improving medical student smoking cessation counseling. *Family Medicine* 2010;**42**(5):350-7.

Taylor 2019 *{published data only}*

Taylor S, Haywood M, Shulruf B. Comparison of effect between simulated patient clinical skill training and student role play on objective structured clinical examination performance outcomes for medical students in Australia. *Journal of Educational Evaluation for Health Professions* 2019;**16**:3.

Vanatta 1996 *{published data only}*

Vannatta JB, Smith KR, Crandall S, Fischer PC, Williams K. Comparison of standardized patients and faculty in teaching medical interviewing. *Academic Medicine* 1996;**71**(12):1360-2.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wagner 2011 *{published data only}*

Wagner JA, PA, Harrington KL. Evaluation of online instruction to improve medical and dental students' communication and counseling skills. *Evaluation & the Health Professions* 2011;**34**(3):383-97.

Walsh 1999 *{published data only}*

Walsh RA, Sanson-Fisher RW, Low A, Roche AM. Teaching medical students alcohol intervention skills: results of a controlled trial. *Medical Education* 1999;**33**(8):559-65.

Weihs 1986 *{published data only}*

Weihs K, Chapados JT. Interviewing skills training - a study. *Social Science and Medicine* 1986;**23**(1):31-4.

Windish 2005 *{published data only}*

Price EG, Windish DM, Magaziner J, Cooper LA. Assessing validity of standardized patient ratings of medical students' communication behavior using the Roter Interaction Analysis System. *Patient Education and Counseling* 2008;**70**(1):3-9.

Windish DM, Price EG, Clever SL, Magaziner JL, Thomas PA. Teaching medical students the important connection between communication and clinical reasoning. *Journal of General Internal Medicine* 2005;**20**(12):1108-13.

Wundrich 2017 *{published data only}*

Wündrich M, Schwartz C, Feige B, Lemper D, Nissen C, Voderholzer U. Empathy training in medical students - a randomized controlled trial. *Medical Teacher* 2017;**39**(10):1096-8.

References to studies excluded from this review

Abraham 2001 *{published data only}*

Abraham A, Cheng TL, Wright JL, Addlestone I, Huang Z, Greenberg L. Assessing an educational intervention to improve physician violence screening skills. *Pediatrics* 2001;**107**(5):E68.

Abraham 2011 *{published data only}*

Abraham A, Hawkins K, Chen R. An educational intervention to improve physician interviewing skills of adolescent patients during the pediatric clerkship. *Journal of Adolescent Health* 2011;**48**(2 Suppl 1):S34-5.

Ahmadzadeh 2019 *{published data only}*

Ahmadzadeh A, Esfahani MN, Ahmadzad-Asl M, Shalbafan M, Shariat SV. Does watching a movie improve empathy? A cluster randomized controlled trial. *Canadian Medical Education Journal* 2019;**10**(4):e4-12.

Aper 2012 *{published data only}*

Aper L, Reniers J, Koole S, Valcke M, Derese A. Impact of three alternative consultation training formats on self-efficacy and consultation skills of medical students. *Medical Teacher* 2012;**34**(7):e500-7.

Betson 1997 *{published data only}*

Betson CL, Fielding R, Wong G, Chung SF, Nestel DF. Evaluation of two videotape instruction programmes on how to break bad news - for Cantonese-speaking medical students in Hong Kong. *Journal of Audiovisual Media in Medicine* 1997;**20**(4):172-7.

Bientzle 2015 *{published data only}*

Bientzle M, Griewatz J, Kimmerle J, Kuppers J, Cress U, Lammerding-Koeppel M. Impact of scientific versus emotional wording of patient questions on doctor-patient communication in an internet forum: a randomized controlled experiment with medical students. *Journal of Medical Internet Research* 2015;**17**(11):268.

Bishop 2016 *{published data only}*

Bishop TW, Gorniewicz J, Floyd M, Tudiver F, Odom A, Zoppi K. Innovative patient-centered skills training addressing challenging issues in cancer communications: using patients' stories that teach. *International Journal of Psychiatry in Medicine* 2016;**51**(4):357-66.

Bittner 2016 *{published data only}*

Bittner A, Bittner J, Jonietz A, Dybowski C, Harendza S. Translating medical documents improves students' communication skills in simulated physician-patient encounters. *BMC Medical Education* 2016;**16**:72.

Blake 2000 *{published data only}*

Blake K, Mann KV, Kaufman DM, Kappelman M. Learning adolescent psychosocial interviewing using simulated patients. *Academic Medicine* 2000;**75**(10 Suppl):S56-8.

Blanch-Hartigan 2012 *{published data only}*

Blanch-Hartigan D. An effective training to increase accurate recognition of patient emotion cues. *Patient Education and Counselling* 2012;**89**(2):274-80.

Bragard 2018 *{published data only}*

Bragard I, Servotte JC, Van Cauwenberghe I, Donneau A, Dardenne N, Tubes R, et al. Breaking bad news in the emergency department: a randomized controlled study of a training using role-play simulation. *Critical Care; Conference: 38th International Symposium on Intensive Care and Emergency Medicine, ISICEM* 2018;**22**(Suppl 1):212-5.

Brown 1980 *{published data only}*

Brown JE, O'Shea JS. Improving medical student interviewing skills. *Pediatrics* 1980;**65**(3):575-8.

Claramita 2006 *{published data only}*

Claramita M, Majoor G. Comparison of communication skills in medical residents with and without undergraduate communication skills training as provided by the Faculty of Medicine of Gadjah Mada University. *Education for Health* 2006;**19**(3):308-20.

Clever 2011 *{published data only}*

Clever SL, Dudas RA, Solomon BS, Yeh HC, Levine D, Bertram A, et al. Medical student and faculty perceptions of volunteer outpatients versus simulated patients in communication skills training. *Academic Medicine* 2011;**86**(11):1437-42.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Craig 1992 *{published data only}*

Craig JL. Retention of interviewing skills learned by first-year medical students: a longitudinal study. *Medical Education* 1992;**26**(4):276-81.

D'Souza 2020 *{published data only}*

D'Souza PC, Rasquinha SL, D'Souza TL, Jain A, Kulkarni V, Pai K. Effect of a single-session communication skills training on empathy in medical students. *Academic Psychiatry* 2020;**44**(3):289-94.

Daetwyler 2010 *{published data only}*

Daetwyler CJ, Cohen DG, Gracely E, Novack DH. eLearning to enhance physician patient communication: a pilot test of "doc.com" and "WebEncounter" in teaching bad news delivery. *Medical Teacher* 2010;**32**(9):e381-90.

Daryazadeh 2020 *{published data only}*

Daryazadeh S, Adibi P, Yamani N, Mollabashi R. Impact of narrative medicine program on improving reflective capacity and empathy of medical students in Iran. *Journal of Educational Evaluation for Health Professions* 2020;**17**(3):1-7. [DOI: https://dx.doi.org/10.3352/jeehp.2020.17.3]

Davis 1992 *{published data only}*

Davis H, Nicholaou T. A comparison of the interviewing skills of first- and final-year medical students. *Medical Education* 1992;**26**(6):441-7.

Dickson 2012 *{published data only}*

Dickson RP, Engelberg RA, Back AL, Ford DW, Curtis JR. Internal medicine trainee self-assessments of end-of-life communication skills do not predict assessments of patients, families, or clinician-evaluators. *Journal of Palliative Medicine* 2012;**15**(4):418-26.

Duke 2015 *{published data only}*

Duke P, Grosseman S, Novack DH, Rosenzweig S. Preserving third year medical students' empathy and enhancing self-reflection using small group "virtual hangout" technology. *Medical Teacher* 2015;**37**(6):566-71.

Ellis 2002 *{published data only}*

Ellis MV, Krengel M, Beck M. Testing self-focused attention theory in clinical supervision: effects on supervisee anxiety and performance. *Medical Teacher* 2002;**49**(1):101-16.

Fernndez-Olano 2008 *{published data only}*

Fernndez-Olano C, Montoya-Fernndez J, Salinas-Sanchez AS. Impact of clinical interview training on the empathy level of medical students and medical residents. *Medical Teacher* 2008;**30**(3):322-4.

Fine 1977 *{published data only}*

Fine VK, Therrien ME. Empathy in the doctor-patient relationship: skill training for medical students. *Journal of Medical Education* 1977;**52**(9):752-7.

Fukuta 2018 *{published data only (unpublished sought but not used)}*

Fukuta J, Morgan J. First-person perspective video to enhance simulation. *Simulation* 2018;**15**:231-5.

Ghofranipour 2018 *{published data only}*

Ghofranipour F, Ghaffarifar S, Ahmadi F, Hosseinzadeh H, Akbarzadeh A. Improving interns' patient-physician communication skills: application of self-efficacy theory, a pilot study. *Cogent Psychology* 2018;**5**(1):1-13.

Gorniewicz 2017 *{published data only}*

Gorniewicz J, Floyd M, Krishnan K, Bishop Thomas W, Tudiver F, Lang F. Breaking bad news to patients with cancer: a randomized control trial of a brief communication skills training module incorporating the stories and preferences of actual patients. *Patient Education and Counseling* 2017;**100**(4):665-6.

Haeseler 2011 *{published data only}*

Haeseler F, Fortin AH, Pfeiffer C, Walters C, Martino S. Assessment of a motivational interviewing curriculum for year 3 medical students using a standardized patient case. *Patient Education and Counseling* 2011;**84**(1):27-30.

Haq 2006 *{published data only}*

Haq I, Fuller J, Dacre J. The use of patient partners with back pain to teach undergraduate medical students. *Rheumatology* 2006;**45**(4):430-4.

Hoffman 2014 *{published data only (unpublished sought but not used)}*

Hoffmann TC, Bennett ST, Tomsett C, DelMar C. Brief training of student clinicians in shared decision making: a single-blind randomized controlled trial. *Journal of General Internal Medicine* 2014;**29**(6):844-9.

Junger 2005 *{published data only}*

Junger J, Schafer S, Roth C, Schellberg D, Friedman M, Nikendei C. Effects of basic clinical skills training on objective structured clinical examination performance. *Medical Education* 2005;**39**(10):1015-20.

Kalet 2005 *{published data only}*

Kalet AL, Janicik R, Schwartz M, Roses, D, Hopkins MA, Riles T. Teaching communication skills on the surgery clerkship. *Medical Education Online* 2005;**10**(16):1-8.

Karanth 2008 *{published data only}*

Karanth K, Veena L, Kumar MV. Implementation and evaluation by formal assessments and term end student feedback of a new methodology of clinical teaching in surgery in small group sessions. *Annals of the Academy of Medicine* 2008;**37**(12):1008-11.

Koponen 2012 *{published data only}*

Koponen J, Pyorala E, Isotalus P. Comparing three experiential learning methods and their effect on medical students' attitudes to learning communication skills. *Medical Teacher* 2012;**34**(3):e198-207.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Koponen 2014 *{published data only}*

Koponen J, Pyorala E, Isotalus P. Communication skills for medical students: results from three experiential methods. *Simulation and Gaming* 2014;**45**(2):235-54.

Kron 2017 *{published data only}*

Kron FW, Fetters MD, Scerbo MW, White CB, Lypson ML, Padilla MA, et al. Using a computer simulation for teaching communication skills: a blinded multisite mixed methods randomized controlled trial. *Paitent Education and Counseling* 2017;**100**(4):748-59.

Lau 2001 *{published data only}*

Lau KCJ, Stewart SM, Fielding R. Preliminary evaluation of "interpreter" role plays in teaching communication skills to medical undergraduates. *Medical Education* 2001;**35**(3):217-21.

Leber 2012 *{published data only}*

Leber M, He C, Akhtar S, Asher S, Bania T, Di C, et al. A comparison of individualized feedback versus standard didactic lecture to teach interpersonal communication skills to emergency medicine residents: a multicenter randomized controlled trial. *Annals of Emergency Medicine* 2012;**60**(4 Suppl 1):S41.

LeBlanc 2009 *{published data only}*

LeBlanc V, Tabak D, Kneebone R, Nestel D, MacRae H, Moulton C. Psychometric properties of an integrated assessment of technical and communication skills. *American Journal of Surgery* 2009;**197**:96-101.

Lee 2004 *{published data only}*

Lee JH, Shin JS Suh DH, Eun HC. Evaluation of the outcome of communication skill training using a role play model in the dermatological clerkship. *Korean Journal of Dermatology* 2004;**42**(11):1440-8.

Leung 2015 *{published data only}*

Leung CG, Russell D, Way DP, Thompson L, Greenberger S, Kman NE. Observation without active participation is an effective method of learning in high fidelity simulation. *Academic Emergency Medicine* 2015;**22**(5 Suppl 1):S193-4.

Malik 2013 *{published data only}*

Malik S, Zaheer R, Bilal M. Impact of movie-based simulation training, with or without conventional verbal demonstration on observed OSPE scores in medical undergraduates: a double control study. *Journal of Ayub Medical College, Abbottabad* 2013;**25**:127-8.

Malik 2014 *{published data only}*

Malik S, Zaheer R, Bilal M. Impact of movie-based simulation training, with or without conventional verbal demonstration, on observed OSPE scores in male & female medical undergraduates: a double control study. *Nepal Journal of Epidemiology* 2014;**4**(2):1.

Marko 2015 *{published data only}*

Marko EK, Buery-Joyner SD, Sheridan MJ, Nieves K, Khoury AN, Dalrymple JL. Structured teaching of early pregnancy loss counseling. *Obstetrics and Gynecology* 2015;**126**(Suppl 4):S1-6.

Moulton 2009 *{published data only (unpublished sought but not used)}*

Moulton C, Tabak D, Kneebone R, Nestel D, MacRae H, LeBlanc VR. Teaching communication skills using the integrated procedural performance instrument (IPPI): a randomized controlled trial. *American Journal of Surgery* 2009;**197**(1):113-8.

Pettit 2017 *{published data only}*

Pettit KE, Turner JS, Kindrat JK, Blythe GJ, Hasty GE, Perkins AJ, et al. Effect of socioeconomic status bias on medical student-patient interactions using an emergency medicine simulation. *AEM Education and Training* 2017;**1**(1):126-31.

Phisalprapa 2013 *{published data only}*

Phisalprapa P, Pandejpong D. Effect of ambulatory medicine tutorial on clinical performance of 5th year medical students. *Journal of the Medical Association of Thailand* 2013;**96**(Suppl 2):S75-81.

Quirk 1982 *{published data only}*

Quirk M, Babineau RA. Teaching interviewing skills to students in clinical years: a comparative analysis of three strategies. *Journal of Medical Education* 1982;**57**(12):939-41.

Reinders 2010 *{published data only}*

Reinders ME, Blankenstein AH, van der Horst HE, Knol DL, Schoonheim PL, van Marwijk HWJ. Does patient feedback improve the consultation skills of general practice trainees? A controlled trial. *Medical Education* 2010;**44**(2):156-64.

Ricciotti 2010 *{published data only}*

Ricciotti H, Hacker M, Flesco L, Dodge L, Huang G. Randomized, controlled trial of a normal pregnancy virtual patient to teach medical students counseling skills. *Journal of Reproductive Medicine* 2010;**55**:498-502.

Saba 2014 *{published data only}*

Saba GW, Chou CL, Satterfield J, Teherani A, Hauer K, Poncelet A, et al. Teaching patient-centered communication skills: a telephone follow-up curriculum for medical students. *Medical Education Online* 2014;**19**:22522.

Scheidt 1986 *{published data only}*

Scheidt PC, Lazoritz S, Ebbeling WL, Figelman AR, Moessner HF, Singer JE. Evaluation of system providing feedback to students on videotaped patient encounters. *Journal of Medical Education* 1986;**61**(7):585-90.

Schwartz 2007 *{published data only}*

Schwartz LR, Fernandez R, Kouyoumjian SR, Jones KA, Compton S. A randomized comparison trial of case-based learning versus human patient simulation in medical student education. *Academic Emergency Medicine* 2007;**14**(2):130-7.

Scott 1976 *{published data only}*

Scott N, Donnelly M, Hess J. Longitudinal investigation of changes in interviewing performance of medical students. *Journal of Clinical Psychology* 2007;**32**(2):424-31.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Servotte 2019 *{published data only}*

Servotte JC, Bragard I, Szyld D, Van Ngoc P, Scholtes B, Van Cauwenberge I, et al. Efficacy of a short role-play training on breaking bad news in the emergency department. *Western Journal of Emergency Medicine* 2019;**20**(6):893-902.

Sokas 1991 *{published data only}*

Sokas RK, Diserens D, Johnston MA. Integrating occupational health into the medicine clerkship using problem-based learning. *Journal of General Internal Medicine* 1991;**6**(5):450-4.

Stillman 1976 *{published data only}*

Stillman PL, Sabers DL, Redfield DL. The use of paraprofessionals to teach interviewing skills. *Pediatrics* 2855;**57**(5):769-74.

Taylor 2018 *{published data only}*

Taylor S, Bobba S, Roome S, Ahmadzai M, Tran D, Vickers D, et al. Simulated patient and role play methodologies for communication skills training in an undergraduate medical program: randomized, crossover trial. *Education for Health* 2018;**31**(1):10-6.

Tolsgaard 2013 *{published data only}*

Tolsgaard MG, Bjorck S, Rasmussen MB, Gustafsson A, Ringsted C. Improving efficiency of clinical skills training: a randomized trial. *Journal of General Internal Medicine* 2013;**28**(8):1072-7.

Walsh 2001 *{published data only}*

Walsh RA, Roche AM, Sanson-Fisher RW, Saunders JB. Interactional skills of students from traditional and non-traditional medical schools before and after alcohol education. *Medical Education* 2001;**35**(3):211-6.

Werner 2013 *{published data only}*

Werner A, Holderried F, Schaffeler N, Weyrich P, Riessen R, Zipfel S, et al. Communication training for advanced medical students improves information recall of medical laypersons in simulated informed consent talks - a randomized controlled trial. *BMC Medical Education* 2013;**13**:15.

Wolf 1987 *{published data only}*

Wolf FM, Woolliscroft JO, Calhoun JG, Boxer GJ. A controlled experiment in teaching students to respond to patients' emotional concerns. *Journal of Medical Education* 1987;**62**(1):25-34.

References to studies awaiting assessment

Ahmadreza 2011 *{published data only}*

Ahmadreza Z, Armindokht S, Maedeh M. Design, implementation and evaluation a tool to strengthen physician-patient communication skills of medical students. *Iranian Journal of Medical Education* 2011;**10**(5):1-8.

Alyami 2019 *{published data only}*

Alyami H, Alawami M, Lyndon M, Alyami M, Coomarasamy C, Henning M, et al. Impact of using a 3D visual metaphor serious game to teach history-taking content to medical students: longitudinal mixed methods pilot study. *JMIR Serious Games* 2019;**7**(3):e13748. [DOI: https://dx.doi.org/10.2196/13748]

Barais 2012 *{published data only}*

Barais M, Attencourt C, Calvez A, Le Floch Pinar T, Pradier EH, Le R. Becoming old makes one empathic: a randomized controlled trial. *European Journal of General Practice* 2012;**18**(3):171.

Caruso 1994 *{published data only}*

Caruso B, Nieman, LZ, Gracely E. Developing and assessing the effectiveness of an HIV sexual history and risk assessment workshop for medical professionals. *Journal of Sex Education and Therapy* 1994;**20**:10-109.

Cherry 2010 *{published data only}*

Cherry G, Fletcher I, O'Sullivan H. The communication skills of medical students: video analysis, OSCE scores and emotional intelligence. *Medical Education Supplement* 2010;**44**:149.

Dubosh 2019 *{published data only}*

Dubosh NM, Hall MM, Novack V, Shafat T, Shapiro NI, Ullman EA. A multimodal curriculum with patient feedback to improve medical student communication: pilot study. *Western Journal of Emergency Medicine* 2019;**21**(1):115-21.

Engler 1981 *{published data only}*

Engler CM. An analysis of medical students' acquisition and retention of interviewing skills. *Dissertation Abstracts International* 1981;**41**:3422-3.

Florenzano 2000 *{published data only}*

Florenzano R, Altuzarra R, Carvajal C, Weil K, Dorr A, Fullerton C, et al. Improving the quality of clinical interview teaching: evaluation of an intervention with medical students. *Revista Medica de Chile* 2000;**128**(3):294-300.

Geoffroy 2020 *{published data only}*

Geoffroy PA, Delyon J, Strullu M, Dinh AT, Duboc H, Zafrani L, et al. Standardized patients or conventional lecture for teaching communication skills to undergraduate medical students: a randomized controlled study. *Psychiatry Investigation* 2020;**17**(4):299-305. [DOI: https://dx.doi.org/10.30773/pi.2019.0258]

Hartigan 2012 *{published data only}*

Hartigan DB. Towards more accurate recognition of patient emotion cues: meta-analysis of training literature and development of an assessment tool and multi-component intervention for clinicians. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 2012;**72**(8-B):4998.

Hermann-Werner 2019b *{published data only}*

Hermann-Werner A, Weber H, Loda T, Keifenheim KE, Erschens R, Molbert SC, et al. "But Dr Google said..." – Training medical students how to communicate with E-patients. *Medical Teacher* 2019;**41**(12):1434-40. [DOI: doi:10.1080/0142159X.2018.1555639]

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Holtedahl 1999 *{published data only}*

Holtedahl KA, Bo B, Hansen AH, Hensrud A, Johansen ML, Prydz P, et al. Student education in a laboratory for general practice skills. *Journal of the Norwegian Medical Association* 1999;**119**(19):2854-7.

Jaury 2018 *{published data only}*

Jaury P, Buffel du Vaure C, Bunge L, Galam E, Vincens M, Catu-Pinault A, et al. Patient relationship training inspired by Balint groups in the 4th year of medical school. Assessment of these PRT groups on clinical empathy by a multisite controlled randomized trial. *Psycho-Oncology* 2018;**12**(1):8-12.

Kaper 2019 *{published data only}***doi:10.3390/ijerph17010081**

Kaper MS, Reijneveld SA, van Es FD, de Zeeuw J, Almansa J, Koot JAR, et al. Effectiveness of a comprehensive health literacy consultation skills training for undergraduate medical students: a randomized controlled trial. *International Journal of Environmemtal Research and Public Health* 2019;**17**(1):1-15.

Koch 2020 *{published data only}*

Koch C, Dreimuller N, Weiskircher J, Deis N, Gaitzsch E, Wagner S, et al. Teaching conflicts of interest and shared decision-making to improve risk communication: a randomized controlled trial. *Journal of General Internal Medicine* 2020;**35**(2):473-80. [DOI: https://dx.doi.org/10.1007/s11606-019-05420-w]

Lai 2020 *{published data only}***https://dx.doi.org/10.1186/s12909-020-02171-9**

Lai MMY, Roberts N, Mohebbi M, Martin J. A randomised controlled trial of feedback to improve patient satisfaction and consultation skills in medical students. *BMC Medical Education* 2020;**20**(1):277. [DOI: https://dx.doi.org/10.1186/s12909-020-02171-9]

Martin 2017 *{published data only}*

Martin, O, Rockenbauch K, Kleinert E, Stöbel-Richter Y. Effectively communicate active listening: comparison of two concepts [Aktives Zuhören effektivvermittelnZwei Konzepte im Vergleich]. *Der Nervenarzt* 2017;**88**(9):1026-35.

Mazori 2019 *{published data only}***https://dx.doi.org/10.22454/FamMed.2019.229662**

MazoriAY, Maron MI, Osterbur MLB, Santos D, Marco VF, Lin J, et al. Enhancing medical student-interpreter collaboration in an urban free clinic. *Family Medicine* 2019;**51**(7):593-7.

Pati 2010 *{published data only}*

Pati S. Enhancing empathy among medical students towards psychosocial needs of cancer patients and their families through micro-skills interview training. *Psycho-Oncology* 2010;**19**:S197.

Qureshi 2020 *{published data only}***https://dx.doi.org/10.1186/s12909-019-1914-2**

Qureshi AA, Zehra T. Simulated patient's feedback to improve communication skills of clerkship students. *BMC Medical Education* 2020;**20**(1):15.

Rae 1973 *{published data only}*

Rae WA, Willis DJ. Effectiveness of discussion, modeling and experiential-simulation in training empathy skills to medical students. *Unknown (ERIC)* 1973;**Unknown**:22.

Smith 2019 *{published data only}***https://dx.doi.org/10.7759/cureus.5765**

Smith C, Likourezos A, Schiller J. Focused teaching improves medical student professionalism and data gathering skills in the emergency department. *Cureus* 2019;**11**(9):e5765.

Sobana 2020 *{published data only}***http://dx.doi.org/10.18311/ajprhc/2019/25135**

Sobana R, Jaiganesh K, Parthasarathy S, Patric TKS, Sunitha P, Dharmara J, et al. Simulation for training in communication skills: active participant vs. active observer - An interventional case control study. *Asian Journal of Pharmaceutical Research and Health Care* 2020;**11**(2-4):63-7.

Spiess 1988 *{published data only}*

Spiess K, Mutschlechner R, Bilek H. Supportive psychotherapy by supervised medical students in different internistic patient groups. *Klinische Wochenschrift* 1988;**66**(Suppl 13):47.

Thompson 2012 *{published data only}*

Thompson L, Prommer E, Lower J. Delivering bad news: improving communication skills training of medical students and residents through the use of second life technology. *Journal of Pain and Symptom Management* 2012;**43**(2):452-3.

Additional references

Abbe 2013

Abbe A, Brandon SE. The role of rapport in investigative interviewing: a review. *Journal of Investigative Psychology and Offender Profiling* 2013;**10**(3):237-49.

Alelwani 2014

Alelwani SM, Ahmed YA. Medical training for communication of bad news: a literature review. *Journal of Education and Health Promotion* 2014;**3**:51. [DOI: 10.4103/2277-9531.134737]

Alliger 1989

Alliger GM, Janak EA. Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology* 1989;**42**(2):331-42.

Al-Mously 2014

Al-Mously N, Mabil NM, Al-Babtain SA, Fouad Abbas MA. Undergraduate medical students' perceptions on the quality of feedback received during clinical rotations. *Medical Teacher* 2014;**36**(S1):s17-23.

Aper 2015

Aper L, Veldhuijzen W, Dornan T, van de Ridder M, Koole S, Derese A, et al. "Should I prioritize medical problem solving or attentive listening?" The dilemmas and challenges that medical students experience when learning to conduct consultations. *Patient Education and Counseling* 2015;**98**(1):77-84.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Artemiou 2014

Artemiou E, Adams CL, Toews L, Violato C, Coe JB. Informing web-based communication curricula in veterinary education: a systematic review of web-based methods used for teaching and assessing clinical communication in medical education. *Journal of Veterinary and Medical Education* 2014;**41**(1):44-54. [DOI: 10.3138/jvme.0913-126R]

Aspegren 1999

Aspegren K. BEME Guide No. 2. Teaching and learning communication skills in medicine - a review with quality grading of articles. *Medical Teacher* 1999;**21**(6):563-70.

Australian Medical Council 2012

Australian Medical Council Limited. Standards for assessment and accreditation of primary medical programs by the Australian Medical Council 2012. www.amc.org.au/joomla-files/images/Accreditation/FINAL-Standards-and-Graduate-Outcome-Statements-20-December-2012.pdf 2012 (accessed prior to 6 October 2016).

Baile 1999

Baile WF, Kudelka AP, Beale EA, Glober GA, Myers EG, Greisinger AJ, et al. Communication skills training in oncology: description and preliminary outcomes of workshops on breaking bad news and managing patient reactions to illness. *Cancer* 1999;**86**:887-97.

Bandura 1977

Bandura A. Self efficacy: toward a unifying theory of behaviour change. *Psychological Review* 1977;**84**:191-215.

Bandura 1986

Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall, 1986.

Batt-Rawden 2013

Batt-Rawden SA, Chisolm MS, Anton B, Flickinger TE. Teaching empathy to medical students: an updated, systematic review. *Academic Medicine* 2013;**88**(8):1171-7. [DOI: 10.1097/ACM.0b013e318299f3e3]

Benbassat 2009

Benbassat J, Baumal R. A proposal for overcoming problems in teaching interviewing skills to medical students. *Advances in Health Sciences Education* 2009;**14**(3):441-50.

Benson 2015

Benson MS, Powell MB. Evaluation of a comprehensive interactive training system for investigative interviewers of children. *Psychology, Public Policy, and Law* 2015;**21**:309-22.

Berkhof 2011

Berkhof M, van Rijssen HJ, Schellart AJM, Anema JR, van der Beek AJ. Effective training strategies for teaching communication skills to physicians: an overview of systematic reviews. *Patient Education and Counseling* 2011;**84**(2):152-62.

Berne 1975

Berne E. What Do You Say After You Say Hello?. New York: Grove Press, 1975.

Bird 1990

Bird J, Cohen-Cole SA. The three function model of the medical interview. *Advances in Psychosomatic Medicine* 1990;**20**:65-88.

Boon 1998

Boon H, Stewart M. Patient-physician communication assessment instruments: 1986 to 1996 in review. *Patient Education and Counseling* 1998;**35**(3):161-76.

Bosse 2015

Bosse HM, Nickel M, Huwendiek S, Schultz JH, Nikendei C. Cost-effectiveness of peer role play and standardized patients in undergraduate communication training. *BMC Medical Education* 2015;**15**(183):6.

Braddock 1999

Braddock CH, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed decision making in the outpatient practice. Time to get back to the basics. *JAMA* 1999;**282**:2313-20.

Brockway 1978

Brockway BS. Evaluating physician competency: what difference does it make? *Evaluation and Program Planning* 1978;**1**:211-20.

Burleson 2010

Burleson BR. The nature of interpersonal communication: a message-centred approach. In: Berger CR, Roloff ME, Roskos-Ewoldsen DR, editors(s). The Handbook of Communication Science. 2nd edition. Newbury Park, CA: Sage Publishers, 2010.

CCCG 2013

Cochrane Consumers and Communication Group. Standard protocol text and additional guidance for review authors. cccrg.cochrane.org 2013 (accessed prior to 6 October 2016).

Cole 2013

Cole SA, Bird J. The Medical Interview: The Three Function Approach. 3rd edition. Philadelphia, PA: Elsevier Health Sciences, 2013.

Collins 2019

Collins K, Carthy N. No rapport, no comment: the relationship between rapport and communication during investigative interviews with suspects. *Journal of Investigative Psychology and Offender Profiling* 2019;**16**(1):18-31.

Cook 2010

Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. *Academic Medicine* 2010;**85**(10):1589-602.

Deveugele 2005

Deveugele M, Derese A, De Maesschalck S, Willems S, Van Driel M, De Maeseneer J. Teaching communication skills to medical students, a challenge in the curriculum? *Patient Education and Counseling* 2005;**58**(3):265-70.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Donovan 1999

Donovan J, Radosevich D. A meta-analytic review of the distribution of practice effect: now you see it, now you don't. *Journal of Applied Psychology* 1999;**84**:795-805.

Duffy 2004

Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R, Buffone N, et al. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. *Academic Medicine* 2004;**79**(6):495-507.

Duke 2013

Duke P, Frankel RM, Reis S. How to integrate the electronic health record and patient-centered communication into the medical visit: a skills-based approach. *Teaching and Learning in Medicine* 2013;**25**(4):358-65.

Engerer 2016

Engerer C, Berberat PO, Dinkel A, Rudolph B, Sattel H, Wuensch A. Integrating 360° behavior-orientated feedback in communication skills training for medical undergraduates: concept, acceptance and students' self-ratings of communication competence. *BMC Medical Education* 2016;**16**:271.

Fiore 1994

Fiore MC, Jaen CR, Baker TB, Bailey WC, Benowitz NL, Curry SJ, et al. Treating Tobacco Use and Dependence: 2008 Update. Clinical Practice Guideline. Rockville, MD: US Department of Health and Human Services, 2008.

Fortin 2012

Fortin AH, Dwamena FC, Frankel RM, Smith RC. Smith's Patient-Centred Interviewing: An Evidence-Based Model. 3rd edition. New York: The McGraw-Hill Companies, 2012.

Galinsky 2000

Galinsky AD, Moskowitz GB. Perspective taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Perspectives in Social Psychology* 2000;**78**:708-24.

Garfield 1978

Garfield SL, Bergin AE. Handbook of Psychotherapy and Behavioir Change. 2nd edition. New York: Wiley, 1978.

Gartmeier 2011

Gartmeier M, Bauer J, Fischer MR, Karsten G, Prenzel M. Modeling and assessment of teachers' professional competence for parent-teacher conversations [Modellierung und Assessmentprofessioneller Gespra¨chsfu¨hrungskompetenz von Lehrpersonen im Lehrer-Elterngespra¨ch]. In: Zlatikin Troitschanskaia O, editors(s). Stationen Empirischer Bildungsforschung. Traditionslinien und Perspektiven. Wiesbaden: VS, 2011:S412-26.

General Medical Council 2015

General Medical Council. Outcomes for graduates. www.gmc-uk.org/education/undergraduate/undergrad_outcomes.asp 2015 (accessed prior to 6 October 2016).

Gilligan 2016

Gilligan C, James EL, Snow P, Outram S, Ward BM, Powell M, et al. Interventions for improving medical students' interpersonal communication in medical consultations. *Cochrane Database of Systematic Reviews* 2016;(11):16. [DOI: 10.1002/14651858.CD012418] [Art. No.: CD012418]

Gregory 2007

Gregory J. Conceptualising consumer engagement: a review of the literature. Australian Institute of Health Policy Studies 2007.

Hargie 2010

Hargie O, Boohan M, McCoy M, Murphy P. Current trends in communication skills training in UK schools of medicine. *Medical Teacher* 2010;**32**(5):385-91.

Hargie 2011

Hargie O. Skilled Interpersonal Communication: Research, Theory and Practice. 5th edition. New York, NY: Routledge, 2011.

Hartley 1999

Hartley P. Interpersonal Communication. 2nd edition. New York, NY: Routledge (Taylor and Francis), 1999.

Health Professions Council of South Africa 2014

Health Professions Council of South Africa. Core competencies for undergraduate students in clinical associate, dentistry and medical teaching and learning programmes in South Africa. www.hpcsa.co.za/PBMedicalDental/Education 2014 (accessed prior to 6 October 2016).

Hermann-Werner 2019

Hermann-Werner A, Loda T, Erschens R, Schneider P, Junne F, Gilligan C, et al. Face yourself! - learning progress and shame in different approaches of video feedback: a comparative study. *BMC Medical Education* 2019;**19**(88):8.

Higgins 2011

Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Hobgood 2005

Hobgood C, Harward D, Newton K, Davis W. The educational intervention "GRIEV ING" improves the death notification skills of residents. *Academic Emergency Medicine* 2005;**12**:296-301.

Hoffman 2004

Hoffman M, Ferri J, Sison C, Roter D, Schapira L, Baile W. Teaching communication skills: an AACE survey of oncology training programs. *Journal of Cancer Education* 2004;**19**(4):220-4.

Hojat 2001

Hojat M, Mangione S, Nasca TJ, Cohen MJM, Gonnella JS, Erdmann JB, et al. The Jefferson Scale of Physician Empathy: development and preliminary psychometric data. *Educational and Psychological Measurement* 2001;**61**:349-65.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ivey 1968

Ivey A, Normington C, Miller C, Morrill D, Haase R. Microcounseling and attending behaviour: an approach to prepracticum counselor training. *Journal of Counseling Psychology Monograph - Supplement* 1968;**15**(5):Part 2.

Ivey 1994

Ivey AE. Intentional Interviewing and Counselling: Facilitating Client Development in a Multicultural Society. 3rd edition. Pacific Grove CA: Brooks/Cole, 1994.

Kagan 1998

Kagan A. Supported conversations for adults with aphasia: methods and resources for training conversation partners. *Aphasiology* 1998;**12**(9):816-30.

Kaufman 2003

Kaufman DM. Applying educational theory in practice. *BMJ* 2003;**326**(7382):213-6.

Keifenheim 2015

Keifenheim KE, Teufel M, Ip J, Speiser N, Leehr EJ, Zipfel S, Herrmann-Werner A. Teaching history taking to medical students: a systematic review. *BMC Medical Education* 2015;**15**(159):1-12. [DOI: 10.1186/s12909-015-0443-x]

Keller 1994

Keller V, Carroll JG. A new model for physician-patient communication. *Patient Education and Counseling* 1994;**23**:131-40.

Kelm 2014

Kelm Z, Womer J, Walter JK, Feudtner C. Interventions to cultivate physician empathy: a systematic review. *BMC Medical Education* 2014;**14**:219. [DOI: 10.1186/1472-6920-14-219]

Kirkpatrick 1996

Kirkpatrick D. Revisiting Kirkpatrick's four-level model. *Training and Development* 1996;**50**(1):54-9.

Kleinman 2006

Kleinman A, Benson P. Anthropology in the clinic: the problem of cultural competency and how to fix it. *PLoS Med [Internet]* 2006;**3**(10):e294.

Kretzschmar 1978

Kretzschmar RM. Evolution of the gynaecology teaching associate: an educational specialist. *American Journal of Obstetrics and Gynaecology* 1978;**131**:367-73.

Kurtz 1996

Kurtz SM, Silverman JD. The Calgary-Cambridge Referenced Observation Guides: an aid to defining the curriculum and organizing the teaching in communication training programmes. *Medical Education* 1996;**30**:83-9.

Kurtz 1998

Kurtz S, Silverman J, Draper J. Teaching and Learning Communication Skills in Medicine. Abingdon, UK: Radcliffe Medical Press, 1998.

Kurtz 2003

Kurtz S, Silverman J, Benson J, Draper J. Marrying content and process in clinical method teaching: enhancing the Calgary–Cambridge guides. *Academic Medicine* 2003;**78**(8):802-9.

Kyaw 2019

Kyaw BM, Posadzki P, Paddock S, Car J, Campbell J, Tudor Car L. Effectiveness of digital education on communication skills among medical students: systematic review and meta-analysis by the digital health education collaboration. *Journal of Medical Internet Research* 2019;**21**(8):e12967. [DOI: 10.2196/12967]

Laidlaw 2009

Laidlaw A, Guild S, Struthers J. Graduate attributes in the disciplines of medicine, dentistry and veterinary medicine: a survey of expert opinions. *BMC Medical Education* 2009;**9**:28.

Lamb 2002

Lamb M, Sternberg K, Orbach Y, Hershkowitz I, Horowitz D, Esplin P. The effects of intensive training and ongoing supervision on the quality of investigative interviews with alleged sex abuse victims. *Applied Developmental Science* 2002;**6**:114-25.

Lamb 2002b

Lamb M, Sternberg K, Orbach Y, Esplin P, Mitchell S. Is ongoing feedback necessary to maintain the quality of investigative interviews with allegedly abused children? *Applied Developmental Science* 2002;**6**:35-41.

Lane 2007

Lane C, Rollnick S. The use of simulated patients and role-play in communication skills training: a review of the literature to August 2005. *Patient Education and Counselling* 2007;**67**(1-2):13-20. [DOI: 10.1016/j.pec.2007.02.011]

Lanken 2015

Lanken PN, Novack DH, Daetwyler C, Gallop R, Landis JR, Lapin J, et al. Efficacy of an internet-based learning module and small-group debriefing on trainees' attitudes and communication skills toward patients with substance use disorders: results of a cluster randomized controlled trial. *Academic Medicine* 2015;**90**(3):345-54. [DOI: 10.1097/ACM.0000000000000506]

Lipkin 1995

Lipkin M Jr, Kaplan C, Clark W, Novack DH. Teaching medical interviewing: the Lipkin model. In: Lipkin M Jr, Putnam SM, Lazare A, editors(s). The Medical Interview: Clinical Care, Education, and Research. 1st edition. New York, NY: Springer, 1995:422-35.

Lumma-Sellenthin 2009

Lumma-Sellenthin A. Talking with patients and peers: medical students' difficulties with learning communication skills. *Medical Teacher* 2009;**31**(6):528-34.

Macdonald 2002

Macdonald G. Transformative unlearning: safety, discernment and communities of learning. *Nursing Inquiry* 2002;**9**(3):170-8.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

MacDonald-Wicks 2012

MacDonald-Wicks L, Levett-Jones T. Effective teaching of communication to health professional undergraduate and postgraduate students: a systematic review. *JBI Library of Systematic Reviews* 2012;**10**(28 Suppl):S172-83.

Maguire 1986

Maguire P, Fairbairn S, Fletcher C. Consultation skills of young doctors: I - benefits of feedback training in interviewing as students persist. *British Medical Journal* 1986;**292**(6535):1573-6.

Maguire 2002

Maguire P, Pitceathly C. Key communication skills and how to acquire them. *BMJ* 2002;**325**(7366):697-700.

Makoul 1998

Makoul G. Communication research in medical education. In: Jackson L, Duffy BK, editors(s). Health Communication Research: A Guide to Developments and Directions. Wesport, CT: Greenwood Press, 1998:17-35.

Makoul 2001a

Makoul G. The SEGUE framework for teaching and assessing communication skills. *Patient Education and Counseling* 2001;**45**(1):23-34.

Makoul 2001b

Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Academic Medicine* 2001;**76**(4):390-3.

Mead 2002

Mead N, Bower P. Patient-centred consultations and outcomes in primary care: a review of the literature. *Patient Education and Counseling* 2002;**48**(1):51-61.

Miller 2002

Miller WR, Rollnick S. Motivational interviewing: preparing people for change. New York: Guilford Press, 2002.

Morse 1992

Morse J, Anderson G, Botter JL, Yonge O, O'Brien B, Solberg SM, et al. Exploring empathy: a conceptual fit for nursing practice? *Image - The Journal of Nursing Scholarship* 1992;**24**:273-80.

Naugle 2000

Naugle KA, Naugle LB, Naugle RJ. Kirkpatrick's evaluation model as a means of evaluating teacher performance. *Education* 2000;**121**(1):135.

NHS 2010

National Health Service (UK). Liberating the NHS: no decision about me, without me - further consultation on proposals to secure shared decision-making. www.gov.uk/government/uploads/system/uploads/attachment_data/file/216980/Liberating-the-NHS-No-decision-about-me-without-me-Government-response.pdf 2010 (accessed prior to 6 October 2016).

Novack 1992

Novack DH, Dube C, Goldstein MG. Teaching medical interviewing: a basic course on interviewing and the physician-patient relationship. *Archives of Internal Medicine* 1992;**152**(9):1814-20.

Ockene 1988

Ockene JK, Quirk ME, Goldberg RJ, Kristeller JL, Donnelly G, Kalan KL, et al. A residents' training program for the development of smoking intervention skills. *Archives of Internal Medicine* 1988;**148**:1039-45.

Peterson 1992

Peterson MC, Holbrook JH, Von Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. *Western Journal of Medicine* 1992;**156**(2):163-5.

Pinto 2012

Pinto RZ, Ferreira ML, Oliveira VC, Franco MR, Adams R, Maher CG, et al. Patient-centred communication is associated with positive therapeutic alliance: a systematic review. *Journal of Physiotherapy* 2012;**58**(2):77-87.

Pollak 2011

Pollak KI, Alexander SC, Tulsky JA, Lyna P, Coffman CJ, Dolor RJ, et al. Physician empathy and listening: associations with patient satisfaction and autonomy. *Journal of the American Board of Family Medicine* 2011;**24**(6):665-72.

Prochaska 1983

Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of Consultations in Clinical Psychology* 1983;**51**(3):390-5.

Rees 2002

Rees CE, Sheard CE, McPherson AC. A qualitative study to explore undergraduate medical students' attitudes towards communication skills learning. *Medical Teacher* 2002;**24**(3):289-93.

RevMan 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration Review Manager (RevMan). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Richardson 2001

Richardson WC, Berwick DM, Bisgard JC, Bristow L, Buck CR, Cassel CK, et al. Crossing the Quality Chasm: A New Health System for the 21st Century. 1st edition. Vol. **1**. Washington DC: National Academies Press, 2001.

Rollnick 1992

Rollnick S, Heather N, Bell A. Negotiating behaviour change in medical settings: the development of brief motivational interviewing. *Journal of Mental Health* 1992;**1**:25-37.

Rosenbaum 2004

Rosenbaum ME, Ferguson KJ, Lobas JG. Teaching medical students and residents skills for delivering bad news: a review

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

of strategies. *Academic Medicine* 2004;**79**(2):107-17. [DOI: 10.1097/00001888-200402000-00002]

Rosenbaum 2013

Rosenbaum M, Axelson R. Curricular disconnects in learning communication skills: what and how students learn about communication during clinical clerkships. *Patient Education and Counseling* 2013;**91**(1):85-90.

Royston 1997

Royston V. How do medical students learn to communicate with patients? A study of fourth-year medical students' attitudes to doctor-patient communication. *Medical Teacher* 1997;**19**(4):257-62.

Ryan 2013

Ryan R, Hill S, Prictor M, McKenzie J, Cochrane Consumers and Communication. Study quality guide. cccrg.cochrane.org/authorresources May 2013 (accessed prior to 6 October 2016).

Schmitz 2017

Schmitz FM, Schnabel KP, Stricker D, Fischer MR, Guttormsen S. Learning communication from erroneous video-based examples: a double-blind randomised controlled trial. *Patient Education and Counseling* 2017;**100**:1203-12.

Schopper 2016

Schopper H, Rosenbaum M, Axelson R. 'I wish someone watched me interview': medical student insight into observation and feedback as a method for teaching communication skills during the clinical years. *BMC Medical Education* 2016;**16**(286):8.

Schulz 1996

Schulz von Thun F. Talking to Each Other. 2nd edition. Vol. **1**. Reinbeck bei Hamburg: Rowohlt Taschen-buch Verlag, 1996.

Schünemann 2011

Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11:. Presenting results and 'Summary of findings' tables. In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Sherwood 1987

Sherwood RD, Kinzer CK, Bransford JD, et al. Some benefits of creating macro-contexts for science instruction: initial findings. *Journal of Research in Science Training* 1987;**24**:417-35.

Silverman 2013

Silverman J, Kurtz SM, Draper J, van Dalen J, Platt FW. Skills for Communicating with Patients. 3rd edition. Oxford (UK): Radcliffe, 2013.

Simpson 1991

Simpson M, Buckman R, Stewart M, Maguire P, Lipkin M, Novack D, et al. Doctor-patient communication: the Toronto consensus statement. *BMJ* 1991;**303**(6814):1385-7.

Smidt 2009

Smidt A, Balandin S, Sigafoos J, Reed VA. The Kirkpatrick model: a useful tool for evaluating training outcomes. *Journal of Intellectual and Developmental Disability* 2009;**34**(3):266-74.

Smith 2007

Smith S, Hanson JL, Tewksbury LR, Christy C, Talib NJ, Harris MA, et al. Teaching patient communication skills to medical students: a review of randomized controlled trials. *Evaluation & the Health Professions* 2007;**30**(1):3-21.

Smith 2009

Smith RM, Powell MB, Lum J. The relationship between job status, interviewing experience, gender, and police officers' adherence to open-ended questions. *Legal and Criminological Psychology* 2009;**14**:51-63. [DOI: 10.1348/135532507X262360]

Stewart 1995a

Stewart M, Belle Brown J, Weston WW, McWhinney IR, McWilliam CL, Freeman TR. Patient-centred Medicine: Transforming the Clinical Method. Thousand Oaks, CA: SAGE, 1995.

Stewart 1995b

Stewart MA. Effective physician-patient communication and health outcomes: a review. *Canadian Medical Association Journal* 1995;**152**(9):1423-33.

Stewart 2001

Stewart M. Towards a global definition of patient centred care: the patient should be the judge of patient centred care. *BMJ* 2001;**322**:444.

Street 2009

Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician–patient communication to health outcomes. *Patient Education and Counseling* 2009;**74**(3):295-301.

Thistlethwaite 1999

Thistlethwaite JE, Jordan JJ. Patient-centred consultations: a comparison of student experience and understanding in two clinical environments. *Medical Education* 1999;**33**(9):678-85.

Thompson 2016

Thompson L, Howes C, McCabe R. Effect of questions used by psychiatrists on therapeutic alliance and adherence. *British Journal of Psychiatry* 2016;**209**(1):40-7.

Towle 2010

Towle A, Bainbridge L, Godolphin W, Katz A, Kline C, Lown B, et al. Active patient involvement in the education of health professionals. *Medical Education* 2010;**44**:64-74.

Uhm 2015

Uhm S, Lee GH, Jin JK, Bak YI, Jeoung YO, Kim CW. Impact of tailored feedback in assessment of communication skills for medical students. *Medical Education Online* 2015;**20**(1):7.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Vanderhaellen 2011

Vanderhallen M, Vervaeke G, Holmberg U. Witness and suspect perceptions of working alliance and interviewing style. *Journal of Investigative Psychology and Offender Profiling* 2011;**8**(2):110-30.

Van Nuland 2005

Van Nuland M, Hannes K, Cools F, Goedhuys J. Educational interventions for improving the communication skills of general practice trainees in the clinical consultation. *Cochrane Database of Systematic Reviews* 2005, Issue 4. Art. No: CD005559. [DOI: 10.1002/14651858.CD005559]

Ventres 2006

Ventres W, Kooienga S, Vuckovic N, Marlin R, Nygren P, Stewart V. Physicians, patients, and the electronic health record: an ethnographic analysis. *Annals of Family Medicine* 2006;**4**(2):124-31.

Verby 1979

Verby J, Holden P, Davis R. Peer reviews of consultations in primary care: the use of audiovisual recordings. *British Medical Journal* 1979;**1**:1686-8.

Walsh 1995

Walsh RA. Medical education about alcohol: review of its role and effectiveness. *Alcohol and Alcoholism* 1995;**30**:689-702.

Weiner 2008

Weiner SJ. Contextual error. In: Kattan MW, editors(s). Encyclopedia of Medical Decision Making. Thousand Oaks, CA: SAGE Publications, 2008:189-202.

Williams 1998

Williams GC, Deci EL. The importance of supporting autonomy in medical education. *Archives of Internal Medicine* 1998;**129**:303-8.

Windover 2014

Windover AK, Boissy A, Rice TW, Gilligan T, Velez VJ, Merlino J. The REDE model of healthcare communication: optimizing relationship as a therapeutic agent. *Journal of Patient Experience* 2014;**1**(1):8-13.

Zolnierek 2009

Zolnierek KB, DiMatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. *Medical Care* 2009;**47**(8):826-34.

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Allen 1990

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** time (rotation) **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NA **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** second year **Setting:** unclear **Percentage recruitment:** NR **Number of subjects randomised:** NR **Number of subjects participating:** - Int: 98 - Ctrl: 90 - Total: 188 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to teach students to deliver a patient-centred smoking cessation protocol **Content:** the workshop consisted of a 2-hour training session on the first day of the family medicine rotation after course orientation. Experimental groups received both standard course orientation and the workshop in a 2-hour extended session. The smoking intervention training session included (1) a lec-

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Allen 1990 *(Continued)*

	ture on the role of the physician in preventive care and on the health effects of smoking, (2) a description of a successful clinic-based smoking intervention programme, (3) a lecture on the theoretical basis of the patient-centred approach to behavioural intervention, (4) a discussion of the patient-centred smoking intervention protocol developed by Ockene and viewing of a videotaped patient-centred encounter, (5) student practice using open-ended questions in the protocol format, and (6) an explanation of the clinical assignment, which involved implementing the protocol in the clinical setting with a patient who smoked **Intervention delivered by:** unclear **Comparison:** standard course orientation session **Theoretical basis:** Ockene protocol (Ockene 1988) **Duration and timing:** 2 hours; 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** OSCE **Assessment timing:** end of year (not clear how long from training) **Primary measures:** rating focused on student ability in 3 skill areas: (1) use of open-ended questions to elicit information from the patient, (2) provision of information about smoking and smoking cessation to the patient, and (3) maintenance of direction of the protocol. Within the first 2 skill areas, students were scored on the first 4 steps of the smoking cessation protocol: patient motivation, past experience with change, barriers to change, and ways of coping. Because of time constraints of the OSCE, students were not evaluated on the last steps of the protocol: plan for change and plan for follow-up. The entire protocol was designed to take 10 to 15 minutes to complete, and the OSCE allowed only 5 minutes per station. In the third skill area, students were evaluated on their ability to explore a topic logically and with consistent direction and to follow the protocol systematically **Other measures:** student self-ratings on attributes thought to predict physician participation in preventive care **Measures assessed by:** trained raters from the Department of Family Medicine
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported Quote: "students are assigned randomly to rotation sequence and clinical site"
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Not adequately reported
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition for OSCE data

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Allen 1990 *(Continued)*

Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred between rotations.
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline outcome measures reported
Other Bias: Baseline characteristics similarity	High risk	Not reported

Alroy 1984

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** class/group **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NA **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** teaching hospital **Percentage recruitment:** NR **Number of subjects randomised:** NR **Number of subjects participating:** - Int: NR - Ctrl: NR - Total: NR (16 to 20 students in 2 groups of 8 to 10 each) **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** Israel
Interventions	**Number of experimental conditions:** 2 **Aim:** to use an interpersonal skills programme to prevent dehumanisation of medical students during their clinical clerkship **Content:** the programme involved 8 weekly meetings of a group (fifth year medical students during their initial clerkship in internal medicine). Each meeting began with previewing of a short (10-minute) videotaped doctor-patient encounter ('trigger film') and was followed by a discussion of the interpersonal skills portrayed in the trigger films. Each session was devoted to 1 trigger film and lasted 1.5 to 2 hours **Intervention delivered by:** NR **Comparison:** usual clinical sessions (no trigger films and no formal instruction on interpersonal skills) **Theoretical basis:** NR **Duration and timing:** up to 16 hours total over 8 weeks **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** 30× 10-second 'observation periods' during an observed patient encounter. Students observed twice at beginning and twice at termination of their clerkship **Assessment timing:** end of clerkship, 2 months after intervention **Primary measures:** bedside manner captured in 10-second observation periods including verbal explanation, questioning, small talk, listening, calming, empathy, encouragement, nodding, smiling/laughing, eye contact, supportive touch, eye evasion, rejection/disrespect, and writing, all coded as

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Alroy 1984 *(Continued)*

	supportive, rejection, or neutral. Also, overall assessments were made of students' behaviour as indifferent vs caring, passive vs active, aggressive vs kind, hesitant vs confident, and insulting vs respectful **Other measures:** none **Measures assessed by:** senior physicians in internal medicine
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-random Quote: "the groups are assigned to clinical departments of the university teaching hospital. For this study two groups of students in two different departments of internal medicine were chosen and designated groups I and 11"
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: the pre-trained observers (assessors) "had no knowledge pertaining to the grouping of the students"
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Information not adequately provided. *N* is not provided in results
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods.
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred between rotations.
Other Bias: Baseline outcome measurement similarity	Low risk	Pre-intervention measures similar across both groups
Other Bias: Baseline characteristics similarity	High risk	Not reported

Bearman 2001

Study characteristics

Methods	**Design:** RCT **Unit of randomisation:** student

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Bearman 2001 *(Continued)*

	Ethics and informed consent: unclear **Intention-to-treat analysis:** NA **Adjustment for clustering:** NA
Participants	**Student level:** NR **Setting:** unclear **Percentage recruitment:** 89.8% **Number of subjects randomised:** 255 **Number of subjects participating:** - Int (narrative VP): 78 - Ctrl (problem-solving VP): 79 - Baseline comparison group: 55 - Total: 255 43 of the 255 participants were excluded from analysis **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** Australia
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' history taking and communication skills **Content:** for 1 of their 8 clinical and communication skills tutorials, students were exposed to a 1-hour virtual patient (VP) tutorial. 'Intervention group' was narrative design (students are presented with a case in which they interact by selecting questions to ask) **Intervention delivered by:** computer (online) **Comparison:** problem-solving design (VP tutorial focused on diagnostic decisions) **Theoretical basis:** NR **Duration and timing:** 1 hour; 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** students conducted a role-play with a `live' simulated patient, who was played by a trained actor. There were 8 different scenarios, all of which were based on psychosocial problems. A video was taken, which was immediately assessed by an evaluator **Assessment timing:** 1 to 3 weeks after intervention **Primary measures:** overall communication skills; use of open-ended questions; identification of the reason for attendance; exploration through response to cues; asking appropriate direct questions; collection of appropriate data; attentive body language; use of appropriate language; verbal skills including listening, facilitation, and silence; response to patient; involvement of patient in consultation; sum of scores for 10 items scored out of 50 **Other measures:** none **Measures assessed by:** general practitioners
Notes	**Conflicts of interest:** not stated **Funding:** this study was supported by the Centre of Medical Informatics, Monash University **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; VP: virtual patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	The baseline group "was not a random allocation". The 2 active treatment/study groups were randomly allocated, but no information was provided on the method of randomisation

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Bearman 2001 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Quote: "the students themselves were not aware of the theoretical difference between the two types of tutorials or of the hypothesis under investigation"
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. Evaluators were general practitioners and were not aware which group students had been assigned to
Incomplete outcome data (attrition bias) All outcomes	Low risk	All attrition and exclusions reported (1 evaluator data and students not understanding). Low attrition (29/284). In addition, "Total exclusions of 43 students were 10% overall, of which approximately half were random due to the removal of an evaluator". One evaluator was excluded (different from others as not a practising GP) and so were the data (n = 21) collected. Two students who "had not understood the instructions or performed significantly badly were also not included in the data"
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Study authors mentioned possible contamination from other activities. Some students had a 3-week gap between tutorial and evaluation
Other Bias: Baseline outcome measurement similarity	Unclear risk	Measures for 'baseline groups' were reported but not for all groups
Other Bias: Baseline characteristics similarity	High risk	Not reported. Baseline characteristics not examined

Berney 2017

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional approval granted; consent unclear **Intention-to-treat analysis:** no **Adjustment for clustering:** NA
Participants	**Student level:** fourth year **Setting:** classroom (university hospital) **Percentage recruitment:** NR **Number of subjects randomised:** 236 **Number of subjects participating:** - Int: 96

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Berney 2017 *(Continued)*

- Ctrl: 140

- Total: 236

Age (mean): 23

Sex: 43% male

Ethnicity: NR

Country: Switzerland

Interventions	**Number of experimental conditions:** 2 **Aim:** the aim of the study was to compare 1-to-1 SP training and individual supervision with small-group SP training and collective supervision as part of an undergraduate CST programme in BBN in oncology **Content:** in the intervention group, each student conducted 2 videotaped interviews with an SP (2 training sessions with an interval of 6 weeks); after each session, the student had a 60-minute individual supervision by a faculty tutor **Intervention delivered by:** 8 tutors who were involved in the basic course in clinical communication for second-year medical students and had extensive experience in clinical supervision; their professional background is clinical psychology or psychiatry. Moreover, they attended a "train-the-trainer" course based on videotaped BBN interviews of students and SPs, focusing on the feedback that would be given to the student on core components of the BBN task **Comparison:** standard curriculum in BBN; 2 teaching sessions in small groups (12 students), with 1 student per session conducting a videotaped interview with an SP, which is observed by the other students and discussed collectively—90-minute feedback—with faculty **Theoretical basis:** NR **Duration and timing:** 2 hours in total (2 × 2-hour sessions separated by 6 weeks) **Fidelity:** tutor training **Adherence:** 3 students did not complete the CST programme to the end and were excluded from analysis
Outcomes	**Assessment:** post-training videotaped interaction with SP **Assessment timing:** NR, but appears to be days/weeks after intervention **Primary measures:** the BBN communication performance of medical students was rated using the checklist of teaching objectives on the basis of the methods developed and validated within the Calgary Cambridge framework for BBN OSCE. The overall impression item is a global rating of the interview on a scale from 1 (very bad overall impression/clear fail) to 5 (very good overall impression/excellent pass). The process skills grid includes 17 items rated on 3 levels (0 = not done/inadequate, 1 = adequate, 2 = good) **Other measures:** verbal behaviour during the videotaped interviews was coded for both student and patient with the Roter Interaction Analysis System (RIAS). The RIAS provides a framework of 37 mutually exclusive categories of communication to which patients' and clinicians' utterances are assigned, and that reflect the content and form of the communication **Measures assessed by:** trained coder viewing videotapes
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** Swiss Cancer Research Foundation/Swiss Cancer League

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Berney 2017 *(Continued)*

Abbreviations: BBN: Breaking Bad News; CST: communication skills training; Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	High risk	Those who did not complete training were excluded
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods. Scales reported at summary level
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No comparison at baseline
Other Bias: Baseline characteristics similarity	High risk	Not reported

Betchart 1984

Study characteristics

Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** unclear **Intention-to-treat analysis:** unclear **Adjustment for clustering:** NA
Participants	**Student level:** first year **Setting:** university campus **Percentage recruitment:** NR **Number of subjects randomised:** 12 **Number of subjects participating:** - Int: 5 - Ctrl: 7 - Total: 12

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Betchart 1984 *(Continued)*

	Age (mean): NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve the interviewing skills of students who demonstrated inadequate skills in a test with proven reliability and validity **Content:** after participation in a required interviewing course (including a videotaped interview reviewed with peers and preceptors), students in the experimental group conducted 4 additional interviews, which were reviewed in tutorial sessions. During each of these interviews, students were observed on a remote television monitor by 1 of 2 trained raters and were assessed on the process and content scales. Immediately following the interviews, the rater reviewed the tape with the student and provided structured feedback about how the student could improve specific weaknesses **Intervention delivered by:** 'trained raters' (qualifications not specified) **Comparison:** usual curriculum **Theoretical basis:** NR **Duration and timing:** 4 feedback sessions **Fidelity:** not fidelity, but interrater reliability using an intraclass correlation technique was .91 and.95 for process and content, respectively **Adherence:** NR
Outcomes	**Assessment:** video-recorded interview rated at the end of the tutoring intervention **Assessment timing:** immediately after intervention **Primary measures:** process and content items: explore concerns, eye contact, posture and deportment, cover major areas, direct the interview, introduction, open-ended questions, simple questions, clarification, transitional statements, facilitation, closing **Other measures:** none **Measures assessed by:** 'assessors' (qualifications not specified)
Notes	**Conflicts of interest:** not stated **Funding:** the research on which this communication is based was supported in part by National Institute of Mental Health grant **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported Quote: "five of these students were randomly assigned to the experimental condition....seven were assigned to the control condition"
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Quote: "the raters were not 'blind' during the tutoring sessions…However, the high degree of interrater and intra-rater reliability that was obtained provides reasonable assurance that the effects found were due to the tutoring itself and not to rater bias"

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Betchart 1984 *(Continued)*

Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition.
Selective reporting (reporting bias)	High risk	Some data reporting is incomplete. The subgroup intervention (5 exp + 7 control) overall score is reported, but detailed scales are reported only for the 5 exp groups. Then, each group (exp and control) is only partially reported against a 'random' sample of 29 classmates (control group) and the average of the class (exp group)
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	Not adequately reported
Other Bias: Baseline characteristics similarity	High risk	Not adequately reported

Blatt 2010

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** ethical approval was given by the institutional review boards of the university with participating students. No information on informed consent **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** third year **Setting:** university campus **Percentage recruitment:** 100% **Number of subjects randomised:** 608 **Number of subjects participating:** - Int: 122 (study 1), 135 (study 3) - Ctrl: 123 (study 1), 123 (study 3) - Total: 245 (study 1), 258 (study 3) **Age (mean):** NR **Sex:** study 1: 44% male (intervention) and 41% male (control). Study 3: 43% male (intervention) and 44% male (control) **Ethnicity:** study 1: African American 30% (intervention) and 41% (control); white 43% (intervention) and 37% (control); other race 26% (intervention) and 21% (control). Study 3: African American 41% (intervention) and 33% (control); white 35% (intervention) and 41% (control); other race 24% (intervention) and 26% (control)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Blatt 2010 *(Continued)*

	Country: USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to induce empathy in students by using a didactic approach **Content:** the students in the intervention groups (study 1 and study 3) received special instructions on perspective taking. Students were also instructed to write a note after seeing the patient; the note had to describe what students imagined the patients were experiencing **Intervention delivered by:** study authors **Comparison:** neutral instructions before clinical skills examination **Theoretical basis:** perspective taking (Galinsky 2000) **Duration and timing:** 1-off intervention, intervention length not reported **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** NR **Assessment timing:** end of third year. **Primary measures:** in study 1, patients reported on students' listening skills, caring, fostering patient participation in care, trust, and patients' overall satisfaction. In study 3, patients reported on professional competence, students' information gathering skills, their listening skills, relationship building, students' ability to explore patient perspective, addressing patient's feelings, meeting patient needs, patient's overall satisfaction **Other measures:** history taking, physical examination **Measures assessed by:** SP
Notes	**Conficts of interest:** not stated **Funding:** these studies were integrated into the standard education programmes of the George Washington School of Medicine and Health Sciences and the Howard University College of Medicine. They received no outside support **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated randomisation Quote: "randomised students, using a computer-generated schedule, to an intervention or control group...stratified the randomisation by gender and race"
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Not adequately reported. No information about medical student blinding. All students in a study received the same cases. Intervention and control groups were asked to write a brief note after completing each case. All students attended pre-examination orientation sessions in small groups; students were assigned to groups according to their intervention status

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Blatt 2010 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	SP assessors were blind to group assignment Quote: "in all studies, the SPs were blind as to whether the students were in the control or intervention group"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition (< 5%) - 9/251 students did not complete the subscale in Study 3
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure for study 1 or 2. Study 3 has a related measure at baseline
Other Bias: Baseline characteristics similarity	Low risk	Differences between control and intervention groups in each study are generally small

Bosse 2012

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** no ethics approval; opt-out opportunity provided **Intention-to-treat analysis:** unclear **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fifth year **Setting:** unclear **Percentage recruitment:** 94.2% **Number of subjects randomised:** 103 **Number of subjects participating:** - Int 1 (SP group): 32 - Int 2 (RP group): 28 - Ctrl: 32 - Total: 97 **Age (mean):** 24.6 **Sex:** 54.4% male **Ethnicity:** NR **Country:** Germany
Interventions	**Number of experimental conditions:** 3 **Aim:** to improve students' scores in objective communication performance using accessible and sustainable methods **Content:** a total of 3 small group training sessions each with 1 tutor and 3 students were conducted on 3 consecutive weeks for all participants of both intervention groups - RP group: peer role-play - SP group: simulated patient role-play

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Bosse 2012 *(Continued)*

	In each session, students worked on 3 training cases and rotated in the roles of physician, parent, and observer in the RP group, and in the roles of physician and observer in the SP group. Each session started with a 10-minute interview. The student in the physician's role then was the first to reflect on the interview, followed by feedback (a) by the student in the parent role (RP group) or structured feedback by the standardised patient (SP group), respectively; (b) by observers (peers) using a structured feedback checklist addressing major medical and interaction issues; and (c) by the tutor with time for a subsequent group discussion and a debriefing Parallel to the course, a weekly seminar covering the key issues addressed in the scenarios was attended well by all 3 groups. Key issues addressed in the seminars relating to medical aspects of underlying diseases and proposed management of the scenarios, as well as communication tasks that were expected to arise in the parent encounter with respective solution All 3 groups additionally received a printout, as well as electronic access to an abstract of each scenario, which summed up these key issues **Intervention delivered by:** RP group: tutors (not described) and peers; SP group: tutors and SPs **Comparison:** usual curriculum (established course contents were maintained identical in all groups and included seminars, problem-based learning, virtual patients, bedside teaching, skills training, and placement in private paediatric practices) **Theoretical basis:** theoretical framework of transactional analysis (Berne 1975), as well as communication theories of Schulz von Thun (Schulz 1996). Also, Calgary Cambridge Guides for assessment (Kurtz 1996) **Duration and timing:** 3 sessions over 3 consecutive weeks **Fidelity:** NR **Adherence:** all students undertook all aspects of the programme
Outcomes	**Assessment:** 6 station OSCEs **Assessment timing:** end of course (not specified) **Primary measures:** Calgary Cambridge domains: understanding patients' perspective, building the relationship, exploring the problem, structuring the consultation **Other measures:** self-efficacy ratings **Measures assessed by:** trained psychologists
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** the study was conducted without financial support of a third party **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; RP: role-play (refers to peer role-play); SP: simulated/standardised patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported Quote: "fifth year students eligible for their rotation in pediatrics were randomly assigned to one of three study groups"
Allocation concealment (selection bias)	High risk	Quote: "a concealment of allocation could not be performed due to the nature of the course and the study design"
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention Quote: "a concealment of allocation or blinding of tutors could not be performed due to the nature of the course and the study design"
Blinding of outcome assessment (detection bias)	Low risk	Assessors were blind to group assignment. Raters were blinded to allocation of students and did not participate in the training (i.e. as tutors)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Bosse 2012 *(Continued)*

All outcomes		
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	6/103 dropped out before training, 5/97 after training but before OSCE. Not balanced across groups Quote: "due to the fact that Heidelberg medical students frequently opt for rotations abroad and also do so at short notice, there was a dropout of six students"
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	Baseline self-assessment of communication similar
Other Bias: Baseline characteristics similarity	Low risk	No significant group differences were found with respect to sex and motivation to study medicine

Bowyer 2006

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** class/group **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NA **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** skills laboratory **Percentage recruitment:** 100% **Number of subjects randomised:** 39 **Number of subjects participating:** - Int: 18 - Ctrl: 21 - Total: 39 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' skills in delivering bad news in clinical scenarios **Content:** after completing a trauma ER scenario in which a patient dies, student's next skills station is informing the patient's wife of her husband's death. The intervention group received training in the form of a didactic lecture about how to deliver bad news to a patient, followed by small-group sessions that allowed students to practice these skills and receive feedback **Intervention delivered by:** unclear **Comparison:** no training before BBN simulation **Theoretical basis:** instruction was based on the SPIKES model developed to teach these communication skills (Baile 1999) **Duration and timing:** unclear duration, 1-off **Fidelity:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Bowyer 2006 *(Continued)*

	Adherence: NR
Outcomes	**Assessment:** BBN simulation with SP **Assessment timing:** immediately after intervention **Primary measures:** 21 items relating to students' appearance, communication skills, and emotional affect. Only 6 listed: the student inspired trust, the student provided guidance, the student prepared me for news, the student assessed my knowledge, the student had a plan to BBN, the student appeared comfortable **Other measures:** none **Measures assessed by:** SP
Notes	**Conficts of interest:** not stated **Funding:** not stated **Abbreviations:** BBN: breaking bad news; Ctrl: control; Int: intervention; ER: emergency room; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated/standardised patient; SPE: standardised patient experience; SPIKES: setting, perception, invitation, knowledge, empathy, summary and strategy model

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-RCT
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	SP assessors were blind to group assignment Quote: "the SPWs were blinded to the student groups"
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	N not provided in results, so attrition is not known
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods.
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	Not measured
Other Bias: Baseline characteristics similarity	Unclear risk	Groups were equal in terms of their self-reported previous breaking bad news training, but no other characteristics were reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Bowyer 2010

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional review board approval and permission from the Dean of students; no informed consent **Intention-to-treat analysis:** NA **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** skills laboratory **Percentage recruitment:** 100% **Number of subjects randomised:** 553 **Number of subjects participating:** - Int 1: 163 - Int 2: 109 - Int 3: 118 - Ctrl: 163 - Total: 553 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 4 **Aim:** to improve students' skills in delivering bad news in clinical scenarios **Content:** after completing a trauma ER scenario in which a patient dies, the student's next skills station is informing the patient's wife of her husband's death Intervention 1: watched a 15-minute video on the SPIKES model just before speaking with the patient's wife. Intervention 2: received a 45-minute didactic lecture on BBN and the SPIKES model, and observed a faculty facilitator informing an SP that she had miscarried using the SPIKES protocol with subsequent discussion Intervention 3: received the same training as intervention 2 as well as the 15-minute video **Intervention delivered by:** 'faculty' and SP, or video only **Control:** no preparation for BBN simulation **Theoretical basis:** instruction was based on the SPIKES model developed to teach these communication skills (Baile 1999). Mixed-reality simulation **Duration and timing:** intervention 1: 15 minutes; intervention 2: 45 minutes; intervention 3: 1 hour; all 1-off **Fidelity:** extensive training and standardisation of SPs **Adherence:** NR
Outcomes	**Assessment:** BBN simulation **Assessment timing:** immediately after intervention **Primary measures:** 21 items relating to students' appearance, communication skills, and emotional affect. Only 6 listed: the student inspired trust, the student provided guidance, the student prepared me for news, the student assessed my knowledge, the student had a plan to BBN, the student appeared comfortable **Other measures:** none **Measures assessed by:** SP
Notes	**Conficts of interest:** not stated **Funding:** not stated

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Bowyer 2010 *(Continued)*

Abbreviations: BBN: breaking bad news; Ctrl: control; ER: emergency room; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated/standardised patient; SPE: standardised patient experience; SPIKES: setting, perception, invitation, knowledge, empathy, summary and strategy model

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	SP assessors were blind to group assignment Quote: "the SPWs were blinded to the student groups"
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	N not provided in results, so attrition is not known
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	Baseline self-assessment of communication similar
Other Bias: Baseline characteristics similarity	High risk	All third-year students, but no other information

Buffel du Vaure 2017

Study characteristics

Methods	**Design:** RCT **Unit of randomisation:** student, rotation (different methods at 2 participating sites) **Ethics and informed consent:** institutional ethics approval and informed consent provided through a website **Intention-to-treat analysis:** no, but post-hoc analysis done including dropouts; no differences were found **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Buffel du Vaure 2017 *(Continued)*

Participants	**Student level:** fourth year **Setting:** classroom (teaching hospital) **Percentage recruitment:** 88% **Number of subjects randomised:** 352 **Number of subjects participating:** - Int: 155 - Ctrl: 144 - Total: 299 **Age (mean):** NR **Sex:** 40% male **Ethnicity:** NR **Country:** France
Interventions	**Number of experimental conditions:** 2 **Aim:** to assess the efficacy of Balint groups in improving empathy among fourth-year medical students **Content:** groups of 12 or 13 students received training over 2 months that included 7 weekly 1.5-hour Balint group sessions. Participants in Balint groups were asked to react to a particularly touching, upsetting, or interesting live clinical situation that involves interpersonal problems, under the supervision of a trained facilitator. This training was not specifically tailored to improve empathy and did not differ from usual Balint group sessions **Intervention delivered by:** facilitators who were accredited as Balint group leaders either by the French Balint Medical Society (Société Médicale Balint France) or the Balint Training Association (Association de Formation Balint) **Comparison:** no specific training **Theoretical basis:** NR **Duration and timing:** 7 1.5-hour sessions over 2 months **Fidelity:** NR **Adherence:** 52.9% attended all 7 sessions
Outcomes	**Assessment:** 2 15-minute OSCE stations **Assessment timing:** 1 week after final intervention session **Primary measures:** CARE Scale: a 10-item patient-rated questionnaire of physician empathy (e.g. “How good was the practitioner at showing care and compassion (seeming genuinely concerned, connecting with you on a human level; not being indifferent or detached)?”). Each item can be scored on a Likert scale from 1 (poor) to 5 (excellent), with a ‘does not apply’ option, leading to a summed score ranging from 10 to 50, with higher score indicating higher levels of empathy **Other measures:** self-rated Jefferson's School Empathy Scale - Medical Student version **Measures assessed by:** SPs
Notes	**Conficts of interest:** declaration of no conflicts of interest

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Buffel du Vaure 2017 *(Continued)*

Funding: this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector

Abbreviations: CARE: Consultation and Relational Empathy; Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: standardised patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	At 1 site, computer-generated random numbers were used. At the other site, methods were not described
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. OSCE observers and data analysts were blinded to group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Sensitivity analysis conducted using baseline data for those lost to follow-up, with no difference in outcome
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No CARE Scale measure at baseline
Other Bias: Baseline characteristics similarity	Low risk	Groups did not differ regarding sociodemographic factors at baseline

Cave 2007

Study characteristics

Methods	**Design:** C-RCT **Unit of randomisation:** class/group **Ethics and informed consent:** ethics unclear; only the results of students who gave their consent were used in the analysis **Intention-to-treat analysis:** no **Adjustment for clustering:** no
Participants	**Student level:** third year **Setting:** skills laboratory **Percentage recruitment:** 91% **Number of subjects randomised:** 396 **Number of subjects participating:**

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Cave 2007 *(Continued)*

	- Int A: 107 - Int B: 128 - Ctrl: 124 - Total: 359 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** UK
Interventions	**Number of experimental conditions:** 3 **Aim:** the aims of this study are 3-fold: (1) to investigate 3 different ways of introducing standard assessment criteria into communication skills teaching, and to determine the effects of these upon performance in the OSCE; (2) to investigate students' ability to assess their own and their peers' communication skills and compare this assessment to assessments of their tutors and simulated patients; (3) to investigate the relationship between students' performance in communication skills teaching and the communication skills OSCE **Content:** Intervention A: during communication skills teaching, every student was given a copy of the standard assessment criteria. This was used to inform the discussion and feedback Intervention B: communication skills teaching sessions were run as 'mini-OSCEs' with simulated patients. Students' performance was graded by themselves, their peers, the tutor, and the actor, using the standard assessment criteria **Intervention delivered by:** 'tutors' (qualifications not specified) **Comparison:** students attended communication skills teaching as usual (including role-play and feedback/discussion and video review) and were told that the standard assessment criteria were available on the medical school website **Theoretical basis:** links between teaching and assessment **Duration and timing:** 4 sessions across the year **Fidelity:** NR **Adherence:** all students took part in teaching, and only 1 missed the assessment
Outcomes	**Assessment:** end of year OSCE with 4 communication stations, each marked out of 20 **Assessment timing:** end of year (within 12 months of intervention) **Primary measures:** performance in the communication skills stations in end of year summative OSCE. Mark sheets included Introduction and information gathering (clearly introduces self and patient, clarifies interview purpose, establishes and maintains rapport, attends to patient concerns, establishes what the patient knows so far and wants to know before proceeding); giving information and decision-making (provides appropriate amount of information in non-medical language, at appropriate pace, checks patients' understanding and concerns, negotiates and agrees on a plan) **Other measures:** scores on the marksheets completed by students, tutors, and simulated patients during teaching **Measures assessed by:** 'assessors' (qualifications not described)
Notes	**Conficts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; SP: simulated/standardised patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomisation was done on a group level and was performed by drawing cards out of a box

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Cave 2007 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Not adequately reported
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Not adequately reported
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	High risk	Same institution, same academic year. Study authors stated: "because of the high chance of contamination between groups, this study could not contain a 'true' (non-intervention) control"
Other Bias: Baseline outcome measurement similarity	Unclear risk	Baseline not done
Other Bias: Baseline characteristics similarity	High risk	Not reported
Recuitment Bias (cluster trials only)	Unclear risk	Not adequately reported
Selective recruitment (cluster trials only)	Unclear risk	Not adequately reported
Loss of clusters (cluster trials only)	Unclear risk	Not adequately reported
Statistical methods (cluster trials only)	High risk	No adjustment for clustering reported in statistical analysis section
Herd effect (cluster trials only)	Unclear risk	Not adequately reported

Chibhall 2005

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional ethical approval and written informed consent **Intention-to-treat analysis:** unclear **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Chibhall 2005 *(Continued)*

Participants	**Student level:** third year **Setting:** skills laboratory **Percentage recruitment:** NR **Number of subjects randomised:** 27 **Number of subjects participating:** - Int: 15 - Ctrl: 12 - Total: 27 **Age (mean):** NR **Sex:** 40.7% male **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' bedside manner (empathy and positive attitude towards patients) through religious awareness training **Content:** general patient supportiveness training (ST) plus religious (RAT) awareness training. Students learned to assess patient psychosocial concerns and to assist in managing those concerns. In the RAT group, in addition to ST, religious issues were presented as part of the psychosocial presentation. Ways of supporting patients emotionally, informationally, and instrumentally were taught regarding the health relevance of positive religious coping, including use of the HOPE spiritual history method **Intervention delivered by:** NR **Comparison:** ST only **Theoretical basis:** HOPE spiritual history method (no reference given) **Duration and timing:** unclear **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** SP rating **Assessment timing:** 1 week after training **Primary measures:** SP rated students' interpersonal behaviours on seven 5-point Likert scales (0 = poor, 4 = excellent). These ratings yielded 2 subscores: affective behaviour (e.g. acknowledging patient distress, showing interest in patient as a person, exuding warmth) and communication behaviour (e.g. using lay language, speaking to patient as an adult without condescension) **Other measures:** none **Measures assessed by:** SP
Notes	**Conficts of interest:** declaration of no conflicts of interest **Funding:** this study was supported by the Lutheran Foundation, St. Louis, MO, and the Marchetti Jesuit Endowment, Saint Louis University, St. Louis, MO **Abbreviations:** Ctrl: control; Int: intervention; RAT: religious awareness training; RCT: randomised controlled trial; SP: standardised patient; ST: supportiveness training

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Participants blind to hypotheses, but unclear if blind to group assignment

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Chibhall 2005 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Not adequately reported
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not adequately reported
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	Not measured
Other Bias: Baseline characteristics similarity	High risk	Not adequately reported

Colletti 2001

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional ethics approval; informed consent unclear **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but clustering effect possible due to study design
Participants	**Student level:** junior year clerkship (year 3) **Setting:** teaching hospital **Percentage recruitment:** NR **Number of subjects randomised:** NR **Number of subjects participating:** - Int 1 (pregnancy loss): 16 - Int 2 (colon cancer): 15 - Ctrl: 17 - Total: 38 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 3 **Aim:** to improve students' skills in delivering bad news in clinical scenarios **Content:** students participated in an encounter with a standardised patient instructor (SPI) in 1 of 2 clinical scenarios Intervention/Clinical scenario 1: discussion of a new diagnosis of rectal cancer Intervention/Clinicial scenario 2: discussion of spontaneous miscarriage SPIs critiqued students' performance after a simulated interview. An information packet was provided before the encounter, with information about the particular SPI experience and about techniques for breaking bad news **Intervention delivered by:** SPI

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Colletti 2001 *(Continued)*

	Comparison: usual care – no SPI experience **Theoretical basis:** NR **Duration and timing:** 1 to 2 hours; 1-off **Fidelity:** SPIs trained and 'standardised' on performance and rating using the 27-item scale **Adherence:** NR
Outcomes	**Assessment:** clinical performance examination (CPE) **Assessment timing:** end of year (time from intervention unclear) **Primary measures:** 27 items assessed in CPE (items not specified) **Other measures:** none **Measures assessed by:** SPI (different from the one who delivered intervention)
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** CPE: clinical performance examination; Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SPI: standardised patient instructor

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-random. Method of randomisation not adequately reported. Randomisation based on rotation group
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Not clear whether SPs conducting assessment were blinded to student allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures seem to match methods. Reported at summary level only
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline outcome measure
Other Bias: Baseline characteristics similarity	High risk	Not reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Daeppen 2012

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** yes, institutional ethics approval and oral consent obtained **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** second-year master's (fifth year equivalent) **Setting:** unclear **Percentage recruitment:** 69.5% **Number of subjects randomised:** 131 **Number of subjects participating:** - Int: 42 - Ctrl: 49 - Total: 91 **Age (mean):** 24.7 **Sex:** 40.7% male **Ethnicity:** NR, but 80.2% French speaking **Country:** Switzerland
Interventions	**Number of experimental conditions:** 2 **Aim:** aims of the training workshop were to understand different MI communication styles (directing, following, and guiding) and to be able to use each style adequately within a medical consultation; to define MI and MI style (evocation, collaboration, autonomy, and empathy); and to acquire core MI skills (ask, listen, and inform) and learn to use them appropriately to explore patient motivation, to elicit and reinforce change talk, to lower resistance, and to guide patients through change **Content:** MI training was conducted in groups of 15 to 20 students and consisted of two 4-hour sessions of practical exercises and role-playing **Intervention delivered by:** NR **Comparison:** no MI workshop before SP experience/OSCE **Theoretical basis:** motivational interviewing principles (Rollnick 1992) **Duration and timing:** 8 hours **Fidelity:** NR **Adherence:** attendance; 6 did not attend training, 3 attended half, 1 attended three-quarters. Data were excluded for all these participants
Outcomes	**Assessment:** students met with 2 standardised patients for 15 minutes each. Standardised patients were a 60-year-old man with a history of severe nicotine dependence hospitalised after a myocardial infarction, and a 50-year-old diabetic woman with an unhealthy diet, problems with medication compliance, and lack of exercise **Assessment timing:** 1 week after training **Primary measures:** MITI 3.0: a reliable and valid behavioural coding system that provides information on use of MI by practitioners as a treatment integrity measure for MI clinical trials. Its components are (1) global scores (capture the global impression, or overall judgement, of the 3 dimensions: empathy, direction, and MI spirit (itself comprising 3 subscales, i.e. evocation, collaboration, and autonomy/support); and (2) behaviour counts, where the coder is asked to tally (but not judge the quality and overall adequacy of) instances of certain interviewer behaviours. Counts are categorised into giving information; MI-adherent behaviours (e.g. asking permission, affirming, emphasising control, supporting) and MI-non-adherent behaviours (e.g. giving advice without permission, confronting, directing), closed question, open question, simple reflection, and complex reflection **Other measures:** several summary scores additionally serve as outcome measures for determining MI competence: per cent MI-adherent (MI-adherent divided by MI-adherent + MI-non-adherent), per cent open questions (open questions divided by total questions), per cent complex reflections (complex reflections divided by total reflections), and reflections/questions ratio (total reflections divided by total questions) **Measures assessed by:** trained master-level students coded taped encounters using MITI 3.0

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Daeppen 2012 *(Continued)*

Notes	**Conficts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; MI: motivational interviewing; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated/standardised patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Students were a priori randomly allocated to groups through a computerised algorithm
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. Coded by blinded RA
Incomplete outcome data (attrition bias) All outcomes	Low risk	Although attrition was high (30.5%), a sensitivity analysis indicated that attrition bias is unlikely to impact the results
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	High risk	Same institution, "the randomised controlled design of our study also exposes a risk of contamination between groups"
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline outcome measures
Other Bias: Baseline characteristics similarity	Low risk	No differences between groups on age, sex, or first language

Edwardsen 2006

Study characteristics

Methods	**Design:** RCT **Unit of randomisation:** class/group **Ethics and informed consent:** ethics approval granted and written informed consent obtained from all participants **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** first year

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Edwardsen 2006 *(Continued)*

	Setting: university campus **Percentage recruitment:** 86% **Number of subjects randomised:** 100 randomised for intervention, 50 randomly selected for evaluation **Number of subjects participating:** - Int: 22 - Ctrl: 21 - Total: 43 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to guide trainees through an inquiry and assessment of intimate partner violence (IPV) **Content:** a 1-hour presentation on IPV was incorporated into the curriculum for a longitudinal first-year course and delivered to the entire first-year class in a large-group session. The presentation included brief general information on IPV, exposure to the partner violence mnemonic SCRAPED, and a model interview of an IPV survivor performed by a faculty member. The large-group didactic session was followed by a 1-hour faculty-facilitated small-group discussion. Trained actors were made available to all small groups and were used for students to practice their interviewing skills. We gave students in the intervention discussion groups a laminated card with a copy of the mnemonic and additional descriptive information about its utility in the clinical interview. One SP interview was performed by a student in each group using the mnemonic as a guide and as a reminder of key issues. The interview was used as the basis for a faculty-facilitated dialogue about the mnemonic and how to use it during the interview. We presented a structured format for using the mnemonic during discussion sessions **Intervention delivered by:** 'faculty facilitators' (skills not specified) **Comparison:** standard educational methods including general discussion about IPV and use of a simulated patient. Provision of mnemonic at end of session **Theoretical basis:** NR **Duration and timing:** 2 hours, 1-off **Fidelity:** SPs underwent training and standardisation **Adherence:** NR
Outcomes	**Assessment:** a videotaped competency-based evaluation using SPs **Assessment timing:** day of training **Primary measures:** communication styles/skills that enabled students to successfully elicit a history of partner abuse and encouraged open discussions about abuse: 1. Was a direct question about abusive behaviours asked? 2. Was an empathic statement provided? 3. Was the possibility of prior abuse explored? 4. Was the partner's behavior discussed? 5. Was safety of the patient addressed? **Other measures:** perceived usefulness of the mnemonic based on a post-evaluation questionnaire and an assessment of the mnemonic **Measures assessed by:** the research team assessed videotaped interviews blinded to the intervention status of interviewees
Notes	**Conflicts of interest:** not stated **Funding:** this project was funded by a development grant from the Department of Emergency Medicine, University of Rochester School of Medicine **Abbreviations:** Ctrl: control; Int: intervention; IPV: intimate partner violence; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated/standardised patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Edwardsen 2006 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Low risk	"To reduce test expectancy and test–retest bias, we informed the students that their interviewing skills for the entire course were being evaluated, not just the IPV session they had just completed"
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group allocations Quote: "the research team assessed the videotaped interviews blinded to the intervention status of the interviewees"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition was balanced between groups and was less than 20%, but reason for attrition was not reported. 3/25 for intervention group, 4/25 for control group
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred. Attempted to reduce bias by advising "students not to discuss the educational sessions before the evaluation phase"
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measures reported
Other Bias: Baseline characteristics similarity	High risk	No baseline measures reported

Eells 2002

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** unclear whether ethics approval was required/granted, but students 'agreed to participate' **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but clustering effect possible due to study design
Participants	**Student level:** second-year **Setting:** university campus **Percentage recruitment:** 76% **Number of subjects randomised:** 104 **Number of subjects participating:** - Int 1 (problem-oriented instruction): 27 - Ctrl 1 (lecture format): 27 - Ctrl 2 (self-study): 23 - Total: 77 **Age (mean):** 24.5 **Sex:** 53.2% male **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 3

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Eells 2002 *(Continued)*

Aim: to improve students' skills in psychiatric interviewing
Content:

Intervention group: 2 sessions of problem-oriented instruction - a 2-hour session, followed 2 weeks later by a 1-hour session. Principal activity involved students role-playing interview situations. Three rounds occurred of role-play following a brief didactic instruction (covering attending behaviours, asking questions, and reflecting feelings, and paraphrasing and summarising). For each round of role-playing, students were arranged in groups of 3. Each round consisted of a series of 3-minute vignettes in which each student shifted between roles (patient, interviewer, observer). Each interviewer was expected to practise the specific skills emphasised in each round, but each was given a different topic or line of inquiry. In the second session, role-plays were 7 minutes long
Intervention delivered by: psychiatrists, clinical psychologists, social workers, pastoral counsellors, art therapist with experience in treating adults in psychotherapy (average rating of 6.8 on 9-point Likert scale of experience)
Control:

Comparison group 1 (**l** ecture format): received 3 hours of lecture material on psychiatric interviewing (same topics as Intervention group but with further elaboration and examples)
Comparison group 2 (self-study): no face-to-face. Self-study with detailed information about interviewing skills was available for their review and study in the course syllabus (this was available to all students)
Theoretical basis: encouraging spontaneous access to problem-solving strategies through problem-oriented instruction (Sherwood 1987). Teaching based on Ivey's interviewer training model (Ivey 1994
Duration and timing: 3 hours (2-hour session and 1-hour session 1 week later)
Fidelity: NR
Adherence: NR

Outcomes

Assessment: 20- to 30-minute interviews of real patients drawn from the community with diagnoses of major depression, bipolar disorder, schizophrenia, or similar
Assessment timing: interview conducted 1 to 3 weeks after training
Primary measures: interview evaluation questionnaire - 18 statements addressing 4 basic categories of interviewing activity: focus on non-verbal aspects of the interview, use of questions/communication skills, use of techniques that encourage patient to continue productive talk, exploring patients' feelings. Also rates overall quality of the interview. All rated on a scale of 1 (below expectations) to 10 (truly exceptional)
Other measures: interview difficulty (based on complexity of patient presentation) and student self-reported preparedness for the interview
Measures assessed by: experienced psychiatrists and other clinicians

Notes

Conficts of interest: not stated

Funding: not stated

Abbreviations: NR: not reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	No information provided

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Eells 2002 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to the student's learning condition
Incomplete outcome data (attrition bias) All outcomes	High risk	More than 20% not included in analysis. Reasons provided, but balance between groups unclear. Of 104 participants randomised, 27 were excluded from analysis for the following reasons: 4× more than 1 year MH interviewing experience, 8× interviewed a child, 15× interview supervisors did not return sufficient data
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No information
Other Bias: Baseline characteristics similarity	High risk	No information

Engerer 2019

Study characteristics	
Methods	**Design:** C-RCT **Unit of randomisation:** class/group **Ethics and informed consent:** institutional ethics approval and oral and written consent **Intention-to-treat analysis:** no **Adjustment for clustering:** no
Participants	**Student level:** third year **Setting:** classroom (teaching hospital) **Percentage recruitment:** NR **Number of subjects randomised:** 69 **Number of subjects participating:** - Int: 34 - Ctrl: 32 - Total: 66 **Age (mean):** 21.9 **Sex:** 24.2% male **Ethnicity:** NR **Country:** Germany

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Engerer 2019 *(Continued)*

Interventions	**Number of experimental conditions:** 2 **Aim:** to investigate the effectiveness of 360° behaviour-oriented feedback on students' observable communication behaviours **Content:** three 90-minute modules, each with a different focus: (1) how to begin a consultation and build up a positive relationship with the patient; (2) how to structure a consultation, and (3) how to handle the patient's emotions. After a brief theoretical introduction of 30 minutes, training takes place in small groups (8 to 9 students) with focus on role-play with standardised patients. All students performed at least once in the role of a GP and conducted an initial contact with a standardised patient according to the case vignettes. Intervention arm participants (CST-behave) were provided with specific behavioural feedback. Key communication skills were introduced in a short introduction and focused on skills to initiate conversation, pick up patient's perception, structure conversation, handle patient's emotions, and end the conversation, as well as on general communication skills (e.g. taking pauses, a global item for communicative competence). These key communication skills were summarised in a memory card and were basic for observation and feedback tasks for fellow students **Intervention delivered by:** authors - from Departments of Psychotherapy and Psychosomatic Medicine **Control:** as per CST, behave but with standard feedback **Theoretical basis:** 360° behaviour-oriented feedback (Engerer 2016 **Duration and timing:** 270 minutes (3 × 90 minutes) **Fidelity:** facilitators were involved with development of new training and delivered training to both groups to even out variability in teaching style **Adherence:** NR
Outcomes	**Assessment:** 5-minute simulated consultation **Assessment timing:** NR **Primary measures:** performance of communication skills was assessed by the validated checklist Com-ON-check. This checklist uses a 5-point Likert scale ranging from 0 (poor) to 4 (excellent) for the following domains: global rating of the consultation, start of conversation, picking up patient's perception, structure of conversation, handling patient's emotions, ending the conversation, general communication skills (i.e. clear wording, appropriate non-verbal communication, using suitable pausing, reinforcing questioning, checking patient's understanding) These items were aggregated in the domain "general communication skills". Standardised patients who were appropriately trained also evaluated observed student performance with a single global item using a 10-cm-long visual analogue scale (33) from poor (left side = 0) to excellent (right side = 10) **Other measures:** none **Measures assessed by:** 2 psychologists blinded to group assignment, and SPs
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** this study was part of a research scholarship programme funded by the Medical Faculty of the Technical University of Munich, Germany. The first author, CE, was the recipient of a scholarship to run this study. The funding body had no influence in the design of the study nor in collection, analysis, and interpretation of data **Abbreviations:** CST: communication skills training; Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial

Risk of bias

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Engerer 2019 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Students registered to groups based on day, then coin-toss used to allocate days to experimental or control
Allocation concealment (selection bias)	Low risk	Students and raters were unaware of the allocation process
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors and SPs blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition (< 5%). Only 3/66 did not complete. Group membership of these 3 not reported
Selective reporting (reporting bias)	Low risk	All a priori described scales reported but only for the intervention group due to baseline imbalance of the outcome measure
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	Baseline measures were imbalanced; as a result, researchers did not compare groups
Other Bias: Baseline characteristics similarity	Low risk	Gender, age, and experience of the groups were comparable at baseline
Recuitment Bias (cluster trials only)	Low risk	All students were in groups that were then randomly assigned to intervention or control
Selective recruitment (cluster trials only)	Low risk	No differences between clusters at baseline
Loss of clusters (cluster trials only)	Low risk	No loss of entire cluster
Statistical methods (cluster trials only)	High risk	No adjustment for clustering reported - trial authors did not treat the study as a cluster trial (although reference is made once to cluster-randomisation)
Herd effect (cluster trials only)	Unclear risk	Unable to determine if herd effect exists

Evans 1989

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Evans 1989 *(Continued)*

	Ethics and informed consent: NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourthyear **Setting:** teaching hospital **Percentage recruitment:** 100% **Number of subjects randomised:** 60 **Number of subjects participating:** - Int: 30 - Ctrl: 30 - Total: 60 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** Australia
Interventions	**Number of experimental conditions:** 2 **Aim:** to develop flexibility in history-taking behaviours, given that different interaction skills are required to obtain a full clinical history from different patients **Content:** the course involved a series of 5 lectures and 3 workshops. Lectures covered (1) patient and doctor characteristics affecting the interaction: consultation variables; patient characteristics, uncertainty, and feelings of loss; health understanding; anxiety and expectations; doctor characteristics; reported difficulties in consultations; students' reported difficulties in history-taking; (2) verbal behaviours: verbal communication problems; verbal structure of the consultation; question typology, uses, and problems; stages of the consultation; (3) non-verbal behaviours; (4) listening skills: listening; attending; following; reflecting; (5) summary: student's tasks in the consultation. Students were given comprehensive notes with relevant theory and research along with a series of illustrative case histories. The skills workshops came after the lectures and consisted of small groups of up to 6 students utilising active training formats (such as discussions, role-plays, videotaping with real and simulated patients, interpersonal process recall) to integrate the theory material with their own and other students' experiences with patients, to help develop specific consulting skills and techniques **Intervention delivered by:** NR **Comparison:** traditional medical education **Theoretical basis:** NR **Duration and timing:** 5 lectures (1 hour each) and 3 workshops (2 hours long) **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** videotaped interactions with SPs were analysed by trained raters, using interaction and 16-item scales derived from Verby 1979 **Assessment timing:** within a week of completion of the workshop series

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Evans 1989 *(Continued)*

Primary measures: beginning the interview, seating arrangement, body posture, empathy, warmth, use of facilitation, maintaining relevance, eliciting psychological concerns patient might have, appropriate use of silence, personal and social issues, verbal/non-verbal leads, question style, clarity, eye contact, any interruptions, closure of interview

Other measures: a follow-up study scored students on their 'diagnostic efficiency', which took into consideration all of the above measures, as well as the degree to which students worked towards a correct diagnosis

Measures assessed by: experimentally naive but trained raters

Notes

Conflicts of interest: not stated

Funding: this study was supported by a grant from the Rowden White Fund, University of Melbourne

Abbreviations: Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition (< 5%) - 2 dropouts, not clear which group they were from
Selective reporting (reporting bias)	Low risk	No protocol evident, but reported measures match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred. Lecture notes returned after intervention lectures, so sharing of content less likely
Other Bias: Baseline outcome measurement similarity	Low risk	Pre- and post-training measurements were conducted. Results showed the 2 groups were very similar in their interview behaviour before any training was given to the experimental group
Other Bias: Baseline characteristics similarity	High risk	Not reported

Evans 1996

Study characteristics

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Evans 1996 *(Continued)*

Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** teaching hospital **Percentage recruitment:** post-hoc analysis of data obtained in the 1989 study **Number of subjects randomised:** 30 **Number of subjects participating:** - Int: 15 videotaped interviews from trained students - Ctrl: 15 videotaped interviews from control group students - Total: 30 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** Australia
Interventions	**Number of experimental conditions:** 2 **Aim:** to evaluate the impact of a training programme in communication skills on subdiagnostic efficiency **Content:** students in the trained group participated in a consulting skills training programme consisting of a series of 5 lectures on communication theory and techniques, then 3 small-group skills workshops attended by groups of 5 students **Intervention delivered by:** first author, who was based at Behavioural Medicine Unit at Austin Hospital **Comparison:** traditional medical training **Theoretical basis:** Brockway Rating Scale (Brockway 1978) **Duration and timing:** 11 hours that comprised 5 lectures (1 hour each) and 3 small group workshops (2 hours long) **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** videotaped interviews were rated for diagnostic efficiency by 2 medical practitioners: a psychiatrist and a GP **Assessment timing:** not reported **Primary measures:** the rating was based on interaction scales derived from Verby et al. Students were rated on their introduction to the interview, communication, problem diagnosis, and how they sum-

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Evans 1996 *(Continued)*

	marised the interview. They were then scored on their 'diagnostic efficiency', which took into consideration all of the above measures **Other measures:** time taken for interviews **Measures assessed by:** 2 medical practitioners: a psychiatrist and a GP
Notes	**Conflicts of interest:** not stated **Funding:** this study was funded by a grant from the Medical Defence Union **Abbreviations:** Ctrl: control; GP: general practitioner; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Allocation concealment not reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	No information provided
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: "observer naivete regarding group membership and random ratings of videotapes from the two groups was ensured"
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	Not reported
Other Bias: Baseline characteristics similarity	Low risk	No differences in sex distribution or mean age

Feddock 2009

Study characteristics

Methods	**Design:** RCT **Unit of randomisation:** class/group

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Feddock 2009 *(Continued)*

	Ethics and informed consent: ethical approval was obtained through the university institutional review board. It is unclear if consent was obtained **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** teaching hospital **Percentage recruitment:** 100% **Number of subjects randomised:** 186 **Number of subjects participating:** - Int: 95 - Ctrl: 91 - Total: 186 **Age (mean):** NR **Sex:** 57% males **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' knowledge and clinical skills for primary care **Content:** during the first 3 days of the Primary Care Internal Medicine clerkship, students participate in five 3- to 4-hour clinically based simulated patient (SP) workshops. Workshops begin with a brief introduction by the faculty preceptor. Following the introduction, students participate in 4 role-plays using SPs to represent different clinical scenarios of that domain. One or 2 students interview and counsel each SP in front of the other 4 to 6 students for about 15 minutes each. Following the SP exercises, the workshop co-ordinator leads the students in a discussion of the nuances of each case and the topic in general. Before integration into the clerkship, the workshop co-ordinator developed 4 SP cases, each with a training guide and checklist. SPs introduced adolescent medicine (AM) to students through different characters. Following these SP role-plays, the AM workshop co-ordinator led a discussion of the issues present in each case. Topics included adolescent development, history taking (rapport building, interviewing techniques, confidentiality, and consent for treatment), and the specific counselling issues raised in each case **Intervention delivered by:** workshop co-ordinator **Comparison:** control group students received an alternative workshop **Theoretical basis:** adult learning theory (no reference given) **Duration and timing:** all sessions were 1-off and were 4 hours long **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** on the final day of the clerkship, students completed multi-station SP examination with an open-ended written exercise corresponding to each station, as well as a 100-item multiple-choice examination. The SP completed the yes/no checklist after encountering each student

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Feddock 2009 *(Continued)*

Assessment timing: end of clerkship (i.e. after 48 weeks)

Primary measures: AM checklist: 18 items on the checklist were related to interpersonal and communication skills (e.g. "I was comfortable being interviewed by this student")

Other measures: AM checklist contained 72 total items including 38 items regarding the patient's social history (e.g. asks if patient is sexually active) and 16 counselling items (e.g. discusses the importance of condom use) for a total of 54 items specific to AM

Measures assessed by: SP

Notes

Conflict of interest: not stated

Funding: this study was supported in part by a Predoctoral Primary Care Internal Medicine Training Grant funded by the Health Resources Services Administration

Abbreviations: AM: adolescent medicine; C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; NR: not reported; SP: simulated patients

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Randomisation based on rotation group
Allocation concealment (selection bias)	Unclear risk	Allocation concealment not reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Not adequately reported
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measures reported
Other Bias: Baseline characteristics similarity	High risk	Baseline academic performance was similar in both groups, with a mean (SD) US Medical Licensing Examination step 1 score for the AM workshop. Groups did differ with respect to sex distribution; the AM workshop group had 52% women compared with only 35% in the control group ($X^2 = 5.06$; P = .03)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Filipetto 2006

Study characteristics	
Methods	**Design:** quasi-C-RCT **Unit of randomisation:** time (rotation) **Ethics and informed consent:** unclear re ethics; signed consent required **Intention-to-treat analysis:** unclear **Adjustment for clustering:** no
Participants	**Student level:** first year **Setting:** community health setting **Percentage recruitment:** 76.7% **Number of subjects randomised:** 159 - Int: 68 - Ctrl: 54 - Total: 122 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to assess the effectiveness of a first-year preceptorship programme concerning interpersonal/communication and data collection skills **Content:** the preceptorship consisted of random assignments of students in the class of 2002 to full-time and volunteer faculty within the Department of Family Medicine. Students were required to spend 5 hours a month for 10 consecutive months shadowing their assigned preceptor. Scheduling of this experience was done at the convenience of both the student and the preceptor but could not take place during regular classroom hours. The goal of the preceptorship was to introduce students to family medicine and family medicine role models in an ambulatory care setting, so that students could observe their preceptors and understand the concepts of family medicine. Students were provided with a 1-hour orientation to the preceptorship. During orientation, students were introduced to the goals. They were instructed that the first-year preceptorship experience was observational only, and participation in formal patient care was discouraged. The orientation, the community medicine course, and the remainder of the first-year curriculum contained no formal instruction on interpersonal/communication skills or data collection skills. Assignments for the preceptorship included submission of a patient log, a preceptorship experience summary, and an essay describing the most memorable patient seen **Intervention delivered by:** 64 full-time and volunteer faculty within the Department of Family Medicine **Comparison:** usual curriculum **Theoretical basis:** NR **Duration and timing:** 5 hours per month for 10 months **Fidelity:** no fidelity, but to ensure consistency among all evaluators, all viewed 16 standardised videotaped interviews of students from a neighbouring medical school. Training regarding use of this assessment tool continued until all faculty evaluators achieved 100% agreement **Adherence:** NR
Outcomes	**Assessment:** in their second year, students were assigned to complete a history on 1 of 5 standardised patients. All interviews were videotaped. SPs were trained by an education specialist, and students were given 15 minutes to complete the history **Primary measures:** 72 items from BCI, grouped as interpersonal communication skills and data collection skills, but items not specified **Other measures:** none **Measures assessed by:** faculty in the Department of Family Medicine, blinded as to class membership, assessed the videos **Assessment timing:** second year – up to 12 months after intervention
Notes	**Conficts of interest:** not stated **Funding:** not stated

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Filipetto 2006 *(Continued)*

Abbreviations: BCI: Brown Interview Checklist; Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-random: allocated to groups by year of university enrolment
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: "on completion of videotaped interviews, all faculty evaluators were blinded as to student class membership. Videotapes were randomly assigned to faculty evaluators"
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures seem to match methods
Other Bias: Contamination	Low risk	Same institution, but participants were from 2 consecutive years. Control group (2001) was assessed before study group (2002). Their contamination is unlikely Quote: "the class of 2001 served as the control, because this group did not participate in a first-year preceptorship. The class of 2002 served as the study group with each of the students participating in a required first-year medical school preceptorship"
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measures reported
Other Bias: Baseline characteristics similarity	High risk	No baseline measures reported
Recuitment Bias (cluster trials only)	Unclear risk	Not reported
Selective recruitment (cluster trials only)	Unclear risk	Not reported
Loss of clusters (cluster trials only)	Unclear risk	Not reported
Statistical methods (cluster trials only)	Unclear risk	Not reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Filipetto 2006 *(Continued)*

Herd effect (cluster trials only)	Unclear risk	Not reported

Foster 2016

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** the study protocol was approved by the institutional review board of the Medical College of Georgia. Informed consent was obtained by using the institutional board-approved consent form **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** first year **Setting:** online text-based interface **Percentage recruitment:** NR **Number of subjects randomised:** 70 **Number of subjects participating:** - Intervention 1 (empathy-feedback VP interaction): 35 - Intervention 2 (back story VP interaction): 18 - Comparison group (control VP interaction): 17 - Total: 70 **Age (mean):** 23.43 **Sex:** 52.86% males **Ethnicity:** *total*: American Indian 1.43%, Asian 34.29%, Black 2.86%, White 61.43%; *back story group*: American Indian 5.56%, Asian 27.78%, Black 0, White 66.67%; *control group*: American Indian 0, Asian 41.18%, Black 0, White 58.82%; *empathy feedback group*: American Indian 0, Asian 34.29%, Black 5.71%, White 60% **Country:** USA
Interventions	**Number of experimental conditions:** 3 **Aim:** to enhance verbal empathy in students' encounters with patients **Content:** Intervention group 1: students interacted with the empathy-feedback virtual patient (VP). Empathy-feedback VP was able to give students immediate feedback about empathic communication Intervention group 2: students interacted with the back story VP. Creation of a back story combining embodied conversational agents and narrative video vignettes. When specific questions are asked of the VP, non-interactive video vignettes are presented, showing scenes of the VP illustrating his or her condition. Researchers created the back story VP using the Sims 3 video game. Each video showed

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Foster 2016 *(Continued)*

	medical students how patients' home life can be greatly impacted by their illness and thus simulates patient shadowing **Intervention delivered by:** NA **Comparison:** control VP provided only a typed interaction with Cynthia Young, without empathy feedback or patient back story **Theoretical basis:** NR **Duration and timing:** the intervention was 1-off; intervention length is not reported **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** empathy coding was performed using ECCS, both in the intervention stage of the study, when assessors coded students' empathic responses during the empathy feedback VP interaction, and in the assessment stage, when assessors coded the primary study outcome - students' empathic responses in interactions with real humans (SPs). A study investigator videotaped and transcribed each live SP-student interaction. Another investigator verified random transcripts for accuracy. Investigators extracted all student responses to pre-determined empathic opportunities within de-identified SP interaction transcripts. To ensure that assessors did not identify the source of transcripts, investigators took measures to label the transcripts in each study group (back story VP, empathy feedback VP, control VP) **Assessment timing:** NR **Primary measures:** students' verbal response to all opportunities to show empathy was coded using the Empathic Communication Coding System **Other measures:** empathy scores by type of empathic opportunity were determined as secondary outcomes **Measures assessed by:** PhD student volunteers and study co-investigators
Notes	**Conflicts of interest:** declaration of no conflicts **Funding:** this work was sponsored in part by a research grant from Arnold P Gold Foundation, awarded to Adriana Foster, MD **Abbreviations:** Ctrl: control; ECCS: Empathic Communication Coding System; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated patient; VP: virtual patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Students were anonymous to assessors in the feedback group
Blinding of outcome assessment (detection bias)	Low risk	Assessors and SPs were blind to group assignment

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Foster 2016 *(Continued)*

All outcomes		
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	High risk	Only statistically different items presented
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline outcome measurement
Other Bias: Baseline characteristics similarity	Low risk	No significant differences in student demographics were noted between groups

Gartmeir 2015

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** NR **Setting:** university campus **Percentage recruitment:** NR **Number of subjects randomised:** 168 (72 from medicine and 96 from teaching) **Number of subjects participating:** - Int 1: 20 in combination - Int 2: 17 in e-learning - Int 3: 17 in role-play - Ctrl: 17 - Total: 71 **Age (mean):** NR **Sex:** 32% males in the overall study (medicine plus teaching groups) **Ethnicity:** NR **Country:** Germany
Interventions	**Number of experimental conditions:** 4

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Gartmeir 2015 *(Continued)*

Aim: to prepare medical students and student teachers for leading professional SDM conversations

Content:

Conditions: e-learning with video cases, role-play, combined condition

Deductive vs inductive instruction: e-learning environment consisted of theory-based instruction in a visually enriched and ergonomically segmented way. The role-play condition represented an inductive and experience-based instructional strategy: after a brief theory-based introduction, students could practise communication skills in role-plays

Exercises: e-learning environment contained several exercises (quizzes or sorting tasks) that aimed at elaborating on and applying theoretical and conceptual basics of communication

Video vs peer models: e-learning featured videos of complete conversations displaying competent professional communication behaviours (video models). In contrast, the role-play condition offered extensive opportunity to observe other participants in role-play exercises (peer models)

Systematic vs unsystematic comparisons: learners encountered and analysed successful and suboptimal communicative behaviours in a systematic vs unsystematic way: learners reflected upon suboptimal video sequences, which were contrasted with clips showing improved versions of the same scenes. In the role-play condition, communicative behaviours that occurred during role-play were analysed in training groups based on videotapes. No systematically planned comparisons were made, and all participants highlighted certain aspects of the conversation in the group discussion

Active communication (yes/no) and feedback (yes/no): during e-learning: learners did not have the opportunity to communicate actively in any way and did not receive individualised feedback. The role-play condition provided many opportunities to practise active communication and to give as well as receive individualised feedback to and from peers and trainers

Intervention delivered by: NR

Comparison: waitlist control group

Theoretical basis: professional conversation competence (Gartmeier 2011)

Duration and timing: the intervention was 1-off, and all training conditions were 300 minutes each

Fidelity: NR

Adherence: NR

Outcomes

Assessment: NR

Assessment timing: NR

Primary measures: building interpersonal relationships, explaining symptoms and their causes, planning possible solutions with joint decision-making with patients, situation-specific communication

Other measures: none

Measures assessed by: assessors (not defined any further)

Notes

Conflicts of interest: not stated

Funding: funded by the German Federal Ministry of Education and Research

Abbreviations: Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SDM: shared decision-making

Risk of bias

Bias	Authors' judgement	Support for judgement

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Gartmeir 2015 *(Continued)*

Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	No information provided
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	No information provided
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition (< 5%) Quote: "three cases (1.8 %) had to be removed because their videos could not be analysed due to technical errors"
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures seem to match methods
Other Bias: Contamination	Low risk	Assessment was done immediately after training, so not much time for contamination to occur Quote: "the ASP/P was conducted directly after the training programme with little time for reflection on what had been learnt"
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline outcome measure
Other Bias: Baseline characteristics similarity	Unclear risk	Stated no differences, but details not reported

Gerber 1985

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NA **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** second year **Setting:** unclear **Percentage recruitment:** 41.2% **Number of subjects randomised:** 170 **Number of subjects participating:** - Int: 29 - Ctrl: 41 - Total: 70 **Age (mean):** NR **Sex:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Gerber 1985 *(Continued)*

	Ethnicity: NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' learning of the male genital-rectal exam **Content:** during Teaching Associate Simulated Patient (TASP) instruction, groups of 4 students interacted with 2 non-physician teaching associates, with 1 assuming the role of a patient and the other an examiner. Physical examination was modelled by the TASP examiner on the TASP patient. Both technical and affective skills were emphasised. Then each student performed the examination until competent, receiving immediate feedback from both TASPs. Information was therefore available on procedures that were externally observable and on those based solely upon internal sensations **Intervention delivered by:** TASPs **Comparison:** staff training (urologists instructed and supervised students who performed the examination on each other) **Theoretical basis:** based on programme initiated by Kretzschmar 1978 **Duration and timing:** 90 minutes; 1-off **Fidelity:** inter-TASP reliability was ensured through intensive training with a urologist **Adherence:** NR
Outcomes	**Assessment:** male genital-rectal examination performed on TASP **Assessment timing:** immediately after learning, then 2 years later **Primary measures:** interpersonal skills relating to communication with the patient during the examination (overall explanation of procedures, explanations of specific procedures, appropriate terminology, questions invited, expectations of physical discomfort described, information about findings provided, concern for patient demonstrated, tissues offered) **Other measures:** specific technical skills associated with the examination **Measures assessed by:** TASP
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; TASP: Teaching Associate Simulated Patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	SP assessors were blind to group assignment Quote: "TASPs blinded as to the students method of training"
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Not adequately reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Gerber 1985 *(Continued)*

Selective reporting (reporting bias)	Low risk	All outcomes reported
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	Not reported
Other Bias: Baseline characteristics similarity	High risk	Not reported

Ho 2008

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** the study protocol was approved by a research ethics committee. No information was provided on informed consent **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** fifth year **Setting:** university campus **Percentage recruitment:** NR **Number of subjects randomised:** 57 **Number of subjects participating:** - Int 1: 15 (basic) - Int 2: 15 (extensive) - Ctrl: 27 - Total: 57 **Age (mean):** 25 **Sex:** 80% male **Ethnicity:** 87% Taiwanese **Country:** Taiwan
Interventions	**Number of experimental conditions:** 3 **Aim:** to examine whether a patient-centred cultural competency curriculum integrated into an Asian clinical clerkship could improve cross-cultural communication skills **Content:** students in both intervention groups received workshops on cultural competency. The focus of the first workshop was knowledge and attitudes and included basic concepts such as culture, health

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ho 2008 *(Continued)*

	disparities, and hidden biases. The second workshop introduced video clips showcasing cross-cultural communication skills Intervention 1 (basic): the basic intervention group received the 2 workshops Intervention 2 (extensive): the extensive intervention group received the 2 workshops and an additional 2-hour practice session that focused on role-play using cross-cultural communication skills **Intervention delivered by:** NR **Comparison:** the control group did not receive workshops on cultural competency **Theoretical basis:** NR **Duration and timing:** workshops lasted 2 hours (the extensive intervention group received 2 extra hours). The intervention was a 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** OSCE **Assessment timing:** NR **Primary measures:** eliciting the patient's perspective, social factors relating to illness (eliciting patient's sources of social support, impact of illness on work, affordability of medication, prescription literacy, access to clinics), basic communication skills **Other measures:** history taking, differential diagnosis **Measures assessed by:** SPs
Notes	**Conflicts of interest:** declaration of no conflicts of interests **Funding:** this research project was supported by National Science Council of Taiwan, R.O.C. M Ho was sponsored by the National Taiwan University to attend the Harvard Macy Institute Program for Educators in Health Professions, which shaped the design of this research **Abbreviations:** NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random number generator used to assign students in both blocks to the control group or the intervention group
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention Quote: "at the beginning of the clerkship, we informed all students that half of them would be given additional instruction in cultural competence"
Blinding of outcome assessment (detection bias) All outcomes	Low risk	SP assessors were blind to group assignment, and students were not informed of the methods used to assess cultural competence
Incomplete outcome data (attrition bias)	Low risk	No attrition

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ho 2008 *(Continued)*

All outcomes		
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No information
Other Bias: Baseline characteristics similarity	Low risk	There were no statistically significant differences in student age, gender, or nationality, nor in prior communication skills training between control and intervention groups

Hobgood 2009

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** class/group **Ethics and informed consent:** the study was approved by the institutional review board of the University of North Carolina School of Medicine. Students were invited via signed consent to allow their data to be used for research purposes **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** fourth year **Setting:** university campus **Percentage recruitment:** 100% **Number of subjects randomised:** 138 **Number of subjects participating:** - Int (Grp 1): 35 - Int (Grp 2): 35 - Control (Grp 3): 68 - Total: 138 **Age (mean):** 28.2 **Sex:** 46% male **Ethnicity:** African American 12%, Caucasian 77%, Other 10%, no response 1% **Country:** USA
Interventions	**Number of experimental conditions:** 3 **Aim:** to improve death notification skills of medical students

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Hobgood 2009 *(Continued)*

	Content: the intervention included important actions relating to death notification. These include correctly identifying the deceased and survivors to be notified, explaining the events of the death and any medical procedures that were used to treat injuries, informing survivors directly about the death, giving survivors enough time to ask questions. It also teaches students how to offer survivors viewing of the body shortly after the notification (if it is available), and offer assistance, information, and follow-up for survivors Conditions: GRIEV_ING education with feedback, GRIEV_ING education with no feedback **Intervention delivered by:** NR **Comparison:** the control group received no baseline self-efficacy questionnaire at baseline and had no contact with a simulated survivor before assessment **Theoretical basis:** educational intervention GRIEV_ING (Hobgood 2005) **Duration and timing:** each session lasted 2 hours. The intervention was a 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** simulated survivors assessed students' competence based on whether they achieved each of the 12 items in the GRIEV_ING mnemonic. To assess interpersonal skills, simulated survivors completed an interpersonal communication questionnaire consisting of nine 5-point Likert items. Total score was calculated as the mean of the 9 items. To assess students' confidence, study authors used a self-efficacy questionnaire that expressed confidence on a 5-point Likert scale for 10 items addressed by the GRIEV_ING mnemonic **Assessment timing:** NR **Primary measures:** death notification competence, interpersonal communication **Other measures:** self-confidence **Measures assessed by:** simulated survivors
Notes	**Conflicts of interest:** not stated **Funding:** Dr Tamayo-Sarver was supported by the Agency for Health Care Research and Quality Training Grant **Abbreviations:** C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; NR: not reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"used a two-tiered balanced-block randomization without replacement to assign students to group"
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias)	Low risk	Assessors were blind to group assignment

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Hobgood 2009 *(Continued)*

All outcomes		
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition
Selective reporting (reporting bias)	Low risk	All outcomes were reported and protocol was available
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	No significant differences
Other Bias: Baseline characteristics similarity	Low risk	Differences are controlled for in analysis

Ishikawa 2010

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NA (no data) **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** pre-clerkship - years 1 and 2 **Setting:** unclear **Percentage recruitment:** NR **Number of subjects randomised:** 106 **Number of subjects participating:** - Int: 67 - Ctrl: 39 - Total: 106 **Age (mean):** NR **Sex:** 75% male **Ethnicity:** NR **Country:** Japan
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' non-verbal communication (NVC) skills in medical interviewing **Content:** educational goals of this program were to teach communication skills to build the doctor–patient relationship; to open the discussion; to gather information; to understand the patient's perspective; and to provide closure, based on the essential elements of patient–physician communication. In the training session, each student conducted a 10-minute medical interview with a simulated patient. After the interview, the facilitator invited comments and feedback from the student him/herself, other students in the group, and the simulated patient, and then the discussion was wrapped up, after participants made suggestions for the next student. As it was not practical to cover all educational goals at once in the feedback for each student, the goals were loosely grouped into 4 steps: communication skills for (1) opening and closing; (2) rapport building; (3) information gathering to make diagnosis; and (4) understanding the patient's perspective. As the session progressed, the facilitator gradually moved the focus of the feedback and suggestions from step (1) to step (4). Comments regarding NVC could be included in the original programme, but this was not a major focus

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ishikawa 2010 *(Continued)*

	In the NVC intervention group, the facilitator provided extra feedback regarding NVC skills in addition to the original training programme. NVC skills targeted were those that were found previously to have significant impact on patient evaluation. As in the original programme, the focus of the feedback on NVC progressed in a step-by-step manner from relatively simple to more complex skills, with these skills considered useful in achieving the goals of the original programme: (1) trunk angle and interpersonal distance; (2) self-touching and non-purposive movements, body leaning, nodding, and gazing; (3) facial expressivity and distribution of gaze; and (4) pacing with the patient and matching voice tone and intonation with verbal contents. **Intervention delivered by:** trained facilitators with experience teaching communication skills **Comparison:** usual programme **Theoretical basis:** based on essential elements of patient–physician communication as suggested by the Bayer–Fetzer Kalamazoo Consensus Statement (Makoul 2001b) **Duration and timing:** 3 hours, 1-off **Fidelity:** contents of the original programme were developed and agreed upon through discussion among the 5 facilitators and had been used for the past 2 years. The NVC intervention was developed by the 2 facilitators in the intervention group. One training session per facilitator was monitored by 1 of the study authors to check the differences between original and intervention programmes. Twenty interviews were double-coded by 2 independent coders to ascertain interrater reliability. All NVC skills were covered in NVC groups, whereas not all of them were mentioned, and the number of comments relating to NVC was obviously lower in control groups. Interrater reliability was found to be adequate (kappa 1⁄4 0.73 on average, range 0.56 to 1.00) **Adherence:** NR
Outcomes	**Assessment:** OSCE rated based on video recording **Assessment timing:** end of pre-clerkship year, 1 month after intervention **Primary measures:** non-verbal communication items included: (1) facial expressivity; (2) amount of gaze toward the patient; (3) distribution of gaze (i.e. whether the student looked at the patient primarily when talking or equally when talking and listening); (4) nodding to facilitate patient's talking; (5) self-touching or non-purposive movements; (6) body lean; (7) trunk angle; (8) patient–physician accordance of speech rate and voice volume; and (9) match of voice tone and intonation with verbal contents **Other measures:** student goals for NVC performance **Measures assessed by:** independent coders (qualifications unclear)
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** this study was supported by a Grant-in-Aid for Young Scientists (B) from the Japanese Ministry of Education, Culture, Sports, Science, and Technology **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; NVC: non-verbal communication; OSCE: objective structured clinical examination; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported.
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Participants and SPs were blind. However only the 3 facilitators of control groups were blinded. Intervention group facilitators were not blind due to the nature of the study Quote: "the students and the three facilitators in the control group as well as the simulated patients at the OSCE were blind to this intervention"

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ishikawa 2010 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Not adequately reported
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data appear complete. No withdrawals or loss to follow-up
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures seem to match methods
Other Bias: Contamination	Unclear risk	Inadequate information provided. Possible contamination across medical student groups
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline outcomes scores
Other Bias: Baseline characteristics similarity	Unclear risk	Stratified by gender at randomisation but no reports of other factors that may make the groups different (background, exposure, etc.)

Kahan 2003

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third- and fourth-year students **Setting:** classroom **Percentage recruitment:** NR **Number of subjects randomised:** 76 **Number of subjects participating:** - Int: 38 - Ctrl: 38 - Total: 76 **Age:** 55% aged between 24 and 25 **Sex:** 45% male **Ethnicity:** NR **Country:** Canada
Interventions	**Number of experimental conditions:** 2 **Aim:** the goal of the workshop was to enable students to diagnose an alcohol problem, distinguish problem drinking from alcohol dependence, provide brief advice to the problem drinker, and outline a management plan for the alcohol-dependent patient (abstinence as a treatment goal, treatment of withdrawal, and use of self-help groups and formal treatment) **Content:** the 3-hour alcohol workshop reviewed assessment and management of problem drinking and alcohol dependence. It consisted of a 1-hour presentation, role-plays, and a demonstration of a clinical interview **Intervention delivered by:** NR **Comparison:** depression workshop using a parallel design **Theoretical basis:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Kahan 2003 *(Continued)*

	Duration and timing: 3 hours, 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** 8 OSCE stations (4 alcohol, 4 depression) **Assessment timing:** immediately following the workshop **Primary measures:** a 5-item global rating scale measuring verbal and non-verbal communication, coherence of the interview, empathy, and overall interview performance **Other measures:** each station had an extensive checklist (between 54 and 86 items) of questions asked (assessment items) or advice and recommendations (management items). For analysis purposes, the checklist was divided into appropriate and inappropriate items **Measures assessed by:** SPs and 'examiners' - qualifications not described
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Participants unaware of topic of study BUT no information on those who delivered workshops
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors and SPs were blind to students' workshop assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods. Reported at summary level only
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred, but both groups received parallel workshops
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline OSCE
Other Bias: Baseline characteristics similarity	Low risk	Distribution of males, females, year of study, and age bracket did not differ significantly between groups

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Kaltman 2018

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional approval and written informed consent **Intention-to-treat analysis:** yes **Adjustment for clustering:** NA
Participants	**Student level:** first year **Setting:** university campus **Percentage recruitment:** 51.3% **Number of subjects randomised:** 99 **Number of subjects participating:** - Int: 60 - Ctrl: 39 - Total: 99 **Age (mean):** 24.5 **Sex:** 45.5% male **Ethnicity:** 77.8% white, 16.2% Asian/Pacific Islander, 2% African American, 2% multi-racial, 1% Latino, 1% other **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to assess whether simulations represent a feasible and acceptable strategy for teaching basic communication skills to first-year medical students, and whether use of simulations was associated with improved communication skills. **Content:** E-learning simulations – after an initial open-ended question and response, students choose their next question or response by first typing in their response and then choosing the response that most closely matches from a list of possibilities. The list contains 8 to 10 pre-programmed responses, which include logical questions or responses. In each list, there are both “good” choices and “bad” choices that early learners typically make. When a student makes a choice, another video is triggered of the patient responding to what the student chooses to ask or say. Students can take multiple paths in each simulation. A standardised patient actor filmed the patient response clips in an examination room while talking directly into the camera, so that the patient appears to be talking directly to the student participant when engaging with the simulation. At critical junctures or at the end of a path, a video of a faculty member giving feedback is triggered based on responses that the student makes. Students were given positive feedback when they went down a productive path (“Great job! You responded directly to the patient's emotion, which made him feel comfortable to share his true reason for coming to see you”), as well as constructive feedback when they did not (“Closed-ended questions follow-up on specific details. It's a bit too early in the interview to focus in that narrowly. At this point, it's better to stay open-ended to get a broader and more complete sense of why the patient came into the office today”) **Intervention delivered by:** online

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Kaltman 2018 *(Continued)*

	Control: compulsory physician-patient communication class, which included traditional instruction methods (e.g. lectures, small group discussions and role-play, observed interviews with actual patients) throughout the academic year **Theoretical basis:** patient-centred interviewing (Fortin 2012) **Duration and timing:** unlimited access to 3 simulations over 3 weeks **Fidelity:** NA - computer-based (pre-programmed) **Adherence:** 52 students (86.7%) in the simulation arm interacted with at least 1 simulation. Once students interacted with 1 simulation, they tended to interact with multiple simulations. Thirty-six students (60%) in the simulation arm interacted with all 3 simulations, 11 students (18.3%) interacted with 2 simulations, and 5 students (8.3%) interacted with only 1 simulation
Outcomes	**Assessment:** 2 OSCE stations **Assessment timing:** OSCEs scheduled at the end of the 3-week block in which students accessed the e-simulations **Primary measures:** coded behaviours included open-ended questions, closed-ended questions, reflections, and empathic responses based on specific communication skills that are part of the patient-centred interviewing model **Other measures:** student-reported communication-related self-efficacy **Measures assessed by:** 3 trained raters who were research assistants in health- or communication-related fields
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NA: OSCE: objective structured clinical examination; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Intervention was computer-based, so blinding of personnel was not relevant; lack of blinding of participants not likely to lead to bias
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. Raters were blinded to the study arm of participants when rating OSCE videos
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Kaltman 2018 *(Continued)*

Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	No significant demographic differences were observed between students in the 2 study arms

Klein 2000

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** time (rotation, year) **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** teaching hospital **Percentage recruitment:** 94% **Number of subjects randomised:** 249 **Number of subjects participating:** - Int: NR - Ctrl: NR - Total: 233 **Age (mean):** NR **Sex:** 53% males **Ethnicity:** NR **Country:** UK
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' attitude towards and their communication skills with cancer patients **Content:** the study was embedded within an interview methods course for third-year medical students. It comprised materials to increase students' awareness of the importance of the doctor-patient relationship and to teach basic clinical interview skills. Cancer patients participated in these sessions to help students practice interview skills. Three blocks of 6-weekly sessions were 1.5 hours long **Intervention delivered by:** tutors - one is a psychologist or a psychiatrist, and the other is a physician or a surgeon or a general practitioner **Comparison:** students were taught with patients with diagnoses other than cancer

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Klein 2000 *(Continued)*

	Theoretical basis: NR **Duration and timing:** 3 blocks of 6-weekly sessions taught in an academic year. Each session lasts 1.5 hours **Fidelity:** detailed teaching protocol was provided to all tutors **Adherence:** NR
Outcomes	**Assessment:** the IRI was used to assess students' interview performance. It assessed students' interview performance (beginning, the process, conclusion, overall performance) by utilising a rating manual that described the criteria to be used to rate each item **Assessment timing:** 2 years following the course (when students were in the fifth year) **Primary measures:** the following measures were rated using the IRI: mentions own name and patient's name during the introduction. Makes eye contact. Student defines purpose of interview. Indicates own status. States time available for interview. Shakes patient's hands. Uses appropriate language and avoids jargon. Maintains eye contact. Avoids putting words into patient's mouth. Avoids being tactless. Indicates to the patient when interview is about to end. Thanks patient. Gives appropriate encouragement. Encourages precision. Does not interrupt inappropriately. Asks one question at a time. Uses suitably brief questions. Clarifies what the patient has said. Makes explicit transitional statements. Avoids repetition. Picks up verbal leads. Controls the interview appropriately. Responds empathically. Shows regard and concern for patient. Assesses impact of symptoms on patient's life. Documents history of present illness. Uses closed-ended questions. Completes global assessment of interview performance **Other measures:** the Attitude Questionnaire was used to assess the attitudes of students towards aspects of treatment and management of cancer patients **Measures assessed by:** simulated patients
Notes	**Conflicts of interest:** not stated **Funding:** from the 'Cancer Research Campaign' **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; IRI: Interview Rating Instrument; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Compiled alphabetically by the Faculty Office at the beginning of the academic year and used to obtain an approximation to random distribution of the 249 students
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	No information but unlikely, particularly given the time lag from intervention to final assessment
Blinding of outcome assessment (detection bias) All outcomes	High risk	One rater was `blind' to the allocation of students to 2 groups (2 raters independently rated the recordings)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Klein 2000 *(Continued)*

Incomplete outcome data (attrition bias) All outcomes	High risk	Study design Figure 1 states third year N = 249. However, post-course measures N = 233 (94% response). In 1 of the 2 groups of third-year students, 89/119 were eligible for fifth year evaluation "The remaining 30 students were excluded from the study because 11 students were completing an intercalated B.Sc. degree; 16 were required to repeat their fourth year, and three students were non-graduating" Of the 89 students eligible to participate in the follow-up evaluation of this study, 54 fifth year students agreed to take part (61% response rate).
Selective reporting (reporting bias)	High risk	Not reported by group. All items of attitude questionnaire reported. Not all items from Interview questionnaire reported "The items with the lowest level of agreement were item 29 (kappa value = 0.33) and item 30 (kappa value = 0.36). Thus, both of these items were excluded from the follow-up analysis"
Other Bias: Contamination	Unclear risk	Inadequate information provided. Possible contamination across medical student groups
Other Bias: Baseline outcome measurement similarity	Low risk	Pre-course attitudes measured "Attitudes before the course were similar: none of the items on the AQ showed a significant difference between the two groups (data not shown)"
Other Bias: Baseline characteristics similarity	Low risk	No significant differences between groups in terms of gender or native language

Knowles 2001

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** class/group **Ethics and informed consent:** unclear **Intention-to-treat analysis:** yes **Adjustment for clustering:** NA, but unit of analysis error may be present
Participants	**Student level:** fourth year **Setting:** teaching hospital **Percentage recruitment:** 93.6% **Number of subjects randomised:** 133 **Number of subjects participating:** - Int: 40 - Ctrl: 92 - Total: 132 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** UK
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' communication skills and application of knowledge in GU medicine; the objective was to test the hypothesis that adding video feedback to traditional methods of undergraduate teaching would achieve a better outcome in terms of acquiring and retaining communication skills and knowledge than traditional methods alone

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Knowles 2001 *(Continued)*

	Content: students were filmed in playing the role of a doctor interviewing a patient (played by an actor) presenting with a GU problem. Structured feedback by a GU physician and an educational psychologist was given a week later **Intervention delivered by:** simulated patients and physicians **Comparison:** usual training without video feedback **Theoretical basis:** NR **Duration and timing:** 2 contacts 1 week apart **Fidelity:** not fidelity, but evaluators viewed 16 standardised videotaped interviews of students from a neighbouring medical school. Training regarding use of this assessment tool continued until all faculty evaluators achieved 100% agreement. To ensure consistency among all evaluators, faculty evaluators viewed 16 standardised videotaped interviews of students from a neighbouring medical school **Adherence:** Ctrl: 85 students attended all or > 85% of teaching sessions - i.e. 7 protocol violations Int: 37 students attended all or > 85% of teaching sessions - i.e. 3 protocol violations
Outcomes	**Assessment:** 6-minute OSCE station **Assessment timing:** 1 week after intervention **Primary measures:** checklist asked assessors to score students from 0 to 16 on performance criteria (professional greeting and thorough information gathering) and interviewing skills (the manner in which information is gathered) **Other measures:** GU knowledge **Measures assessed by:** 'examiners' (qualifications not described)
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** not stated **Abbreviations:** Ctrl: control; GU: genitourinary; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Students were assigned to their groups at the start of the academic year by the central registry in a random fashion. The method of assignment was to have every third group (of 12 to 15 students) allocated to the intervention group. Every other 2 groups out of 3 acted as a control group
Allocation concealment (selection bias)	Low risk	No individual taking part in this study was associated with the assignment process of students to their groups
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Not adequately reported
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: "examiners were blind to which groups were in the intervention group or control group"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition and balanced between groups. 7.5% in each group did not attend all teaching sessions. All attended OSCE
Selective reporting (reporting bias)	Low risk	All outcomes reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Knowles 2001 *(Continued)*

Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline outcomes scores
Other Bias: Baseline characteristics similarity	Low risk	No significant differences in age and sex between intervention and control groups (figures not reported)

Lee 2015

Study characteristics	
Methods	**Design:** C-RCT **Unit of randomisation:** time (rotation, year) **Ethics and informed consent:** Tufts University Health Sciences Campus Institutional Review Board **Intention-to-treat analysis:** NR **Adjustment for clustering:** no
Participants	**Student level:** third year **Setting:** unclear **Percentage recruitment:** 72% **Number of subjects randomised:** 125 **Number of subjects participating:** - Int: 60 - Ctrl: 59 - Total: 119 **Age (mean):** 27.14 (intervention) and 27.49 (control) **Sex:** 56% males (intervention) and 53% males (control) **Ethnicity:** white 69% (intervention) and 75% (control), African American 6% (intervention) and 3% (control), South Asian 9% (intervention) and 10% (control), East Asian 15% (intervention) and 12% (control) **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** use of an online cross-cultural communication module to facilitate cross-cultural communication and increase students' understanding of patients' perspectives improve clinical outcomes **Content:** the teaching module comprised cross-cultural communication and health disparities. Students practiced "Problem-Affect-Concern-Treatment" (PACT) questions with their patients and wrote about these experiences. They asked patients the following questions: (1) What do you think the Problem is? (2) How is this problem Affecting the rest of your life? (3) What Concerns you the most about this problem? (4) Do you feel this Treatment plan will help you? **Intervention delivered by:** NA

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lee 2015 *(Continued)*

	Comparison: standard curriculum **Theoretical basis:** Arthur Kleinman's explanatory models approach (Kleinman 2006) **Duration and timing:** 1-hour-long online teaching modules **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** final week of Family Medicine clerkship **Assessment timing:** NR **Primary measures:** PACT questions that included the following: (1) What do you think the Problem is? (2) How is this problem Affecting the rest of your life? (3) What Concerns you the most about this problem? (4) Do you feel this Treatment plan will help you? **Other measures:** NR **Measures assessed by:** faculty members who were blinded to randomisation
Notes	**Conflicts of interest:** not stated **Funding:** Tufts Health Care Institute provided funding to support development of the online learning module discussed in this paper. The collaborative evaluation was partially funded through an HRSA grant **Abbreviations:** C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; PACT: Problem-Affect-Concern-Treatment

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Block randomisation, but method of randomisation unclear Quote: "randomised by assigning numbers to each student's name before allocating their numbers to either group"
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment "Faculty members who were blinded to group allocation, assessed/scored participants"
Incomplete outcome data (attrition bias) All outcomes	Low risk	6/125 had incomplete data (< 5% attrition), group membership unclear
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures reported seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lee 2015 *(Continued)*

Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline outcomes scores
Other Bias: Baseline characteristics similarity	Low risk	No significant differences in age and sex between intervention and control groups (figures not reported)
Recuitment Bias (cluster trials only)	Low risk	All students were in groups that were randomly assigned to intervention or control
Selective recruitment (cluster trials only)	Unclear risk	Block-level baseline data not provided
Loss of clusters (cluster trials only)	Low risk	No loss of entire clusters
Statistical methods (cluster trials only)	High risk	No adjustment for clustering reported - trial authors did not treat the study as a cluster trial (although reference is made once to cluster-randomisation)
Herd effect (cluster trials only)	Unclear risk	Unable to determine if herd effect exists

Legg 2005

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** students **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** sixth year **Setting:** teaching hospital **Percentage recruitment:** NR **Number of subjects randomised:** 24 **Number of subjects participating:** - Int: 11 - Ctrl: 10 - Total: 21 **Age (mean):** 25 years and 4 months **Sex:** 38.10% males **Ethnicity:** White 38%, Black 43%, Indian 14%, Colored 5% **Country:** South Africa

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Legg 2005 *(Continued)*

Interventions	**Number of experimental conditions:** 2 **Aim:** to improve medical students' ability to communicate with adults with aphasia **Content:** the workshop introduces students to supported communication techniques and provides students an opportunity to practice the techniques through activities such as discussion, role-play, and reflection. Students were shown video sequences of medical professionals interviewing an adult with aphasia, before and after training in Supported Conversations for Adults with Aphasia. The video sequences provided students with discussion points on behaviours that hindered and facilitated communication between doctors and adults with aphasia. This was followed by students role-playing a case interview, followed by discussion on students' thoughts and feelings during role-play. Students considered how they could maximise communication with aphasic patients and were introduced to the Supported Conversations for Adults with Aphasia techniques. This was followed by another role-play; they were then provided with informal feedback. To guide students in the future use of Supported Conversations for Adults with Aphasia, they were provided with a booklet with pictographic resources relating to the history-taking interview **Intervention delivered by:** NR **Comparison:** theoretical information about aphasia. **Theoretical basis:** Supported Conversations for Adults with Aphasia (Kagan 1998) **Duration and timing:** 1-off intervention with 4-hours-long training **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** NR **Assessment timing:** pre- and post-training (timings not reported) **Primary measures:** establishing and developing rapport; understanding patients' perspectives; involving patients in the interview; exploring problems; identifying patients' reasons for consultation; structuring the consultation. Patients' overall perceptions of the interview; patients' ability to provide information; patients' satisfaction with the encounter **Other measures:** students' overall perception of the interview; their ability to obtain relevant information; quality of the interaction; students' abilities to acknowledge and reveal aphasic patients' competence; students verification of patients' responses **Measures assessed by:** speech pathologists
Notes	**Conflicts of interest:** not stated **Funding:** University of Cape Town Medical Faculty **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"randomised by assigning numbers to each student's name before allocating their numbers to either group"
Allocation concealment (selection bias)	Unclear risk	Not adequately reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Legg 2005 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	Low risk	Participants (students, patients, and raters) blinded
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: "blinding of patients and raters served to reduce bias"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition (3/24)
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	The control group completed both first and second interviews before the experimental group conducted their first interview to ensure that there was no "spill over" of information from training sessions of the experimental group
Other Bias: Baseline outcome measurement similarity	Low risk	No significant differences between experimental and control groups' pre-training performances measured by (M)SCA. MCCOG baseline had significant differences in "providing structure to the consultation and developing rapport" However as these differences favoured the control group, this was considered acceptable
Other Bias: Baseline characteristics similarity	Low risk	No significant baseline differences between the 2 groups

Levenkron 1990

Study characteristics	
Methods	**Design:** quasi-C-RCT **Unit of randomisation:** time (year) **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** no
Participants	**Student level:** second year **Setting:** university campus **Percentage recruitment:** 100% **Number of subjects randomised:** 183 **Number of subjects participating:** - Int: 91 - Ctrl: 92

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Levenkron 1990 *(Continued)*

	- Total: 183 **Age (mean):** not reported **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to use PI to teach students counselling skills for promoting risk factor modification and lifestyle behaviour change **Content:** as part of the unit, "Interviewing the Healthy Patient", students in group 2 received a homework reading assignment on the role of the physician in conducting clinical health promotion, as well as a how-to guide on risk factor-related behaviour change. Each student then performed the PI exercise as the primary source of instruction instead of attending a lecture and small group discussion. The PI exercise required 2 hours of student contact time divided and comprised student orientation, Pl interview, tutorial feedback, and debriefing **Intervention delivered by:** PI **Comparison:** class lecture and small discussions instead of individual PI exercise **Theoretical basis:** NR **Duration and timing:** Two x two hour sessions **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** in the subsequent fall semester for each study group **Assessment timing:** 6 months **Primary measures:** University of Rochester Risk Factor Interview Scale (URRFIS) with the following subscales: relationship building, general interview technique, positive focus, instigating change **Other measures:** NR **Measures assessed by:** PI
Notes	**Conflicts of interest:** not stated **Funding:** Charles A. Dana Foundation, Inc., and the National Heart, Lung, and Blood Institute **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; PI: patient Instructor; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-random. Allocated to groups by year of university enrolment
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Levenkron 1990 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	No information provided
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	No information provided
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Control group was assessed before the intervention group; therefore contamination bias is unlikely. Consecutive academic years - 1985 traditional teaching (control group) and 1986 intervention (group)
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	No baseline measure
Recuitment Bias (cluster trials only)	Low risk	Consecutive cohorts of students
Selective recruitment (cluster trials only)	Low risk	Consecutive cohorts of students BUT quasi-randomised
Loss of clusters (cluster trials only)	Low risk	No loss of entire cluster
Statistical methods (cluster trials only)	Unclear risk	Adjustment for clustering not reported
Herd effect (cluster trials only)	Unclear risk	Unable to determine if herd effect exists

Lim 2011

Study characteristics	
Methods	**Design:** quasi-C-RCT **Unit of randomisation:** year **Ethics and informed consent:** institutional approval and informed consent **Intention-to-treat analysis:** NR, but any missing data were automatically adjusted for in each analysis **Adjustment for clustering:** no
Participants	**Student level:** fifth year **Setting:** unclear **Percentage recruitment:** 97% **Number of subjects randomised:** 153

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lim 2011 *(Continued)*

	Number of subjects participating: - Int: 77 - Ctrl: 72 - Total: 149 **Age (mean):** NR **Sex:** 46% male **Ethnicity:** NR **Country:** New Zealand
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' competence in medical consultation skills, particularly for promoting risk factor modification and lifestyle behaviour change **Content:** the Psychological Medicine module runs 6 times a year for successive small groups of 12 to 14 students. The module includes a seminar about brief intervention, motivational interviewing with examples of how doctors and patients can use motivational interviewing, brief intervention to discuss substance use/abuse issues, and a role-play workshop. The role-play workshop was conducted using a pre-set teaching plan and pre-tested case scenarios. The 2010 intervention group also had a 1-hour actor-facilitated teaching innovation on ‘‘How to act-in-role’’ - a brief introduction to some acting skills and methods that focused on enhancing participants’ capacity to connect with their patients, listen to what they are saying, observe their body language, pick up interpersonal cues, and improve their interpersonal and interactive skills. During the course of the workshop, participants used these skills in a series of role-playing in 5 training scenarios, during which time they would take the role of both patient and clinician **Intervention delivered by:** 2 from a team of 4 addiction health practitioners who are also experienced tutors in brief intervention, and actor facilitator for intervention group **Comparison:** psychological medicine module without "How to act-in-role" component **Theoretical basis:** Jefferson Scale of Physician Empathy (JSPE) (Hojat 2001) **Duration and timing:** 1 hour, 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** OSCE **Assessment timing:** end of module (time unclear) **Primary measures:** BECCI score; 11 items rated on a 5-point Likert scale (0, not at all; 4, to a great extent) used to assess medical students’ competence in consultations about behaviour change; OSCE performance as measure of competence in clinical communication **Other measures:** Jefferson Scale of Physician Empathy (JSPE), a 20-item self-report scale to measure empathy specifically among medical students, residents, and clinicians **Measures assessed by:** tutor (OSCE rating) and Independent rater (BECCI score)
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** funding from the Committee for the Advancement of Learning and Teaching (CALT) **Abbreviations:** BECCI: Behaviour Change Counselling Index; Ctrl: control; Int: intervention; JSPE: Jefferson Scale of Physician Empathy; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-random. Allocated to groups by year of university enrolment
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lim 2011 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	Low risk	It was not possible to blind tutors to knowledge of cohort exposure to the intervention, but students themselves were unaware of which educational exposure they had received. Differences in student self-rating between control and intervention cohorts and correlations between tutors' and students' self-ratings provide some reassurance that tutor awareness was not an important bias
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: "OSCE marked by independent examiner who was blinded. Blinding of the tutors and student assessors to the students who were or were not participants. This was important to ensure integrity of the research and prevent assessment bias"
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition. Any missing data were automatically adjusted for in each analysis
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods. Reported as overall scores
Other Bias: Contamination	Low risk	Control group was assessed before intervention group; therefore contamination bias is unlikely Quote: "year 2009 students served as the control group while the year 2010 students were the intervention group in this study"
Other Bias: Baseline outcome measurement similarity	Low risk	Students in both groups did not significantly differ in baseline empathy scores
Other Bias: Baseline characteristics similarity	Unclear risk	Appear to be more females in the intervention group, but significance of difference is not reported. No other characteristics are reported
Recuitment Bias (cluster trials only)	Low risk	Students were in groups, and groups were allocated to intervention or control group by year of enrolment
Selective recruitment (cluster trials only)	Low risk	Empathy scores equivalent at baseline
Loss of clusters (cluster trials only)	Low risk	No loss of entire clusters
Statistical methods (cluster trials only)	High risk	Trial authors do not treat the study as a C-RCT, and therefore do not adjust for clustering
Herd effect (cluster trials only)	Unclear risk	Unable to determine if herd effect exists

Liu 2016

Study characteristics	
Methods	**Design:** RCT (cross-over study) **Unit of randomisation:** student

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Liu 2016 *(Continued)*

	Ethics and informed consent: the study was approved by the University of New South Wales Research Ethics Committee. Students signed an online consent form **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** second year **Setting:** university campus **Percentage recruitment:** 62% **Number of subjects randomised:** 268 **Number of subjects participating:** - Int: 59 - Ctrl: 107 - Total: 166 **Age (mean):** NR **Sex:** Int 36% male; Ctrl 48% male **Ethnicity:** NR **Country:** Australia
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve clinical communication skills of medical students **Content:** the intervention group completed the EQClinic during weeks 1 to 5. The telehealth communication skills training platform (EQClinic) comprises 5 components: an online training component, a personal calendar, a real-time interaction component, a non-verbal behaviour detector, and a feedback generator. The training component provides videos and documents that help students and SPs familiarise themselves with the platform. The personal calendar component allows students and SPs to make bookings. Once appointments are confirmed by the automated messaging system, videoconferencing enables a student and an SP to have a teleconference. Online assessments were available for students. SPs evaluated students' performance immediately after the teleconsultation, and students conducted a self-assessment using the same form. The EQClinic then automatically analysed the video recordings by utilising audio processing and computer vision techniques. This detects non-verbal behaviours such as head movements, facial expressions, body movements, voice properties, and speech patterns. The feedback includes computer-generated non-verbal behaviour feedback as well as comment feedback from SPs **Intervention delivered by:** computer-based platform **Control:** students completed EQClinic training during weeks 8 to 11 **Theoretical basis:** NR **Duration and timing:** the intervention was a 1-off.**I**nterview and assessment components were 40 minutes for students, and 25 to 30 minutes for SPs **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Liu 2016 *(Continued)*

	Assessment timing: NR **Primary measures:** Student-Patient Observed Communication Asessment (SOCA) including **r**apport building, information gathering, providing structure to the interview, understanding the patient's needs, and total score **Other measures:** for the purpose of comparison in this review, only data from week 6 to 7 was used here, enabling comparison of the intervention with usual curriculum. The study was a cross-over trial, so further data are available to compare the 2 arms **Measures assessed by:** SP
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** this project was funded by the Brain and Mind Centre at The University of Sydney Australia and the Australian Government. RAC is supported by the Australian Research Council **Abbreviations:** NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated random number sequence used
Allocation concealment (selection bias)	Low risk	No information on timing, but all participants ultimately had the same opportunities due to cross-over design, so any knowledge of allocation is unlikely to impact outcomes
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding not possible, but due to cross-over design, lack of blinding unlikely to impact outcomes
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: "tutors assessing students were blinded to condition allocation"
Incomplete outcome data (attrition bias) All outcomes	High risk	High attrition and appears to be imbalanced between groups Quote: "a third limitation is the relatively low proportion of students (30%) who completed all components of the study. While the sample was still appropriate for the statistical tests conducted, future investigations will benefit from exploring the reason"
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lorin 2006

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** student, then rotation **Ethics and informed consent:** institutional approval and informed consent **Intention-to-treat analysis:** NR, but no apparent attrition **Adjustment for clustering:** NA
Participants	**Student level:** fourth year **Setting:** classroom (teaching hospital) **Percentage recruitment:** 97% **Number of subjects randomised:** 109 **Number of subjects participating:** - Int: 53 - Ctrl: 53 - Total: 106 **Age (mean):** **Sex: 51**% male **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve fourth-year medical students' ICU communication skills **Content:** The intervention group participated in a formal teaching session held with an actor and experienced critical care educators in the medical ICU conference room, where family meetings regularly occur. The intervention group received a lecture on a framework for communication, discussed guidelines on how to use this framework, then practised applying this with an actor. The role-playing session focused on practising the guidelines and consisted of interventions and teaching points regarding communication skills. The setting and the actor's responses were frequently manipulated to evaluate the student's reactions, and the session could be placed on "pause" for questions by students **Intervention delivered by:** experienced critical care educators **Control:** the control group of students did not receive a teaching session about a clinician-family communication **Theoretical basis:** NR **Duration and timing:** 90 minutes, 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** video-recorded OSCE **Assessment timing:** last week of 4-week rotation **Primary measures:** each videotaped session was graded in 4 domains: (a) presenting introduction; (b) gathering information; (c) imparting information; and (d) setting goals and expectations. Point scores were determined for each individual question, a summation of each of the 4 sections, and a total score that varied from 0 to 24, with higher scores indicating better performance **Other measures:** none **Measures assessed by:** an investigator blinded to students' intervention assignment
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** not stated **Abbreviations:** Ctrl: control; ICU: intensive care unit; Int: intervention; NR: not reported; OSCE: objective structures clinical examination

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lorin 2006 *(Continued)*

Random sequence generation (selection bias)	High risk	Quasi-RCT. Alternate blocks allocated to intervention and control groups
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. Each videotaped session was scored by an investigator blinded to students' intervention assignments
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match method. Reported at summary level only
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred between rotations
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	No significant differences in group characteristics

LoSasso 2017

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional review board approval and informed consent by interview **Intention-to-treat analysis:** unclear **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** classroom (teaching hospital) **Percentage recruitment:** NR **Number of subjects randomised:** 70 **Number of subjects participating:** - Int: 38 - Ctrl: 32

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

LoSasso 2017 *(Continued)*

	- Total: 70 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to examine whether a specific intervention to teach proper use of the EMR in a clinical setting could help improve students' empathic engagement **Content:** the intervention group underwent an additional 1-hour training session on EMR-specific communication skills on their first day of orientation. That intervention session began with a discussion of the positives and negatives of EMR use, illustrated through 3 video clips in which the EMR is used poorly, moderately, and well. Then, the SALTED mnemonic - Set-up, Ask, Listen, Type, Exceptions, and Documentation - was introduced, and the technique it teaches for proper use of EMRs in clinical settings was discussed. Students then practised applying the SALTED technique in 4 brief role-playing scenarios **Intervention delivered by:** NR **Control:** regular clinic training on EMR use **Theoretical basis:** SALTED mnemonic - Set-up, Ask, Listen, Type, Exceptions, and Documentation **Duration and timing:** 1-off 1-hour session **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** 15-minute recorded SP encounter **Assessment timing:** 6 weeks after intervention (end of clerkship) **Primary measures:** 6-item version of the JSPPPE. Items were answered on a 5-point Likert-type scale (poor = 1, fair = 2, good = 3, very good = 4, excellent = 5), with higher score indicating more empathic engagement (possible range of scores: 1 to 5). An example item is, "Did the student seem concerned about me and my family?" The communication skills scale consisted of 10 items related to communication and interpersonal skills answered on a 5-point Likert-type scale (poor = 1, fair = 2, good = 3, very good = 4, excellent = 5). A sample item is "Clarity of Questions: asked clear questions, 1 question at a time. Avoided leading questions." The history-taking skills scale consisted of 3 items relating to the SALTED technique for EMR use answered on a similar 5-point Likert-type scale. An example item is, "How well did the student set up the room with the computer–patient–physician triangle in mind?" **Other measures:** student self-reported empathy using JSE **Measures assessed by:** trained faculty members and SPs
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** study authors report that no funding was received **Abbreviations:** Ctrl: control; EMR: electronic medical record; Int: intervention; JSE: Jefferson Scale of Empathy; JSPPPE: Jefferson Scale of Patient Perceptions of Physician Empathy; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated patient
Risk of bias	

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

LoSasso 2017 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors and SPs were blind to group assignment "The SP and faculty raters were blinded to whether students were in the intervention or control group"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Appears to be no attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match method. Scales reported at group level
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	No significant differences were found between intervention and control groups on gender, Medical College Admission Test (MCAT) scores, and United States Medical Licensing Examination Step 1 scores

Lupi 2012

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional review board approval; implied consent **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** skills laboratory **Percentage recruitment:** 100% **Number of subjects randomised:** 106 **Number of subjects participating:** - Int: 61 (50 analysed) - Ctrl: 46 (45 analysed) - Total: 106 **Age (mean):** NR **Sex:** 59% male

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lupi 2012 *(Continued)*

	Ethnicity: NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to provide students the opportunity to discuss and practice options counselling, and to learn a systematic approach to the related ethical issue of conscientious refusal **Content:** the workshop consisted of 3 parts designed to first illustrate the connection between ethical lapses and communication skills, then facilitate private values clarification, and finally provide opportunity for practice of communication skills. Trigger skits are enacted by faculty or shown on video; the first involves a doctor who refuses to discuss the option of abortion with a young mother faced with unplanned pregnancy, the second involves a resident who tries to influence a pregnant teenager against pregnancy continuation. Discussion focuses on the impact of the moral bias on communication skills. Students consider their own responses to other moral dilemmas and role-play these in trios **Intervention delivered by:** tutors (qualifications not described) **Control:** no workshop before SPE/OSCE **Theoretical basis:** learning objectives for the ethics section were drawn from the American College of Obstetricians and Gynaecologists Ethics Committee opinion on conscientious refusal in reproductive medicine **Duration and timing:** 1.5 hours, 1-off **Fidelity:** NR **Adherence:** all students took part – a required but ungraded course element
Outcomes	**Assessment:** blinded assessor rating of recorded OSCE **Assessment timing:** unclear (after workshop) **Primary measures:** effect of workshop on communication skills: student communicated verbal empathy and support, student communicated non-verbal empathy and support, overall rating of student's communication skills. How would you rate your satisfaction with this encounter with the student? How would you rate the student's ability to develop rapport with the patient? **Other measures:** none **Measures assessed by:** assessors (qualifications not detailed)
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** this study was supported by a grant from the Arsht Ethics and Community Program of the University of Miami and with resources from the Association of Professors of Gynecology and Obstetrics/Solvay Educational Scholars and Leaders Program **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; SP: simulated/standardised patient; SPE: standardised patient experience; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Allocated to groups by computer-generated random numbers table
Allocation concealment (selection bias)	Unclear risk	Not adequately reported. No information on timing of allocation nor methods of concealment
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors (reviewer, observer, and data analyst) were blind to group assignment

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Lupi 2012 *(Continued)*

Incomplete outcome data (attrition bias) All outcomes	Unclear risk	There appears to be differential attrition between groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Control group was assessed before intervention group; therefore contamination bias is unlikely Quote: "student's participated in the standardised patient exercise the morning before the workshop (control group), or 1 week after the workshop" (intervention group)
Other Bias: Baseline outcome measurement similarity	Unclear risk	Differences between groups at baseline not reported
Other Bias: Baseline characteristics similarity	Low risk	No differences in characteristics at baseline

Maguire 1977

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** groups of students **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** not reported. **Setting:** teaching hospital **Percentage recruitment:** **Number of subjects randomised:** 30 (each group contained 10 participants) **Number of subjects participating:** - Int 1: 10 - Int 2: 10 - Ctrl: 10 - Total: 30 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** England

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Maguire 1977 *(Continued)*

Interventions	**Number of experimental conditions:** 3 **Aim:** to utilise television demonstrations of history-taking, and to practise with simulated patients to equip medical students with essential history-taking skills **Content:** didactic mode:**i**n the first seminar, students were given a printed handout, which described the questions to be asked when taking a history and the sequence to be followed. They were then shown a videotape of an interview conducted by the teacher, which demonstrated these points. This was followed by a videotape of a senior medical student's interview, chosen because it exemplified important and common omissions. Students were then invited to point out what these were. In the second week, each student was asked to interview a simulated patient for 12 minutes to give them practice in obtaining a history of the patient's present problems. They were told that these interviews would be videotaped for assessment purposes, but they were not informed that patients were simulated. The final seminar concentrated on teaching the students how to interview. It began with the presentation of a second handout, which described how to begin an interview, obtain accurate and relevant data, and end within the agreed time limit. This was followed by a videotape, in which the 'teacher' interview used in the first seminar was re-shown, but with the addition of introductory and termination sequences. The medical student tape was then re-shown, with the same additional sequences. Students were asked to compare the student's performance with the teacher's, and the complete history-taking procedure was then reviewed. Although some discussion was encouraged in this mode, it was stressed that the recommended procedure must be followed if interviews were to be effective Discovery mode: the teacher, the timing, and the length of the 2 seminars and practice session were the same as in the didactic mode; however, the first seminar began with a request to students to suggest the questions that should be asked when a history of the present illness is obtained. Once they had formulated a sequence of questions, the first 'medical student-what questions to ask' tape was shown, and the group was invited to compare it with their model. The first 'teacher tape' followed, and further discussion was actively encouraged. After this, the first handout was presented, and students were asked to comment on it in the light of their own model. In the second week, students each interviewed a simulated patient under the same conditions as in the didactic mode. In the second seminar, the same sequence was followed as in the first, but while using the 'how to conduct an interview' version of the medical student and teacher tapes. Each of these tapes was stopped at 3 points: after the introduction; during the main body of the interview; and after the final sequence to provoke discussion. The second handout was then distributed and compared with the model that the students had derived. In this mode, the model was offered as just 1 possible procedure that could be followed **Intervention delivered by:** faculty **Comparison:** traditional mode - teacher demonstration and student practice with discussion **Theoretical basis:** NR **Duration and timing:** one 90-minute session per week for 3 weeks **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** NR **Assessment timing:** 1 week following all 3 courses **Primary measures:** History-taking techniques used assessed on 5-point likert scales indicating the extent to which students use desired or undesired techniches (e.g. ' picking up verbal leads', 'using medical jargon'). Overall score for techniques used is reported. **Other measures:** Ratings of the videotape recordings in terms of the amount of accurate and relevant information obtained. **Measures assessed by:** simulated patients
Notes	**Conflicts of interest:** not stated

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Maguire 1977 *(Continued)*

Funding: British Medical Association and British Life Assurance Trust

Abbreviations: Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Randomised alphabetically into 5 groups, then 3 groups were randomly assigned. Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear whether assessors were blinded to group allocation
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	No information regarding attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No pre-test measure for experimental groups
Other Bias: Baseline characteristics similarity	High risk	Not reported

Maguire 1978

Study characteristics

Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** teaching hospital

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Maguire 1978 *(Continued)*

Percentage recruitment: total number of students not reported

Number of subjects randomised: 48

Number of subjects participating:

- Int 1 (video replay feedback): 12

- Int 2 (audio replay feedback): 12

- Int 3 (rating scale feedback): 12

- Ctrl: 12

- Total: 48

Age (mean): NR

Sex: NR

Ethnicity: NR

Country: England

Interventions	**Number of experimental conditions:** 4 **Aim:** to equip medical students' with essential history-taking skills **Content:** all students received training in history taking, using traditional methods. This required students to interview patients followed by discussion of patients' histories during seminars or Ward Rounds. Intervention group students then received feedback from tutors by television replays, audio replays, or rating of students' practice interviews Intervention group 1 (video feedback): students in the first group watched a video replay of their interviews with tutors. Tutors discussed students' performance in relation to the standard set in a previously discussed handout and points relevant to psychiatric history taking Intervention group 2 (audio feedback): in the second group, feedback was given in sound only Intervention group 3 (rating scale feedback): in the third group, students received feedback unaided by either video or audiotape. However, the tutor did watch the video replay before meeting with each student. Following this, tutors met with students, who were asked to reflect on their interviews Students in the feedback groups were seen by the same tutor. Feedback students conducted practice interviews 1 and 2 weeks after the first interview **Intervention delivered by:** tutors **Comparison:** traditional apprenticeship method **Theoretical basis:** NR **Duration and timing:** 1-off intervention. 45 minutes for each feedback group **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** NR **Assessment timing:** 3 weeks, and follow-up 5 years later **Primary measures:** 24 items of behaviour (contained in 3 subsections, namely, greeting and seating, self-introduction, and orienting the patient) and 10 items on closing the interview (in 3 subsections, namely, summarising, checking accuracy, and concluding) were assessed on a 2-point scale (0 to 1) according to whether each one (such as a doctor introducing himself) was present. The 11 main inter-

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Maguire 1978 *(Continued)*

	viewing skills (clarifying statements, using open questions, noticing verbal cues, inquiring about psychological problems, preventing repetition, keeping patients to the point, providing verbal and visual encouragement, getting precise information, using brief questions, reducing use of jargon, and avoiding use of jargon) were each assessed on a 5-point (0 to 4) scale. Overall ratings were also assessed on 5-point scales to indicate the degree of self-assurance, warmth, empathy, and competence shown throughout each interview **Other measures:** written history **Measures assessed by:** trained psychologist
Notes	**Conflicts of interests:** not stated **Funding:** Nuffield Foundation **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blinded to group allocation Quote: "ratings were carried out independently of the tutors by raters who were blind to the training method to which the students had been assigned"
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Unclear information regarding attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	The 4 groups differed little in their history taking skills before the feedback programmes began (Tables 2 to 5). The only exception was control students' superiority over the television group on their written histories (Table 5)
Other Bias: Baseline characteristics similarity	High risk	Not reported

Mason 1988

Study characteristics

Methods	**Design:** RCT

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Mason 1988 *(Continued)*

	Unit of randomisation: students **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** NR **Percentage recruitment:** NR **Number of subjects randomised:** 60 **Number of subjects participating:** - Int: 45 - Ctrl:15 - Total: 60 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 4 **Aim:** to develop and test self-instruction that was designed to improve and reinforce the basic communication skills learned in medical school. To compare the effectiveness of 2 types of self-instruction in improving communication skills **Content:** Instructional videotape only: the instructional videotape covered physical setting of the interview, professional manner during the interview, interpersonal relationship, information gathering skills, information providing skills, and closure of the interview. Topics were introduced by an on-camera narrator, using appropriate graphics. This was followed by a vignette demonstrating a desirable aspect of a typical doctor-patient interview Instructional videotape and videotape of pre-intervention interview *:* besides watching instructional videotapes, group 2 students also self-critiqued their pre-intervention videotapes Videotape of pre-intervention interview only: this group of students only conducted a self-critique of their pre-intervention videotapes **Intervention delivered by:** NA, as instructions were delivered through videotapes **Comparison:** students did not participate in educational interventions and received no additional instruction in communication skills **Theoretical basis:** NR **Duration and timing:** 1-off intervention. Instructional videotape was 41 minutes long **Fidelity:** NR **Adherence:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Mason 1988 *(Continued)*

Outcomes	**Assessment:** at the end of the intervention period (August to October 1987) **Assessment timing:** NR **Primary measures:** the study measured students' interpersonal relationships, information gathering skills, ability to provide information, and ability to terminate interviews **Other measures:** no other measures reported **Measures assessed by:** staff from the Office of Medical Education
Notes	**Conflicts of interests:** not stated **Funding:** the study reported here was supported in part by a grant from the National Fund for Medical Education **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information on method of randomisation
Allocation concealment (selection bias)	Unclear risk	No information on timing of allocation nor methods of concealment
Blinding of participants and personnel (performance bias) All outcomes	High risk	Intervention was computer-based, so blinding of personnel was not relevant; participants unable to be blinded due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment "Members of the Office of Medical Education staff were trained as raters, and they were blind as to the students' group assignment"
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition but not explicitly reported
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	ANCOVA performed in statistical analysis to control for varying levels of interviewing skills before the intervention
Other Bias: Baseline characteristics similarity	Low risk	No significant differences were found in background and demographic variables of members of the 8 groups participating in the study

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Meirovich 2016

Study characteristics	
Methods	**Design:** C-RCT **Unit of randomisation:** tutors **Ethics and informed consent:** there is no report on ethical approval. Written informed consent was obtained from all participants **Intention-to-treat analysis:** NR **Adjustment for clustering:** no
Participants	**Student level:** first 3 years **Setting:** university campus **Percentage recruitment:** NR **Number of subjects randomised:** 55 **Number of subjects participating:** - Int: NR - Ctrl: NR - Total: 32 **Age (mean):** 22.9 **Sex:** 54% male **Ethnicity:** NR **Country:** Israel
Interventions	**Number of experimental conditions:** 2 **Aim:** to enhance medical students' empathy and patient-centeredness by using LC tutoring **Content:** students in the experimental group were tutored by tutors who were undergoing LC pedagogical mentoring. Students interviewed patients, health personnel, and health professionals in diverse environments such as medical, clinical, and community settings. Each session was followed by debriefing, feedback, and free group discussions, during which students reflected on their behaviours and attitudes. They also discussed how various environments and approaches impacted professionalism and the doctor-patient relationship **Intervention delivered by:** tutors **Comparison:** tutors were not coached in an LC manner **Theoretical basis:** LC model (Williams 1998) **Duration and timing:** weekly for 3 years, 1 semester per year **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** at the end of the third year, using a Hebrew translation of the Roter Interaction Analysis System **Assessment timing:** End of third year

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Meirovich 2016 *(Continued)*

	Primary measures: percentages of open-ended and closed-ended questions, information gathering skills, relationship building, specific skills in communication such as positive talk, negative talk, psychosocial talk, patient-centred communication, and doctor-centred communication **Other measures:** no other measures reported **Measures assessed by:** NR
Notes	**Conflicts of interests:** declaration of no conflicts of interest **Funding:** not stated **Abbreviations:** C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; LC: learner-centred; NR: not reported

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	Blinding of personnel not possible as tutors were randomised to receive training or not. Blinding of students may have been possible
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear whether assessors were blinded to group allocation. No information regarding assessors
Incomplete outcome data (attrition bias) All outcomes	High risk	Some attrition was apparent (32/55 participated in assessment) but group-level rates are not reported and reasons for attrition are not reported
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Students received the same teaching, only tutor training differed
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	No significant differences were found in background and demographic variables of members of the 8 groups participating in the study
Recuitment Bias (cluster trials only)	Unclear risk	Timing of group allocation/allocation of groups to intervention or control is not clear
Selective recruitment (cluster trials only)	Unclear risk	Timing and baseline characteristics is not clear
Loss of clusters (cluster trials only)	Low risk	No loss of entire clusters

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Meirovich 2016 *(Continued)*

Statistical methods (cluster trials only)	High risk	No adjustment for clustering reported
Herd effect (cluster trials only)	Unclear risk	Unable to determine if herd effect exists

Moreland 1973

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** second year **Setting:** teaching hospital **Percentage recruitment:** 47% **Number of subjects randomised:** 24 **Number of subjects participating:** - Int: 12 - Ctrl: 12 - Total: 24 **Age (mean):** NR **Sex:** 100% males **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to effectively teach communication skills to medical students **Content:** students were taught 6 basic skills (i.e. attending behaviour, minimal activity, open-ended questions, paraphrases, reflections of feeling, and summarisation). During every session, students read a manual describing and watched a videotaped model depicting the skill to be taught on a given day. Students then conducted videotaped interviews, and supervisors and fellow students used the videotaped interviews to reinforce appropriate methods used and to critique poor execution **Intervention delivered by:** psychiatric resident. **Comparison:** students in the control group did not receive instructions in the behaviour they were meant to model, and their interviews were not videotaped to be used for critique later **Theoretical basis:** micro-counselling (Ivey 1968)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Moreland 1973 *(Continued)*

	Duration and timing: 1-off intervention. Videotaped interviews were 6 to 7 minutes long **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** post-training interview **Assessment timing:** 9 weeks **Primary measures:** open-ended questions, closed-ended questions, reflection of feeling, summarisation **Other measures:** no other measures reported **Measures assessed by:** clinical psychology graduate students
Notes	**Conflicts of interest:** not stated **Funding:** supported by an Office of Education grant, Department of Health, Education, and Welfare, at the Regional Office in Boston, Massachusetts **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. Raters were independent and were blind to student group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	Baseline measures collected, but between-groups differences not presented
Other Bias: Baseline characteristics similarity	High risk	Not reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Morrow 2009

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** ethics approval was provided by the University of Texas Southwestern Institutional Review Board. Students signed consent forms, allowing their data to be included in analysis **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** first year **Setting:** university campus. **Percentage recruitment:** 85% **Number of subjects randomised:** 17 **Number of subjects participating:** - Int: 9 - Ctrl: 8 - Total: 17 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to teach first-year medical students to integrate EHR into communication with patients in ambulatory encounters **Content:** both intervention and control groups received instructions on and practised EHR-specific communication skills. This included 45-minute role-plays and practical exercises with the use of Epic test patient. The intervention group received additional EHR-specific communication skills training. Via role-play, 1 of the study authors demonstrated good and poor EHR-specific communication skills. An example of poor skills included physicians turning their back to the patient and typing for long periods without looking at/talking to the patient. Students then generated EHR-specific communication skills that they encountered. In a mini-lecture, the faculty elaborated on Ventres' thematic areas affecting EHR-specific communication. The faculty prompted students to categorise skills into 1 of 4 thematic areas: geographical, relational, educational, and structural **Intervention delivered by:** study authors (no further information) **Comparison:** control group did not receive EHR-specific communication skills **Theoretical basis:** 4 thematic areas influencing EHR use in physician-patient encounters (Ventres 2006) **Duration and timing:** 1-off intervention with 90 minutes per session **Fidelity:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Morrow 2009 *(Continued)*

	Adherence: NR
Outcomes	**Assessment:** evaluation-and-feedback session **Assessment timing:** NR **Primary measures:** introducing self, establishing rapport, asking open-ended questions at the beginning, maintaining eye contact, being supportive and concerned, using transitional statement, avoiding medical jargon, allowing patient to speak without interruption, summarising history, giving patient opportunity to ask questions, appearing poised, professional, and confident **Other measures:** EHR-specific communication skills: introducing self before turning to the computer, adjusting chair to meet patient's eye level, alerting patient before turning to the computer for longer periods of time, introducing the computer in the patient/computer/doctor triad, adjusting screen to ensure patient can see it easily, showing patient weight gain and/or vital signs, graphing patient's weight or showing flow sheets or trends about health, visually sharing EHR information on the screen to include patient rather than keeping patient away from computer work, inquiring whether patient needs a copy of the data, accessing any other online education material for patient **Measures assessed by:** simulated patients
Notes	**Conflicts of interest:** not stated **Funding:** University of Texas Academy of Health Science Education **Abbreviations:** Ctrl: control; EHR: electronic health record; Int: intervention; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment (but they could not be blinded to the nature of the intervention)
Incomplete outcome data (attrition bias) All outcomes	Low risk	3 students lost to follow-up, but unclear which groups lost from and at what stage
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Morrow 2009 *(Continued)*

Other Bias: Baseline characteristics similarity	High risk	Not reported

Mounsey 2006

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** ethical approval was granted by the Institutional Review Board at the University of Virginia. Informed consent to videotaping was requested from all students **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** university campus **Percentage recruitment:** 100% **Number of subjects randomised:** 130 **Number of subjects participating:** - Int: 47 - Ctrl: 46 - Total: 93 **Age (mean):** 25.8 (intervention), 25.7 (control) **Sex:** 51.1% males (intervention), 69.6% males (control) **Ethnicity:** white 72.3% (intervention), 60.9% (control), African American 6.4% (intervention), 6.5% (control), Asian 21.3% (intervention), 28.3% (control), other 0 (intervention), 4.3% (control) **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to determine whether using standardised patients to teach motivational interviewing to medical students would be more effective than using student role-plays **Content:** third-year medical students who had completed a 1-month family medicine clerkship participated in the study. The clerkship consisted of a series of workshops including 1 on MI. Following the workshop, students received written information describing the study and evaluation of the final interview. For intervention and control groups, SPs and students portrayed smokers in the pre-contemplation stage of change **Intervention delivered by:** NR **Comparison:** students practised interviewing with fellow medical students **Theoretical basis:** motivational interviewing (Miller 2002; Rollnick 1992)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Mounsey 2006 *(Continued)*

	Duration and timing: 1-off intervention with students practising MI during their 4-week ambulatory clerkship **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** on the last day of the 4-week ambulatory clerkship **Assessment timing:** after 4 weeks **Primary measures:** empathy, open-ended questions **Other measures:** MI spirit (autonomy, evocation, and collaboration), MI adherence (permission request, affirmation, emphasis of control and support), MI non-adherence (advice, confrontation, and direction), types of questions used (open or closed), number of reflections made **Measures assessed by:** trained masked evaluator
Notes	**Conflicts of interest:** declaration of no conflict of interest **Funding:** this project was funded by a grant from the Health Resources and Services Administration **Abbreviations:** Ctrl: control; Int: intervention; MI: motivational interviewing; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated randomisation "...randomised by the clerkship administrator, using computer-generated random numbers"
Allocation concealment (selection bias)	Low risk	Consent obtained before randomisation
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors (SPs, primary coder and principal investigator) were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Mounsey 2006 *(Continued)*

Other Bias: Baseline characteristics similarity	Low risk	There were no statistically significant differences between the 2 groups in age, sex, and ethnicity, although the control group had a considerably greater proportion of male students

Nomura 2017

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional ethics approval and written informed consent **Intention-to-treat analysis:** NA **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** classroom **Percentage recruitment:** 94.3% **Number of subjects randomised:** 116 **Number of subjects participating:** - Int: 58 - Ctrl: 58 - Total: 116 **Age (mean):** NR **Sex:** 44% male **Ethnicity:** NR **Country:** Japan
Interventions	**Number of experimental conditions:** 2 **Aim:** to confirm the non-inferiority of objective educational outcomes of medical interview training using CYPT compared with FLT, and to explore qualitatively the educational benefits of CYPT for undergraduate communication training **Content:** in the medical interview training module, students take 4 sessions consisting of a lecture on basic interviewing skills (1.5 hours), a large-group demonstration of a medical interview session (3 hours), a structured small-group role-play of a medical interview (3 hours), and a similar small group medical interview session with actors playing the role of patients (3 hours). The present study was conducted during the third small-group structured role-play session (3 hours) out of 4 sessions in the medical interview training module. For the CYPT group, teaching was led by a group of 6 student tutors from year 5 **Intervention delivered by:** 6 tutors from year 5 **Comparison:** faculty-led teaching – students taught by 6 physicians **Theoretical basis:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Nomura 2017 *(Continued)*

	Duration and timing: 3 hours, 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** 10-minute OSCE under supervision of Common Achievement Test Organisation (CATO) **Assessment timing:** 1 week after training **Primary measures:** OSCE score 0 to 100 **Other measures:** none **Measures assessed by:** OSCE is managed by the government-run CATO
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** no external funding was received **Abbreviations:** CATO: Common Achievement Test Organisation; Ctrl: control; CYPT: cross-year peer tutoring; FLT: faculty-led training; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated permuted block randomisation
Allocation concealment (selection bias)	Unclear risk	Not adequately reported. No information on timing of allocation nor methods of concealment
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. OSCE conducted by external government-run Common Achievment Test Organaisation (CATO)
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods, but limited detail provided
Other Bias: Contamination	Low risk	No protection against contamination, but as the teaching session was the same, the risk is low
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Unclear risk	Appear to be no differences in participant characteristics between groups

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ockene 2016

Study characteristics	
Methods	**Design:** C-RCT **Unit of randomisation:** university **Ethics and informed consent:** institutional ethics approval; consent unclear **Intention-to-treat analysis:** yes **Adjustment for clustering:** yes
Participants	**Student level:** third year **Setting:** teaching hospital **Percentage recruitment:** 71.2% **Number of subjects randomised:** 1503 **Number of subjects participating:** - Int: 565 - Ctrl: 531 - Total: 1096 **Age (mean):** 23.3 **Sex:** 55.6% male **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to develop, implement, and evaluate a comprehensive MME intervention for teaching effective delivery of tobacco dependence treatment to medical students **Content:** components took place in the first or third year of medical school. These included (1) a 4-hour self-paced Web-based course for MS1s covering all 5As, with particular emphasis on Assist and Arrange and pharmacotherapy counselling; (2) a role-play classroom demonstration and practice implemented in tandem with the Web course; and (3) a clerkship booster session during the third-year family or internal medicine clerkship rotation that reinforced prior tobacco dependence treatment instruction. MME students, therefore, participated in an enhanced clerkship rotation, in which students and preceptors were strongly recommended to work together with patients who smoked. MME schools were encouraged to continue providing their traditional tobacco treatment education concurrent with MME implementation **Intervention delivered by:** online and clerkship preceptors from the family or internal medicine clerkship who attended a 30- to 60-minute individual or group academic detailing (AD) session reviewing tobacco dependence treatment guidelines, as well as how to model, observe, instruct, and provide critical feedback to students in the specific 5A and pharmacotherapy counselling skill set for tobacco treatment **Comparison:** traditional TE. This curriculum over the 4 years of medical school was reported to be ≤4 hours in duration and varying in content. Traditional teaching was largely didactic, interspersed among basic science and behavioural counselling classes in medical school years 1 and 2. Some schools also included tobacco case studies as part of their behavioural counselling role-plays taught to first or second-year students **Theoretical basis:** NR **Duration and timing:** duration unclear, spread over 3 years **Fidelity:** each MME component was provided to MME schools and their faculty instructors with standardised written guidelines, instructions, objectives, and discussion points to ensure that students would receive the intervention in a similar manner and within a specified time frame (e.g. during MS1 year). ADs were first trained by researchers in an 8-hour training session and received follow-up supervision. Before the teaching of preceptors, ADs practised their sessions with 'practice participants', and all sessions were audiotaped. Each AD completed fidelity checklists after each session with preceptors to ensure that all pertinent topics were covered **Adherence:** study co-ordinators at each school tracked each student's participation in the role-play and booster MME component (Yes/No). The study-tracked total MME intervention dose received per student was computed as the total number of components completed (0 through 3). Each school's AD also recorded the attendance of preceptors in their AD sessions. Third-year students at MME and TE schools were also asked whether they had participated in a Web-based course and role-play classroom demonstration regarding tobacco dependence treatment. In addition, students completed a series of

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ockene 2016 *(Continued)*

	questions to assess the extent of the clerkship preceptors' provision of tobacco training instruction for students. Students rated on a 4-point scale (1 = Strongly disagree to 4 = Strongly agree) how much they agreed with 7 statements related to their preceptor providing them with clear instruction, feedback, and training regarding tobacco dependence counselling. The self-reported total intervention dose was computed as the sum of having received each of the MME components (range 0 to 3)
Outcomes	**Assessment:** video-recorded OSCE station **Assessment timing:** third year post intervention**Primary measures:** 4 items on 'communication skills' **Other measures:** detailed OSCE checklist based on 5As **Measures assessed by:** 'blinded and trained raters'
Notes	**Conflicts of interest:** Sean P. David has stock partnership in Base Health. All other study authors declare that they do not have a conflict of interest **Funding:** the study was supported by an investigator-initiated grant from the National Cancer Institute (NCI) **Abbreviations:** AD: academic detailer; C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; MME: multi-modal education; MS1: first-year medical student; NR: not reported; OSCE: objective structured clinical examination; TE: traditional education

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment "Three blinded and trained raters viewed the video recording of the OSCE and completed the checklist"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition is substantial but appears to be balanced between groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Control group taught and assessed 2 years before intervention group
Other Bias: Baseline outcome measurement similarity	Low risk	Baseline scores not reported, but adjustment made appropriately
Other Bias: Baseline characteristics similarity	Low risk	No MME vs TE differences in baseline demographics or smoking history were noted
Recuitment Bias (cluster trials only)	Low risk	Randomisation of entire schools unrelated to student enrolment

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ockene 2016 *(Continued)*

Selective recruitment (cluster trials only)	Low risk	Schools were pair-matched before randomisation
Loss of clusters (cluster trials only)	Low risk	No clusters lost
Statistical methods (cluster trials only)	Low risk	Adjusted for effects of clustering "In the second stage, the ten post-intervention school means were regressed on intervention assignment and the corresponding pre-intervention school mean. This approach adjusted for the MME and TE comparison for baseline school means while accounting for within-school correlation"
Herd effect (cluster trials only)	Unclear risk	Outcomes not reported by cluster (school), unable to assess

Ozcakar 2009

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** ethical approval was given by the ethics committee of Dokuz Eylul University Faculty of Medicine. Students provided their consent to participate in the study **Intention-to-treat analysis:** NR **Adjustment for clustering:** NR
Participants	**Student level:** second year **Setting:** unclear **Percentage recruitment:** 36% **Number of subjects randomised:** 52 **Number of subjects participating:** - Int: 27 - Ctrl: 25 - Total: 52 **Age (mean):** 20.80 (intervention), 20.88 (control) **Sex:** 67% males (intervention), 56% males (control) **Ethnicity:** NR **Country:** Turkey
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve clinical skills of medical students by using videotaped consultation training followed by feedback

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ozcakar 2009 *(Continued)*

	Content: in the first year, the curriculum consisted of basic perspectives and characteristics of interview and history taking. Students are taught skills based on systems in the second year. Skills from the first 2 years are synthesised in the third year. Students in the study were observed during their first interviews. Interviews were videotaped for students in the study group. All students then assessed themselves using the checklist that the assessor had used to assess their interviews. Study group students were given verbal feedback after students and assessors had watched the video recording of their interviews. Control group students received only verbal feedback **Intervention delivered by:** NR **Comparison:** only verbal feedback **Theoretical basis:** NR **Duration and timing:** 1-off intervention with videotape and feedback session taking 2 weeks **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** 15 days after the first interviews **Assessment timing:** 15 days **Primary measures:** relationship building through greeting the patient, comforting, and determining the level of communication; students' facilitating skills and use of communication skills; use of open- and closed-ended questions; ability to summarise the history and ability to structure the interview; ability to gain the patient's perspective **Other measures:** ability to obtain patient's past medical history, present health condition, family history, personal history, and review of systems **Measures assessed by:** family medicine department staff
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table used
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	High risk	Asessors not blind to students' group allocation
Incomplete outcome data (attrition bias)	Low risk	No attrition

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ozcakar 2009 *(Continued)*

All outcomes		
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	Significance of time 1 measures not reported, but differences appear to be low
Other Bias: Baseline characteristics similarity	Low risk	There were no significant differences in the distribution of gender between the 2 groups ($Chi^2 = 0.624$, $P = 0.43 > 0.05$) Mean ages of students were 20.80 for the study group and 20.88 for the control group, with no significant differences between groups ($P = 0.95 > 0.05$)

Palmer 2018

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** exemption status granted by institutional review board **Intention-to-treat analysis:** yes **Adjustment for clustering:** NA
Participants	**Student level:** first year **Setting:** online and classroom **Percentage recruitment:** 100% **Number of subjects randomised:** 50 **Number of subjects participating:** - Int: 24 - Ctrl: 26 - Total: 50 **Age (mean):** 27.6 **Sex:** 50% male **Ethnicity:** 94% white **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to investigate whether students who learned online would do as well as their in-class peers in assessment and treatment of a standardised patient with antidepressant-induced sexual dysfunction

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Palmer 2018 *(Continued)*

	Content: An online module consisting of an expanded version of traditional lecture presentation relating to interaction between antidepressants and sexual dysfunction, case studies depicting these concepts, and comprehension checks after informational slides **Intervention delivered by:** online **Comparison:** a traditional lecture, which included 2 faculty psychiatrists (1 lecturer plus a course director who provided additional clinical experience) delivering the same slides and case studies as the online version **Theoretical basis:** NR **Duration and timing:** 1-off, length NR **Fidelity:** NR (online, so consistent between participants) **Adherence:** NR
Outcomes	**Assessment:** video-recorded SP interview **Assessment timing:** within the 8-week course **Primary measures:** assessment included 7 domains: (1) introduction (introduce self, define role in patient's care, sit down, begin with open-ended question, or clarify that this is a follow-up visit), (2) verbal empathy (use hopeful or supportive remarks, use continuer phrases, repeat important information, match patient's tone and mood), (3) non-verbal empathy (eye contact, leaning in or smiling as appropriate, facial expression reflective of patient mood), (4) basic skills (mostly open-ended questions, 1 question at a time, gender-neutral terminology, appropriate sexual health terminology), (5) depression follow-up (guilt, mood, sleep, appetite, concentration, self-harm ideation, medication adverse effects, questions about changes in desire, arousal, and orgasm), (6) closure (summarise, ask for questions), and (7) sensitivity (appropriate tone and demeanour) SPs were surveyed after the interview to assess how they perceived the care that they received. A satisfaction survey, adapted from the Accreditation Council for Graduate Medical Education Core Competency Assessment toolkit [18], gauged the communication skills and delivery of the student, as well as requesting the patient's overall rating of the student's interaction and willingness to have the student as a provider in the future **Other measures:** none **Measures assessed by:** faculty raters and SPs
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** this work was supported by an Endowment for Educational Research Award from Mayo Clinic (Rochester, MN). The funding source had no role in the design of the study, collection or analysis of data, or interpretation of findings **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: standardised patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Students were randomised through the random number generator function in Microsoft Excel
Allocation concealment (selection bias)	Unclear risk	Not adequately reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Palmer 2018 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding not possible due to the nature of the intervention, but lack of blinding unlikely to influence outcomes (online vs large-group lecture)
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to students' group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Students received the same content, only mode of delivery differed
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	No significant differences

Papadakis 1997

Study characteristics	
Methods	**Design:** C-RCT **Unit of randomisation:** group **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** no
Participants	**Student level:** first year **Setting:** university campus **Percentage recruitment:** 60% **Number of subjects randomised:** 72 **Number of subjects participating:** - Grp 1: 35 - Grp 2: 37 - Total: 72 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Papadakis 1997 *(Continued)*

	Country: USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to teach smoking cessation techniques to medical students by using SPs and role-play **Content:** all first-year students received lectures on the health effects of smoking and about how physicians can help patients in quitting smoking. These students were then randomised into 2 groups; 1 group practised counselling with SPs **Intervention delivered by:** faculty and SP **Comparison:** comparison was between group 1 students who practised smoking cessation techniques vs SPs and group 2 students who practised smoking cessation techniques via peer role-play **Theoretical basis:** NR **Duration and timing:** 1-off intervention. Each appointment with SPs was 30 minutes long **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** following the lectures **Assessment timing:** 1 week **Primary measures:** students' introduction of self, empathy, tone of voice, use of language that patients understand, demonstration of interest in how patients' personal lives affected smoking, use of appropriate reassuring/supportive statements, allowing patients to ask questions, making eye contact **Other measures:** cognitive skills measured by whether students obtained smoking history from patients; inquired about patients' intention, motivation, and obstacles to quit; student-linked specific symptoms to smoking; explaining expected benefits of smoking cessation; setting a quit date; writing a prescription for quitting; discussing common potential withdrawal symptoms; providing information to patients on where to obtain material; discussing use of pharmacological agents for cessation; making a follow-up appointment soon after quit date **Measures assessed by:** SPs
Notes	**Conflicts of interest:** not stated **Funding:** University of California, San Francisco, Instructional Improvement Grant **Abbreviations:** C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; NR: not reported; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Low risk	All students randomised after initial lectures
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Papadakis 1997 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	SP assessors were blind to group assignment. Assessed by SPs who were not made aware of which group students had been assigned to
Incomplete outcome data (attrition bias) All outcomes	High risk	Possible attrition bias as only a proportion of participants assigned to the assessment participated in it. It seems only a portion of participants were invited to complete post-test assessment
Selective reporting (reporting bias)	High risk	Detailed data not presented
Other Bias: Contamination	Low risk	Contamination between groups was unlikely as the 2 groups received the same informational lecture, then Group 1 (intervention) conducted the practise role-play exercise with an SP, while Group 2 (control) conducted the role-play exercise in student pairs
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported
Recuitment Bias (cluster trials only)	Low risk	All students were in groups that were then randomly assigned to intervention or control
Selective recruitment (cluster trials only)	Low risk	Randomisation occurred for all at the same point
Loss of clusters (cluster trials only)	Low risk	No clusters lost
Statistical methods (cluster trials only)	High risk	No adjustment reported to account for clustering of groups
Herd effect (cluster trials only)	Unclear risk	Outcomes not reported by cluster

Perera 2010

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** students **Ethics and informed consent:** ethical approval was obtained from the Research and Ethics Committee of the International Medical University. Informed consent is not reported **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** semester 1 **Setting:** university campus

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Perera 2010 *(Continued)*

Percentage recruitment: 94%

Number of subjects randomised: NR

Number of subjects participating:

- Int: 97

- Ctrl: 93

- Total: 190

Age (mean): NR

Sex: NR

Ethnicity: NR

Country: Malaysia

Interventions	**Number of experimental conditions:** 2 **Aim:** to teach smoking cessation techniques to medical students by using SPs and role-play **Content:** during a CS session, after role-play and feedback from SPs, students in the experimental group were given a structured self-assessment tool with open- and closed-ended questions. This questionnaire helped students reflect on their performance to identify gaps. Further to this, peers provided feedback to students on any additional points missed with the same log sheet. Following this, facilitators addressed any further gaps in feedback **Intervention delivered by:** teachers **Comparison:** students in this group did not undergo self-assessment and peer feedback **Theoretical basis:** self-evaluation and feedback **Duration and timing:** 1-off intervention **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** OSCE **Assessment timing:** 2 weeks after final learning. **Primary measures:** rapport building skills, listening skills **Other measures:** use of appropriate language, interview style, interview structure **Measures assessed by:** teachers who instructed both groups
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; CS: communication skill; Int: intervention; NR: not reported; OSCE: objective structured clinical examination; SP: simulated patient; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quote: "we divided the students in the semester one medical undergraduate program (n = 202) into experimental and control groups using their university

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Perera 2010 *(Continued)*

		entrance academic grades (the latest pre-study performance data) to ensure equivalence between the two groups, although academic grades may not be an important determinant factor"
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Control group facilitators and students were blinded, but experimental group facilitators and students were not blinded
Blinding of outcome assessment (detection bias) All outcomes	High risk	Whilst study authors state that "assessors were blinded to the identity of the group of students", assessors comprised "teachers who had instructed both the experimental and the control groups". Therefore assessors/teachers who participated in teaching of the experimental group may have recognised some students as belonging to the experimental group, and bias in assessment is a possibility. Students in both groups were equally distributed among assessors to ensure uniformity and to reduce bias
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition low and likely balanced, but reasons unclear, and initial group sizes unclear
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred Quote: "although the students in the intervention group were requested to maintain confidentiality, it is possible that they conveyed some information about the intervention to the control group"
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	Gender distribution and mean age between the 2 groups were not significantly different

Pirdehghan 2018

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** unclear re ethics approval; signed informed consent obtained **Intention-to-treat analysis:** yes **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** years 6 and 7 **Setting:** classroom

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Pirdehghan 2018 *(Continued)*

	Percentage recruitment: NR **Number of subjects randomised:** 60 **Number of subjects participating:** - Int: 30 - Ctrl: 30 - Total: 60 **Age (mean):** 25 **Sex:** 47.1% male **Ethnicity:** NR **Country:** Iran
Interventions	**Number of experimental conditions:** 2 **Aim:** to measure the effect of CST through role-playing experience of medical students to examine their communication skills improvement **Content:** topics included theoretical training relating to the doctor-patient relationship in a patient-centred model, and counselling skills with a focus on the trans-theoretical model. Communication skills practice was in the form of role-playing based on a hypothetical patient **Intervention delivered by:** community medicine specialists **Comparison:** NR **Theoretical basis:** NR **Duration and timing:** 18 hours over 3 days **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** 20-minute SP interaction **Assessment timing:** 1 week and 3 months after intervention **Primary measures:** communication skills assessment tool including 17 items scored 1 to 10: introduction, put at ease, empathy, confidence, respect, listening, transitions, explanations, non-judgemental, open questions, 1 question at a time, invited questions, invited patient to express reason for attending, eye contact, language (no jargon), comfortable distance, closure **Other measures:** none **Measures assessed by:** external rater
Notes	**Conflicts of interest:** not stated **Funding:** the study was funded by the Monitoring Health Research Centre at Shahid Sadughi University of Medical Science **Abbreviations:** CST: communication skills training; Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated patient
Risk of bias	

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Pirdehghan 2018 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random numbers table used
Allocation concealment (selection bias)	Unclear risk	Not adequately reported. No information on timing of allocation nor methods of concealment
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors (external observers) were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	High risk	All a priori scales reported, but detailed items for intervention group only (comparisons at summary level)
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	No significant differences at baseline
Other Bias: Baseline characteristics similarity	Low risk	No significant differences

Poole 1980

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fifth year **Setting:** university campus **Percentage recruitment:** NR **Number of subjects randomised:** NR **Number of subjects participating:**

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Poole 1980 *(Continued)*

	- Int: 25 - Ctrl: 20 - Total: 45 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** unclear
Interventions	**Number of experimental conditions:** 2 **Aim:** to train medical students to increase their empathy **Content:** the intervention used the Tune-In Empathy Training Workshop to train medical students to increase their level of empathy **Intervention delivered by:** NA (audiotape-led) **Comparison:** control group students did not participate in the empathy training programme **Theoretical basis:** NR **Duration and timing:** 1-off intervention. Each audiotaped session was 1.5 to 2 hours long **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** NR **Assessment timing:** NR **Primary measures:** students' level of empathy **Other measures:** no other measures reported **Measures assessed by:** trained rater
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Subjects in this study were 45 fifth year medical students, of whom 25 were experimental and 20 were control. The experimental group was randomly selected from those students who had participated in the empathy training programme as part of their second-year course in behavioural science. Not randomly selected for second-year participation. No information on randomisation methods at this step
Allocation concealment (selection bias)	Unclear risk	Not adequately reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Poole 1980 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were independent and blind to student group membership
Incomplete outcome data (attrition bias) All outcomes	Low risk	Random selection of potential participants selected for follow-up, but no attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported

Roche 1996

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** rotation (students randomly allocated to groups, and groups allocated to condition) **Ethics and informed consent:** NR **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fifth year **Setting:** teaching hospital **Percentage recruitment:** 89% **Number of subjects randomised:** 210 **Number of subjects participating:** - Int 1: 28 - Int 2: 32 - Int 3: 30 - Ctrl: 37 - Total: 187 (23 did not complete pre-test assessment) **Age (mean):** 23 **Sex:** 62% male **Ethnicity:** NR **Country:** Australia
Interventions	**Number of experimental conditions:** 4 **Aim:** to examine the relative effectiveness of 4 different educational programmes in teaching smoking cessation skills to undergraduate medical students **Content:**

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Roche 1996 *(Continued)*

	Intervention 1: received a 1-hour lecture on smoking and were provided with pre-reading and instructions to produce a 10-minute audiotaped doctor–patient interview on smoking cessation. Faculty feedback on students' performances in the audiotaped interviews was later provided through 2-hour-long small- group tutorial sessions Intervention 2: in addition to the 1-hour lecture and pre-reading package on smoking, students participated in role-plays of smoking intervention interviews over a 2-hour period. Students worked in triads and alternatively role-played doctor and patient; they also assumed the role of observer. Peer feedback on intervention techniques was provided by the observer in each role-play and was supervised by faculty staff. Students were encouraged to rate each other's performances Intervention 3: received a 1-hour lecture on smoking and a pre-reading package. Each student in this group also role-played smoking cessation interviews that were videotaped. Feedback was given on students' performances in these videotaped interviews through 2-hour-long small-group sessions, in which faculty tutors provided individual feedback on the videotaped interviews **Intervention delivered by:** faculty staff (skills not specified) **Comparison:** control group received a standard 3-hour didactic teaching presentation on prevalence, physiological and other effects, and the public health implications of smoking. Lecture content encouraged the participation of doctors in smoking cessation. This group received no specific lectures on intervention techniques and/or counselling skills for smoking and was not involved with smoking-related group work or role-plays **Theoretical basis:** evidence for various educational approaches described in background **Duration and timing:** all were 1-off 3-hour sessions **Fidelity:** SPs were trained/standardised **Adherence:** NR
Outcomes	**Assessment:** 10-minute video-recorded interaction with SP viewed and rated independently **Assessment timing:** end of term – up to 10 weeks after intervention **Primary measures:** category 15 of the rating scale consisted of 18 items that examined interactional skills. These included opening the consultation, empathy, non-verbal interaction, language, question style, control of the interview, strategies to effect change, concluding the interview, and attempts to summarise **Other measures:** interview content was measured in sections 1 to 14 of the rating schedule: (1) assessment of patient's smoking status (1 item); (2) elicitation of details of patient's consumption levels (3 items); (3) assessment of patient's motivation to quit (2 items); (4) assessment of patient's awareness of risks (2 items); (5) description of benefits of quitting (19 items); (6) expression of personal desire for patient to quit (1 item); (7) counselling on anticipated problems (2 items); (8) expression of confidence in ability to quit (5 items); (9) establishment of a quit contract (5 items); (10) addressing self-exemptions (1 item); (11) offer of behavioural tips (19 items); (12) advice on use of Nicorette gum (3 items); (13) offer of written materials (1 item); (14) provision of information on referral/follow-up (2 items). Students also completed a knowledge and attitudes survey **Measures assessed by:** independent raters who had no prior knowledge of the content area
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Students were allocated to a control group or to 1 of 3 experimental groups on the basis of block randomisation by term
Allocation concealment (selection bias)	Unclear risk	Allocation concealment not reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Roche 1996 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: "five independent raters were recruited and trained as a group, using a single gold standard, to rate the students' videotaped performances using a rating schedule. Raters were selected who had no prior knowledge of the content area to minimize rater bias"
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was moderate. Many students did not show for post-test assessment due to non-compulsory nature, but mostly balanced across all groups (31% control, 37% video, 27% peer, 40% audio)
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods. Reported at group level
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred between rotations
Other Bias: Baseline outcome measurement similarity	Low risk	Some differences at baseline, but they were adjusted for In analyses and were not likely to impact results
Other Bias: Baseline characteristics similarity	High risk	Characteristics not reported by group

Roche 1997

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** rotation (students randomly allocated to groups and groups allocated to condition) **Ethics and informed consent:** unclear **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fifth year **Setting:** teaching hospital **Percentage recruitment:** 94% **Number of subjects randomised:** 123 **Number of subjects participating:** - Int: 57 - Ctrl: 58 - Total: 115 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** Australia
Interventions	**Number of experimental conditions:** 2 **Aim:** to examine the relative effectiveness of 2 different educational programmes to teach senior medical students brief intervention skills for alcohol problems

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Roche 1997 *(Continued)*

	Content: 1 hour of formal didactic lecturing was provided presenting, in summary form, the same content as that delivered to the traditional teaching group covering theoretical components of the principles and practice of early and brief interventions. This was followed by a small-group tutorial session, with each group having approximately 8 students. In the tutorials, students received feedback on their videotaped interviews (pre-test) completed earlier in the week. Both traditional and interactive groups received a set of comprehensive reading materials, which outlined the burden of illness associated with alcohol use, the potential for intervention by medical practitioners, components of a brief intervention, and strategies for instigating behaviour change **Intervention delivered by:** tutors from the academic staff and from the College of General Practitioners' Family Medicine Programme training scheme **Comparison:** traditional teaching - 3-hour lecture on the principles and practice of early and brief interventions **Theoretical basis:** principles of brief intervention **Duration and timing:** 3 hours, 1-off **Fidelity:** simulated patient training conducted and consistency ensured **Adherence:** NR
Outcomes	**Assessment:** 10-minute video-recorded interaction with SP viewed and rated independently **Assessment timing:** 8 to 10 weeks after intervention (end of rotation) **Primary measures:** total score for 18 items that examined interactional skills. These included opening the consultation, empathy, non-verbal interaction, language, question style, control of the interview, strategies to effect change, concluding the interview, and attempts to summarise **Other measures:** 14 'sections' on components of the interview relating to alcohol and brief intervention techniques. These items examined whether the student satisfactorily accomplished the following tasks: assessed the patient's drinking status (1 item), level of consumption (3 items), alcohol dependence (12 items), problems associated with alcohol (3 items), and perception of his or her drinking (3 items); educated the patient about safe drinking (3 items); highlighted the association between excess alcohol consumption and health (7 items); advised the patient to modify his or her consumption (6 items); outlined the benefits of cutting down (19 items); expressed confidence in the patient's ability to reduce consumption (1 item); countered the patient's self-exemptions (1 item); addressed concerns about changing drinking behaviour (2 items); offered behavioural tips for reduction (25 items); and offered written materials and a follow-up visit (3 items). One section on interactional skills (opening the consultation, empathy, non-verbal interaction, language, question style, control of the interview, strategies to effect change, concluding the interview, and attempts to summarise) **Measures assessed by:** 4 independent raters who were naive to the study design and had no prior knowledge of the content area
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Randomised to experimental groups by rotation term
Allocation concealment (selection bias)	Unclear risk	Allocation concealment not reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Roche 1997 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment Quote: "raters were independent of study and had no knowledge of study design"
Incomplete outcome data (attrition bias) All outcomes	Low risk	Minimal attrition, equivalent between groups
Selective reporting (reporting bias)	High risk	Individual rating items not reported
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	No significant differences between groups on pre-teaching scores
Other Bias: Baseline characteristics similarity	High risk	Not reported

Ruesseler 2017

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** class/group **Ethics and informed consent:** local ethics committee approval and written informed consent obtained **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but checked for differences between study weeks - no differences
Participants	**Student level:** fourth year **Setting:** classroom **Percentage recruitment:** 85% **Number of subjects randomised:** 106 **Number of subjects participating:** - Int: 49 - Ctrl: 51 - Total: 100 (complete data) **Age (mean):** 24 **Sex:** 36% male **Ethnicity:** NR **Country:** Germany

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ruesseler 2017 *(Continued)*

Interventions	**Number of experimental conditions:** 2 **Aim:** to evaluate the educational benefits of video-assisted feedback vs oral feedback in building communicative competencies in the surgical context **Content:** the unit focuses on the specific problem-oriented approach to a patient in terms of taking a patient history and obtaining informed consent before surgery. Students first repeat and enlarge their existing communicative competencies in an interactive workshop under the supervision of 2 trained peer tutors in a small group (maximum of 6 students) using specific clinical examples (120 minutes). In the subsequent 90 minutes, students practise taking patient histories and obtaining informed consent before surgery in defined role-plays. One student plays the medical expert, and another plays the patient (using a standardised description of patient role) Video feedback: each role-play was videotaped and reviewed immediately after participant performance, inspired by the agenda-led, outcome-based guidelines. All participants were encouraged to reflect on the performance based on various observer tasks and were to define a maximum of 4 specific and realistic aspects of improvement before giving their feedback. Subsequent feedback was provided as described for the oral feedback group. Feedback took a total of approx. 5 to 10 minutes for each role-play. Tutors completed a minimum of 3 of the defined role-plays (1 for patient history, 1 for informed consent, and 1 as requested by the student group) and subsequent feedback within the training unit **Intervention delivered by:** trained peer tutors **Comparison:** Oral feedback: among the oral feedback group, students received 5 to 10 minutes of feedback on each role-play guided by the peer tutor and inspired by the ALOBA guidelines. All members of the group were instructed to define a maximum of 4 specific and realistic aspects of improvement before giving their feedback. The feedback process starts with the 'medical expert' reflecting on his own performance and what could be improved. The 'patient' then describes the patient perspective and what could be improved. The 'observers' give their feedback based on their specific tasks (e.g. non-verbal communication). Finally, the tutor concludes the feedback with constructive comments and a summary of what went well and what could be improved. The tutors completed 4 defined role-plays (2 for patient history, 2 for informed consent) and subsequent feedback within the training unit **Theoretical basis:** NR **Duration and timing:** 210 minutes, 1-off **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** two 5-minute OSCE stations **Assessment timing:** immediately after intervention **Primary measures:** students were rated on a 2-part checklist at both OSCE stations, where part 'B' of the checklist assessed verbal and non-verbal skills comprising B1 introduction, B2 intelligibility, B3 questioning, B4 impression and impact, and B5 interaction. As for taking a patient's history, this part specified structured information gathering (e.g. beginning and duration of symptoms). Scores for each item ranged from 0 unsatisfactory to 3 excellent **Other measures:** part 'A' of the checklist assessed content: specified indication for surgery, choice of procedure, general and specific risks, and postoperative treatment/follow-up for 'obtaining informed consent' **Measures assessed by:** trained examiners
Notes	**Conflicts of interest:** declaration of no conflicts of interest

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Ruesseler 2017 *(Continued)*

Funding: this study was funded by the German Federal Ministry of Education and Research as part of the joint research project, "Practical clinical competence - a joint program to improve training in surgery"

Abbreviations: Ctrl: control; Int: intervention; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Low risk	Group allocation occurred independently of and before the study.
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	Six students excluded due to incomplete data. Group allocation not provided
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	High risk	The study was conducted in 3 training weeks over a total period of 7 weeks. Thus, it is possible that students in later training weeks received information about the OSCE checklist contents, which were evaluated. Because this was a voluntary study without effect on student certifications, we considered this possibility to be marginal. This was borne out by our study results, showing no significant differences between the 3 training weeks
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	No significant differences

Schmitz 2018

Study characteristics

Methods

Design: RCT

Unit of randomisation: student

Ethics and informed consent: the study was deemed exempt from formal review by the independent cantonal ethics committee; participants provided written informed consent

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Schmitz 2018 *(Continued)*

	Intention-to-treat analysis: NR **Adjustment for clustering:** NA
Participants	**Student level:** fourth year **Setting:** online/classroom (teaching hospital) **Percentage recruitment:** 80% **Number of subjects randomised:** 91 **Number of subjects participating:** - Int 1: 30 - Int 2: 10 - Ctrl: 27 - Total: 83 **Age (mean):** 23.3 **Sex:** 42% male **Ethnicity:** NR **Country:** Switzerland
Interventions	**Number of experimental conditions:** 3 **Aim:** to clarify which learning intervention with a given worked example, presented in 3 formats - text, video, and video with hints - optimally prepares undergraduate medical students for an SP task involving breaking bad news **Content:** after a communication skills role-play session, students received online access to a learning environment that provided the materials they needed to prepare individually for a breaking bad news scenario. The preparatory learning environment was a web-based learning module prepared in German. It introduced the SPIKES protocol for delivering bad news to patients. The learning module provided the same instructions and learning content to all participants; the only difference was the experimentally varied presentation format of the embedded worked example. The worked example embedded in the learning module provided the same content to all 3 groups: a physician delivering a diagnosis of lung cancer, following the SPIKES protocol Intervention 1 (video example): dialogue of the doctor-patient interaction presented as a video Intervention 2 (video with hints): the same video was enriched with written hints, appearing to the right. These emphasised the SPIKES steps and their underlying principles. Three hints referred to EMOTIONS, each summarising 1 physician response to an emotional cue or concern. A single hint was provided to denote each of the other SPIKES steps. These hints referred only to information already conveyed in the learning module **Intervention delivered by:** online **Comparison:** same online module with a text only worked example - the clinical interview was a written dialogue, with alternating utterances from the physician and the patient **Theoretical basis:** previous studies conducted by the research group (Schmitz 2017) **Duration and timing:** length unclear, 1-off **Fidelity:** NR **Adherence:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Schmitz 2018 *(Continued)*

Outcomes	**Assessment:** 15-minute SP encounter simulating breaking bad news **Assessment timing:** 1 week after intervention **Primary measures:** a self-developed 5-point scale with 6 items, each addressing a SPIKES step and the published 5-point glBAS. Each participant's final SPIKES and glBAS scores were calculated as grand means, based on assessors' ratings **Other measures:** relevant responses were coded as using a space-providing or space-reducing technique, following the VERONA coding protocol **Measures assessed by:** trained Master's level psychology students
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** no external funding was received for this study **Abbreviations:** Ctrl: control; gIBAS: global Breaking Bad News Assessment Scale; Int: intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: standardised patient; SPIKES: Setting, Perception, Invitation, Knowledge, Emotion, Strategy, and Summary

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Low risk	Participants and assessors were not told about hypotheses nor alternative interventions until data were collected and performance ratings were complete
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Online; participants unaware of alternative interventions
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blinded to group assignment
Incomplete outcome data (attrition bias) All outcomes	High risk	imbalanced exclusion of data due to SP problems in 1 group
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Contamination between groups unlikely as both control and intervention groups received the same content in their online learning module. It was the method of presenting the information that differed between groups (e.g. text example, video example, video example with hints)
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	No significant differences

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Schwartz 2010

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** class/group **Ethics and informed consent:** granted exempt status from institutional review board, but written informed consent obtained **Intention-to-treat analysis:** no (but sensitivity analysis performed) **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** 2 medical centres (teaching hospitals) where students undertake sub-internships **Percentage recruitment:** 82.2% **Number of subjects randomised:** - Int: 94 - Ctrl: 95 - Total: 189 **Number of subjects participating:** - Int: 65 - Ctrl: 59 - Total: 124 (17 withdrew before assessment; 48 did not present for assessment) **Age (mean):** 28.8 **Sex:** 52.3% male **Ethnicity:** 33.9% white **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to evaluate an educational intervention designed to increase physicians' skills in identifying patient context and to decrease the rate of contextual errors **Content:** students were taught to apply 6 concepts of assessment and management to contextualise patient care by listening to the narrative surrounding the patient's complaints, identifying contextual red flags, and, when identified, formulating a "contextual differential" from 10 contextual domains: exploring contextual assumptions for each domain in their differential; asking questions to narrow the differential (contextual assessment); characterising the relevant patient context; and merging contextual assessment findings with information about clinical state, research evidence, and patient preferences to formulate a contextually appropriate plan of care. During the first session, students discussed and applied basic concepts to a written case describing contextual issues. During the second session, students applied basic concepts to another written case and then described and applied the concepts to their own patients on service. During the final 2 sessions, students applied and developed their knowledge and skills by interviewing patients with potential contextual issues at the bedside, with 1 of the study authors (S.J.W.) serving as a guide and role model **Intervention delivered by:** physicians **Comparison:** no intervention – observation only. **Theoretical basis:** contextualising patient care (Weiner 2008) **Duration and timing:** 4 × 1-hour sessions over 4 weeks (4 hours total) **Fidelity:** all sessions delivered by 1 of the study authors **Adherence:** attendance. 26 attended 4 teaching sessions; 25 attended 3 teaching sessions; 12 attended 2 teaching sessions; 2 attended 1 teaching session
Outcomes	**Assessment:** assessment consisted of 4 standardised patient encounters (cases A, B, C, and D) for each student, with each standardised patient presenting a different variant (of 4 variants) of each case. Sub-internships each month were randomly assigned to 1 of the 16 possible permutations of case and variant In all cases and in all variants, the standardised patient presented a typical clinical scenario but mentioned 2 red flags that could indicate a biomedically or contextually atypical diagnosis. In baseline variants, patients presented no symptoms of the atypical diagnosis if students probed these red flags. In biomedical variants, student probing of the biomedical red flag led the patient to present further symp-

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Schwartz 2010 *(Continued)*

	toms of the medically atypical diagnosis; similarly, in contextual variants, probing the contextual red flag led the patient to present further evidence of the contextually atypical situation. In the biomedical/contextual variants, each red flag, if probed, led to presentation of the combination of evidence for both biomedically and contextually atypical situations. Development and validation of cases and variants used a panel of expert internists to ensure that not identifying the atypical diagnosis would lead to a plan of care that would be unequivocally inappropriate for the patient **Assessment timing:** 3 to 10 days after intervention **Primary measures:** whether the student probed for the biomedical red flag, whether the student probed for the contextual red flag, whether the student's management plan addressed problem(s) presented in the case and variant (information gathering) **Measures assessed by:** assessor
Notes	**Conflicts of interest:** Drs Schwartz and Weiner reported being owners of a company that provides management consulting services to clinicians and institutions interested in collecting customer service and performance data using unannounced standardised patient methods. They have not to date received any consulting fees, honoraria, contracts, or other payments. No other study authors reported disclosures **Funding:** this project was funded in part by a National Board of Medical Examiners (NBME) Edward J. Stemmler, MD, Medical Education Research Fund grant **Abbreviations:** Ctrl: control; Int: intervention; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Students were allocated to rotation by a lottery, then allocation to intervention or control was based on site, alternating by month
Allocation concealment (selection bias)	High risk	Although participants were assigned to sites using a quasi-random process independent of this trial, assignment of site to intervention or control was alternated. Although this reduced the risk of contamination, it may increase the risk that participants could predict their allocation to trial groups
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding not possible due to the nature of the intervention, but given lack of delivery to the control group, not likely to impact. Students told they were validating new standardised patient assessments without discussion of assessments or study hypotheses
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessor blind to student allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Moderate attrition, but reasons and numbers equivalent between groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Group allocation to intervention or control was randomised monthly to reduce contamination risk, but potential for contamination between students at the same site is unclear
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Schwartz 2010 *(Continued)*

Other Bias: Baseline characteristics similarity	Low risk	Groups were demographically similar regarding sex, age, race, academic year, and clinical knowledge

Seim 1995

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** class/group **Ethics and informed consent:** unclear**Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** teaching hospital **Percentage recruitment:** NR **Number of subjects randomised:** 70 **Number of subjects participating:** - Int: 30 - Ctrl: 30 - Total: 60 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to determine whether multi-modal learning is more effective than a unimodal approach (distribution of algorithm) in teaching 4-year medical students to use patient-centred counselling skills with a patient in relation to smoking cessation **Content:** the 2.5-hour formal training session began with slides showing epidemiological evidence of the relationship of cigarette smoking to chronic disease morbidity and mortality and the benefits of cessation. The patient-centred approach was introduced, and the smoking cessation algorithm was distributed. A videotape with 4 vignettes illustrating the patient-centred approach was viewed. Students then participated in doctor-patient role-playing in groups of 2 or 3; students played both doctor and patient (switched roles) in groups of 2 students with an observer (a teacher) giving feedback. In groups of 3, students switched roles between doctor, patient, and observer **Intervention delivered by:** NR **Comparison:** brief training - received the smoking cessation algorithm and approximately 5 minutes of instructions about how to incorporate the patient-centred approach into clinical activities **Theoretical basis:** patient-centred approach and positive self-efficacy (Bandura 1977) **Duration and timing:** 2.5 hours, 1-off **Fidelity:** the same researcher played SP role for all assessments **Adherence:** the reviewers, who were blinded, noted the consistency of the simulated patient's presentations to students during the audiotaped reviews, and also noted that he offered no information without prompting by the student doctor
Outcomes	**Assessment:** SP interaction audio-recorded and rated by blinded assessor - interrater reliability was 83.7% **Assessment timing:** 4 to 6 weeks after intervention **Primary measures:** smoking cessation counselling skills; eliciting information, providing information, and eliciting/responding to affect relating to 6 content areas of the counselling protocol **Other measures:** none **Measures assessed by:** trained research associates
Notes	**Conflicts of interest:** not stated

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Seim 1995 *(Continued)*

Funding: this article was supported in part by a Health and Human Services Grant

Abbreviations: Ctrl = control; Int = intervention; NA: not applicable; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Limited information on method of randomisation. Students were randomised to rotations, and researchers had no control over that - then rotations were randomised for the study. Quasi-random design
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition low and balanced between groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported

Shaddheau 2015

Study characteristics

Methods	**Design:** C-RCT **Unit of randomisation:** rotation (each block of students rotating through the Women's Care clerkship was randomised to 1 of the 2 groups) **Ethics and informed consent:** study received exempt status by the Colorado Multiple Instructional Review Board. Unclear whether students gave informed consent **Intention-to-treat analysis:** NA **Adjustment for clustering:** no
Participants	**Student level:** third year **Setting:** teaching hospital **Percentage recruitment:** 100% **Number of subjects randomised:** 135

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Shaddheau 2015 *(Continued)*

	Number of subjects participating: - Int: 75 - Ctrl: 60 - Total: 135 **Age (mean):** range 20 to 40+ (could calculate mean) **Sex:** 52.6% male **Ethnicity:** 72.6% Caucasian overall **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to determine if a 1-hour NPI improves student performance when caring for a standardised patient with an unintended pregnancy **Content:** the NPI consisted of small groups of students (55) spending 1 hour with a patient-facilitator who volunteered to discuss her experience with options counselling, her decision-making process, and termination of her pregnancy. Students could then ask the patient-facilitator questions about any aspect of her experience. Clerkship directors and staff were not present during the sessions. We required patient-facilitators to have had a recent experience with options counselling, to be willing to discuss their personal experiences with small groups of medical students, and to participate in a 2-hours-long training session. The NPI group was exposed to both the standard curriculum and the NPI **Intervention delivered by:** patient facilitators **Comparison:** standard third-year clerkship curriculum. During the study period, the family planning curriculum included in the clerkship consisted of 2 didactic sessions (contraception and abortion), a small-group ethics session (focused on conscientious refusal and legal aspects of abortion), and a simulation session (hands-on experience with manual vacuum uterine aspiration and intrauterine device insertion) **Theoretical basis:** NR **Duration and timing:** 1 hour, 1-off **Fidelity:** Women's Care course director and research assistant conducted a 2-hours-long training session **Adherence:** NR
Outcomes	**Assessment:** options counselling OSCE with standardised patient **Assessment timing:** end of clerkship, approximately 6 weeks **Primary measures:** proportion of participants achieving 'excellence' on the OSCE checklist. Specific measures included whether results were delivered to the patient in a neutral fashion, establishing trust with the patient, and discomfort in delivering test results **Other measures:** student performance of specific components of options counselling, standardised patient-perceived comfort with options counselling, student-perceived performance, and impressions of NPI patient-facilitators **Measures assessed by:** standardised patient
Notes	**Conflicts of interest:** not stated **Funding:** this study received funding from the Society of Family Planning and The University of Colorado Academy of Medical Educators, Rymer Small Grants Program **Abbreviations:** C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; NA: not applicable; NPI: novel patient interaction; NR: not reported; OSCE: objective structured clinical examination; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Shaddheau 2015 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	SP assessors were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Groups were demographically similar regarding sex, age, and religiosity. Intervention group had significantly higher percentage of Caucasian students compared to control group
Recuitment Bias (cluster trials only)	Low risk	All students were already allocated to rotations before rotations were randomised. All students participated
Selective recruitment (cluster trials only)	Unclear risk	Unclear at what point randomisation took place
Loss of clusters (cluster trials only)	Low risk	No clusters lost
Statistical methods (cluster trials only)	High risk	No adjustment for clustering reported - trial authors did not treat the study as a cluster trial and therefore did not adjust for clustering
Herd effect (cluster trials only)	Unclear risk	Outcomes not reported by cluster (rotation), unable to assess

Shapiro 2009

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** ethics approval granted through university, and written informed consent obtained **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA
Participants	**Student level:** first year **Setting:** unclear **% recruitment:** NR **Number of subjects randomised:** 95

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Shapiro 2009 *(Continued)*

	Number of subjects participating: - Int: 38 - Ctrl: 41 - Total: 75 (13 dropped out and 2 did not receive the intervention) **Age (mean):** 23.1 **Sex:** 39% male **Ethnicity:** 67% Canadian born **Country:** Canada
Interventions	**Number of experimental conditions:** 2 **Aim:** to assess the effectiveness of TCom in improving first-year medical students' communication skills. The programme strives to accomplish 3 goals: first, to increase the ability of students to interact effectively with medical patients; second, to strengthen their skills in eliciting, understanding, and utilising the various types of psychological information available in a doctor-patient relationship; and, third, to enhance their curiosity, tolerance, and comfort in dealing with a variety of patients and with different kinds of symptoms, emotions, attitudes, and behaviours **Comparison:** TCom offers first-year medical students the opportunity to meet weekly with patients on a 1-to-1 basis for 4 months while receiving group supervision from a faculty psychiatrist **Intervention delivered by:** real patients and psychiatrists **Control:** waitlist **Theoretical basis:** NR **Duration and timing:** weekly sessions for 3 months **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** participants interviewed 2 SPs at pre-intervention and post-intervention times and at the end of the academic year **Assessment timing:** post-intervention (within academic year) **Primary measures:** communication skills during interviews were rated by SPs using the ISRS. There are 7 items, with a Likert scale of 1 (strongly disagree) to 7 (strongly agree). Examples include the following: "the doctor wanted to understand how I saw things" and "the doctor just took no notice of some things that I thought or felt" **Other measures:** participants completed the SAICQ, a 40-item scale, with a Likert scale of 1 (I'm poor at this) to 5 (I'm extremely good at this). The SPIR, designed to assess participants' expressed empathy based on their written open-ended responses to a series of 24 statements made by hypothetical patients, was administered. Responses to these items were rated by trained external raters according to a manual that describes in detail how to classify responses into disengaging and engaging sets **Measures assessed by:** SP
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** funding was received from the Medical Research Council of Canada (now the Canadian Institutes of Health Research) and the Association of Canadian Medical Colleges. **Abbreviations:** Ctrl: control; Int: intervention; ISRS: Interpersonal Skills Rating Scale; NA: not applicable; NPI: novel patient interaction; NR: not reported; RCT: randomised controlled trial; SAICQ: Self-Assessment of Interpersonal Competence Questionnaire; SP: simulated patient; SPIR: Staff-Patient Interaction Rating Scale; TCom: Therapeutic Communication Program

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Shapiro 2009 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding not possible, but waitlist control design not likely to influence outcomes
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Not clear whether SPs were blinded to students' group allocation. External raters of students' written responses were, but this measure was not used in this review
Incomplete outcome data (attrition bias) All outcomes	High risk	Some dropouts and imbalance in dropout rates
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Control group was assessed before the intervention group; therefore contamination is unlikely Quote: "the study used a repeated measures design with a waiting list control group: students were randomly assigned to groups starting the educational intervention in either September (N = 38) or February (N = 41), with the latter being used as a control for the former. Communication skills were assessed at the pre- and post-intervention times and at the end of the academic year from the perspectives of student, standardized patient and external rater"
Other Bias: Baseline outcome measurement similarity	Low risk	Groups were equivalent across all measures at baseline
Other Bias: Baseline characteristics similarity	Low risk	No significant differences between control and intervention group demographics

Siassakos 2010

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** class/group **Ethics and informed consent:** ethical approval granted and informed consent obtained **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** maternity unit of large teaching hospital **Percentage recruitment:** 100% **Number of participants randomised:** - Int: 12 - Ctrl: 12 - Total: 24 **Number of participants participating:** - Int: 11 - Ctrl: 9 - Total: 20 **Age:** NR **Sex:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Siassakos 2010 *(Continued)*

	Country: UK
Interventions	**Number of experimental conditions:** 2 **Aim:** to investigate whether students' communication skills during simulated emergencies, as assessed with the PPS, are improved by rehearsals that include patient-actors **Content:** all participants attended an SIM on management of SD that included practical demonstration using a high-fidelity mannequin. All information given to participants verbally was derived from national guidelines. For HYB (intervention group), the actor was integrated with a part-task trainer; she was kneeling behind the pelvic model. Training materials include specific instructions to the patient-actor on how to interact with learners and when to provide verbal cues. In addition to the standardised script, the manual contains instructions on how to provide feedback to learners about aspects of communication that are perceived as good or as in need of improvement **Intervention delivered by:** unclear (actor/tutor) **Comparison:** small-group training with no communication teaching showed again how to manage SD, having had already attended the initial SIM. An MS PowerPoint presentation was used that focused on theory and practical management and was developed from standardised evidence-based training materials. PowerPoint slides contained images depicting correct management of SD and which errors to avoid, based on findings from 450 simulations **Theoretical basis:** NR **Duration and timing:** 2 sessions within 8-week rotation **Fidelity:** standardised training materials used **Adherence:** NR
Outcomes	**Assessment:** participants from both groups were asked to manage an SD simulation scenario while communicating with a patient-actor, again kneeling behind the pelvic model. A healthcare professional played the role of the partner. Communication skills were assessed by an experienced assessor blinded to the intervention group **Assessment timing:** 3 to 10 days after intervention **Primary measures:** communication skills throughout physical procedure (relationship building; planning and explaining), assessed using PPS - For each drill, the patient-actor assessed the quality of communication using three 5-point Likert scales (5, strongly agree; 4, agree; 3, neutral; 2, disagree; 1, strongly disagree): communication – "I felt well informed due to good communication"; respect – "I felt I was treated with respect at all times"; safety – "I felt safe at all times" **Measures assessed by:** SP **Others:** all participants also completed a survey to assess their reaction to the training intervention. Force applied was also measured using a transducer
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; HYB: hybrid; Int = intervention; NR = not reported; PPS: Patient Perception Scale; RCT: randomised controlled trial; SD: shoulder dystocia; SIM: simulation; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomised by electronic random numbers generator
Allocation concealment (selection bias)	Unclear risk	Not adequately reported. Unclear re-timing of randomisation
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Siassakos 2010 *(Continued)*

Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors blinded to student group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition low and balanced between groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported

Snow 2016

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional approval and consent **Intention-to-treat analysis:** no **Adjustment for clustering:** NA
Participants	**Student level:** fifth year **Setting:** teaching hospital (classroom) **Percentage recruitment:** 65% **Number of subjects randomised:** 88 **Number of subjects participating:** - Int: 45 - Ctrl: 43 - Total: 88 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** UK, England
Interventions	**Number of experimental conditions:** 2 **Aim:** to explore the effect on student performance of viewing video of untrained laywomen simply describing their experiences of investigations for cervical abnormalities **Content:** online modules contained identical interactive online written and visual content about cervical screening and clinical management of cervical abnormalities. Each one also included a series of video clips featuring case histories of 2 patients. In the experimental group (''patient module''), the clips were of the 2 patients themselves describing their personal experience of the procedures. The modules were identical, apart from the perspectives given in the video clips **Intervention delivered by:** online modules including recorded patient stories **Comparison:** ''doctor module'' - the clips showed a female doctor recounting patients' case histories and outlining relevant procedures from a medical viewpoint **Theoretical basis:** patient and public involvement in education (Towle 2010) **Duration and timing:** 20 minutes, 1-off **Fidelity:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Snow 2016 *(Continued)*

	Adherence: completed module
Outcomes	**Assessment:** OSCE set in general practice with task to see patient who has had an abnormal smear and has been invited for colposcopy **Assessment timing:** immediately after viewing module (same day) **Primary measures:** DISQ adapted for students (includes items based on students' ability to explore the patient's concerns, explain accurately what a colposcopy is and how it is performed, and answer the patient's questions about time scales for receiving results). 12 DISQ item scores were averaged to calculate an overall ISI, expressed as a percentage. Global OSCE score was given by the examiner **Other measures:** MCQ and student reported experiences **Measures assessed by:** OSCE assessed by examiner on a global scale and by SP using the DISQ
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** this project was funded by the Nuffield Oxford Hospital Trust **Abbreviations:** Ctrl: control; DISQ: Doctors' Interpersonal Skills Questionnaire; Int: intervention; ISI: Interpersonal Skills Index; MCQ: multiple choice question; NA: not applicable; NPI: novel patient interaction; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Randomised by researcher (who was independent of students, and had not previously met them) consecutively assigning them to A or B with no knowledge of how the modules differed. This researcher played no further part
Allocation concealment (selection bias)	Low risk	Students and researcher conducting allocation had no knowledge of differences between modules, and researcher did not know students
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Intervention was computer-based, so blinding of personnel was not relevant; lack of blinding of participants not likely to lead to bias
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors and SPs were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Low risk	Very low attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Both control and intervention groups were online modules, and assessment occurred immediately after completion of the module. Therefore, very unlikely that contamination could occur
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Solomon 2004

Study characteristics	
Methods	**Design:** quasi-C-RCT **Unit of randomisation:** rotation **Ethics and informed consent:** unclear (students 'offered an opportunity to participate') **Intention-to-treat analysis:** NA **Adjustment for clustering:** no
Participants	**Student level:** third year **Setting:** teaching hospital **Percentage recruitment:** NR **Number of subjects randomised:** NR **Number of subjects participating:** - Int 1 (4 weeks after PDM module): 5 - Ctrl: 17 - Total: 47 (only intervention 1 and control comparison included here) **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 3 **Aim:** to improve students' skills in incorporating PDM elements in informed decision-making consultations with patients **Content:** the family practice clerkship at CHM implemented an educational module to enhance skills in common preventive services and to teach students how to inform and involve patients concerning those services. Students are trained to use a variation of a PDM model developed by Braddock and colleagues. The model includes 7 elements of informed decision-making: (1) discussion of the patient's role in decision-making; (2) discussion of the clinical issue or nature of the decision; (3) discussion of alternatives; (4) discussion of the pros and cons of the alternatives; (5) discussion of uncertainties associated with the decision; (6) assessment of the patient's understanding; and (7) exploration of patient preference. Specific decisions are also categorised into "basic", "intermediate", and "complex". PDM is presented and modelled by faculty, discussed in small groups, and practised using web-based case simulations. Students must pass an evaluative SPE on shared decision-making and preventive services before completion of the clerkship Intervention 1: completion of PDM 4 weeks before SPE Intervention 2: completion of PDM 12 weeks before SPE **Intervention delivered by:** unclear ('faculty') **Comparison:** no PDM completion before SPE **Theoretical basis:** Braddock's model of informed decision-making (Braddock 1999) **Duration and timing:** unclear **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** the SPE station is structured as an office appointment for a man in his early twenties with a first-time seizure. Before meeting the SP, students are given a "patient file" providing the results of a history and physical examination. To ensure adequate knowledge on the part of students, they are provided with a fact sheet about seizures. Students are given approximately 5 minutes to review the material. Each student then spent up to 20 minutes discussing treatment options and the implications of the seizure with the SP while observed by a faculty member **Assessment timing:** 4 or 12 weeks after PDM module **Primary measures:** 11-item rating scale used in evaluation; PDM scale is based on items 1 to 4, 6, and 9: (1) appropriately greets patient; (2) establishes the purpose of the encounter; (3) clarifies roles of patient and physician in decision-making and gains patient's permission to continue the discussion; (4) presents issues in terms the patient can comprehend; (5) reviews available test results; (6) discusses alternatives (treatment vs no treatment) and their pros and cons; (7) discusses safety issues relating to new seizures; (8) counsels re: driving restriction under Michigan state law; (9) explores patient preferences; (10) ensures patient understands the issues; (11) responds appropriately to affect

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Solomon 2004 *(Continued)*

Other measures: none
Measures assessed by: unclear ('faculty')

Notes

Conflicts of interest: not stated

Funding: this study was funded in part by a grant from the Division of Medicine, Bureau of Health Professions, Health Services and Resources Administration

Abbreviations: C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; NPI: novel patient interaction; PDA: personal digital assistant; PDM: participatory decision-making; SPE: simulated patient experience

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-RCT. Randomised by rotation
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Because faculty members rating students did not know a particular student's rotation schedule, and students were not informed at the time of the SPE of the specifics of the study, the design is essentially double-blinded
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blinded to group assignment
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	No information on recruitment rates or attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported
Recuitment Bias (cluster trials only)	Low risk	All students were in groups that were then randomly assigned to intervention or control
Selective recruitment (cluster trials only)	Low risk	Rotation structure provided a natural experiment, with lack of allocation concealment; not likely to pose risk; not adequately reported
Loss of clusters (cluster trials only)	Low risk	No loss of entire clusters
Statistical methods (cluster trials only)	High risk	No adjustment for clustering - study not treated as cluster trial (although reference is made once to cluster-randomisation)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Solomon 2004 *(Continued)*

Herd effect (cluster trials only)	Unclear risk	Unable to determine if herd effect exists

Spollen 2010

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** institutional review board approval and participants 'agreed' to participate. **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** second year **Setting:** university campus **Percentage recruitment:** NA, first 40 volunteers (of 156 students) were recruited **Number of subjects randomised:** 40 **Number of subjects participating:** - Int: 17 - Ctrl: 18 - Total: 35 **Age (mean):** 28.1 Sex: 63.0% male **Ethnicity:** all Caucasian apart from 1 Asian **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve students' skills in behaviour change counselling **Content:** the workshop employed the use of both didactic instruction and role-play. At the start of the workshop, students were asked to divide into groups of 2 or 3 and to complete a brief role-play exercise in 'how not to' perform behaviour change counselling, using the 'Persuasion Exercise'. Following this, students received a didactic presentation on the theory and practice of behavioural change counselling. Concepts were divided into 3 sections: 'Develop an Agenda', 'Explore Behaviour and Thoughts about Change', and 'Negotiate a Change Plan'. After each section was reviewed, a brief role-play allowed the small group of students to practice their skills. After completion of all 3 sections and accompanying role-plays, a complete behavioural change counselling role-play was performed by each of the students **Intervention delivered by:** a psychiatrist trained in behaviour change counselling methods **Comparison:** delayed access to workshop **Theoretical basis:** motivational interviewing theory (Miller 2002) **Duration and timing:** 2 hours, 1-off **Fidelity:** NR **Adherence:** all students participated in the workshop
Outcomes	**Assessment:** recorded SPE **Assessment timing:** 2 days after workshop **Primary measures:** BECCI measures: invite patient to talk about behaviour change, demonstrate sensitivity talking about other issues, encourage patient to talk about current behaviour or status quo, encourage patient to talk about change, ask questions to elicit how patient thinks and feels about the topic, use empathic listening statements, use summaries to bring together what the patient says, acknowledge challenges about the behaviour that the patient faces, provide information sensitive to patient concerns and understanding, actively convey respect for patient choice about behaviour change, practitioner and patient exchange ideas about how patient could change current behaviour. BECCI practitioner score

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Spollen 2010 *(Continued)*

	Other measures: attitudes (measured as responses to video stimulus), knowledge (measured using 6 MCQs and skills (using the BECCI)) **Measures assessed by:** 3 blinded assessors who had minimal or no prior contact with students, and who were blinded to group assignments and timing of the SPE (pre/post), reached consensus for each recording
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** not stated **Abbreviations:** BECCI: Behaviour Change Counselling Index; Ctrl: control; Int: intervention; MCQ: multiple choice question; NR: not reported; RCT: randomised controlled trial; SP: simulated/standardised patient; SPE: standardised patient experience

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Not blinded, but all students received the workshop, so lack of blinding unlikely to impact delivery
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors and SPs blind to student group assignment and had minimal previous contact with students
Incomplete outcome data (attrition bias) All outcomes	Low risk	No attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	No differences between groups at baseline
Other Bias: Baseline characteristics similarity	High risk	Not reported

Steifel 2013

Study characteristics

Methods	**Design:** RCT **Unit of randomisation:** student

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Steifel 2013 *(Continued)*

	Ethics and informed consent: unclear **Intention-to-treat analysis:** no **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fourth year **Setting:** classroom **Percentage recruitment:** 96% **Number of subjects randomised:** 124 **Number of subjects participating:** 119 - Int: 34 - Ctrl: 31 - Total: 65 in evaluation **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** Switzerland
Interventions	**Number of experimental conditions:** 2 **Aim:** (1) to evaluate how individual vs group training in breaking bad news is perceived by students, and (2) to gain information on students' communication skills and training needs **Content:** individual training with students conducting an interview with an SP. After conducting the videotaped interview, each student visualised it during the following days, before meeting with faculty during a 1-hour supervision of the student. The scenarios consisted of breaking bad news to a middle-aged woman diagnosed with stage 1 breast cancer and to a middle-aged man with stomach cancer. Discussion of videotaped interviews focused on (1) structure of the interview, (2) exchange of information, (3) response to emotions, and (4) relational aspects of communication and specific elements concerning breaking bad news **Intervention delivered by:** 'faculty' – 2 study authors **Comparison:** group training consisted of two 2-hour sessions with 12 students. During each session, 1 of the students conducted a videotaped interview with an SP (conducting group), which was observed by other students (observing group) and then was discussed with faculty while the video was replayed. The same scenarios were used for both group and individual training. Discussion during group training focused on the same elements as during individual training **Theoretical basis:** NR **Duration and timing:** 2 sessions (up to 4 hours total) **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** videotaped SP interview **Assessment timing:** unclear

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Steifel 2013 *(Continued)*

	Primary measures: students' communication skills in breaking bad news using the RIAS - 42 mutually exclusive and exhaustive coding categories reflecting the content and form of the medical interaction; categories reflect content - exchanges about medical condition, treatments, lifestyle behaviours, and psychosocial issues - and form - informative, interrogative, persuasive, affective, and process-oriented statements - of the medical interaction **Other measures:** students' perceptions of group and individual training **Measures assessed by:** a coder trained in the RIAS
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** Fonds d'Innovation Pedagogique (FIP) of the University of Lausanne **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial; RIAS: Roter Interaction Analysis System; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Unclear whether coder was blinded
Incomplete outcome data (attrition bias) All outcomes	High risk	Data not available for all who participated in training and apparent differences between groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported

Stolz 2012

Study characteristics

Methods	**Design:** RCT

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Stolz 2012 *(Continued)*

	Unit of randomisation: student **Ethics and informed consent:** the study was exempt from ethics approval; students were invited to take part and were instructed about the protocol, design, and objectives of the study **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** online and classroom **Percentage recruitment:** 92% **Number of subjects randomised:** 140 **Number of subjects participating:** - Group A: 35 - Group B: 30 - Group C: 31 - Group D: 33 - Total: 129 **Age (mean):** NR **Sex:** 32.6% male **Ethnicity:** NR **Country:** Switzerland
Interventions	**Number of experimental conditions:** 4 **Aim:** to evaluate the efficiency of a self-directed web-based learning module on SCC skills **Content:** between 2 and 4 weeks after gaining access to the web-based learning module on smoking cessation, each group was given 1 of the following 2-hour supervised smoking cessation education settings - Group A - "web-based training" using the newly designed web-based learning module Group B - "lecture" - instruction through a faculty preceptor including demonstration of video material followed by group discussion Group C - "patient-centred counselling through role-play in small groups" - 1 medical student acted as a smoker and another acted as the physician, following a standardised case description Group D - supervised "interaction with real patients" - counselling encounters with real patients who smoke and took part in a smoking cessation intervention programme for university hospital employees **Intervention delivered by:** online and ‘faculty’ **Comparison:** 4 groups – comparative effectiveness **Theoretical basis:** NR **Duration and timing:** 2 hours, 1-off **Fidelity:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Stolz 2012 *(Continued)*

	Adherence: the amount of time applied by students for self-directed study of the website was documented. Students randomised to lecture, role-playing, and critique interaction with real patients' groups spent a median time of 30 minutes on the web-based module, and students randomised to the web tool used it for a total of 40 minutes (P = 0.03)
Outcomes	**Assessment:** 12-item OSCE **Assessment timing:** within 6 weeks of teaching sessions **Primary measures:** SCC skills **Other measures:** theoretical knowledge (20 MCQs), student satisfaction, student self-assessment **Measures assessed by:** a respiratory or primary care physician and a didactic assistant trained in smoking cessation
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** University Hospital Basel & Swiss National Foundation & European Respiratory Society School Educational Research Grant **Abbreviations:** MCQ: multiple choice question; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; SCC = smoking cessation counselling

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-based block randomisation scheme
Allocation concealment (selection bias)	Low risk	Randomisation occurred after all had access to online module
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	No information about whether evaluators were aware of group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition; not different between groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	High risk	Not reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Strayer 2010

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** class/group **Ethics and informed consent:** institutional review board approval; informed consent unclear **Intention-to-treat analysis:** yes **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** teaching hospital **Percentage recruitment:** 92% **Number of subjects randomised:** 122 **Number of subjects participating:** - Int: 53 - Ctrl: 63 - Total: 122 **Age (mean):** 25.8 **Sex:** 52% male **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to determine if a PDA-based tool can improve medical student SCC **Content:** the workshop consisted of lecture-discussion reviewing SCC techniques, including the 5 A's, principles of MI, stages of change assessment, and stages of change-guided interventions, as well as role-play practice with peers. Students in both groups received a paper-based summary of MI techniques relating to SCC following the workshop. Students in the intervention group also had the E-SMOKE-I.T. software loaded onto their required PDA. Following the MI workshop, they received 15 minutes of instruction from the study co-ordinator on the contents, organisation, and use of E-SMOKE-I.T. software. The E-SMOKE-I.T. tool was designed to operationalise the 5 A's, stages of change, and MI and was adapted from earlier versions designed for practising physicians. The software helps users determine patients' stages of change, provides scripted motivational interviews targeted to their stage, and makes relevant health behaviour and stage-based interventions immediately accessible **Intervention delivered by:** unclear **Comparison:** students in the comparison group only received the paper-based summary of MI techniques **Theoretical basis:** 5 A's, stages of change, motivational Interviewing (Fiore 1994; Prochaska 1983) **Duration and timing:** 1-off session, duration NR **Fidelity:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Strayer 2010 *(Continued)*

	Adherence: NR
Outcomes	**Assessment:** SP interview **Assessment timing:** end of clerkship (1 month after intervention) and end of the academic year **Primary measures:** SCC behaviours; to be classified as having met the criteria for "correct" SCC assistance", students had to (1) assess patients' readiness to quit smoking, (2) advise patients to quit smoking, including personalising risks, and (3) assist patients with smoking cessation through stage-appropriate counselling. This represents 3 of the 5 A's **Other measures:** comfort and knowledge; satisfaction with the PDA tool, usability, and barriers to use **Measures assessed by:** 2 independent raters
Notes	**Conflicts of interest:** not stated **Funding:** American Cancer Society and Department of Health & Human Services awards **Abbreviations:** Ctrl: control; Int: intervention; MI: motivational interviewing; NA: not applicable; NR: not reported; PDA: personal digital assistant; RCT: randomised controlled trial; SCC: smoking cessation counselling; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Randomised students by paired clerkship blocks
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. Independent raters blinded to the nature of the intervention
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition, balanced across groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measure
Other Bias: Baseline characteristics similarity	Low risk	No significant differences between control and intervention groups except that the control group reported increased pre-intervention use of stages of change

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Taylor 2019

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** time (year) **Ethics and informed consent:** institutional review board approval and informed consent obtained **Intention-to-treat analysis:** yes **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** second year **Setting:** classroom (university campus) **Percentage recruitment:** NR **Number of subjects randomised:** NR **Number of subjects participating:** - Int: 148 - Ctrl: 182 - Total: 330 **Age (mean):** 19.94 **Sex:** NR **Ethnicity:** NR **Country:** Australia
Interventions	**Number of experimental conditions:** 2 **Aim:** to determine whether a volunteer SP method translates into improved OSCE performance in communication skills compared to RP **Content:** each student participated in 3 campus-based clinical skills sessions per course, with the last session of each course dedicated to developing communication skills utilising SP interactions relevant to the course content **Intervention delivered by:** NR **Comparison:** peer role-play interaction **Theoretical basis:** NR **Duration and timing:** 3 sessions in each of 7 courses across first 2 years **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** 6 station OSCEs with an identical mix of stations across both years. Each station therefore required candidates to take a history (clinical communication skills), perform an examination or procedural skill (examination/procedural skills), and provide an explanation or advice to the patient (generic communication skills)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Taylor 2019 *(Continued)*

Assessment timing: End of year 2 - not clear how long from intervention delivery

Primary measures: each criterion at each OSCE station was graded as fail (F), borderline (P-), clear pass (P), or pass with distinction (P+):

(1) Initiate and end the consultation: greet patient, introduce self, outline agenda, seek permission to proceed, thank the patient, and offer help with repositioning, dressing, etc.

(2) Listen attentively, engage patient, and maintain respect: allow patient to use his or her own words without premature interruption, use open and closed questions, reflect important feelings, pick up verbal and non-verbal cues, display sensitivity to patient's needs, respect boundaries, and gain patient's trust

(3) Elicit a relevant clinical history: establish reason for presentation, course, and nature of symptoms; summarise patient's symptoms to check understanding

(4) Elicit a psychosocial history: ask patient about relevant family, social support, cultural and lifestyle factors, employment issue, as appropriate

(5) Gather relevant past medical and family history: ask about past personal and family history, as well as specific risk factor history when appropriate

(6) Communicate with patient and ensure patient comfort when conducting a physical examination/skill: explain to patient what is being done, provide suitable instructions, ensure the patient's privacy and comfort

(7) Summarise case findings: should use medical jargon, identify patient's key concerns and reason for presenting, and summarise relevant history and examination findings

Other measures: (8) Perform technically competent physical examination or skill (1): correct positioning of patient, adept with equipment, competent approach to examination

(9) Perform technically competent physical examination or skill (2): correct positioning of patient, adept with equipment, competent approach to examination

Measures assessed by: clinicians familiar with the examination format who had received identical OSCE assessment training in the form of reading material and a verbal briefing

Notes

Conflicts of interest: declaration of no conflicts of interest

Funding: no funding received

Abbreviations: Ctrl: control; Int: intervention; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; RP: role-play; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Not randomised - based on year of enrolment BUT all data collection analysis was retrospective
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Retrospective study, so those participating were unaware of the study at the time
Blinding of outcome assessment (detection bias)	Low risk	Retrospective study - data were collected separately from the study itself

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Taylor 2019 *(Continued)*

All outcomes		
Incomplete outcome data (attrition bias) All outcomes	High risk	Only those with complete data were included
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods
Other Bias: Contamination	Low risk	Control group was assessed before the intervention group; therefore contamination bias is unlikely Quote: "the cohort of students who engaged in RP (*CONTROL)* enrolled onto the medicine program in 2013 and took an OSCE at the end of their second pre-clinical year in 2014; similarly, the SP cohort of students (*INTERVENTION)* enrolled in 2015 and were examined in 2016"
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measures
Other Bias: Baseline characteristics similarity	Low risk	Significant difference in UMAT score only, but unlikely to influence results (would reduce intervention effect if anything)

Vanatta 1996

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** class/group **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** first year **Setting:** classroom (university campus) **Percentage recruitment:** all students participated, but complete data available for 78% **Number of subjects randomised:** 154 **Number of subjects participating:** - Group A: 43 - Group B: 41 - Group C: 36 - Total: 120 **Age (mean):** NR **Sex:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Vanatta 1996 *(Continued)*

	Ethnicity: NR **Country:** USA
Interventions	**Number of experimental conditions:** 3 **Aim:** to compare the interviewing skills of first-year medical students receiving feedback primarily from standardised patients (SPs) with skills of students receiving feedback primarily from faculty **Content:** Group A: students were instructed by faculty, practised interviewing with SPs, and received feedback primarily from SPs. Students' peers also gave feedback. Faculty were absent for the practice and feedback Group B: students were instructed by faculty, practised interviewing with SPs, and received feedback primarily from faculty with added feedback from peers. SPs were absent during feedback **Intervention delivered by:** 'faculty' (n = 16) who were MDs and full-time clinician educators (10 from family medicine, 3 each from internal medicine and paediatrics) **Comparison:** Group C: students practised role-play (i.e. no practice with SPs) and received feedback from faculty and peers. Group C was included to detect any differences between students practising with SPs and those practicing role-plays **Theoretical basis:** NR **Duration and timing:** 2 × 4-hour sessions **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** videotaped SP interaction **Assessment timing:** within 2 weeks of the intervention **Primary measures:** open-ended Q's, empathy skills, facilitative cues measured using the ACIRS and a modified RIAS **Other measures:** none **Measures assessed by:** trained laypersons
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** ACIRS: Arizona Clinical Interview Rating Scale; NR: not reported; RCT: randomised controlled trial; RIAS: Rotor Interactional Analysis System; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Vanatta 1996 *(Continued)*

Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. Raters were blind to study group and to whether an interview occurred pre- or post-intervention
Incomplete outcome data (attrition bias) All outcomes	Low risk	Some data loss and incomplete data, but no apparent differences across groups
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Students received the same intervention; difference was who provided feedback
Other Bias: Baseline outcome measurement similarity	Unclear risk	Only interrater reliability presented (not group differences in performance)
Other Bias: Baseline characteristics similarity	High risk	Not reported

Wagner 2011

Study characteristics	
Methods	**Design:** quasi-C-RCT **Unit of randomisation:** time (year of enrolment) **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NR **Adjustment for clustering:** no
Participants	**Student level:** first year **Setting:** online (university campus) **Percentage recruitment:** 100% **Number of subjects randomised:** 164 **Number of subjects participating:** - Int: 76 - Ctrl: 88 - Total: 164 **Age (mean):** NR **Sex:** NR

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wagner 2011 *(Continued)*

	Ethnicity: NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to evaluate effects of online, interactive video modules on students' data gathering, behavioural counselling, and communication skills with SPs **Content:** several online, interactive video modules were developed by medical and dental faculty and staff in the Faculty Instructional Technology Services at UConn. The curriculum was sequenced so that modules that demonstrate specific skills were assigned before the clinical skills assessment that measured those same skills. Each module consisted of an approximately 5-minute video lecture in which the viewer was provided information about the patient and the videotaped interviewer's task, followed by a 20-minute demonstration of a clinical encounter between a skilled interviewer and an SP. Skilled interviewers were chosen to demonstrate effective skills but were encouraged to not be "perfect" gold standard. Both faculty and fourth-year students who had previously demonstrated good communication skills were used to demonstrate competence. Modules demonstrated specific skills sets for which faculty felt there was a need for more expert modelling: smoking cessation counselling with an adult who expressed readiness to quit and a well child history for a 15-month-old. After viewing the modules, students rated the videotaped interviewer using (a) a case-specific content checklist, and (b) the MIRS. After rating the videotaped interviewer with the MIRS and the content checklist, students were able to view concordance between their rating and a criterion established by consensus among faculty **Intervention delivered by:** online (recorded consultations conducted by faculty and fourth-year students) **Comparison:** students from the 2009/2010 graduating classes (collected in 2007) were not assigned the modules. They attended a didactic lecture **Theoretical basis:** social cognitive theory (Bandura 1986) **Duration and timing:** 2 modules of approximately 25 minutes each **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** live encounter with SP **Assessment timing:** within 1 to 2 weeks of completing the modules **Primary measures:** history taking, behavioural counselling, and communication skills (organisation, transition statements, avoiding jargon, non-verbal facilitation, encouraging questions) measured using the MIRS; each item is rated on a 5-point Likert-type scale with anchors at 1, 3, and 5. Total score is the mean of all items, with higher scores indicating better skills. The MIRS is a revised version of the Arizona Clinical Interview Rating Scale **Other measures:** the 15- to 25-item content checklist of specific pre-determined pieces of information that the student must gather from or communicate to the SP. Items on the checklist are scored by the SP as "yes" or "no". The total content checklist score is calculated as % of item completed by the student, with higher scores indicating better performance **Measures assessed by:** SP
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** Kaiser Permanente Endowment **Abbreviations:** Ctrl: control; Int: intervention; MIRS: Master Interview Rating Scale; NR: not reported; RCT: randomised controlled trial; SP: standardised patient

Risk of bias

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wagner 2011 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-experimental. Performance data from a cohort without intervention compared to a cohort assigned to intervention
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	SP assessors were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Attrition not reported
Selective reporting (reporting bias)	Low risk	No protocol evident. Measures seem to match methods
Other Bias: Contamination	Low risk	Data from the control group (2007) were collected before data from the intervention group (2008); therefore contamination is unlikely
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measures
Other Bias: Baseline characteristics similarity	High risk	Not reported
Recuitment Bias (cluster trials only)	Unclear risk	Unclear at what stage, if at all, students were asked to participate
Selective recruitment (cluster trials only)	Low risk	Consecutive cohorts of students, and all students participated
Loss of clusters (cluster trials only)	Low risk	No clusters lost
Statistical methods (cluster trials only)	High risk	Study authors do not treat the study as a C-RCT and therefore do not adjust
Herd effect (cluster trials only)	Unclear risk	Unable to assess

Walsh 1999

Study characteristics	
Methods	**Design:** quasi-RCT **Unit of randomisation:** student, then rotation

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Walsh 1999 *(Continued)*

	Ethics and informed consent: institutional ethics committee approval and written informed consent **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** fifth (final) year **Setting:** classroom (teaching hospital) **Percentage recruitment:** 76% **Number of subjects randomised:** 55 **Number of subjects participating:** - Int: 27 - Ctrl: 26 - Total: 55 **Age (mean):** 25.2 **Sex:** 25.5% male **Ethnicity:** 76.4% Australian born **Country:** Australia
Interventions	**Number of experimental conditions:** 2 **Aim:** to examine the relative effectiveness of 2 different methods of teaching alcohol brief intervention skills to medical students: an experiential method including videotaped practice and tutor feedback, and a lecture method including a videotaped demonstration **Content:** both groups participated in a 1-hour lecture and demonstration. A 15-minute videotape was shown depicting an intervention that modelled the recommended steps. In addition, during the week following this session, intervention students were asked to make a videotape (maximum 20 minutes) of themselves providing counselling to a simulated patient (usually another student) using a scenario contained in the written package. During a 1.5-hour small-group session, a tutor used these videotapes to help the group provide critical feedback on students' performance. A rating scale in the handout was used as a guide **Intervention delivered by:** Tutors who were members of the faculty staff with medical or behavioural science backgrounds and were chosen primarily because of their expertise in interactional skills rather than alcohol intervention. **Comparison:** 1-hour lecture and demonstration with no videotaped practice and feedback component **Theoretical basis:** traditional vs experiential teaching methods (Walsh 1995). Study authors created their own framework based on experience as well as guidelines proposed in the literature on communication for end of life care (http://epec.net/EPEC/webpages/index.cfm) **Duration and timing:** 3 hours across 2 sessions and video development **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** videotaped OSCE role-play with simulated patient **Assessment timing:** 4 weeks after intervention

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Walsh 1999 *(Continued)*

	Primary measures: alcohol-related information gathering and explanation and planning (89 items), and overall interview performance (18 items) **Other measures:** self-reported alcohol knowledge and attitudes **Measures assessed by:** 3 independent raters who had demonstrated consistency and observational accuracy in another study
Notes	**Conflicts of interest:** not stated **Funding:** this study was funded by a grant from the Committee on Alcohol and Drug Education in Medical Schools **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quasi-RCT. Two blocks allocated - randomisation not clear
Allocation concealment (selection bias)	High risk	Allocation concealment not reported, but due to quasi-RCT design, it is likely 'high risk'
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment. Raters were independent and were blinded to participants' group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition low (but some differences between groups)
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods. Reported at summary level only
Other Bias: Contamination	Low risk	Contamination possible but not likely to be important: groups were largely physically separate for the study period; second, the control group knew that its programme was delayed until after the post-test and was not withheld completely, and, third, students in their busy final year are unlikely to attach considerable importance to research involving teaching and assessment methods with which they were already familiar
Other Bias: Baseline outcome measurement similarity	Low risk	No significant differences at baseline
Other Bias: Baseline characteristics similarity	Low risk	Differences not reported as significant

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Weihs 1986

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** unclear **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** first year **Setting:** teaching hospital **Percentage recruitment:** 100% **Number of subjects randomised:** 32 **Number of subjects participating:** - Int: 16 - Ctrl: 16 - Total: 32 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to study the effect of a structured 10-week course on development of interviewing skills of first-year medical students **Content:** skills were explained didactically and were demonstrated by the instructor interviewing a patient during the first 2 sessions. Skills were then practised by group members interviewing one another, and feedback was given by other group members. The skills of attending, responding to content, and responding to feeling were emphasised during skills practice. Instruction was designed for mastery of elementary skills before work on more advanced skills. When students could consistently use the more basic skills, they were invited to practise the skill of personalising. Students then interviewed patients on general medical services for the next 8 weeks about their reason for hospitalisation and their situation at the time of the interview. Students were encouraged to learn about the medical problem as well as the personal implications of the illness for the patient. At the group meeting at the end of each session, student colleagues offered feedback about interviewer behaviour as well as about interview content **Intervention delivered by:** 2 family physicians **Comparison:** patient interviews were followed by discussions among all group members about the content of the interview and its impact on the interviewer. Specific interviewing skills were addressed as they arose in discussion. No specific model for skills acquisition was used **Theoretical basis:** Carkhuff's Developmental Helping Model (Garfield 1978) **Duration and timing:** 3-hour weekly sessions for 10 weeks

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Weihs 1986 *(Continued)*

	Fidelity: NR **Adherence:** NR
Outcomes	**Assessment:** 15-minute videotaped SP interview **Assessment timing:** final session of 10-week programme **Primary measures:** interviewing skills (attending, responding to content and feeling, personalising, initiating) **Other measures:** none **Measures assessed by:** 3 graduate students in a doctoral programme in counselling psychology
Notes	**Conflicts of interest:** not stated **Funding:** not stated **Abbreviations:** Ctrl: control; Int: intervention; NR: not reported; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Assigned to group by random numbers table
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors were blind to group assignment
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Attrition not reported
Selective reporting (reporting bias)	High risk	Only pooled measures reported in full
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Low risk	No significant differences
Other Bias: Baseline characteristics similarity	High risk	Not reported

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Windish 2005

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** groups **Ethics and informed consent:** the Johns Hopkins Institutional Review Board approved the study protocol. No information on informed consent was provided **Intention-to-treat analysis:** not reported. **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** second year **Setting:** university campus **Percentage recruitment:** 100% **Number of subjects randomised:** 121 **Number of subjects participating:** - Int: 59 - Ctrl: 61 - Total: 120 **Age (mean):** 24.4 (intervention), 24.1 (control) **Sex:** 51% male (intervention), 50% male (control) **Ethnicity:** 58% Caucasians (intervention), 54% Caucasians (control) **Country:** USA
Interventions	**Number of experimental conditions:** 2 **Aim:** to improve medical students' communication and clinical reasoning and their appreciation of how these skills interrelate in medical practice **Content:** AIME taught students communication, clinical reasoning, and the connection between them with the aid of self-reflection, group discussion, videotaped encounters, role-play, and feedback. Researchers developed a Communication Skills Observation Guide modelled after the Calgary Cambridge Observation Guide with questions corresponding to the Three Function Model of interviewing. Students used the guide to observe for and comment on communication skills during role-plays. Role-play cases contained communication barriers to allow students to work through communication challenges. Clinical reasoning instruction focused on developing a patient-specific problem list and differential diagnosis. Problem lists included signs and symptoms of disease, past medical history, family history, psychosocial history, and patient preferences for care. Differential diagnosis generation included potential diagnoses in each organ system and disease categories using the mnemonic VINDICATE (vascular, infectious, neoplastic, drug related, inflammatory, collagen vascular, allergic/autoimmune, traumatic, and endocrine) **Intervention delivered by:** sessions were facilitated by faculty members **Comparison:** control group received identical instructions later in the year **Theoretical basis:** Three Function Model of the Medical Interview (Bird 1990) **Duration and timing:** 6-week intervention with each weekly session lasting 3 hours

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Windish 2005 *(Continued)*

	Fidelity: a 2-hour faculty development session was held 1 week before the start of the intervention. This session allowed faculty members **Adherence:** NR
Outcomes	**Assessment:** NR **Assessment timing:** NR **Primary measures:** rapport building skills, data gathering skills **Other measures:** patient education/counselling skills, skills to list correct problems and generate differential diagnoses, diagnostic thinking inventory **Measures assessed by:** SP
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** study authors received an unrestricted educational grant through the Program for Outpatient Education in Medicine of the Division of General Internal Medicine at the Johns Hopkins University School of Medicine, which helped support curricular evaluation. At the time of curriculum inception, Dr. Windish was a research fellow on an Institution Research Service Award training grant supported by the Health Resources and Services Administration. Dr. Price was a research fellow on a training grant supported by the National Institutes of Health in Behavioral Research in Heart and Vascular Disease **Abbreviations:** AIME: An Integrated Medical Encounter; C-RCT: cluster-randomised controlled trial; Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not adequately reported.
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	Unclear risk	No information provided
Blinding of outcome assessment (detection bias) All outcomes	Unclear risk	Analysis conducted in blinded fashion, but blinding of SPs assessing was not mentioned
Incomplete outcome data (attrition bias) All outcomes	Low risk	Low attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred. Efforts made to separate intervention from usual curriculum (i.e. faculty not teaching in other areas of the curriculum delivered the intervention); however contamination between students was still possible

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Windish 2005 *(Continued)*

Other Bias: Baseline outcome measurement similarity	Low risk	Adjustments made for identified differences
Other Bias: Baseline characteristics similarity	Low risk	Study authors found a difference in previous health professional training, but all analyses are adjusted for this variable

Wundrich 2017

Study characteristics	
Methods	**Design:** RCT **Unit of randomisation:** student **Ethics and informed consent:** local ethics committee approval; informed consent unclear **Intention-to-treat analysis:** NR **Adjustment for clustering:** NA, but unit of analysis error may be present due to group nature of intervention delivery
Participants	**Student level:** third year **Setting:** classroom (teaching hospital) **Percentage recruitment:** NR **Number of subjects randomised:** 158 **Number of subjects participating:** - Int: NR - Ctrl: NR - Total: 158 **Age (mean):** NR **Sex:** NR **Ethnicity:** NR **Country:** Germany
Interventions	**Number of experimental conditions:** 2 **Aim:** to investigate whether empathy can be taught to medical students in psychiatry and psychotherapy via empathy skills training with SPs **Content:** empathy skills training consisted of an introduction course on empathy and empathy skills training. In the introduction course, participants reflected on their own experience as patients, and scientific information on the role of empathy in the physician–patient relationship was provided. The domains of the empathy construct described above were used to derive concrete learning objectives that students received as a checklist. Two interviews between the instructor and an SP were presented, followed by a discussion on which empathic behaviours could be identified. Skills training consisted of 2 sessions (2.25 hours each). Each student passed through 4 different stations with simulated psychiatric patients. After each encounter, the student self-rated his/her behaviour. Subsequently, feedback on empathic behavior was given by the SP and by a student who had observed the encounter. At the end of each session, important strategies for empathetic behaviour were discussed and summarised

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wundrich 2017 *(Continued)*

	Intervention delivered by: NR **Comparison:** the control group attended a seminar on an arbitrary psychiatric topic and also participated in small-group discussions on medical history to ensure comparable duration and intensity of teaching **Theoretical basis:** 4-dimensional model of empathy (Morse 1992) **Duration and timing:** three 2-hour sessions over 3 weeks **Fidelity:** NR **Adherence:** NR
Outcomes	**Assessment:** 4 10-minute OSCE stations **Assessment timing:** end of 3-week course **Primary measures:** 11 aspects of empathy-related communication skills as well as general interview techniques 1. Active listening 2. Understanding the situation 3. Understanding the problems 4. Understanding feelings 5. Explanation (of illness, drugs, and so on) 6. Shared decision-making **Other measures:** student self-assessment **Measures assessed by:** SP and examiners
Notes	**Conflicts of interest:** declaration of no conflicts of interest **Funding:** this work was supported by Deutsche Forschungsgemeinschaft **Abbreviations:** Ctrl: control; Int: intervention; NA: not applicable; NR: not reported; OSCE: objective structured clinical examination; RCT: randomised controlled trial; SP: simulated patient

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation not reported
Allocation concealment (selection bias)	Unclear risk	Not adequately reported
Blinding of participants and personnel (performance bias) All outcomes	High risk	Blinding not possible due to the nature of the intervention
Blinding of outcome assessment (detection bias) All outcomes	Low risk	Assessors and SPs were blind to group assignment

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wundrich 2017 *(Continued)*

Incomplete outcome data (attrition bias) All outcomes	Low risk	No apparent attrition
Selective reporting (reporting bias)	Low risk	No protocol evident. Reported measures match methods. Reported at summary level only
Other Bias: Contamination	Unclear risk	Same institution, so contamination may have occurred
Other Bias: Baseline outcome measurement similarity	Unclear risk	No baseline measures
Other Bias: Baseline characteristics similarity	Low risk	NR, but stated that adjusting outcomes for demographic variables had no effect

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Abraham 2001	Patient population - children or adolescents
Abraham 2011	Patient population - children or adolescents
Ahmadzadeh 2019	Outcomes - self-report only (no observation of communication)
Aper 2012	Outcomes - no observation of communication
Betson 1997	Study design - no effort to randomise
Bientzle 2015	Study design - not an interventional study
Bishop 2016	Study design - reports on conference abstract and qualitative data
Bittner 2016	Study design - not randomised
Blake 2000	Patient population - children or adolescents
Blanch-Hartigan 2012	Outcomes - no observed communication outcomes
Bragard 2018	Participants - medical trainees and residents
Brown 1980	Study design - not randomised
Claramita 2006	Participants - residents
Clever 2011	Study design - not an RCT
Craig 1992	Study design - no effort to randomise
D'Souza 2020	Outcomes - self-report only (no observation of communication)
Daetwyler 2010	Participants - interns

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Study	Reason for exclusion
Daryazadeh 2020	Outcomes - narrative reflection (no observation of communication)
Davis 1992	Study design - not an intervention study
Dickson 2012	Study design - observational
Duke 2015	Study design - no control group
Ellis 2002	Participants - not medical students
Fernndez-Olano 2008	Participants - unable to separate, also self-reported outcomes only
Fine 1977	Study design - not an RCT
Fukuta 2018	Outcomes - communication outcome mentioned as part of non-technical skills but not reported numerically. Contacted study authors with no response
Ghofranipour 2018	Participants - medical interns
Gorniewicz 2017	Participants - not just medical students (includes Nursing and Pharmacy). Contacted study authors with no response
Haeseler 2011	Study design - no attempt to randomise
Haq 2006	Outcomes - no specific communication outcomes
Hoffman 2014	Participants - unable to separate medical students from other health professional students
Junger 2005	Outcomes - unable to separate or identify communication outcomes
Kalet 2005	Study design - not an RCT
Karanth 2008	Outcomes - no specific communication outcomes
Koponen 2012	Outcomes - student attitudes only
Koponen 2014	Outcomes - student attitudes only
Kron 2017	Outcomes - interprofessional communication
Lau 2001	Outcomes - student self-assessment
Leber 2012	Participants - residents
LeBlanc 2009	Participants - residents and medical students - unable to separate outcome data
Lee 2004	Study design - no control
Leung 2015	Outcomes - no communication outcomes
Malik 2013	Outcomes - no communication outcomes
Malik 2014	Outcomes - no communication outcomes - appears to be the same study as Malik 2013
Marko 2015	Study design - no attempt at randomisation

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Study	Reason for exclusion
Moulton 2009	Participants - students and residents - data could not be separated
Pettit 2017	Study design - no intervention
Phisalprapa 2013	Outcomes - no communication outcomes
Quirk 1982	Outcomes - no clear communication outcomes
Reinders 2010	Participants - GP trainees
Ricciotti 2010	Outcomes - specific communication outcomes not available
Saba 2014	Outcomes - specific communication outcomes not available
Scheidt 1986	Patients - children or adolescents
Schwartz 2007	Outcomes - no specific communication outcomes
Scott 1976	Study design - no control group
Servotte 2019	Participants - mix of residents and students (data not separated)
Sokas 1991	Outcomes - no communication outcomes
Stillman 1976	Patients - children or adolescents
Taylor 2018	Study design - cross-over trial, no control
Tolsgaard 2013	Outcomes - no communication outcomes
Walsh 2001	Study design - no attempt to randomise
Werner 2013	Outcomes - measured information recall rather than communication skills
Wolf 1987	Outcomes - student-reported only

GP: general practitioner.
RCT: randomised controlled trial.

Characteristics of studies awaiting classification *[ordered by study ID]*

Ahmadreza 2011

Methods	Quasi-experimental intervention study
Participants	Medical students during internship period
Interventions	Tool for guiding physicians to communicate with patients
Outcomes	Physician-patient communication assessed
Notes	Cannot locate full text/not in English

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Alyami 2019

Methods	Longitudinal mixed methods study
Participants	Third year medical students
Interventions	Serious game to teach history taking content
Outcomes	OSCE scores
Notes	Identified in 2020 search update

Barais 2012

Methods	RCT
Participants	Trainees and medical students
Interventions	'Participants had to walk through a course and fulfil a pill box according to a medical prescription'
Outcomes	Empathy towards the elderly
Notes	Conference abstract only

Caruso 1994

Methods	RCT
Participants	Medical students
Interventions	HIV workshop
Outcomes	2 station OSCEs assessed information gathering and sensitive communication
Notes	Cannot locate full text

Cherry 2010

Methods	RCT
Participants	First-year medical students
Interventions	7-month emotional intelligence training course
Outcomes	Emotional intelligence and OSCE communication skills scores
Notes	Conferene abstract only

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Dubosh 2019

Methods	RCT
Participants	Fourth-year medical students
Interventions	Multi-modal communication skills curriculum
Outcomes	Communication assessment tool completed by patients with whom students interact in clerkship
Notes	Identified in 2020 search update

Engler 1981

Methods	Unable to assess
Participants	Unable to assess
Interventions	Unable to assess
Outcomes	Unable to assess
Notes	Cannot locate full text

Florenzano 2000

Methods	Comparison study - unclear whether randomisation occurred
Participants	Medical students in 3 Chilean medical schools
Interventions	Interactive workshop with role-plays, vignettes, and videotaped feedback
Outcomes	Interviews assessed for opening, problem exploration, non-verbal facilitation, interpersonal patient reaction, and closing
Notes	Article in Spanish

Geoffroy 2020

Methods	RCT
Participants	Third-year medical students
Interventions	Role-play with simulated patients
Outcomes	Tutor rating of students' communication skills demonstrated through tutorials
Notes	Identified in 2020 search update

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Hartigan 2012

Methods	Students were randomly assigned to receive the comprehensive training condition, a control group with none of these training components, or 1 of 4 training conditions in which participants received consciousness-raising, instruction, practice alone, or practice with feedback
Participants	Undergraduate medical students
Interventions	Comprehensive training included raising awareness about the importance of emotion cues in healthcare interactions, providing instruction on how to increase emotion cue recognition accuracy, and practising emotion recognition while receiving feedback
Outcomes	Emotion cue recognition ability
Notes	Dissertation abstract - cannot locate full text

Hermann-Werner 2019b

Methods	Cross-over RCT
Participants	Fourth and fifth year medical students
Interventions	Blended learning intervention including videos, self-study, and role-play
Outcomes	Simulated patient encounter - SPs rate students' empathy
Notes	Identified in 2020 search update

Holtedahl 1999

Methods	RCT
Participants	Fourth-year medical students
Interventions	Teaching was performed in small groups of 1 GP instructor and 6 students and consisted of 4 weekly 3-hour sessions
Outcomes	Evaluation of practical skills and communication skills
Notes	Publication not in English

Jaury 2018

Methods	Unable to assess
Participants	Unable to assess
Interventions	Unable to assess
Outcomes	Unable to assess
Notes	Article in French, no abstract, identified in 2020 search update

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Kaper 2019

Methods	RCT
Participants	Second-year medical students
Interventions	Health literacy intervention
Outcomes	Self-report validated using observed health literacy assessment
Notes	Identified in 2020 search update

Koch 2020

Methods	RCT
Participants	Fourth and fifth year medical students
Interventions	Shared decision-making curriculum (15 hours)
Outcomes	Risk communication performance in a video-observed structured clinical examination
Notes	Identified in 2020 search update

Lai 2020

Methods	RCT
Participants	Medical students in their first clinical year
Interventions	Additional written feedback on patient satisfaction combined with guided self-reflection
Outcomes	Observed consultation rated by tutor
Notes	Identified in 2020 search update

Martin 2017

Methods	Unable to assess
Participants	Unable to assess
Interventions	Unable to assess
Outcomes	Unable to assess
Notes	Article not in English

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Mazori 2019

Methods	Quasi-RCT
Participants	Third-year medical students
Interventions	Lesson on working with interpreters
Outcomes	Rating of encounter with patient and interpreter
Notes	Identified in 2020 search update

Pati 2010

Methods	Quasi-experimental study
Participants	17 final year medical students
Interventions	Self-instructional vs faculty-supervised training
Outcomes	Appropriate use of 4 interview microskills, patient and family member interview ratings, and number of psychosocial needs recognised on dictation reports
Notes	Conference abstract only

Qureshi 2020

Methods	RCT
Participants	Final year medical students
Interventions	Simulated patient feedback
Outcomes	Observation of 4 OSCE stations
Notes	Identified in 2020 search update

Rae 1973

Methods	RCT
Participants	First-year medical students
Interventions	Empathy training groups (discussion, modelling, experiential-simulation)
Outcomes	Empathy skills
Notes	Full text in ERIC pending restoration

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Smith 2019

Methods	RCT
Participants	Fourth-year medical students
Interventions	One-on-one preceptor shift
Outcomes	OSCE observation and rating of communication skills
Notes	Identified in 2020 search update

Sobana 2020

Methods	RCT
Participants	Medical students
Interventions	Active participation compared to observation
Outcomes	Rating of OSCE performance
Notes	Identified in 2020 search update

Spiess 1988

Methods	Unable to assess
Participants	Unable to assess
Interventions	Unable to assess
Outcomes	Unable to assess
Notes	Article not in English

Thompson 2012

Methods	RCT
Participants	Medical students and residents
Interventions	Medical students and medical residents were observed acting as the physician in a clinical scenario during which they communicated bad news to a simulated patient, both before and after one of the following arms: (1) no additional training, (2) participation in a Second Life scenario using avatars, (3) attending a didactic session
Outcomes	Unclear - 'communication skills'
Notes	Conference poster abstract - no publication found

ERIC: Education Resource Information Center.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

GP: general practitioner.
OSCE: objective structured clinical examination.
RCT: randomised controlled trial.
SP: simulated patient.

DATA AND ANALYSES

Comparison 1. Communication intervention versus control or usual care

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Overall communication skills	18	1356	Std. Mean Difference (IV, Random, 95% CI)	0.92 [0.53, 1.31]
1.1.1 Assessor rating	13	959	Std. Mean Difference (IV, Random, 95% CI)	1.21 [0.69, 1.74]
1.1.2 SP rating	5	397	Std. Mean Difference (IV, Random, 95% CI)	0.27 [-0.07, 0.60]
1.2 Subgroup_Overall communication skills_control only	13	1057	Std. Mean Difference (IV, Random, 95% CI)	1.06 [0.54, 1.58]
1.2.1 Assessor rating	9	702	Std. Mean Difference (IV, Random, 95% CI)	1.59 [0.84, 2.33]
1.2.2 SP rating	4	355	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.16, 0.52]
1.3 Sensitivity_RoB_Overall communication skills	14	1035	Std. Mean Difference (IV, Random, 95% CI)	1.14 [0.62, 1.66]
1.3.1 Assessor rating	10	676	Std. Mean Difference (IV, Random, 95% CI)	1.58 [0.81, 2.34]
1.3.2 SP rating	4	359	Std. Mean Difference (IV, Random, 95% CI)	0.34 [-0.03, 0.71]
1.4 Sensitivity_cluster 5%_Overall communication skills	18	1322	Std. Mean Difference (IV, Random, 95% CI)	1.02 [0.58, 1.46]
1.4.1 Assessor rating	13	925	Std. Mean Difference (IV, Random, 95% CI)	1.39 [0.78, 2.00]
1.4.2 SP rating	5	397	Std. Mean Difference (IV, Random, 95% CI)	0.29 [-0.03, 0.61]
1.5 Sensitivity_cluster 10%_Overall communication skills	18	1322	Std. Mean Difference (IV, Random, 95% CI)	1.03 [0.58, 1.48]
1.5.1 Assessor rating	13	925	Std. Mean Difference (IV, Random, 95% CI)	1.41 [0.78, 2.03]

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

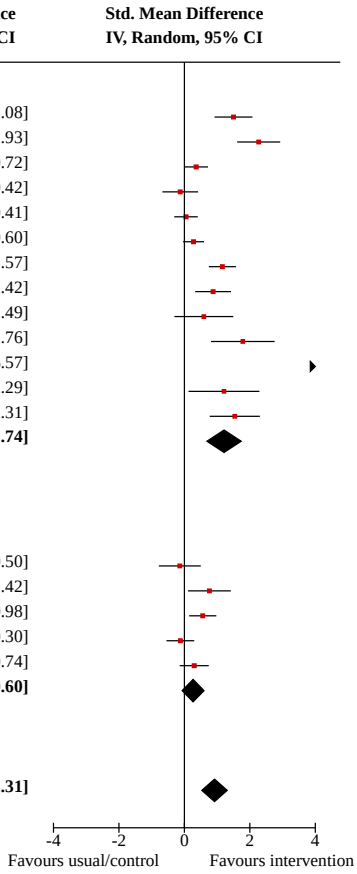
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.5.2 SP rating	5	397	Std. Mean Difference (IV, Random, 95% CI)	0.31 [-0.00, 0.63]
1.6 Sensitivity_cluster 20%_Overall communication skills	18	1322	Std. Mean Difference (IV, Random, 95% CI)	1.05 [0.59, 1.51]
1.6.1 Assessor rating	13	925	Std. Mean Difference (IV, Random, 95% CI)	1.43 [0.79, 2.07]
1.6.2 SP rating	5	397	Std. Mean Difference (IV, Random, 95% CI)	0.34 [0.03, 0.64]
1.7 Empathy	6	831	Std. Mean Difference (IV, Random, 95% CI)	0.64 [0.23, 1.05]
1.7.1 Assesor rating	4	251	Std. Mean Difference (IV, Random, 95% CI)	0.98 [0.61, 1.35]
1.7.2 SP rating	2	580	Std. Mean Difference (IV, Random, 95% CI)	0.09 [-0.07, 0.26]
1.8 Relationship building/Rapport	9	834	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.15, 0.51]
1.8.1 Assessor rating	5	456	Std. Mean Difference (IV, Random, 95% CI)	0.03 [-0.19, 0.26]
1.8.2 SP perception/satisfaction/rapport	4	378	Std. Mean Difference (IV, Random, 95% CI)	0.39 [-0.33, 1.12]
1.9 Sensitivity_cluster 5%_Relationship building/Rapport	9	800	Std. Mean Difference (IV, Random, 95% CI)	0.24 [-0.07, 0.56]
1.9.1 Assessor rating	5	422	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.09, 0.30]
1.9.2 SP perception/satisfaction/rapport	4	378	Std. Mean Difference (IV, Random, 95% CI)	0.40 [-0.29, 1.10]
1.10 Sensitivity_cluster 10%_Relationship building/Rapport	9	800	Std. Mean Difference (IV, Random, 95% CI)	0.25 [-0.07, 0.57]
1.10.1 Assessor rating	5	422	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.09, 0.30]
1.10.2 SP perception/satisfaction/rapport	4	378	Std. Mean Difference (IV, Random, 95% CI)	0.42 [-0.25, 1.09]
1.11 Sensitivity_cluster 20%_Relationship building/Rapport	9	800	Std. Mean Difference (IV, Random, 95% CI)	0.26 [-0.05, 0.58]

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.11.1 Assessor rating	5	422	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.09, 0.30]
1.11.2 SP perception/satisfaction/rapport	4	378	Std. Mean Difference (IV, Random, 95% CI)	0.44 [-0.20, 1.08]
1.12 Information gathering about patient perspectives/concerns	5	405	Std. Mean Difference (IV, Random, 95% CI)	1.07 [0.61, 1.54]

Analysis 1.1. Comparison 1: Communication intervention versus control or usual care, Outcome 1: Overall communication skills

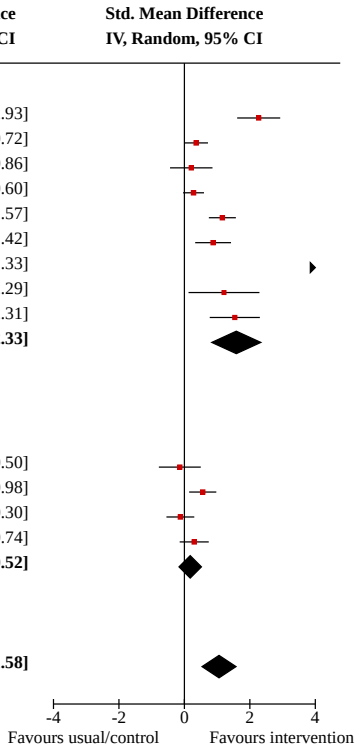
Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.1.1 Assessor rating						
Bosse 2012	1.4999	0.295	28	32	5.7%	1.50 [0.92 , 2.08]
Evans 1996	2.27	0.3356	30	30	5.5%	2.27 [1.61 , 2.93]
Filipetto 2006	0.36	0.1838	68	54	6.2%	0.36 [-0.00 , 0.72]
Gartmeir 2015	-0.1267	0.2783	54	17	5.8%	-0.13 [-0.67 , 0.42]
Lee 2015	0.05	0.1834	60	59	6.2%	0.05 [-0.31 , 0.41]
Liu 2016	0.2792	0.1629	59	107	6.3%	0.28 [-0.04 , 0.60]
Lorin 2006	1.16	0.2105	53	53	6.1%	1.16 [0.75 , 1.57]
Lupi 2012	0.8773	0.2792	50	45	5.8%	0.88 [0.33 , 1.42]
Maguire 1977	0.5942	0.4593	10	10	4.8%	0.59 [-0.31 , 1.49]
Maguire 1978	1.7848	0.4961	12	12	4.6%	1.78 [0.81 , 2.76]
Pirdehghan 2018	5.444	0.5753	30	30	4.2%	5.44 [4.32 , 6.57]
Solomon 2004	1.2095	0.5526	5	16	4.3%	1.21 [0.13 , 2.29]
Spollen 2010	1.54	0.3906	17	18	5.2%	1.54 [0.77 , 2.31]
Subtotal (95% CI)			**476**	**483**	**70.8%**	**1.21 [0.69 , 1.74]**
Heterogeneity: Tau² = 0.80; Chi² = 144.06, df = 12 (P < 0.00001); I² = 92%						
Test for overall effect: Z = 4.54 (P < 0.00001)						
1.1.2 SP rating						
Colletti 2001	-0.14	0.3267	21	17	5.6%	-0.14 [-0.78 , 0.50]
Ho 2008	0.7643	0.3337	15	27	5.5%	0.76 [0.11 , 1.42]
Hobgood 2009	0.56	0.2118	35	68	6.1%	0.56 [0.14 , 0.98]
Shaddheau 2015	-0.123	0.2169	75	60	6.1%	-0.12 [-0.55 , 0.30]
Shapiro 2009	0.3	0.2265	38	41	6.0%	0.30 [-0.14 , 0.74]
Subtotal (95% CI)			**184**	**213**	**29.2%**	**0.27 [-0.07 , 0.60]**
Heterogeneity: Tau² = 0.08; Chi² = 8.94, df = 4 (P = 0.06); I² = 55%						
Test for overall effect: Z = 1.56 (P = 0.12)						
Total (95% CI)			**660**	**696**	**100.0%**	**0.92 [0.53 , 1.31]**
Heterogeneity: Tau² = 0.60; Chi² = 165.36, df = 17 (P < 0.00001); I² = 90%						
Test for overall effect: Z = 4.64 (P < 0.00001)						
Test for subgroup differences: Chi² = 8.92, df = 1 (P = 0.003), I² = 88.8%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.2. Comparison 1: Communication intervention versus control or usual care, Outcome 2: Subgroup_Overall communication skills_control only

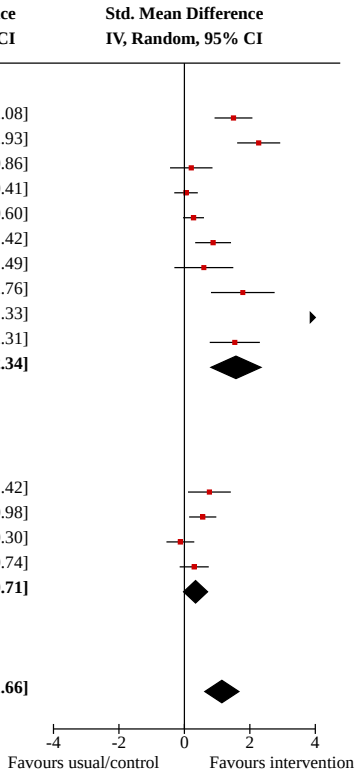
Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.2.1 Assessor rating						
Evans 1996	2.27	0.3356	30	30	7.8%	2.27 [1.61 , 2.93]
Filipetto 2006	0.36	0.1838	68	54	8.5%	0.36 [-0.00 , 0.72]
Gartmeir 2015	0.21	0.3309	20	17	7.8%	0.21 [-0.44 , 0.86]
Liu 2016	0.2792	0.1629	59	107	8.6%	0.28 [-0.04 , 0.60]
Lorin 2006	1.16	0.2105	53	53	8.4%	1.16 [0.75 , 1.57]
Lupi 2012	0.8773	0.2792	50	45	8.1%	0.88 [0.33 , 1.42]
Pirdehghan 2018	9.5	0.9336	30	30	4.2%	9.50 [7.67 , 11.33]
Solomon 2004	1.2095	0.5526	5	16	6.4%	1.21 [0.13 , 2.29]
Spollen 2010	1.54	0.3906	17	18	7.4%	1.54 [0.77 , 2.31]
Subtotal (95% CI)			**332**	**370**	**67.1%**	**1.59 [0.84 , 2.33]**
Heterogeneity: Tau² = 1.16; Chi² = 132.30, df = 8 (P < 0.00001); I² = 94%						
Test for overall effect: Z = 4.15 (P < 0.0001)						
1.2.2 SP rating						
Colletti 2001	-0.14	0.3267	21	17	7.8%	-0.14 [-0.78 , 0.50]
Hobgood 2009	0.56	0.2118	35	68	8.4%	0.56 [0.14 , 0.98]
Shaddheau 2015	-0.123	0.2169	75	60	8.4%	-0.12 [-0.55 , 0.30]
Shapiro 2009	0.3	0.2265	38	41	8.3%	0.30 [-0.14 , 0.74]
Subtotal (95% CI)			**169**	**186**	**32.9%**	**0.18 [-0.16 , 0.52]**
Heterogeneity: Tau² = 0.06; Chi² = 6.38, df = 3 (P = 0.09); I² = 53%						
Test for overall effect: Z = 1.02 (P = 0.31)						
Total (95% CI)			**501**	**556**	**100.0%**	**1.06 [0.54 , 1.58]**
Heterogeneity: Tau² = 0.79; Chi² = 155.69, df = 12 (P < 0.00001); I² = 92%						
Test for overall effect: Z = 4.00 (P < 0.0001)						
Test for subgroup differences: Chi² = 11.22, df = 1 (P = 0.0008), I² = 91.1%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.3. Comparison 1: Communication intervention versus control or usual care, Outcome 3: Sensitivity_RoB_Overall communication skills

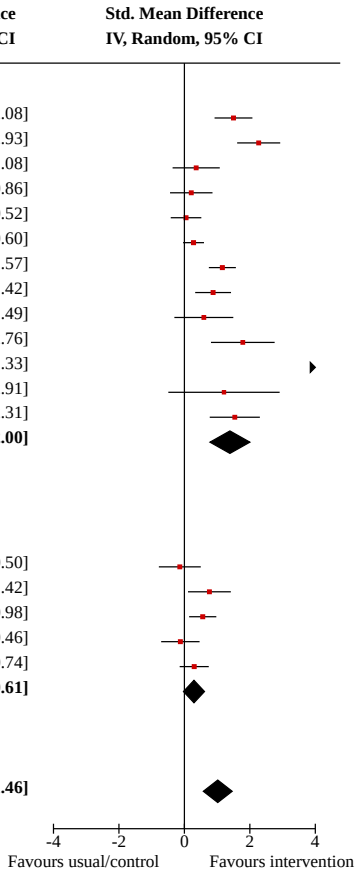
Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.3.1 Assessor rating						
Bosse 2012	1.4999	0.295	28	32	7.5%	1.50 [0.92 , 2.08]
Evans 1996	2.27	0.3356	30	30	7.3%	2.27 [1.61 , 2.93]
Gartmeir 2015	0.21	0.3309	20	17	7.3%	0.21 [-0.44 , 0.86]
Lee 2015	0.05	0.1834	60	59	7.9%	0.05 [-0.31 , 0.41]
Liu 2016	0.2792	0.1629	59	107	8.0%	0.28 [-0.04 , 0.60]
Lupi 2012	0.8773	0.2792	50	45	7.5%	0.88 [0.33 , 1.42]
Maguire 1977	0.5942	0.4593	10	10	6.6%	0.59 [-0.31 , 1.49]
Maguire 1978	1.7848	0.4961	12	12	6.4%	1.78 [0.81 , 2.76]
Pirdehghan 2018	9.5	0.9336	30	30	4.1%	9.50 [7.67 , 11.33]
Spollen 2010	1.54	0.3906	17	18	7.0%	1.54 [0.77 , 2.31]
Subtotal (95% CI)			**316**	**360**	**69.4%**	**1.58 [0.81 , 2.34]**
Heterogeneity: Tau² = 1.35; Chi² = 149.16, df = 9 (P < 0.00001); I² = 94%						
Test for overall effect: Z = 4.04 (P < 0.0001)						
1.3.2 SP rating						
Ho 2008	0.7643	0.3337	15	27	7.3%	0.76 [0.11 , 1.42]
Hobgood 2009	0.56	0.2118	35	68	7.8%	0.56 [0.14 , 0.98]
Shaddheau 2015	-0.123	0.2169	75	60	7.8%	-0.12 [-0.55 , 0.30]
Shapiro 2009	0.3	0.2265	38	41	7.7%	0.30 [-0.14 , 0.74]
Subtotal (95% CI)			**163**	**196**	**30.6%**	**0.34 [-0.03 , 0.71]**
Heterogeneity: Tau² = 0.08; Chi² = 7.23, df = 3 (P = 0.06); I² = 59%						
Test for overall effect: Z = 1.82 (P = 0.07)						
Total (95% CI)			**479**	**556**	**100.0%**	**1.14 [0.62 , 1.66]**
Heterogeneity: Tau² = 0.86; Chi² = 164.79, df = 13 (P < 0.00001); I² = 92%						
Test for overall effect: Z = 4.31 (P < 0.0001)						
Test for subgroup differences: Chi² = 8.16, df = 1 (P = 0.004), I² = 87.7%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.4. Comparison 1: Communication intervention versus control or usual care, Outcome 4: Sensitivity_cluster 5%_Overall communication skills

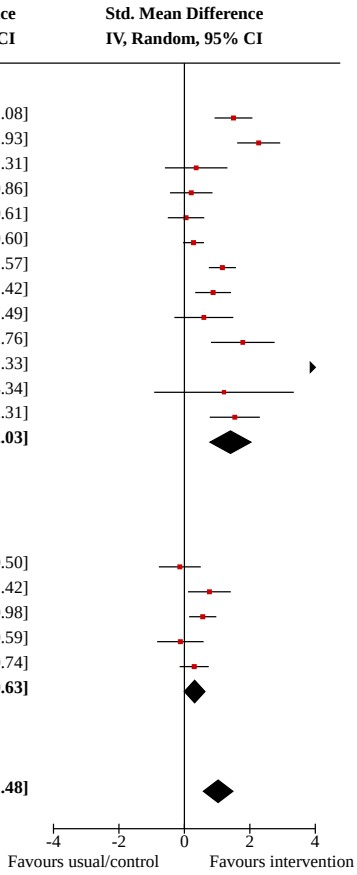
Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.4.1 Assessor rating						
Bosse 2012	1.4999	0.295	28	32	5.9%	1.50 [0.92 , 2.08]
Evans 1996	2.27	0.3356	30	30	5.8%	2.27 [1.61 , 2.93]
Filipetto 2006	0.36	0.3676	68	54	5.6%	0.36 [-0.36 , 1.08]
Gartmeir 2015	0.21	0.3309	20	17	5.8%	0.21 [-0.44 , 0.86]
Lee 2015	0.05	0.2384	60	59	6.2%	0.05 [-0.42 , 0.52]
Liu 2016	0.2792	0.1629	59	107	6.4%	0.28 [-0.04 , 0.60]
Lorin 2006	1.16	0.2105	53	53	6.3%	1.16 [0.75 , 1.57]
Lupi 2012	0.8773	0.2792	50	45	6.0%	0.88 [0.33 , 1.42]
Maguire 1977	0.5942	0.4593	10	10	5.2%	0.59 [-0.31 , 1.49]
Maguire 1978	1.7848	0.4961	12	12	5.0%	1.78 [0.81 , 2.76]
Pirdehghan 2018	9.5	0.9336	30	30	3.1%	9.50 [7.67 , 11.33]
Solomon 2004	1.2095	0.8675	5	16	3.3%	1.21 [-0.49 , 2.91]
Spollen 2010	1.54	0.3906	17	18	5.5%	1.54 [0.77 , 2.31]
Subtotal (95% CI)			**442**	**483**	**70.0%**	**1.39 [0.78 , 2.00]**
Heterogeneity: Tau² = 1.07; Chi² = 146.99, df = 12 (P < 0.00001); I² = 92%						
Test for overall effect: Z = 4.48 (P < 0.00001)						
1.4.2 SP rating						
Colletti 2001	-0.14	0.3267	21	17	5.8%	-0.14 [-0.78 , 0.50]
Ho 2008	0.7643	0.3337	15	27	5.8%	0.76 [0.11 , 1.42]
Hobgood 2009	0.56	0.2118	35	68	6.3%	0.56 [0.14 , 0.98]
Shaddheau 2015	-0.123	0.2993	75	60	5.9%	-0.12 [-0.71 , 0.46]
Shapiro 2009	0.3	0.2265	38	41	6.2%	0.30 [-0.14 , 0.74]
Subtotal (95% CI)			**184**	**213**	**30.0%**	**0.29 [-0.03 , 0.61]**
Heterogeneity: Tau² = 0.06; Chi² = 7.24, df = 4 (P = 0.12); I² = 45%						
Test for overall effect: Z = 1.79 (P = 0.07)						
Total (95% CI)			**626**	**696**	**100.0%**	**1.02 [0.58 , 1.46]**
Heterogeneity: Tau² = 0.76; Chi² = 167.84, df = 17 (P < 0.00001); I² = 90%						
Test for overall effect: Z = 4.54 (P < 0.00001)						
Test for subgroup differences: Chi² = 9.78, df = 1 (P = 0.002), I² = 89.8%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.5. Comparison 1: Communication intervention versus control or usual care, Outcome 5: Sensitivity_cluster 10%_Overall communication skills

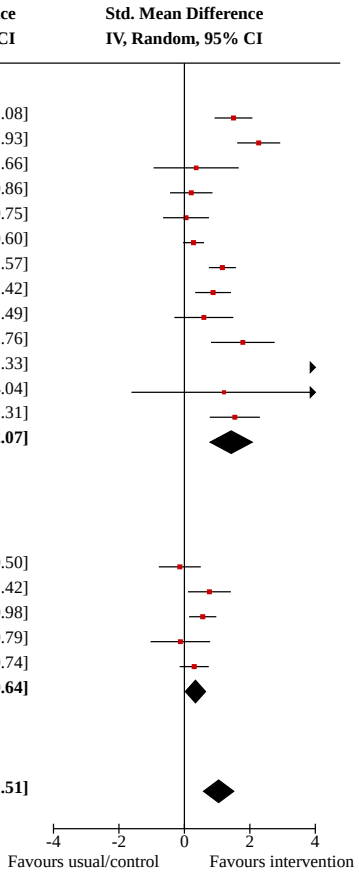
Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.5.1 Assessor rating						
Bosse 2012	1.4999	0.295	28	32	6.0%	1.50 [0.92 , 2.08]
Evans 1996	2.27	0.3356	30	30	5.9%	2.27 [1.61 , 2.93]
Filipetto 2006	0.36	0.4871	68	54	5.2%	0.36 [-0.59 , 1.31]
Gartmeir 2015	0.21	0.3309	20	17	5.9%	0.21 [-0.44 , 0.86]
Lee 2015	0.05	0.2843	60	59	6.1%	0.05 [-0.51 , 0.61]
Liu 2016	0.2792	0.1629	59	107	6.5%	0.28 [-0.04 , 0.60]
Lorin 2006	1.16	0.2105	53	53	6.4%	1.16 [0.75 , 1.57]
Lupi 2012	0.8773	0.2792	50	45	6.1%	0.88 [0.33 , 1.42]
Maguire 1977	0.5942	0.4593	10	10	5.3%	0.59 [-0.31 , 1.49]
Maguire 1978	1.7848	0.4961	12	12	5.1%	1.78 [0.81 , 2.76]
Pirdehghan 2018	9.5	0.9336	30	30	3.2%	9.50 [7.67 , 11.33]
Solomon 2004	1.2095	1.0886	5	16	2.7%	1.21 [-0.92 , 3.34]
Spollen 2010	1.54	0.3906	17	18	5.6%	1.54 [0.77 , 2.31]
Subtotal (95% CI)			**442**	**483**	**69.8%**	**1.41 [0.78 , 2.03]**
Heterogeneity: Tau² = 1.11; Chi² = 142.60, df = 12 (P < 0.00001); I² = 92%						
Test for overall effect: Z = 4.41 (P < 0.0001)						
1.5.2 SP rating						
Colletti 2001	-0.14	0.3267	21	17	5.9%	-0.14 [-0.78 , 0.50]
Ho 2008	0.7643	0.3337	15	27	5.9%	0.76 [0.11 , 1.42]
Hobgood 2009	0.56	0.2118	35	68	6.3%	0.56 [0.14 , 0.98]
Shaddheau 2015	-0.123	0.3622	75	60	5.7%	-0.12 [-0.83 , 0.59]
Shapiro 2009	0.3	0.2265	38	41	6.3%	0.30 [-0.14 , 0.74]
Subtotal (95% CI)			**184**	**213**	**30.2%**	**0.31 [-0.00 , 0.63]**
Heterogeneity: Tau² = 0.05; Chi² = 6.52, df = 4 (P = 0.16); I² = 39%						
Test for overall effect: Z = 1.95 (P = 0.05)						
Total (95% CI)			**626**	**696**	**100.0%**	**1.03 [0.58 , 1.48]**
Heterogeneity: Tau² = 0.78; Chi² = 162.88, df = 17 (P < 0.00001); I² = 90%						
Test for overall effect: Z = 4.49 (P < 0.00001)						
Test for subgroup differences: Chi² = 9.41, df = 1 (P = 0.002), I² = 89.4%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

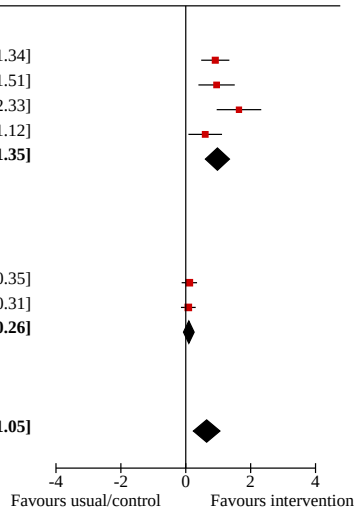
Analysis 1.6. Comparison 1: Communication intervention versus control or usual care, Outcome 6: Sensitivity_cluster 20%_Overall communication skills

Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.6.1 Assessor rating						
Bosse 2012	1.4999	0.295	28	32	6.2%	1.50 [0.92 , 2.08]
Evans 1996	2.27	0.3356	30	30	6.0%	2.27 [1.61 , 2.93]
Filipetto 2006	0.36	0.6635	68	54	4.4%	0.36 [-0.94 , 1.66]
Gartmeir 2015	0.21	0.3309	20	17	6.0%	0.21 [-0.44 , 0.86]
Lee 2015	0.05	0.3576	60	59	5.9%	0.05 [-0.65 , 0.75]
Liu 2016	0.2792	0.1629	59	107	6.6%	0.28 [-0.04 , 0.60]
Lorin 2006	1.16	0.2105	53	53	6.5%	1.16 [0.75 , 1.57]
Lupi 2012	0.8773	0.2792	50	45	6.2%	0.88 [0.33 , 1.42]
Maguire 1977	0.5942	0.4593	10	10	5.4%	0.59 [-0.31 , 1.49]
Maguire 1978	1.7848	0.4961	12	12	5.2%	1.78 [0.81 , 2.76]
Pirdehghan 2018	9.5	0.9336	30	30	3.3%	9.50 [7.67 , 11.33]
Solomon 2004	1.2095	1.4423	5	16	1.9%	1.21 [-1.62 , 4.04]
Spollen 2010	1.54	0.3906	17	18	5.8%	1.54 [0.77 , 2.31]
Subtotal (95% CI)			**442**	**483**	**69.6%**	**1.43 [0.79 , 2.07]**
Heterogeneity: Tau² = 1.15; Chi² = 138.64, df = 12 (P < 0.00001); I² = 91%						
Test for overall effect: Z = 4.37 (P < 0.0001)						
1.6.2 SP rating						
Colletti 2001	-0.14	0.3267	21	17	6.1%	-0.14 [-0.78 , 0.50]
Ho 2008	0.7643	0.3337	15	27	6.0%	0.76 [0.11 , 1.42]
Hobgood 2009	0.56	0.2118	35	68	6.5%	0.56 [0.14 , 0.98]
Shaddheau 2015	-0.123	0.4642	75	60	5.4%	-0.12 [-1.03 , 0.79]
Shapiro 2009	0.3	0.2265	38	41	6.4%	0.30 [-0.14 , 0.74]
Subtotal (95% CI)			**184**	**213**	**30.4%**	**0.34 [0.03 , 0.64]**
Heterogeneity: Tau² = 0.04; Chi² = 5.85, df = 4 (P = 0.21); I² = 32%						
Test for overall effect: Z = 2.15 (P = 0.03)						
Total (95% CI)			**626**	**696**	**100.0%**	**1.05 [0.59 , 1.51]**
Heterogeneity: Tau² = 0.80; Chi² = 158.35, df = 17 (P < 0.00001); I² = 89%						
Test for overall effect: Z = 4.46 (P < 0.00001)						
Test for subgroup differences: Chi² = 9.05, df = 1 (P = 0.003), I² = 88.9%						



Analysis 1.7. Comparison 1: Communication intervention versus control or usual care, Outcome 7: Empathy

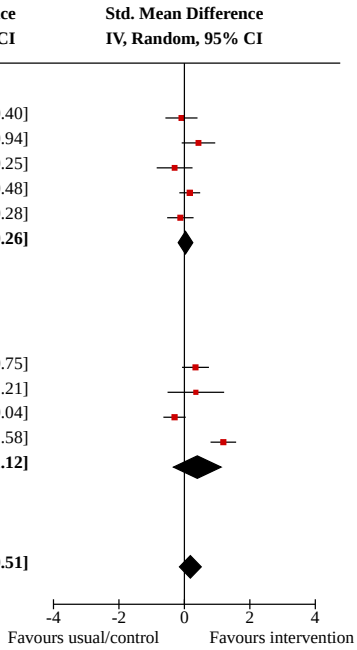
Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.7.1 Assesor rating						
Daeppen 2012	0.91	0.2213	42	49	16.9%	0.91 [0.48 , 1.34]
Evans 1989	0.95	0.2856	28	27	15.0%	0.95 [0.39 , 1.51]
Poole 1980	1.64	0.3504	25	20	13.2%	1.64 [0.95 , 2.33]
Seim 1995	0.6	0.2643	30	30	15.7%	0.60 [0.08 , 1.12]
Subtotal (95% CI)			**125**	**126**	**60.8%**	**0.98 [0.61 , 1.35]**
Heterogeneity: Tau² = 0.07; Chi² = 5.66, df = 3 (P = 0.13); I² = 47%						
Test for overall effect: Z = 5.16 (P < 0.00001)						
1.7.2 SP rating						
Bowyer 2010	0.1109	0.121	118	163	19.5%	0.11 [-0.13 , 0.35]
Buffel du Vaure 2017	0.0789	0.1158	155	144	19.7%	0.08 [-0.15 , 0.31]
Subtotal (95% CI)			**273**	**307**	**39.2%**	**0.09 [-0.07 , 0.26]**
Heterogeneity: Tau² = 0.00; Chi² = 0.04, df = 1 (P = 0.85); I² = 0%						
Test for overall effect: Z = 1.13 (P = 0.26)						
Total (95% CI)			**398**	**433**	**100.0%**	**0.64 [0.23 , 1.05]**
Heterogeneity: Tau² = 0.21; Chi² = 34.58, df = 5 (P < 0.00001); I² = 86%						
Test for overall effect: Z = 3.09 (P = 0.002)						
Test for subgroup differences: Chi² = 18.21, df = 1 (P < 0.0001), I² = 94.5%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

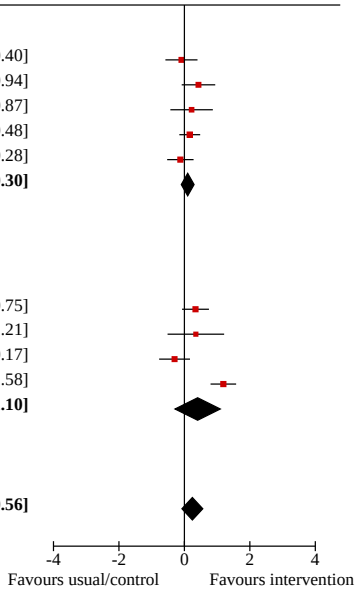
Analysis 1.8. Comparison 1: Communication intervention versus control or usual care, Outcome 8: Relationship building/Rapport

Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.8.1 Assessor rating						
Bosse 2012	-0.09	0.2501	32	32	10.9%	-0.09 [-0.58 , 0.40]
Evans 1989	0.43	0.2614	30	30	10.7%	0.43 [-0.08 , 0.94]
Gartmeir 2015	-0.2989	0.2793	54	17	10.3%	-0.30 [-0.85 , 0.25]
Liu 2016	0.1641	0.1624	59	107	12.7%	0.16 [-0.15 , 0.48]
Lupi 2012	-0.123	0.2057	50	45	11.8%	-0.12 [-0.53 , 0.28]
Subtotal (95% CI)			**225**	**231**	**56.5%**	**0.03 [-0.19 , 0.26]**
Heterogeneity: Tau² = 0.01; Chi² = 5.18, df = 4 (P = 0.27); I² = 23%						
Test for overall effect: Z = 0.30 (P = 0.77)						
1.8.2 SP perception/satisfaction/rapport						
Hobgood 2009	0.3392	0.2094	35	68	11.8%	0.34 [-0.07 , 0.75]
Legg 2005	0.35	0.441	11	10	7.3%	0.35 [-0.51 , 1.21]
Shaddheau 2015	-0.3003	0.1742	75	60	12.5%	-0.30 [-0.64 , 0.04]
Windish 2005	1.19	0.1994	59	60	12.0%	1.19 [0.80 , 1.58]
Subtotal (95% CI)			**180**	**198**	**43.5%**	**0.39 [-0.33 , 1.12]**
Heterogeneity: Tau² = 0.48; Chi² = 31.68, df = 3 (P < 0.00001); I² = 91%						
Test for overall effect: Z = 1.06 (P = 0.29)						
Total (95% CI)			**405**	**429**	**100.0%**	**0.18 [-0.15 , 0.51]**
Heterogeneity: Tau² = 0.20; Chi² = 41.18, df = 8 (P < 0.00001); I² = 81%						
Test for overall effect: Z = 1.07 (P = 0.28)						
Test for subgroup differences: Chi² = 0.86, df = 1 (P = 0.35), I² = 0%						



Analysis 1.9. Comparison 1: Communication intervention versus control or usual care, Outcome 9: Sensitivity_cluster 5%_Relationship building/Rapport

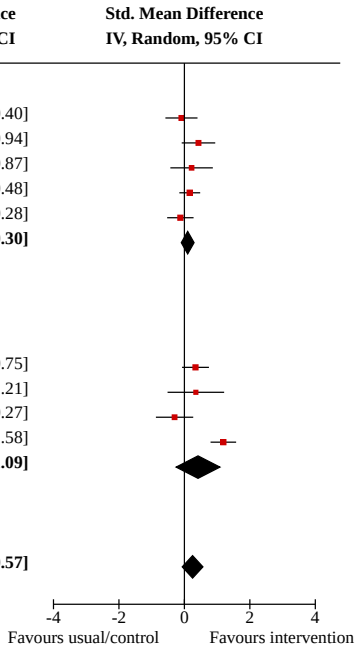
Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.9.1 Assessor rating						
Bosse 2012	-0.09	0.2501	32	32	11.2%	-0.09 [-0.58 , 0.40]
Evans 1989	0.43	0.2614	30	30	10.9%	0.43 [-0.08 , 0.94]
Gartmeir 2015	0.22	0.331	20	17	9.3%	0.22 [-0.43 , 0.87]
Liu 2016	0.1641	0.1624	59	107	13.2%	0.16 [-0.15 , 0.48]
Lupi 2012	-0.123	0.2057	50	45	12.2%	-0.12 [-0.53 , 0.28]
Subtotal (95% CI)			**191**	**231**	**56.9%**	**0.10 [-0.09 , 0.30]**
Heterogeneity: Tau² = 0.00; Chi² = 3.63, df = 4 (P = 0.46); I² = 0%						
Test for overall effect: Z = 1.01 (P = 0.31)						
1.9.2 SP perception/satisfaction/rapport						
Hobgood 2009	0.3392	0.2094	35	68	12.2%	0.34 [-0.07 , 0.75]
Legg 2005	0.35	0.441	11	10	7.2%	0.35 [-0.51 , 1.21]
Shaddheau 2015	-0.3003	0.2404	75	60	11.4%	-0.30 [-0.77 , 0.17]
Windish 2005	1.19	0.1994	59	60	12.4%	1.19 [0.80 , 1.58]
Subtotal (95% CI)			**180**	**198**	**43.1%**	**0.40 [-0.29 , 1.10]**
Heterogeneity: Tau² = 0.42; Chi² = 23.75, df = 3 (P < 0.0001); I² = 87%						
Test for overall effect: Z = 1.15 (P = 0.25)						
Total (95% CI)			**371**	**429**	**100.0%**	**0.24 [-0.07 , 0.56]**
Heterogeneity: Tau² = 0.17; Chi² = 33.58, df = 8 (P < 0.0001); I² = 76%						
Test for overall effect: Z = 1.50 (P = 0.13)						
Test for subgroup differences: Chi² = 0.69, df = 1 (P = 0.41), I² = 0%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

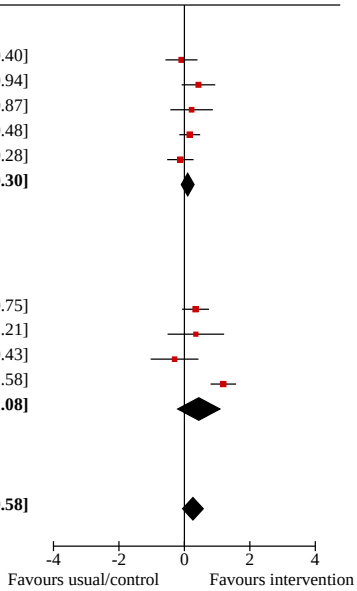
Analysis 1.10. Comparison 1: Communication intervention versus control or usual care, Outcome 10: Sensitivity_cluster 10%_Relationship building/Rapport

Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.10.1 Assessor rating						
Bosse 2012	-0.09	0.2501	32	32	11.3%	-0.09 [-0.58 , 0.40]
Evans 1989	0.43	0.2614	30	30	11.0%	0.43 [-0.08 , 0.94]
Gartmeir 2015	0.22	0.331	20	17	9.4%	0.22 [-0.43 , 0.87]
Liu 2016	0.1641	0.1624	59	107	13.4%	0.16 [-0.15 , 0.48]
Lupi 2012	-0.123	0.2057	50	45	12.4%	-0.12 [-0.53 , 0.28]
Subtotal (95% CI)			**191**	**231**	**57.6%**	**0.10 [-0.09 , 0.30]**
Heterogeneity: Tau² = 0.00; Chi² = 3.63, df = 4 (P = 0.46); I² = 0%						
Test for overall effect: Z = 1.01 (P = 0.31)						
1.10.2 SP perception/satisfaction/rapport						
Hobgood 2009	0.3392	0.2094	35	68	12.3%	0.34 [-0.07 , 0.75]
Legg 2005	0.35	0.441	11	10	7.2%	0.35 [-0.51 , 1.21]
Shaddheau 2015	-0.3003	0.2909	75	60	10.3%	-0.30 [-0.87 , 0.27]
Windish 2005	1.19	0.1994	59	60	12.6%	1.19 [0.80 , 1.58]
Subtotal (95% CI)			**180**	**198**	**42.4%**	**0.42 [-0.25 , 1.09]**
Heterogeneity: Tau² = 0.38; Chi² = 20.06, df = 3 (P = 0.0002); I² = 85%						
Test for overall effect: Z = 1.22 (P = 0.22)						
Total (95% CI)			**371**	**429**	**100.0%**	**0.25 [-0.07 , 0.57]**
Heterogeneity: Tau² = 0.17; Chi² = 31.81, df = 8 (P = 0.0001); I² = 75%						
Test for overall effect: Z = 1.55 (P = 0.12)						
Test for subgroup differences: Chi² = 0.79, df = 1 (P = 0.37), I² = 0%						



Analysis 1.11. Comparison 1: Communication intervention versus control or usual care, Outcome 11: Sensitivity_cluster 20%_Relationship building/Rapport

Study or Subgroup	SMD	SE	Intervention Total	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.11.1 Assessor rating						
Bosse 2012	-0.09	0.2501	32	32	11.5%	-0.09 [-0.58 , 0.40]
Evans 1989	0.43	0.2614	30	30	11.2%	0.43 [-0.08 , 0.94]
Gartmeir 2015	0.22	0.331	20	17	9.5%	0.22 [-0.43 , 0.87]
Liu 2016	0.1641	0.1624	59	107	13.8%	0.16 [-0.15 , 0.48]
Lupi 2012	-0.123	0.2057	50	45	12.7%	-0.12 [-0.53 , 0.28]
Subtotal (95% CI)			**191**	**231**	**58.7%**	**0.10 [-0.09 , 0.30]**
Heterogeneity: Tau² = 0.00; Chi² = 3.63, df = 4 (P = 0.46); I² = 0%						
Test for overall effect: Z = 1.01 (P = 0.31)						
1.11.2 SP perception/satisfaction/rapport						
Hobgood 2009	0.3392	0.2094	35	68	12.6%	0.34 [-0.07 , 0.75]
Legg 2005	0.35	0.441	11	10	7.2%	0.35 [-0.51 , 1.21]
Shaddheau 2015	-0.3003	0.3728	75	60	8.6%	-0.30 [-1.03 , 0.43]
Windish 2005	1.19	0.1994	59	60	12.8%	1.19 [0.80 , 1.58]
Subtotal (95% CI)			**180**	**198**	**41.3%**	**0.44 [-0.20 , 1.08]**
Heterogeneity: Tau² = 0.34; Chi² = 16.44, df = 3 (P = 0.0009); I² = 82%						
Test for overall effect: Z = 1.34 (P = 0.18)						
Total (95% CI)			**371**	**429**	**100.0%**	**0.26 [-0.05 , 0.58]**
Heterogeneity: Tau² = 0.16; Chi² = 30.22, df = 8 (P = 0.0002); I² = 74%						
Test for overall effect: Z = 1.63 (P = 0.10)						
Test for subgroup differences: Chi² = 0.98, df = 1 (P = 0.32), I² = 0%						

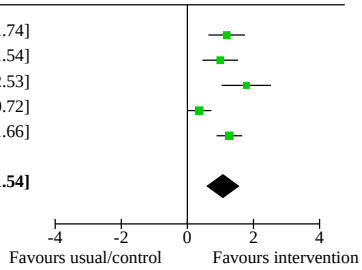


Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.12. Comparison 1: Communication intervention versus control or usual care, Outcome 12: Information gathering about patient perspectives/concerns

Study or Subgroup	Intervention Mean	Intervention SD	Intervention Total	Usual curriculum/control Mean	Usual curriculum/control SD	Usual curriculum/control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	85.72	3.51	28	79.64	6.07	32	19.3%	1.19 [0.64 , 1.74]
Evans 1989	4.5	1.82	30	3	1.05	30	19.6%	1.00 [0.46 , 1.54]
Ho 2008	3.27	0.8	15	1.85	0.77	27	15.8%	1.78 [1.04 , 2.53]
Lee 2015	2.48	0.98	60	2.1	1.11	59	22.8%	0.36 [-0.00 , 0.72]
Schwartz 2010	0.9	0.1614	65	0.62	0.2686	59	22.4%	1.27 [0.88 , 1.66]
Total (95% CI)			**198**			**207**	**100.0%**	**1.07 [0.61 , 1.54]**

Heterogeneity: Tau² = 0.22; Chi² = 18.33, df = 4 (P = 0.001); I² = 78%
Test for overall effect: Z = 4.49 (P < 0.00001)
Test for subgroup differences: Not applicable



Comparison 2. Communication intervention 1 versus communication intervention 2 mode of delivery

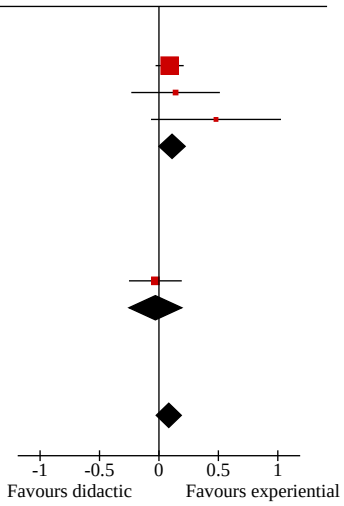
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 Overall communication skills	4	1578	Std. Mean Difference (IV, Random, 95% CI)	0.08 [-0.02, 0.19]
2.1.1 Assessor rating	3	1266	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.00, 0.22]
2.1.2 SP rating	1	312	Std. Mean Difference (IV, Random, 95% CI)	-0.03 [-0.25, 0.19]
2.2 Sensitivity_cluster 5%_Overall communication skills	4	1578	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.00, 0.22]
2.2.1 Assessor rating	3	1266	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.00, 0.22]
2.2.2 SP rating	1	312	Std. Mean Difference (IV, Random, 95% CI)	-0.03 [-0.69, 0.63]
2.3 Sensitivity_cluster 10%_Overall communication skills	4	1578	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.00, 0.22]
2.3.1 Assessor rating	3	1266	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.00, 0.22]
2.3.2 SP rating	1	312	Std. Mean Difference (IV, Random, 95% CI)	-0.03 [-0.93, 0.87]
2.4 Sensitivity_cluster 20%_Overall communication skills	4	1578	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.00, 0.22]
2.4.1 Assessor rating	3	1266	Std. Mean Difference (IV, Random, 95% CI)	0.11 [-0.00, 0.22]

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.4.2 SP rating	1	312	Std. Mean Difference (IV, Random, 95% CI)	-0.03 [-1.29, 1.23]
2.5 Empathy	3	421	Std. Mean Difference (IV, Random, 95% CI)	-0.13 [-0.68, 0.43]
2.5.1 Assessor rating	2	149	Std. Mean Difference (IV, Random, 95% CI)	-0.18 [-1.40, 1.04]
2.5.2 SP report	1	272	Std. Mean Difference (IV, Random, 95% CI)	-0.13 [-0.37, 0.11]
2.6 Relationship building/rapport	3	176	Std. Mean Difference (IV, Random, 95% CI)	0.02 [-0.33, 0.38]
2.6.1 Assessor rating	2	126	Std. Mean Difference (IV, Random, 95% CI)	0.16 [-0.20, 0.52]
2.6.2 SP perception/satisfaction/rapport	1	50	Std. Mean Difference (IV, Random, 95% CI)	-0.35 [-0.96, 0.26]

Analysis 2.1. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 1: Overall communication skills

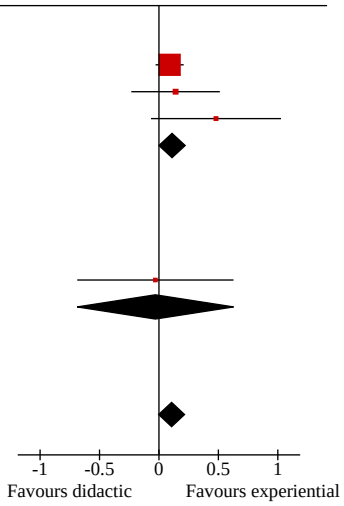
Study or Subgroup	SMD	SE	Experiential/video Total	Didactic Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 Assessor rating						
Ockene 2016	0.09	0.0603	571	531	67.1%	0.09 [-0.03 , 0.21]
Roche 1997	0.14	0.1901	56	55	7.8%	0.14 [-0.23 , 0.51]
Walsh 1999	0.48	0.279	27	26	3.7%	0.48 [-0.07 , 1.03]
Subtotal (95% CI)			**654**	**612**	**78.6%**	**0.11 [-0.00 , 0.22]**
Heterogeneity: Tau² = 0.00; Chi² = 1.89, df = 2 (P = 0.39); I² = 0%						
Test for overall effect: Z = 1.96 (P = 0.05)						
2.1.2 SP rating						
Wagner 2011	-0.03	0.1133	151	161	21.4%	-0.03 [-0.25 , 0.19]
Subtotal (95% CI)			**151**	**161**	**21.4%**	**-0.03 [-0.25 , 0.19]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.26 (P = 0.79)						
Total (95% CI)			**805**	**773**	**100.0%**	**0.08 [-0.02 , 0.19]**
Heterogeneity: Tau² = 0.00; Chi² = 3.12, df = 3 (P = 0.37); I² = 4%						
Test for overall effect: Z = 1.54 (P = 0.12)						
Test for subgroup differences: Chi² = 1.23, df = 1 (P = 0.27), I² = 18.6%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

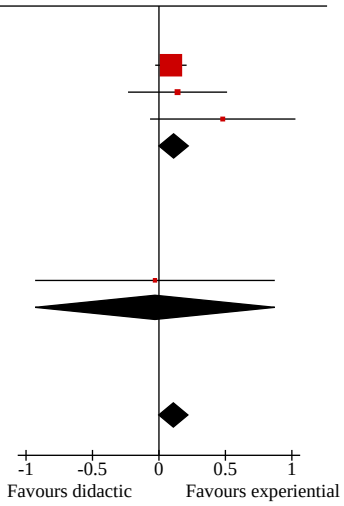
Analysis 2.2. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 2: Sensitivity_cluster 5%_Overall communication skills

Study or Subgroup	SMD	SE	Experiential/video Total	Didactic Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 Assessor rating						
Ockene 2016	0.09	0.0603	571	531	84.8%	0.09 [-0.03 , 0.21]
Roche 1997	0.14	0.1901	56	55	8.5%	0.14 [-0.23 , 0.51]
Walsh 1999	0.48	0.279	27	26	4.0%	0.48 [-0.07 , 1.03]
Subtotal (95% CI)			**654**	**612**	**97.3%**	**0.11 [-0.00 , 0.22]**
Heterogeneity: Tau² = 0.00; Chi² = 1.89, df = 2 (P = 0.39); I² = 0%						
Test for overall effect: Z = 1.96 (P = 0.05)						
2.2.2 SP rating						
Wagner 2011	-0.03	0.3354	151	161	2.7%	-0.03 [-0.69 , 0.63]
Subtotal (95% CI)			**151**	**161**	**2.7%**	**-0.03 [-0.69 , 0.63]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.09 (P = 0.93)						
Total (95% CI)			**805**	**773**	**100.0%**	**0.11 [-0.00 , 0.22]**
Heterogeneity: Tau² = 0.00; Chi² = 2.06, df = 3 (P = 0.56); I² = 0%						
Test for overall effect: Z = 1.92 (P = 0.06)						
Test for subgroup differences: Chi² = 0.17, df = 1 (P = 0.68), I² = 0%						



Analysis 2.3. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 3: Sensitivity_cluster 10%_Overall communication skills

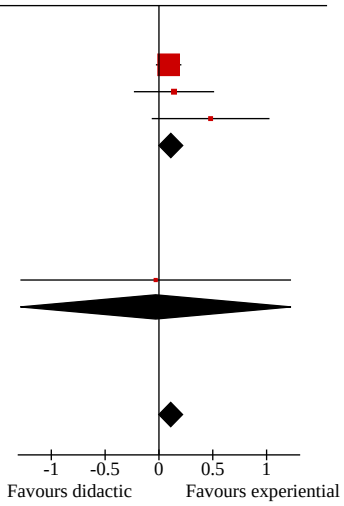
Study or Subgroup	SMD	SE	Experiential/video Total	Didactic Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.3.1 Assessor rating						
Ockene 2016	0.09	0.0603	571	531	85.9%	0.09 [-0.03 , 0.21]
Roche 1997	0.14	0.1901	56	55	8.6%	0.14 [-0.23 , 0.51]
Walsh 1999	0.48	0.279	27	26	4.0%	0.48 [-0.07 , 1.03]
Subtotal (95% CI)			**654**	**612**	**98.5%**	**0.11 [-0.00 , 0.22]**
Heterogeneity: Tau² = 0.00; Chi² = 1.89, df = 2 (P = 0.39); I² = 0%						
Test for overall effect: Z = 1.96 (P = 0.05)						
2.3.2 SP rating						
Wagner 2011	-0.03	0.46	151	161	1.5%	-0.03 [-0.93 , 0.87]
Subtotal (95% CI)			**151**	**161**	**1.5%**	**-0.03 [-0.93 , 0.87]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.07 (P = 0.95)						
Total (95% CI)			**805**	**773**	**100.0%**	**0.11 [-0.00 , 0.22]**
Heterogeneity: Tau² = 0.00; Chi² = 1.99, df = 3 (P = 0.58); I² = 0%						
Test for overall effect: Z = 1.94 (P = 0.05)						
Test for subgroup differences: Chi² = 0.09, df = 1 (P = 0.76), I² = 0%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

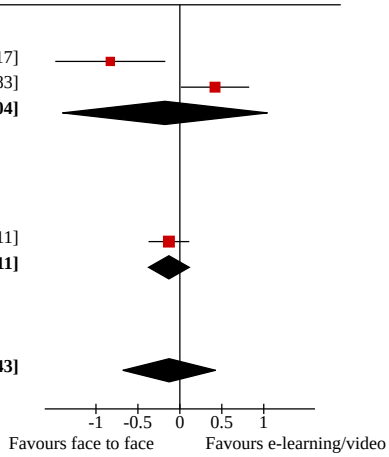
Analysis 2.4. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 4: Sensitivity_cluster 20%_Overall communication skills

Study or Subgroup	SMD	SE	Experiential/video Total	Didactic Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.4.1 Assessor rating						
Ockene 2016	0.09	0.0603	571	531	86.5%	0.09 [-0.03 , 0.21]
Roche 1997	0.14	0.1901	56	55	8.7%	0.14 [-0.23 , 0.51]
Walsh 1999	0.48	0.279	27	26	4.0%	0.48 [-0.07 , 1.03]
Subtotal (95% CI)			**654**	**612**	**99.2%**	**0.11 [-0.00 , 0.22]**
Heterogeneity: Tau² = 0.00; Chi² = 1.89, df = 2 (P = 0.39); I² = 0%						
Test for overall effect: Z = 1.96 (P = 0.05)						
2.4.2 SP rating						
Wagner 2011	-0.03	0.6413	151	161	0.8%	-0.03 [-1.29 , 1.23]
Subtotal (95% CI)			**151**	**161**	**0.8%**	**-0.03 [-1.29 , 1.23]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.05 (P = 0.96)						
Total (95% CI)			**805**	**773**	**100.0%**	**0.11 [-0.00 , 0.22]**
Heterogeneity: Tau² = 0.00; Chi² = 1.94, df = 3 (P = 0.58); I² = 0%						
Test for overall effect: Z = 1.95 (P = 0.05)						
Test for subgroup differences: Chi² = 0.05, df = 1 (P = 0.83), I² = 0%						



Analysis 2.5. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 5: Empathy

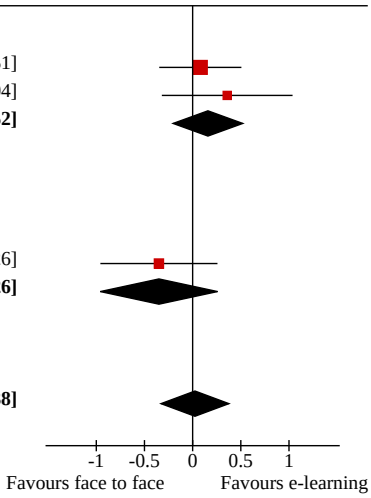
Study or Subgroup	SMD	SE	E-learning/video Total	Face to face Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.5.1 Assessor rating						
Palmer 2018	-0.8292	0.3345	24	26	26.5%	-0.83 [-1.48 , -0.17]
Kaltman 2018	0.4189	0.2079	60	39	34.4%	0.42 [0.01 , 0.83]
Subtotal (95% CI)			**84**	**65**	**60.9%**	**-0.18 [-1.40 , 1.04]**
Heterogeneity: Tau² = 0.70; Chi² = 10.04, df = 1 (P = 0.002); I² = 90%						
Test for overall effect: Z = 0.28 (P = 0.78)						
2.5.2 SP report						
Bowyer 2010	-0.1304	0.1239	163	109	39.1%	-0.13 [-0.37 , 0.11]
Subtotal (95% CI)			**163**	**109**	**39.1%**	**-0.13 [-0.37 , 0.11]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.05 (P = 0.29)						
Total (95% CI)			**247**	**174**	**100.0%**	**-0.13 [-0.68 , 0.43]**
Heterogeneity: Tau² = 0.19; Chi² = 10.92, df = 2 (P = 0.004); I² = 82%						
Test for overall effect: Z = 0.45 (P = 0.65)						
Test for subgroup differences: Chi² = 0.01, df = 1 (P = 0.94), I² = 0%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.6. Comparison 2: Communication intervention 1 versus communication intervention 2 mode of delivery, Outcome 6: Relationship building/rapport

Study or Subgroup	SMD	SE	E-learning Total	Face to face Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.6.1 Assessor rating						
Liu 2016	0.0797	0.2175	33	59	48.7%	0.08 [-0.35 , 0.51]
Gartmeir 2015	0.3587	0.3461	17	17	23.2%	0.36 [-0.32 , 1.04]
Subtotal (95% CI)			**50**	**76**	**72.0%**	**0.16 [-0.20 , 0.52]**
Heterogeneity: Tau² = 0.00; Chi² = 0.47, df = 1 (P = 0.49); I² = 0%						
Test for overall effect: Z = 0.86 (P = 0.39)						
2.6.2 SP perception/satisfaction/rapport						
Palmer 2018	-0.35	0.31	24	26	28.0%	-0.35 [-0.96 , 0.26]
Subtotal (95% CI)			**24**	**26**	**28.0%**	**-0.35 [-0.96 , 0.26]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.13 (P = 0.26)						
Total (95% CI)			**74**	**102**	**100.0%**	**0.02 [-0.33 , 0.38]**
Heterogeneity: Tau² = 0.02; Chi² = 2.46, df = 2 (P = 0.29); I² = 19%						
Test for overall effect: Z = 0.13 (P = 0.89)						
Test for subgroup differences: Chi² = 1.99, df = 1 (P = 0.16), I² = 49.8%						



Comparison 3. Communication intervention 1 versus communication intervention 2 feedback approach

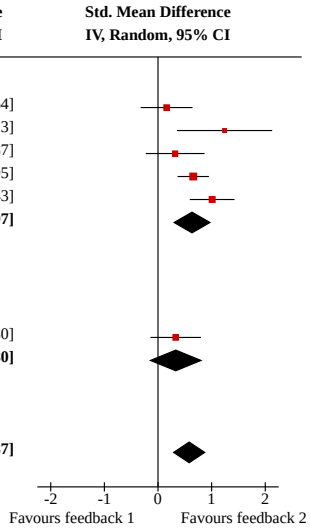
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 Overall communication skills	6	502	Std. Mean Difference (IV, Random, 95% CI)	0.58 [0.29, 0.87]
3.1.1 Assessor rating	5	432	Std. Mean Difference (IV, Random, 95% CI)	0.63 [0.30, 0.97]
3.1.2 SP rating	1	70	Std. Mean Difference (IV, Random, 95% CI)	0.33 [-0.14, 0.80]
3.2 Sensitivity_RoB_Overall communication skills	5		Std. Mean Difference (IV, Random, 95% CI)	0.56 [0.17, 0.96]
3.2.1 Assessor rating	4		Std. Mean Difference (IV, Random, 95% CI)	0.64 [0.14, 1.14]
3.2.2 SP rating	1		Std. Mean Difference (IV, Random, 95% CI)	0.33 [-0.14, 0.80]
3.3 Sensitivity_cluster 5%_Overall communication skills	6	502	Std. Mean Difference (IV, Random, 95% CI)	0.59 [0.31, 0.88]
3.3.1 Assessor rating	5	432	Std. Mean Difference (IV, Random, 95% CI)	0.65 [0.32, 0.97]
3.3.2 SP rating	1	70	Std. Mean Difference (IV, Random, 95% CI)	0.33 [-0.14, 0.80]

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.4 Sensitivity_Cluster 10%_Overall communication skills	6	502	Std. Mean Difference (IV, Random, 95% CI)	0.60 [0.32, 0.88]
3.4.1 Assessor rating	5	432	Std. Mean Difference (IV, Random, 95% CI)	0.66 [0.34, 0.98]
3.4.2 SP rating	1	70	Std. Mean Difference (IV, Random, 95% CI)	0.33 [-0.14, 0.80]
3.5 Sensitivity_Cluster 20%_Overall communication skills	6	502	Std. Mean Difference (IV, Random, 95% CI)	0.61 [0.33, 0.89]
3.5.1 Assessor rating	5	432	Std. Mean Difference (IV, Random, 95% CI)	0.68 [0.36, 0.99]
3.5.2 SP rating	1	70	Std. Mean Difference (IV, Random, 95% CI)	0.33 [-0.14, 0.80]

Analysis 3.1. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 1: Overall communication skills

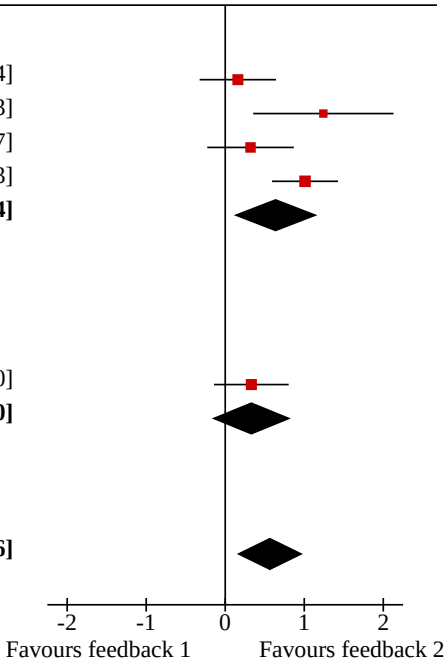
Study or Subgroup	SMD	SE	Intervention 1_feedback 1 Total	Intervention 2_feedback 2 Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.1.1 Assessor rating						
Engerer 2019	0.16	0.2467	34	32	16.8%	0.16 [-0.32 , 0.64]
Maguire 1978	1.2414	0.4529	12	12	8.0%	1.24 [0.35 , 2.13]
Ozcakar 2009	0.32	0.2795	27	25	14.9%	0.32 [-0.23 , 0.87]
Perera 2010	0.6572	0.1491	97	93	24.0%	0.66 [0.36 , 0.95]
Ruesseler 2017	1.01	0.2129	49	51	19.1%	1.01 [0.59 , 1.43]
Subtotal (95% CI)			**219**	**213**	**82.8%**	**0.63 [0.30 , 0.97]**
Heterogeneity: Tau² = 0.08; Chi² = 9.89, df = 4 (P = 0.04); I² = 60%						
Test for overall effect: Z = 3.71 (P = 0.0002)						
3.1.2 SP rating						
Hobgood 2009	0.3306	0.2408	35	35	17.2%	0.33 [-0.14 , 0.80]
Subtotal (95% CI)			**35**	**35**	**17.2%**	**0.33 [-0.14 , 0.80]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.37 (P = 0.17)						
Total (95% CI)			**254**	**248**	**100.0%**	**0.58 [0.29 , 0.87]**
Heterogeneity: Tau² = 0.07; Chi² = 11.29, df = 5 (P = 0.05); I² = 56%						
Test for overall effect: Z = 3.93 (P < 0.0001)						
Test for subgroup differences: Chi² = 1.06, df = 1 (P = 0.30), I² = 5.4%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

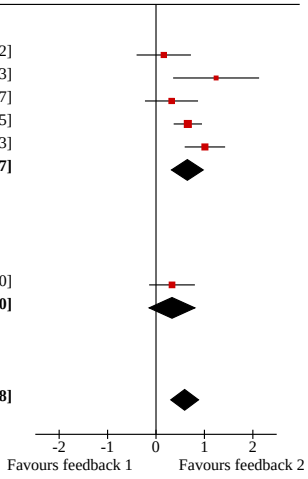
Analysis 3.2. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 2: Sensitivity_RoB_Overall communication skills

Study or Subgroup	SMD	SE	Weight	Std. Mean Difference IV, Random, 95% CI
3.2.1 Assessor rating				
Engerer 2019	0.16	0.2467	21.8%	0.16 [-0.32 , 0.64]
Maguire 1978	1.2414	0.4529	12.3%	1.24 [0.35 , 2.13]
Ozcakar 2009	0.32	0.2795	20.0%	0.32 [-0.23 , 0.87]
Ruesseler 2017	1.01	0.2129	23.8%	1.01 [0.59 , 1.43]
Subtotal (95% CI)			**77.8%**	**0.64 [0.14 , 1.14]**
Heterogeneity: Tau² = 0.17; Chi² = 9.87, df = 3 (P = 0.02); I² = 70%				
Test for overall effect: Z = 2.50 (P = 0.01)				
3.2.2 SP rating				
Hobgood 2009	0.3306	0.2408	22.2%	0.33 [-0.14 , 0.80]
Subtotal (95% CI)			**22.2%**	**0.33 [-0.14 , 0.80]**
Heterogeneity: Not applicable				
Test for overall effect: Z = 1.37 (P = 0.17)				
Total (95% CI)			**100.0%**	**0.56 [0.17 , 0.96]**
Heterogeneity: Tau² = 0.12; Chi² = 11.00, df = 4 (P = 0.03); I² = 64%				
Test for overall effect: Z = 2.81 (P = 0.005)				
Test for subgroup differences: Chi² = 0.77, df = 1 (P = 0.38), I² = 0%				



Analysis 3.3. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 3: Sensitivity_cluster 5%_Overall communication skills

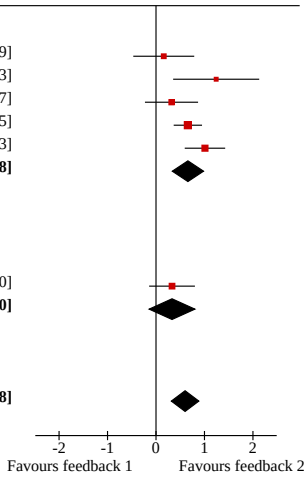
Study or Subgroup	SMD	SE	Intervention 1_feedback 1 Total	Intervention 2_feedback 2 Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.3.1 Assessor rating						
Engerer 2019	0.16	0.2862	34	32	14.7%	0.16 [-0.40 , 0.72]
Maguire 1978	1.2414	0.4529	12	12	7.9%	1.24 [0.35 , 2.13]
Ozcakar 2009	0.32	0.2795	27	25	15.1%	0.32 [-0.23 , 0.87]
Perera 2010	0.6572	0.1491	97	93	25.0%	0.66 [0.36 , 0.95]
Ruesseler 2017	1.01	0.2129	49	51	19.7%	1.01 [0.59 , 1.43]
Subtotal (95% CI)			**219**	**213**	**82.4%**	**0.65 [0.32 , 0.97]**
Heterogeneity: Tau² = 0.07; Chi² = 8.88, df = 4 (P = 0.06); I² = 55%						
Test for overall effect: Z = 3.87 (P = 0.0001)						
3.3.2 SP rating						
Hobgood 2009	0.3306	0.2408	35	35	17.6%	0.33 [-0.14 , 0.80]
Subtotal (95% CI)			**35**	**35**	**17.6%**	**0.33 [-0.14 , 0.80]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.37 (P = 0.17)						
Total (95% CI)			**254**	**248**	**100.0%**	**0.59 [0.31 , 0.88]**
Heterogeneity: Tau² = 0.06; Chi² = 10.47, df = 5 (P = 0.06); I² = 52%						
Test for overall effect: Z = 4.05 (P < 0.0001)						
Test for subgroup differences: Chi² = 1.17, df = 1 (P = 0.28), I² = 14.3%						



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

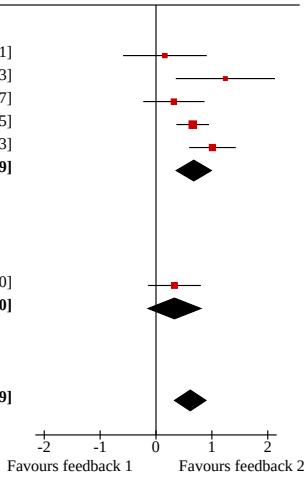
Analysis 3.4. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 4: Sensitivity_Cluster 10%_Overall communication skills

Study or Subgroup	SMD	SE	Intervention 1_feedback 1 Total	Intervention 2_feedback 2 Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.4.1 Assessor rating						
Engerer 2019	0.16	0.3207	34	32	13.0%	0.16 [-0.47 , 0.79]
Maguire 1978	1.2414	0.4529	12	12	7.9%	1.24 [0.35 , 2.13]
Ozcakar 2009	0.32	0.2795	27	25	15.3%	0.32 [-0.23 , 0.87]
Perera 2010	0.6572	0.1491	97	93	25.8%	0.66 [0.36 , 0.95]
Ruesseler 2017	1.01	0.2129	49	51	20.1%	1.01 [0.59 , 1.43]
Subtotal (95% CI)			**219**	**213**	**82.1%**	**0.66 [0.34 , 0.98]**
Heterogeneity: Tau² = 0.07; Chi² = 8.25, df = 4 (P = 0.08); I² = 51%						
Test for overall effect: Z = 4.01 (P < 0.0001)						
3.4.2 SP rating						
Hobgood 2009	0.3306	0.2408	35	35	17.9%	0.33 [-0.14 , 0.80]
Subtotal (95% CI)			**35**	**35**	**17.9%**	**0.33 [-0.14 , 0.80]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.37 (P = 0.17)						
Total (95% CI)			**254**	**248**	**100.0%**	**0.60 [0.32 , 0.88]**
Heterogeneity: Tau² = 0.06; Chi² = 9.95, df = 5 (P = 0.08); I² = 50%						
Test for overall effect: Z = 4.14 (P < 0.0001)						
Test for subgroup differences: Chi² = 1.26, df = 1 (P = 0.26), I² = 20.9%						



Analysis 3.5. Comparison 3: Communication intervention 1 versus communication intervention 2 feedback approach, Outcome 5: Sensitivity_Cluster 20%_Overall communication skills

Study or Subgroup	SMD	SE	Intervention 1_feedback 1 Total	Intervention 2_feedback 2 Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.5.1 Assessor rating						
Engerer 2019	0.16	0.3824	34	32	10.3%	0.16 [-0.59 , 0.91]
Maguire 1978	1.2414	0.4529	12	12	7.9%	1.24 [0.35 , 2.13]
Ozcakar 2009	0.32	0.2795	27	25	15.6%	0.32 [-0.23 , 0.87]
Perera 2010	0.6572	0.1491	97	93	27.1%	0.66 [0.36 , 0.95]
Ruesseler 2017	1.01	0.2129	49	51	20.7%	1.01 [0.59 , 1.43]
Subtotal (95% CI)			**219**	**213**	**81.6%**	**0.68 [0.36 , 0.99]**
Heterogeneity: Tau² = 0.06; Chi² = 7.46, df = 4 (P = 0.11); I² = 46%						
Test for overall effect: Z = 4.21 (P < 0.0001)						
3.5.2 SP rating						
Hobgood 2009	0.3306	0.2408	35	35	18.4%	0.33 [-0.14 , 0.80]
Subtotal (95% CI)			**35**	**35**	**18.4%**	**0.33 [-0.14 , 0.80]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.37 (P = 0.17)						
Total (95% CI)			**254**	**248**	**100.0%**	**0.61 [0.33 , 0.89]**
Heterogeneity: Tau² = 0.05; Chi² = 9.33, df = 5 (P = 0.10); I² = 46%						
Test for overall effect: Z = 4.28 (P < 0.0001)						
Test for subgroup differences: Chi² = 1.43, df = 1 (P = 0.23), I² = 29.9%						



Comparison 4. Communication intervention 1 versus communication intervention 2 SP vs peer

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.1 Overall communication skills	4	637	Std. Mean Difference (IV, Random, 95% CI)	0.17 [-0.33, 0.67]
4.2 Subgroup_Overall communication skills_SP only	3	573	Std. Mean Difference (IV, Random, 95% CI)	0.20 [-0.41, 0.82]

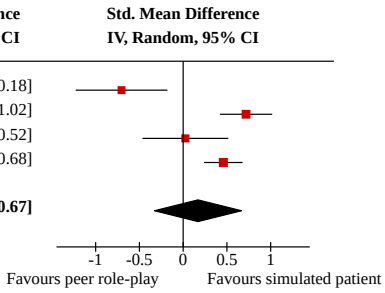
Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.3 Sensitivity_RoB_Overall communication skills	2	124	Std. Mean Difference (IV, Random, 95% CI)	-0.33 [-1.05, 0.38]
4.4 Sensitivity_Cluster 5%_Overall communication skills	4	637	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.44, 0.68]
4.5 Sensitivity_Cluster 10%_Overall communication skills	4	637	Std. Mean Difference (IV, Random, 95% CI)	0.09 [-0.49, 0.68]
4.6 Sensitivity_Cluster 20%_Overall communication skills	4	637	Std. Mean Difference (IV, Random, 95% CI)	0.06 [-0.55, 0.67]

Analysis 4.1. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 1: Overall communication skills

Study or Subgroup	SMD	SE	Simulated (real) patient Total	Peer role-play Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	-0.7043	0.2672	32	28	22.3%	-0.70 [-1.23 , -0.18]
Levenkron 1990	0.7188	0.1526	91	92	26.7%	0.72 [0.42 , 1.02]
Stolz 2012	0.0256	0.2501	33	31	23.0%	0.03 [-0.46 , 0.52]
Taylor 2019	0.4591	0.1121	148	182	28.0%	0.46 [0.24 , 0.68]
Total (95% CI)			**304**	**333**	**100.0%**	**0.17 [-0.33 , 0.67]**

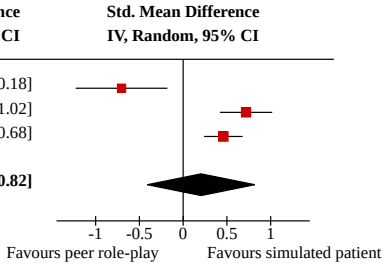
Heterogeneity: Tau² = 0.22; Chi² = 23.90, df = 3 (P < 0.0001); I² = 87%
Test for overall effect: Z = 0.67 (P = 0.50)
Test for subgroup differences: Not applicable



Analysis 4.2. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 2: Subgroup_Overall communication skills_SP only

Study or Subgroup	SMD	SE	Simulated (real) patient Total	Peer role-play Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	-0.7043	0.2672	32	28	29.6%	-0.70 [-1.23 , -0.18]
Levenkron 1990	0.7188	0.1526	91	92	34.5%	0.72 [0.42 , 1.02]
Taylor 2019	0.4591	0.1121	148	182	35.9%	0.46 [0.24 , 0.68]
Total (95% CI)			**271**	**302**	**100.0%**	**0.20 [-0.41 , 0.82]**

Heterogeneity: Tau² = 0.26; Chi² = 21.66, df = 2 (P < 0.0001); I² = 91%
Test for overall effect: Z = 0.65 (P = 0.51)
Test for subgroup differences: Not applicable

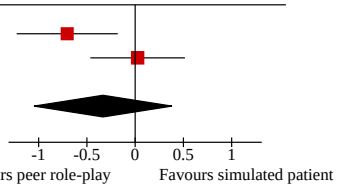


Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 4.3. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 3: Sensitivity_RoB_Overall communication skills

Study or Subgroup	SMD	SE	Simulated (real) patient Total	Peer role-play Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	-0.7043	0.2672	32	28	49.2%	-0.70 [-1.23 , -0.18]
Stolz 2012	0.0256	0.2501	33	31	50.8%	0.03 [-0.46 , 0.52]
Total (95% CI)			**65**	**59**	**100.0%**	**-0.33 [-1.05 , 0.38]**

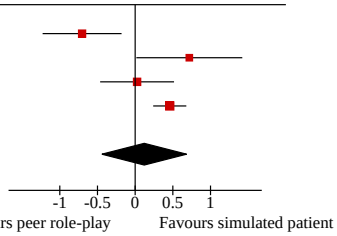
Heterogeneity: Tau² = 0.20; Chi² = 3.98, df = 1 (P = 0.05); I² = 75%
Test for overall effect: Z = 0.91 (P = 0.36)
Test for subgroup differences: Not applicable



Analysis 4.4. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 4: Sensitivity_Cluster 5%_Overall communication skills

Study or Subgroup	SMD	SE	Simulated (real) patient Total	Peer role-play Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	-0.7043	0.2672	32	28	24.4%	-0.70 [-1.23 , -0.18]
Levenkron 1990	0.7188	0.3586	91	92	20.9%	0.72 [0.02 , 1.42]
Stolz 2012	0.0256	0.2501	33	31	25.1%	0.03 [-0.46 , 0.52]
Taylor 2019	0.4591	0.1121	148	182	29.6%	0.46 [0.24 , 0.68]
Total (95% CI)			**304**	**333**	**100.0%**	**0.12 [-0.44 , 0.68]**

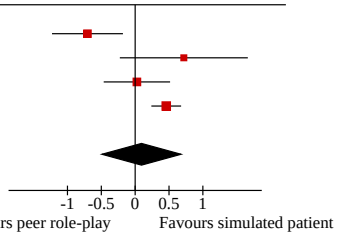
Heterogeneity: Tau² = 0.27; Chi² = 18.65, df = 3 (P = 0.0003); I² = 84%
Test for overall effect: Z = 0.42 (P = 0.67)
Test for subgroup differences: Not applicable



Analysis 4.5. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 5: Sensitivity_Cluster 10%_Overall communication skills

Study or Subgroup	SMD	SE	Simulated (real) patient Total	Peer role-play Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	-0.7043	0.2672	32	28	25.6%	-0.70 [-1.23 , -0.18]
Levenkron 1990	0.7188	0.4822	91	92	17.5%	0.72 [-0.23 , 1.66]
Stolz 2012	0.0256	0.2501	33	31	26.2%	0.03 [-0.46 , 0.52]
Taylor 2019	0.4591	0.1121	148	182	30.8%	0.46 [0.24 , 0.68]
Total (95% CI)			**304**	**333**	**100.0%**	**0.09 [-0.49 , 0.68]**

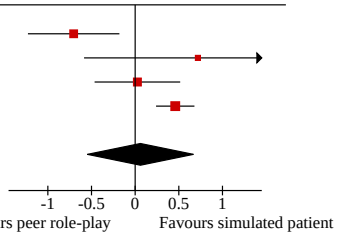
Heterogeneity: Tau² = 0.28; Chi² = 17.96, df = 3 (P = 0.0004); I² = 83%
Test for overall effect: Z = 0.31 (P = 0.75)
Test for subgroup differences: Not applicable



Analysis 4.6. Comparison 4: Communication intervention 1 versus communication intervention 2 SP vs peer, Outcome 6: Sensitivity_Cluster 20%_Overall communication skills

Study or Subgroup	SMD	SE	Simulated (real) patient Total	Peer role-play Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bosse 2012	-0.7043	0.2672	32	28	26.9%	-0.70 [-1.23 , -0.18]
Levenkron 1990	0.7188	0.6653	91	92	13.2%	0.72 [-0.59 , 2.02]
Stolz 2012	0.0256	0.2501	33	31	27.6%	0.03 [-0.46 , 0.52]
Taylor 2019	0.4591	0.1121	148	182	32.3%	0.46 [0.24 , 0.68]
Total (95% CI)			**304**	**333**	**100.0%**	**0.06 [-0.55 , 0.67]**

Heterogeneity: Tau² = 0.28; Chi² = 17.53, df = 3 (P = 0.0005); I² = 83%
Test for overall effect: Z = 0.19 (P = 0.85)
Test for subgroup differences: Not applicable



Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ADDITIONAL TABLES

Table 1. Features of included studies

	Features of included studies				
Study design	**RCT (n = 55)**	**Quasi-RCT (n = 9)**	**C-RCT (n = 7)**	**Quasi C-RCT (n = 5)**	
	Allen 1990; Bearman 2001; Berney 2017; Betchart 1984; Blatt 2010; Bosse 2012; Bowyer 2010; Buffel du Vaure 2017; Chibhall 2005; Daeppen 2012; Edwardsen 2006; Eells 2002; Evans 1989; Evans 1996; Feddock 2009; Foster 2016; Gartmeir 2015; Gerber 1985; Ho 2008; Hobgood 2009; Ishikawa 2010; Kahan 2003; Kaltman 2018; Knowles 2001; Legg 2005; Liu 2016; LoSasso 2017; Lupi 2012; Maguire 1977; Maguire 1978; Mason 1988; Moreland 1973; Morrow 2009; Mounsey 2006; Nomura 2017; Ozcakar 2009; Palmer 2018; Perera 2010; Pirdehghan 2018; Poole 1980; Roche 1996; Roche 1997; Ruesseler 2017; Schmitz 2018; Siassakos 2010; Shapiro 2009; Snow 2016; Spollen 2010; Steifel 2013; Stolz 2012; Strayer 2010; Vanatta 1996; Weihs 1986; Windish 2005; Wundrich 2017	Alroy 1984; Bowyer 2006; Colletti 2001; Klein 2000; Lorin 2006; Schwartz 2010; Seim 1995; Taylor 2019; Walsh 1999	Cave 2007; Engerer 2019; Lee 2015; Meirovich 2016; Ockene 2016; Papadakis 1997; Shaddheau 2015	Filipetto 2006; Levenkron 1990; Lim 2011; Solomon 2004; Wagner 2011	
Unit of randomisation	**Student**	**Student followed by group**	**Group (class or rotation)**	**Time (year or rotation)**	**Other**
	RCT (n = 42) Bearman 2001; Berney 2017; Betchart 1984; Blatt 2010; Bosse 2012; Bowyer 2010; Chibhall 2005; Daeppen 2012; Eells 2002; Evans 1989; Evans 1996; Foster 2016; Gartmeir 2015; Gerber 1985; Ho 2008; Hobgood 2009; Ishikawa 2010; Kahan 2003; Kaltman 2018; Legg 2005; Liu 2016; LoSasso 2017; Lupi 2012; Maguire 1978; Mason 1988; Moreland 1973; Morrow 2009; Mounsey 2006; Nomura 2017; Ozcakar 2009; Palmer 2018; Perera 2010; Pirdehghan 2018; Poole 1980; Schmitz 2018; Shapiro 2009; Snow 2016; Spollen 2010; Steifel 2013; Stolz 2012; Weihs 1986; Wundrich 2017 Quasi-RCT (n = 3) Bowyer 2006; Colletti 2001; Klein 2000	RCT (n = 13) Allen 1990; Buffel du Vaure 2017; Edwardsen 2006; Feddock 2009; Knowles 2001; Maguire 1977; Roche 1996; Roche 1997; Ruesseler 2017; Siassakos 2010; Strayer 2010; Vanatta 1996; Windish 2005 Quasi-RCT (n = 5) Alroy 1984; Lorin 2006; Schwartz 2010 (student followed by site); Seim 1995; Walsh 1999	C-RCT (n = 5) Cave 2007; Engerer 2019; Lee 2015; Papadakis 1997; Shaddheau 2015	Quasi-C-RCT (n = 5) Filipetto 2006; Levenkron 1990; Lim 2011; Solomon 2004; Wagner 2011	Tutor: Meirovich 2016 University: Ockene 2016

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 1. Features of included studies *(Continued)*

Comparison	Control comparison (n = 33)	Alternate mode of delivery (n=16)	Alternate style of feedback (n=9)	Alternate role-play approach (peer vs simulated patient; n=7)	Other alternate structure (n=22)
	Usual care (n = 10) Bosse 2012; Bowyer 2006; Bowyer 2010; Eells 2002; Evans 1989; Filipetto 2006; Knowles 2001; Lee 2015; Legg 2005; Seim 1995 Control (no specific communication training; n = 16) Allen 1990; Alroy 1984; Buffel du Vaure 2017; Colletti 2001; Daeppen 2012; Ho 2008; Hobgood 2009; Lorin 2006; Mason 1988; Poole 1980; Roche 1996; Schwartz 2010; Shaddheau 2015; Solomon 2004; Wundrich 2017 Pirdehghan 2018 (assumed no intervention - this is not explicitly stated in the text) Waitlist control (n = 7) Evans 1989; Evans 1996; Gartmeir 2015; Lupi 2012; Shapiro 2009; Spollen 2010; Windish 2005	Bowyer 2010; Eells 2002; Gartmeir 2015; Kaltman 2018; Liu 2016; Ockene 2016; Palmer 2018; Roche 1997; Stolz 2012; Strayer 2010; Wagner 2011; Walsh 1999 Alternate tutors: Gerber 1985; Meirovich 2016; Nomura 2017; Snow 2016	Cave 2007; Engerer 2019; Hobgood 2009; Maguire 1978; Moreland 1973; Ozcakar 2009; Perera 2010; Ruesseler 2017; Vanatta 1996	Bosse 2012; Levenkron 1990; Mounsey 2006; Papadakis 1997; Stolz 2012; Taylor 2019; Vanatta 1996	Bearman 2001; Berney 2017; Betchart 1984; Blatt 2010; Bosse 2012; Chibhall 2005; Feddock 2009; Foster 2016; Ho 2008; Ishikawa 2010; Kahan 2003; Klein 2000; Lim 2011; LoSasso 2017; Maguire 1977; Mason 1988; Moreland 1973; Morrow 2009; Schmitz 2018; Siassakos 2010; Steifel 2013; Weihs 1986
Frequency	**1-off (n = 47)**	**Multi-week programme (n = 26)**	**Unclear (n = 2)**	**Other**	
	Allen 1990; Bearman 2001; Blatt 2010; Bowyer 2006; Bowyer 2010; Colletti 2001; Edwardsen 2006; Feddock 2009; Foster 2016; Gartmeir 2015; Gerber 1985; Ho 2008; Hobgood 2009; Ishikawa 2010; Kahan 2003; Legg 2005; Levenkron 1990; Lim 2011; Liu 2016; Lorin 2006; LoSasso 2017; Lupi 2012; Maguire 1977; Maguire 1978; Mason 1988; Moreland 1973; Morrow 2009; Mounsey 2006; Nomura 2017; Ockene 2016; Ozcakar 2009; Palmer 2018; Papadakis 1997; Perera 2010; Poole 1980; Roche 1996; Roche 1997; Ruesseler 2017; Schmitz 2018; Seim 1995; Shaddheau 2015; Snow 2016; Spollen 2010; Steifel 2013; Stolz 2012; Strayer 2010; Wagner 2011 One-off interventions lasted for durations of 20 minutes to 5 hours, with an average of 2 hours across the 35 studies that specified the duration	Alroy 1984; Berney 2017; Betchart 1984; Bosse 2012; Buffel du Vaure 2017; Cave 2007; Daeppen 2012; Eells 2002; Engerer 2019; Evans 1989; Evans 1996; Filipetto 2006; Klein 2000; Knowles 2001; Lee 2015; Meirovich 2016; Pirdehghan 2018; Schwartz 2010; Shapiro 2009; Siassakos 2010; Taylor 2019; Vanatta 1996; Walsh 1999; Weihs 1986; Windish 2005; Wundrich 2017 These programmes ranged from 21 ses-	Chibhall 2005; Solomon 2004	Unlimited access to online module over 3 weeks (Kaltman 2018)	

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 1. Features of included studies *(Continued)*

		sions spread across 2 years (Taylor 2019), 5 hours per week for 10 months (Filipetto 2006), to 2 sessions 2 weeks apart (Berney 2017; Knowles 2001; Walsh 1999). Interventions were frequently contained within rotations of 6 to 8 weeks			
Student year level	**First year (n = 14)**	**Second year (n = 11)**	**Third year (n = 18)**	**Fourth year (n = 17)**	**Fifth year and above (n = 10)**
Unspecified in (n = 4): Bearman 2001; Colletti 2001; Gartmeir 2015; Ockene 2016	Betchart 1984; Edwardsen 2006; Filipetto 2006; Foster 2016; Kaltman 2018; Meirovich 2016*; Morrow 2009; Palmer 2018; Papadakis 1997; Perera 2010; Shapiro 2009; Vanatta 1996; Wagner 2011; Weihs 1986 *Years 1 to 3 The year level was not specified, but students were in the 'pre-clerkship' period of their training in Ishikawa 2010	Allen 1990; Daeppen 2012; Eells 2002; Gerber 1985; Levenkron 1990; Liu 2016; Moreland 1973; Ozcakar 2009; Spollen 2010; Taylor 2019; Windish 2005	Blatt 2010; Bowyer 2006; Bowyer 2010; Cave 2007; Chibhall 2005; Engerer 2019; Feddock 2009; Kahan 2003; Klein 2000; Lee 2015; LoSasso 2017; Lupi 2012; Mounsey 2006; Shaddheau 2015; Solomon 2004; Stolz 2012; Strayer 2010; Wundrich 2017	Alroy 1984; Berney 2017; Buffel du Vaure 2017; Evans 1989; Evans 1996; Hobgood 2009; Knowles 2001; Lorin 2006; Mason 1988; Maguire 1977; Nomura 2017; Ruesseler 2017; Schmitz 2018; Schwartz 2010; Seim 1995; Siassakos 2010; Steifel 2013 The year level was not specified, but students were recruited during a psychiatry clerkship in Maguire 1978	Bosse 2012; Ho 2008; Legg 2005*; Lim 2011; Pirdehghan 2018**; Poole 1980; Roche 1996; Roche 1997; Snow 2016; Walsh 1999 *Sixth year **Sixth and seventh years
Mode of delivery	**Face-to-face (n = 57)**	**Online (n = 8)**	**Video (n = 1)**	**Face-to-face and online (n = 6)**	**Face-to-face including video (n = 24)**
	Allen 1990; Alroy 1984; Berney 2017; Betchart 1984; Blatt 2010; Bosse 2012; Bowyer 2006; Bowyer 2010; Buffel du Vaure 2017; Cave 2007; Chibhall 2005; Colletti 2001; Daeppen 2012; Edwardsen 2006; Eells 2002; Engerer 2019; Evans 1989; Evans 1996; Feddock 2009; Filipetto 2006; Gart-	Bearman 2001; Foster 2016; Kaltman 2018; Lee 2015; Palmer 2018; Schmitz 2018; Snow 2016; Wagner 2011	Mason 1988	Gartmeir 2015; Liu 2016; Ockene 2016; Solomon 2004; Stolz 2012; Strayer 2010	Used as learning resource/model or demonstration: Allen 1990; Alroy 1984; Bowyer 2010;

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 1. Features of included studies *(Continued)*

	meir 2015; Gerber 1985; Ho 2008; Hobgood 2009; Ishikawa 2010; Kahan 2003; Klein 2000; Knowles 2001; Legg 2005; Levenkron 1990; Lim 2011; Lorin 2006; LoSasso 2017; Lupi 2012; Maguire 1977; Maguire 1978; Meirovich 2016; Moreland 1973; Morrow 2009; Mounsey 2006; Nomura 2017; Ozcakar 2009; Papadakis 1997; Perera 2010; Pirdehghan 2018; Poole 1980; Roche 1996; Roche 1997; Ruesseler 2017; Schwartz 2010; Seim 1995; Shaddheau 2015; Shapiro 2009; Siassakos 2010; Spollen 2010; Steifel 2013; Taylor 2019; Vanatta 1996; Walsh 1999; Weihs 1986; Windish 2005; Wundrich 2017			Daeppen 2012; Ho 2008; Kahan 2003; Klein 2000; Maguire 1977; Maguire 1978; Mason 1988; Schmitz 2018; Seim 1995; Wagner 2011; Windish 2005 Video-recording of role-play for feedback purposes: Berney 2017; Betchart 1984; Cave 2007; Kaltman 2018; Knowles 2001; Ozcakar 2009; Roche 1996; Roche 1997; Ruesseler 2017 Both: Moreland 1973
Intervention components	**Group workshops (n = 48 studies)**	**Lectures (n = 27 studies)**	**Role-play (n = 56 studies)**	
	69 interventions in Alroy 1984; Blatt 2010; Bowyer 2006; Bowyer 2010; Buffel du Vaure 2017; Cave 2007; Chibhall 2005; Daeppen 2012; Edwardsen 2006; Eells 2002; Engerer 2019; Evans 1989; Evans 1996; Feddock 2009; Gerber 1985; Ho 2008; Hobgood 2009; Kahan 2003; Klein 2000; Legg 2005; Levenkron 1990; Lim 2011; Lorin 2006; LoSasso 2017; Lupi 2012; Maguire 1977; Maguire 1978; Meirovich 2016; Morrow 2009; Mounsey 2006; Nomura 2017; Palmer 2018; Pirdehghan 2018; Poole 1980; Roche 1996; Roche 1997; Ruesseler 2017; Schwartz 2010; Seim 1995; Shaddheau 2015; Solomon 2004; Spollen 2010; Taylor 2019; Vanatta 1996; Walsh 1999; Weihs 1986; Windish 2005; Wundrich 2017	36 interventions in Allen 1990; Bosse 2012; Bowyer 2006; Bowyer 2010; Edwardsen 2006; Eells 2002; Evans 1989; Evans 1996; Hobgood 2009; Kahan 2003; Levenkron 1990; Lim 2011; Lorin 2006; Morrow 2009; Nomura 2017; Palmer 2018; Papadakis 1997; Roche 1996; Roche 1997; Seim 1995; Spollen 2010; Stolz 2012; Strayer 2010; Vanatta 1996; Walsh 1999; Weihs 1986; Windish 2005	n = 87 interventions in Allen 1990; Berney 2017; Betchart 1984; Blatt 2010; Bosse 2012; Bowyer 2006; Cave 2007; Chibhall 2005; Colletti 2001; Daeppen 2012; Edwardsen 2006; Eells 2002; Engerer 2019; Evans 1989; Evans 1996; Feddock 2009; Gartmeir 2015; Gerber 1985; Ho 2008; Ishikawa 2010; Kahan 2003; Klein 2000; Knowles 2001; Legg 2005; Levenkron 1990; Lim 2011; Liu 2016; Lorin	

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 1. Features of included studies *(Continued)*

			2006; LoSasso 2017; Lupi 2012; Maguire 1977; Maguire 1978; Mason 1988; Moreland 1973; Morrow 2009; Mounsey 2006; Nomura 2017; Ockene 2016; Ozcakar 2009; Papadakis 1997; Perera 2010; Pirdehghan 2018; Roche 1996; Roche 1997; Ruesseler 2017; Schwartz 2010; Seim 1995; Shapiro 2009; Siassakos 2010; Spollen 2010; Steifel 2013; Stolz 2012; Strayer 2010; Taylor 2019; Vanatta 1996; Walsh 1999; Weihs 1986; Windish 2005; Wundrich 2017		
Role-play with	**Simulated patients (n = 33 studies)**	**Peers (n = 16 studies)**	**Real patients (n = 12 studies)**	**Virtual patients/online (n = 3 studies)**	
	50 interventions in Berney 2017; Betchart 1984; Blatt 2010; Bosse 2012; Cave 2007; Chibhall 2005; Colletti 2001; Edwardsen 2006; Enger-er 2019; Evans 1996; Feddock 2009; Gartmeir 2015; Gerber 1985; Hobgood 2009; Ishikawa 2010; Knowles 2001; Levenkron 1990; Lorin 2006; Liu 2016; Maguire 1977; Mason 1988; Mounsey 2006; Ozcakar 2009; Papadakis 1997; Perera 2010; Pirdehghan 2018; Schwartz 2010; Siassakos 2010; Steifel 2013; Taylor 2019; Vanatta 1996; Windish 2005; Wundrich 2017	20 interventions in Allen 1990; Bosse 2012; Eells 2002; Hobgood 2009; Lim 2011; Lupi 2012; Mounsey 2006; Papadakis 1997; Roche 1996; Ruesseler 2017; Seim 1995; Spollen 2010; Stolz 2012; Strayer 2010; Taylor 2019; Walsh 1999	Interviews with real patients in 13 interventions in Gerber 1985; Klein 2000; Maguire 1978; Meirovich 2016; Moreland 1973; Schwartz 2010; Shapiro 2009; Stolz 2012; Weihs 1986 Presentations from real patients: Shaddheau 2015 Videos involving real patients: Legg	Five interventions in Bearman 2001; Foster 2016; Kaltman 2018	Not specified in eight studies: Bowyer 2006; Daeppen 2012; Ho 2008; Kahan 2003; Legg 2005; LoSasso 2017; Morrow 2009; Ockene 2016

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 1. Features of included studies *(Continued)*

	2005; Snow 2016

C-RCT: cluster-randomised controlled trial.
RCT: randomised controlled trial.

APPENDICES

Appendix 1. MEDLINE search strategy

1. students medical/

2. ((student* or graduate* or undergraduate* or postgraduate*) adj3 (medical or medicine)).ti,ab,kw.

3. education medical/

4. education medical undergraduate/

5. clinical clerkship/

6. education medical graduate/

7. (medic* adj3 (education or school* or course* or curricul*)).ti,ab,kw.

8. or/1-7

9. communication/

10. (communicat* adj3 (skill* or competen*)).tw.

11. ((patient or client or family) adj (cent?red or focus?ed or tailored)).tw.

12. patient centered care/

13. interpersonal relations/

14. interpersonal.tw.

15. cultural competency/

16. (cultur* adj3 (competenc* or understanding or knowledg* or sensitiv* or aware* or respons* or appropriate* or acceptab* or safe* or humility)).ti,ab,kw.

17. (intercultural* or inter-cultural* or transcultural* or trans-cultural* or cross-cultural* or crosscultural*).ti,ab,kw.

18. empathy/

19. (therapeutic alliance or empath* or bad news or listening skill*).tw.

20. physician patient relations/

21. ((physician or doctor or gp or general practitioner) adj1 (patient or client)).tw.

22. ((patient* or client*) adj3 (interact* or relations* or deal* with or rapport)).tw.

23. (relations* adj3 build*).ti,ab,kw.

24. exp "referral and consultation"/

25. office visits/

26. interviews as topic/

27. negotiating/

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

28. ((ask* adj3 question*) or questioning or explain* or discuss or discussing or closure).ti,ab,kw.

29. (information adj3 gather*).ti,ab,kw.

30. trust/

31. (verbal or nonverbal or non-verbal or smiling or negotiat* or trust or hope or friendl* or warmly or cultural* or spiritual* or comforting or supportive*).tw.

32. (consult* or interview* skill*).tw.

33. ((shar* or join* or concordan* or participat*) adj3 decision making).tw.

34. or/26-33

35. (patient* or client* or skill* or competen*).tw.

36. 34 and 35

37. ((medical or clinical) adj encounter*).tw.

38. exp medical history taking/

39. (history taking or anamnesis).tw.

40. or/9-25,36-39

41. 8 and 40

42. randomized controlled trial.pt.

43. controlled clinical trial.pt.

44. randomized.ab.

45. placebo.ab.

46. drug therapy.fs.

47. randomly.ab.

48. trial.ab.

49. groups.ab.

50. or/42-49

51. 41 and 50

Appendix 2. Embase search strategy

1. medical student/

2. ((student* or graduate* or undergraduate* or postgraduate*) adj3 (medical or medicine)).ti,ab,kw.

3. exp medical education/

4. (medic* adj3 (education or school* or course* or curricul*)).ti,ab,kw.

5. or/1-4

6. exp interpersonal communication/

7. (communicat* adj3 (skill* or competen*)).tw.

8. ((patient or client or family) adj (cent?red or focus?ed or tailored)).tw.

9. human relation/

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

10. interpersonal.tw.

11. cultural competence/

12. (cultur* adj3 (competenc* or understanding or knowledg* or sensitiv* or aware* or respons* or appropriate* or acceptab* or safe* or humility)).ti,ab,kw.

13. (intercultural* or inter-cultural* or transcultural* or trans-cultural* or cross-cultural* or crosscultural*).ti,ab,kw.

14. empathy/

15. (therapeutic alliance or empath* or bad news or listening skill*).tw.

16. doctor patient relation/

17. ((physician or doctor or gp or general practitioner) adj1 (patient or client)).tw.

18. ((patient* or client*) adj3 (interact* or relations* or deal* with or rapport)).tw.

19. (relations* adj3 build*).ti,ab,kw.

20. exp consultation/

21. office visit*.ti,ab,kw.

22. interview/

23. ((ask* adj3 question*) or questioning or explain* or discuss or discussing or closure).ti,ab,kw.

24. (information adj3 gather*).ti,ab,kw.

25. trust/

26. (verbal or nonverbal or non-verbal or smiling or negotiat* or trust or hope or friendl* or warmly or cultural* or spiritual* or comforting or supportive*).tw.

27. (consult* or interview* skill*).tw.

28. ((shar* or join* or concordan* or participat*) adj3 decision making).tw.

29. or/22-28

30. (patient* or client* or skill* or competen*).tw.

31. 29 and 30

32. ((medical or clinical) adj encounter*).tw.

33. anamnesis/

34. (history taking or anamnesis).tw.

35. or/6-21,31-34

36. 5 and 35

37. randomized controlled trial/

38. controlled clinical trial/

39. single blind procedure/ or double blind procedure/

40. crossover procedure/

41. random*.tw.

42. placebo*.tw.

43. ((singl* or doubl*) adj (blind* or mask*)).tw.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

44. (crossover or cross over or factorial* or latin square).tw.

45. (assign* or allocat* or volunteer*).tw.

46. or/37-45

47. 36 and 46

48. limit 47 to medline

49. 47 not 48

50. limit 49 to yr="2017 -Current"

Appendix 3. PsycINFO search strategy

1. medical students/

2. ((student* or graduate* or undergraduate* or postgraduate*) adj3 (medical or medicine)).ti,ab,id.

3. medical education/

4. (medic* adj3 (education or school* or course* or curricul*)).ti,ab,id.

5. or/1-4

6. communication/

7. exp interpersonal communication/

8. communication skills/

9. communication skills training/

10. cultural sensitivity/

11. empathy/

12. client centered therapy/

13. interpersonal interaction/

14. therapeutic processes/

15. (communicat* adj3 (skill* or competen*)).ti,ab,id.

16. ((patient or client or family) adj (cent?red or focus?ed or tailored)).ti,ab,id.

17. interpersonal.ti,ab,id.

18. (cultur* adj3 (competenc* or understanding or knowledg* or sensitiv* or aware* or respons* or appropriate* or acceptab* or safe* or humility)).ti,ab,hw,id.

19. (intercultural* or inter-cultural* or transcultural* or trans-cultural* or cross-cultural* or crosscultural*).ti,ab,hw,id.

20. (therapeutic alliance or empath* or bad news or listening skill*).ti,ab,id.

21. ((physician or doctor or gp or general practitioner) adj1 (patient or client)).ti,ab,id.

22. ((patient* or client*) adj3 (interact* or relations* or deal* with or rapport)).ti,ab,id.

23. (relations* adj3 build*).ti,ab,id.

24. professional consultation/

25. ((ask* adj3 question*) or questioning or explain* or discuss or discussing or closure).ti,ab,hw,id.

26. (information adj3 gather*).ti,ab,id.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

27. (verbal or nonverbal or non-verbal or smiling or negotiat* or trust or hope or friendl* or warmly or cultural* or spiritual* or comforting or supportive*).ti,ab,hw,id.

28. (consult* or interview* skill*).ti,ab,id.

29. ((shar* or join* or concordan* or participat*) adj3 decision making).ti,ab,id.

30. or/25-29

31. (patient* or client* or skill* or competen*).ti,ab,hw,id.

32. 30 and 31

33. ((medical or clinical) adj encounter*).ti,ab,id.

34. (history taking or anamnesis).ti,ab,id.

35. or/6-24,32-34

36. 5 and 35

37. random*.ti,ab,hw,id.

38. trial*.ti,ab,hw,id.

39. controlled stud*.ti,ab,hw,id.

40. placebo*.ti,ab,hw,id.

41. ((singl* or doubl* or trebl* or tripl*) and (blind* or mask*)).ti,ab,hw,id.

42. (cross over or crossover or factorial* or latin square).ti,ab,hw,id.

43. (assign* or allocat* or volunteer*).ti,ab,hw,id.

44. treatment effectiveness evaluation/

45. mental health program evaluation/

46. exp experimental design/

47. "2100".md.

48. or/37-47

49. 36 and 48

50. limit 49 to yr="2017 -Current"

Appendix 4. CINAHL search strategy

#	Query
S41	s40
S40	s29 and s39
S39	S30 or S31 or S32 or S33 or S34 or S35 or S36 or S37 or S38
S38	TI (singl* or doubl* or tripl* or trebl*) and TI (blind* or mask*)
S37	AB (singl* or doubl* or tripl* or trebl*) and AB (blind* or mask*)

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

(Continued)

S36	AB (random* or trial or placebo*) or TI (random* or trial or placebo*)
S35	MH Quantitative Studies
S34	MH Placebos
S33	MH Random Assignment
S32	MH Clinical Trials+
S31	PT Clinical Trial
S30	"randomi?ed controlled trial" or PT randomized controlled trial
S29	s3 and s28
S28	s4 or s5 or s6 or s7 or s8 or s9 or s10 or s11 or s12 or s13 or s14 or s15 or s25 or s26 or s27
S27	"history taking" or anamnesis
S26	(medical or clinical) N1 encounter*
S25	s23 and s24
S24	patient* or client* or skill* or competen*
S23	s16 or s17 or s18 or s19 or s20 or s21 or s22
S22	(shar* or join* or concordan* or participat*) N3 "decision making"
S21	consult* or "interview* skill*"
S20	verbal or nonverbal or "non-verbal" or smiling or negotiat* or trust or hope or friendl* or warmly or cultural* or spiritual* or comforting or supportive*
S19	information N3 gather*
S18	(ask* N3 question*) or questioning or explain* or discuss or discussing or closure
S17	MH negotiation
S16	MH interviews
S15	MH office visits
S14	MW consultation
S13	relations* N3 build*
S12	(patient* or client*) N3 (interact* or relations* or deal* with or rapport)
S11	(physician or doctor or gp or general practitioner) N1 (patient or client)
S10	"therapeutic alliance" or empath* or "bad news" or "listening skill*"

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

(Continued)

S9	intercultural* or "inter-cultural*" or transcultural* or "trans-cultural*" or "cross-cultural*" or cross-cultural*
S8	cultur* N3 (competenc* or understanding or knowledg* or sensitiv* or aware* or respons* or appropriate* or acceptab* or safe* or humility)
S7	(MH "Interpersonal Relations") or (MH "Patient Centered Care")
S6	(patient or client or family) N1 (cent#red or focus#ed or tailored)
S5	communicat* N3 (skill* or competen*)
S4	(MH communication) or (MH "Trust")
S3	s1 or s2 or (MH "Education, Medical") or (MH "Students, Medical")
S2	medic* N3 (education or school* or course* or curricul*)
S1	(student* or graduate* or undergraduate* or postgraduate*) AND (medical or medicine)

Appendix 5. ERIC search strategy

(ab((student* OR graduate* OR undergraduate* OR postgraduate*) NEAR/3 (medical OR medicine)) AND (ab(communicat* OR trust* OR (patient centred) OR (patient centered)) OR ab(cultur* OR intercultural* OR "inter-cultural*" OR transcultural* OR "trans-cultural*" OR "cross-cultural*" OR crosscultural* NEAR/3 (competenc* OR understanding OR knowledg* OR sensitiv* OR aware* OR respons* OR appropriate* OR acceptab* OR safe* OR humility)) OR ab((ask* NEAR/3 question*) OR questioning OR explain* OR discuss OR discussing OR closure))) AND noft(random*)

Appendix 6. CENTRAL search strategy

1. students medical/

2. ((student* or graduate* or undergraduate* or postgraduate*) adj3 (medical or medicine)).ti,ab,kw.

3. education medical/

4. education medical undergraduate/

5. clinical clerkship/

6. education medical graduate/

7. (medic* adj3 (education or school* or course* or curricul*)).ti,ab,kw.

8. or/1-7

9. communication/

10. (communicat* adj3 (skill* or competen*)).tw.

11. ((patient or client or family) adj (cent?red or focus?ed or tailored)).tw.

12. patient centered care/

13. interpersonal relations/

14. interpersonal.tw.

15. cultural competency/

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

16. (cultur* adj3 (competenc* or understanding or knowledg* or sensitiv* or aware* or respons* or appropriate* or acceptab* or safe* or humility)).ti,ab,kw.

17. (intercultural* or inter-cultural* or transcultural* or trans-cultural* or cross-cultural* or crosscultural*).ti,ab,kw.

18. empathy/

19. (therapeutic alliance or empath* or bad news or listening skill*).tw.

20. physician patient relations/

21. ((physician or doctor or gp or general practitioner) adj1 (patient or client)).tw.

22. ((patient* or client*) adj3 (interact* or relations* or deal* with or rapport)).tw.

23. (relations* adj3 build*).ti,ab,kw.

24. exp "referral and consultation"/

25. office visits/

26. interviews as topic/

27. negotiating/

28. ((ask* adj3 question*) or questioning or explain* or discuss or discussing or closure).ti,ab,kw.

29. (information adj3 gather*).ti,ab,kw.

30. trust/

31. (verbal or nonverbal or non-verbal or smiling or negotiat* or trust or hope or friendl* or warmly or cultural* or spiritual* or comforting or supportive*).tw.

32. (consult* or interview* skill*).tw.

33. ((shar* or join* or concordan* or participat*) adj3 decision making).tw.

34. or/26-33

35. (patient* or client* or skill* or competen*).tw.

36. 34 and 35

37. ((medical or clinical) adj encounter*).tw.

38. exp medical history taking/

39. (history taking or anamnesis).tw.

40. or/9-25,36-39

41. 8 and 40

42. randomized controlled trial.pt.

43. controlled clinical trial.pt.

44. randomized.ab.

45. placebo.ab.

46. drug therapy.fs.

47. randomly.ab.

48. trial.ab.

49. groups.ab.

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

50. or/42-49

51. 41 and 50

52. limit 51 to yr="2017 -Current"

WHAT'S NEW

Date	Event	Description
10 February 2021	Amended	New acknowledgements added

HISTORY

Protocol first published: Issue 11, 2016
Review first published: Issue 2, 2021

CONTRIBUTIONS OF AUTHORS

All review authors have contributed to the design and conduct of the review. Conor Gilligan has co-ordinated the review process by being involved in each step of the review: screening, data extraction, data entry, analysis, interpretation, and write-up. All review authors participated in initial screening of titles and abstracts, and Conor Gilligan, Erica James, Pamela Harvey, Marita Lynagh, Martine Powell, Bernadette Ward, and Chris Lonsdale screened full-text articles. Erica James, Pamela Harvey, Marita Lynagh, Smriti Nepal, Sari Dewi, and Hayley Croft extracted data from included studies. Chris Lonsdale, Bernadette Ward, Conor Gilligan, and Dominique Rich conducted 'Risk of bias' analysis. Marita Lynagh, Bernadette Ward, and Conor Giligan completed GRADE analysis. Conor Gilligan, Martine Powell, and Pamela Harvey wrote the Discussion section, and all review authors assisted with writing and editing of the final review. Jonathan Silverman provided general advice on the review and assisted with interpretation of results.

DECLARATIONS OF INTEREST

- Conor Gilligan: none known
- Martine Powell: none known
- Marita C Lynagh: none known
- Bernadette M Ward: none known
- Chris Lonsdale has received consultancy fees from commercial agencies on health promotion products. His research is supported by grants from government and philanthropic agencies
- Pam Harvey: none known
- Erica L James: none known
- Sari Dewi: none known
- Smriti Nepal: none known
- Dominique Rich: none known
- Hayley Croft: none known
- Jonathan Silverman has worked at the forefront of the development of communication skills, teaching in undergraduate medical education for the last 20 years, and recently retired from the School of Clinical Medicine, University of Cambridge, where he led the undergraduate communication programme. He receives royalties from two books that he wrote about healthcare communication, and he travels as president of the European Association for Communication in Healthcare.

SOURCES OF SUPPORT

Internal sources

- University of Newcastle, Other

  Small internal grants supported the employment of research assistants who helped with data extraction from included studies, as well as support for statistical consultation.

External sources

- No sources of support supplied

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Throughout the conduct of the review, several minor modifications were made to planned methodological processes.

In the protocol, we listed five main outcomes (relationship building, information gathering, planning and explaining, specific communication tasks, and simulated or real patient appraisals). Due to the volume of outcomes reported and heterogeneity across studies, we added further groupings of the outcomes to include empathy and overall communication skills, and for these outcomes, we separated outcomes judged by expert assessors from those judged by simulated patients.

The protocol also stated that when communication interventions were delivered as part of a larger complex intervention, we would include only studies that compared interpersonal communication as part of a complex intervention with the same complex intervention without interpersonal communication elements. In fact, in a small number of cases, the communication elements were part of a complex intervention such as those including physical examination. When the intervention was compared with an appropriate control or comparison and communication outcomes were reported separately, these studies were included.

The protocol stated that when multiple primary outcomes were identified within a category, we would rank the reported intervention effect estimates and select the median effect estimate. In fact, we selected the outcome most homogenous with others to justify pooling.

In the protocol, we stated that we would pool results only when heterogeneity was sufficiently low. Although statistical heterogeneity emerges as high in our meta-analyses, homogeneity across comparisons and outcomes was regarded as sufficient to warrant pooling.

The protocol stated that blinding of outcome assessment would be assessed for each outcome measure. As most outcomes reported were generated from a single assessment (e.g. separate components of an OSCE assessment checklist), blinding was assessed at the individual study level rather than at the outcome level.

The protocol stated that we would analyse dichotomous data based on the number of events and the number of people assessed in the intervention and comparison groups, and we would use these data to calculate the risk ratio (RR) and the 95% confidence interval (CI). Given that most outcomes were reported using continuous data, dichotomous outcomes reporting on the same measures as those expressed using continuous data were converted to continuous outcomes by calculating the odds ratio (OR) and the confidence interval (CI) for each study, converting these to standardised mean difference (SMD) and standard error (SE), and entering them into a comparison using generic inverse variance (GIV) in RevMan.

The protocol stated that we would use intention-to-treat analysis when possible, and we would impute missing data when required. In fact, most studies did not use intention-to-treat analysis, and many failed to clearly present attrition, so data were entered as analysed in the original studies.

We stated that for cluster trials, we would obtain estimates of the ICC by contacting authors of included studies, or we would impute them by using estimates from external sources. Only one of the cluster trials included was regarded as such by the study authors, with appropriate adjustments made (Ockene 2016). This trial differed from others, as it randomised students within 10 different universities to participate in the intervention or control group. Most other trials randomised students by group or allocated a year-based cohort in the case of cluster-quasi-RCTs, so the ICC from Ockene 2016 could not be applied. Instead, sensitivity analyses were conducted on all meta-analyses that included a cluster-RCT or a cluster-quasi-RCT, with the design effect calculated based on inflation of 5%, 10%, and 20% (design effect = 1 + (M-1)*ICC, where M is the mean cluster size and ICC is estimated as 0.05, 0.1, and 0.2; inflation is performed as standard error multiplied by the square root of the design effect).

We did not conduct subgroup analyses based on student year level, as the nature of outcome measurement meant that assessments would be calibrated for expectations relating to each year level and use of post-intervention comparison between groups rather than change from baseline renders the student year level irrelevant to demonstration of improvement in communication skills.

NOTES

This protocol is based on standard text and guidance provided by Cochrane Consumers and Communication (CCCG 2013).

INDEX TERMS

Medical Subject Headings (MeSH)

*Communication; Education, Medical [*methods]; *Empathy; Information Management [education]; *Interpersonal Relations; Medical History Taking; Non-Randomized Controlled Trials as Topic; Patient Satisfaction; Patient Simulation; Randomized Controlled Trials as Topic; Role Playing; *Students, Medical

MeSH check words

Humans

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.