Teaching clinically experienced physicians communication skills. A review of evaluation studies

R L Hulsman,[1] *W J G Ros,*[2] *J A M Winnubst*[2] *& J M Bensing*[3,4]

Context Interest in the teaching of communication skills in medical schools has increased since the early seventies but, despite this growing interest, relatively limited curricular time is spent on the teaching of communication skills. The limited attention to the teaching of these skills applies even more to the physicians' clinical years, when attention becomes highly focused on biomedical and technical competence. Continuing training after medical school is necessary to refresh knowledge and skills, to prohibit decline of performance and to establish further improvements.

Objective This review provides an overview of evaluation studies of communication skills training programmes for clinically experienced physicians who have finished their undergraduate medical education. The review focuses on the training objectives, the applied educational methods, the evaluation methodology and instruments, and training results.

Methods CD-ROM searches were performed on MedLine and Psychlit, with a focus on effect-studies dating from 1985.

Results Fifteen papers on 14 evaluation studies were located. There appears to be some consistency in the aims and methods of the training programmes. Course effect measurements include physician self-ratings, independent behavioural observations and patient outcomes. Most of the studies used inadequate research designs. Overall, positive training effects on the physicians' communication behaviour are found on half or less of the observed behaviours. Studies with the most adequate designs report the fewest positive training effects.

Conclusion Several reasons are discussed to explain the limited findings. Future research may benefit from research methods which focus on factors that inhibit and facilitate the physicians' implementation of skills into actual behaviours in daily practice.

Keywords Continuing medical education, *methods; *communication; hospital medical staff, *education; evaluation studies; meta-analysis.

Medical Education 1999;33:655–668

Introduction

Doctor–patient communication is an important tool in health care, some consider it even the most important tool.[1–3] It is the fundamental instrument by which physician and patient relate to each other and attempt to achieve therapeutic goals.[4] However, many patients experience problems in this aspect of medical care, as can be derived from data on patient complaints about health care.[5]

In medical schools the attention to training in communication skills has increased since the early seventies, when medical schools started to develop course programmes on communication skills.[6,7] Since the early nineties all medical schools have offered some kind of communication skills training. Yet, training in communication skills is often treated as a minor subject of little significance in medical schools. Generally, less than 5% of the curriculum time is spent on training in communication skills.[8–10] During the residential years, when the physicians specialize in a particular medical discipline, even less attention is given to training in communication skills.[11–13] Also, the vast majority of continuing medical education (CME) programmes in medicine focus on technologic and biomedical aspects of medical care, instead of communication skills.[14]

[1]Academic Medical Centre, Department of Medical Psychology [2]University Utrecht, Faculty of Medicine; [3]University Utrecht, Faculty of Social Sciences; [4]NIVEL, Netherlands Institute of Primary Health Care

Correspondence: R. L. Hulsman, Academic Medical Centre, Department of Medical Psychology J4, P.O. Box 22660, 1100 DD Amsterdam, The Netherlands

Continuous training in communication skills during their clinical years has several functions and benefits for physicians. First of all, adequate performance is the result of a dynamic process of acquiring new knowledge and skills and integrating these into the existing expertise.[15] This cyclical process involves repeated periodic learning efforts. Postgraduate education is an effective way to refresh the previously acquired skills and to develop new techniques. Repeated learning efforts prohibit the decay of good skills, and help to improve skill performance.

Furthermore, postgraduate education is necessary to adjust the level of medical care to the constantly changing needs of society. These changing needs require a continuous update and improvement of the standards of medical care. For example, with the rise of consumerism and patient empowerment, patients are becoming more and more autonomous and want to play an active role in medical decision-making.[12] Physicians have to be trained to adapt to these changing roles.

The aim of this review is to present an overview of efforts made in the past decade to improve the communication skills of physicians in their clinical years. Reviewed are studies in which the effects of these training programmes are evaluated. The review will focus on the kind of skills the participants are taught and the type of educational methods applied in teaching the skills. Thereafter, the different evaluation methodologies will be examined in terms of research design applied and the kind of instruments used to measure training effects. Finally, the results of the training effects will be reviewed.

Methods and procedure

Studies on communication skills training were located by using the following procedures. First, CD-ROM searches were performed on MedLine and Psychlit. The focus was on effect-studies dating from 1985 up to 1998. Key words used in the retrieval of research literature were: interpersonal communication, communication (skills), interview skills, counselling, physician–patient relations, (medical) education (graduate, postgraduate, continuing), internship and residency, training and evaluation (studies).

Second, reference lists of the articles found in the CD-ROM searches were inspected for undetected papers. Reviews and descriptive papers on training of communication skills which contained no evaluation were discarded. Also excluded were non-English articles, letters and one-page news items.

This review focuses on physicians with clinical experience, including graduates (residents) as well as postgraduates. Studies directed at undergraduates and studies directed at health professionals, excluding physicians, were not included in this review.

Results

In the search, 15 papers[14,16–29] on 14 studies* were found which met the selection criteria. Descriptions of the training and evaluation methods of the studies are summarized in Tables 1a and 1b. The training effects are summarized in Tables 2a and 2b.

1 Participants

Eight studies were directed at graduates (Table 1a) and six at postgraduates (Table 1b). In both graduate and postgraduate studies, the only medical disciplines represented are general practice and internal medicine. This is remarkable since adequate doctor–patient communication is important in all medical disciplines. In Faulkner's study[18] the participants were members of a multidisciplinary group not only of physicians but also of nurses, consultants and other health professionals. In the Gask study[19] the majority of the participating general practitioners were medical trainers themselves. In the later Gask study,[21] both graduate and postgraduate physicians participated.

The average number of subjects in the studies is 38, ranging from 6 to 80. The average size of intervention groups is 20, ranging from 6 to 40 participants. Information about response rates is given in only three graduate and two postgraduate studies.[17,22,24,26,29] Response rates of nearly 100% are reported for recruitment of graduates.[22,26] Since graduates can be approached in medical schools, their recruitment is relatively uncomplicated. Often the effect study is performed in the context of a required course programme on communication skills. Recruitment of postgraduates is less successful; reported initial responses were 16–20%.[17,24] Apparently, postgraduate physicians are reluctant to participate in communication skills training.

The drop-out rate is reported in six studies.[14,20,21,24,27–29] Average drop-out rate in these studies was 14%, ranging from 3 to 22%. Two studies explicitly mentioned the absence of drop-outs.[22,26] In the remaining six studies this aspect is not reported.[16–19,23,25]

2 Training objectives

All reviewed training programmes are directed at enhancing the communication skills of the participants.

* One study by Smith *et al.* is described in two separate papers.[27,28]

Table 1a Training and evaluation methods used in graduate studies

		Training			Evaluation						
						Observation					
Study	Participants	Objectives	Method	Duration	Design	Patients	N/phys.	Recording	Instrument	Physician self-rating	Patient outcomes
Sideris *et al.* 1986[25]	Int. Med. $n = 25 + 25$[a]	information behaviours	lecture modelling discussion	4×1 hour ? period	post-test control group pt: 1–3 month follow-up	real	40 pre, 39 post	audio	specific: 9 cat.*	communication behaviour	assessment of communication behaviour follow-up: compliance
Gask *et al.* 1988[20]	Gen. Pract. $n = 14$	receptive behaviours information behaviours affective behaviours	manual modelling discussion video-feedback	18×2 hours 6 month period	pre-post-test	real	5 pre, 5 post	video	specific: 14 cat.	detection of psychiatric disturbance	GHQ[d]
Gask *et al.* 1989[21]	Gen.Pract. $n = 20$	receptive behaviours information behaviours affective behaviours	lecture (video) modelling role-play direct-feedback video-feedback	$6 \times$? hours 2 month period	pre-post-test	simulated	1 pre, 1 post	video	specific: 15 cat.	–	satisfaction (simulated patients)
Putnam *et al.* 1988[22]	Int.Med. 2nd year $n = 11 + 8$[a]	receptive behaviours information behaviours	manual lecture modelling role-play audio-feedback	1.5 hrs group 2.5 hrs indiv. ? period	pre-post-test control group	real	6 pre, 8 post	audio	standard: 5 cat. 2 pt. cat.	–	satisfaction compliance health status
Roter *et al.* 1990[23]	Int.Med. 1st year $n = 3 \times 16$[b]	receptive behaviours information behaviours (patient management)	manual role-play discussion direct-feedback written-feedback video-feedback	? hours 1 month period	post-test control group	simulated	1 post	video	standard: 22 cat.	patient management (qualitative)	satisfaction (simulated patients)

Table 1a (*contd.*)

		Training			Evaluation						
						Observation					
Study	Participants	Objectives	Method	Duration	Design	Patients	N/phys.	Recording	Instrument	Physician self-rating	Patient outcomes
Smith *et al.* 1991[26]	Int.Med. 1st, 2nd, 3rd year n = 28 + 20[a]	receptive behaviours affective behaviours (patient management)	manual lecture modelling role-play audio-feedback	? hours 1 month period	pre-post-test control group –15 month follow-up (n = 13)	–	–	–	–	psychosocial: – knowledge – attitudes – skills self-rating	–
Smith *et al.* 1995a,[27] 1995b[28]	Int.Med. Fam.Pract. 1st year n = 15 + 11[a]	receptive behaviours information behaviours affective behaviours (patient management)	lecture modelling role-play	? hours 1 month period	pre-post-test randomized control group	simulated real	–	video audio	–	(1995a): psychosocial: – knowledge – attitudes	(1995b): – satisfaction
Smith *et al.* 1998[29]	Int.Med. Fam.Pract. 1st year n = 31 + 32[a]	receptive behaviours information behaviours affective behaviours (patient management)	manual lecture modelling role- play direct-feedback audio-feedback	96 hours 1 month period	pre-post-test randomized control group	simulated real	1-4 pre, 1-4 post (349 sim. 238 real)	video audio	specific: 15 cat.	psychosocial: – knowledge – attitudes	satisfaction GHQ[d] health status

*number of observation categories.

Table 1b Training and evaluation methods used in post graduate studies

		Training			Evaluation						
						Observation					
Study	Participants	Objectives	Method	Duration	Design	Patient	Numbers	Recording	Instrument	Physician self-rating	Patient outcomes
Bensing *et al.* 1985[16]	Gen. Pract. $n = 6$	receptive behaviours affective behaviours	?	?	pre-post-test	real	± 17 pre, ± 13 post	video	standard: 32 cat	–	–
Gask *et al.* 1987[19]	Gen. Pract. $n = 10$	receptive behaviours affective behaviours (patient management)	manual modelling discussion video-feedback	18 × 2 hours 6 month period	pre-post-test	simulated	1 pre, 1 post	video	specific: 7 cat	recognition of PSP[d]	–
Evans *et al.* 1987[17]	Gen. Pract. $n = 40$	information behaviours	manual lecture discussion	2 × 3 hours ? period	pre-post-test	–	–	–	–	–	10 pre, 10 post patient-questionnaires: – anxiety – satisfaction
Levinson *et al.* 1993[14]	Gen. Pract. Int. Med. $n = 15 + 15$[a]	receptive behaviours information behaviours affective behaviours	lecture discussion	4.5 hours ? period	pre-post-test randomized control group pre-post-test	real	5 pre, 5 post	audio	standard: 34 + 9 cat 2 pt	–	–
	Gen. Pract. Int. Med. $n = 20$		role-play direct feedback video-feedback	2.5 days ? period					cat	–	–
Faulkner *et al.* 1995[18]	Health prof. $n = 4 \times 20$[b]	information behaviours	lecture discussion role-play	± 9 hours 2 day period	test during training	simulated	1 post	video	specific: 6 cat	–	–
Roter *et al.* 1995[24]	Gen. Pract. Int. Med $n = 23 + 24$[a,c]	1: *problem-defining* receptive behaviours affective behaviours	manual lecture discussion	2 × 4 hours 2 week period	post-test randomized control group pts: 2 wks, 3 & 6 month follow-up	real simulated	± 9 1	audio	specific: 8 cat	recognition of PSP[d] patient management	GHQ[d]
	Gen. Pract. Int. Med. $n = 22 + 24$[a,c]	2: *emotion-handling* receptive behaviours affective behaviours	role-play direct feedback video-feedback								

[a] First n is the size of the intervention group, second n is the size of the control-group.
[b] Separate workshops are described together.
[c] The control group is the same for both interventions.
[d] PSP = psychosocial problems, GHQ = General Health Questionnaire.
[e] Effect is measured qualitatively.
[f] First digit is the number of significant results, second digit is the total number of observation categories.

Table 2a Results of training in graduate studies

Study	Observations[f]					
	Receptive behaviours	Information behaviours	Affective behaviours	Other effects	Physician ratings	Patient-outcomes
Sideris *et al.* 1986[25]	–	0/9 + total score	–	–	+ sumscore – individual items	+ compliance
Gask *et al.* 1988[20]	1/5	2/3	3/3	0/2 pat. cat.	+ detection psych. disturbance	(not analysed)
Gask *et al.* 1989[21]	3/4	4/5	1/5	–	–	+ satisfaction (1/5 items)
Putnam *et al.* 1988[22]	1/4	1/1	–	more patient exposition	–	– satisfaction – compliance – health status
Roter *et al.* 1990[23]	2/4	0/2	–	–	–	+ satisfaction
Smith *et al.* 1991[26]	–	–	–	–	+ knowledge + skills-rating + attitude follow-up: + attitudes – skills-rating	–
Smith *et al.* 1995a[27], 1995b[28]	not reported	not reported	not reported	–	(1995a): + 7/15 scales + self-confidence	(1995b): + satisfaction (2/5 scales)
Smith *et al.* 1998[29]	real pt: 1/4 sim pt: 6/8	real pt: 1/2 sim pt: 3/6	real pt: 1/4 sim pt: 7/9	+ overall rating	+ knowledge + attitudes (9/15)	– satisfaction – GHQ – health status

Table 2b Results of training in postgraduate studies

Study	Observations					
	Receptive behaviours	Information behaviours	Affective behaviours	Other effects	Physician ratings	Patient-outcomes
Bensing *et al.* 1985[16]	4/14	3/7	3/5	– discussion of PSP[e]	–	–
Gask *et al.* 1987[19]	3/5	?/3 not reported	–	–	smn; recognition of PSP[d,e]	–
Evans *et al.* 1987[12]	–	–	–	–	–	+ general satisfaction (1/17 items) + higher trait anxiety + lower state anxiety
Levinson *et al.* 1993[14]	no effect 2/?	no effect 0/?	no effect 0/?	–	–	–
				+ less pt distress + less affect	–	–
Faulkner *et al.* 1995[18]	–	–	–	inappropriate design	–	–
Roter *et al.* 1995[24]	3/4	–	1/3	–	+ recogn. of PSP[d]	
	2/4	–	2/3	–	+ lower GHQ-score	

+ effect is statistically significant
– effect is not significant
smn; significant effects on some variables
– non-applicable

However, each training programme has more specific objectives, which are not the same in all studies. It is remarkable that the training objectives are not always clearly described in the studies.[14,19,20,25,27,28] Sometimes the objectives had to be derived from the evaluation methods and result sections.

In order to provide a concise overview, the various training objectives were classified into three basic categories: *receptive behaviours*, *information behaviours*, and *interpersonal and affective behaviours*. A cumulative list of the meanings of and corresponding behaviours of achieving these basic objectives is found in Table 3. A certain overlap between the categories cannot be excluded. Some behaviours can have different functions and may well fit in more than one of the three categories. However, efforts were made to allocate each behaviour to the most appropriate category.

Receptive behaviours concern general facilitative behaviours that create conditions for the patient to participate actively in the interaction. It stresses non-directive and active listening, and stimulating patients to tell their story and disclose their concerns.

Information behaviours are directed at improving the effectiveness and efficiency of the information exchange. This is carried out for different reasons such as to reduce anxiety and improve satisfaction and compliance of patients. A well-informed patient can take a more equal position in the interaction and participate meaningfully in medical decision-making.

Table 3 Methods of achieving the basic training objectives

1. Receptive behaviours
   - open-ended questions
   - asking patient's opinion
   - verbal encouragement
   - active listening
   - picking up verbal and non-verbal cues
   - respectful silence
   - no interruptions
2. Information behaviours
   - providing information
   - patient education
   - counselling
   - distressing information, bad news
   - non-technical terms
3. Interpersonal and affective behaviours
   - relationship building
   - express mutuality
   - elicit patient concerns, psychosocial problems and emotions
   - reflections
   - personal warmth and confidence
   - empathy
   - expressing concern

The third aspect, *interpersonal and affective behaviours*, concerns the establishment of a good interpersonal relationship and creating an open and facilitative atmosphere, in which the patient feels free to speak. These behaviours are closely related to *receptive behaviours*, but focus more particularly on the affective domain of the interaction.

The objective of *receptive behaviours* is found in 11 studies, *information behaviours* is found in 10 studies, and improving *interpersonal and affective behaviours* is found in nine studies (Table 1). In five studies all three areas of communication are covered.[14,20,21,27–29] Six studies focus on two objectives which always include *receptive behaviours*. The remaining three studies focus on the single aspect of *providing information*.[17,18,25] The Faulkner study is specifically dedicated to the disclosure of bad news.[18]

3 Educational methods

In the reviewed training programmes, five kinds of educational methods are used: instruction, modelling, skill practice, feedback and discussions on communication skills. Generally the training programmes provide a balance between cognitive learning and experiential learning.

Instruction is part of the training in all but one* of the studies reviewed; nine studies provided *lectures* in a group with an instructor; eight studies provided *manuals* with written information about doctor–patient communication (Table 1). Gask *et al.*[21] used an self-instructional videotape. Five studies provided instruction both by manuals and lectures.[17,22,24,26,29]

Modelling gives the participants clear examples, positive as well as negative, of the use and performance of the target skills.[30] The use of *modelling* is explicitly mentioned in eight studies (Table 1). In one of these studies life models are applied.[29] In some studies models are presented on audiotape[22,25] and in others on videotape.[19–21] In two studies the kind of modelling applied is not described explicitly.[26–28] Modelled examples are mostly derived from daily practice, sometimes that of the participants themselves.[25,27,28]

In nine studies, the learners could *practise* skills in role-plays, while interacting with the other participants[21,22] or simulated patients.[18,23,24,29] In three studies the kind of applied role-play is unclear.[14,26–28] The interaction may be observed by an instructor and an audience of colleagues, and may be recorded on audio- or videotape.[14,18,24] Role-play offers the

* The use of instruction is not explicitly mentioned in the long course of Levinson *et al.*[21] This course has a learner-centred curriculum.

participants a safe situation to practise skills in and in which they are allowed to make mistakes. This helps participants dealing with problems which, if handled inexpertly, could be very distressing or damaging for real patients, such as breaking bad news.[31]

Feedback on the behaviour of the participants is provided in nine studies, based on role-play,[14] on interaction with real patients[19,20,22,26] or both.[21,23,24,29] Feedback may be provided in group sessions[14,19–21,24] or by the trainer in individual sessions.[22,26] Direct-feedback is provided immediately following the interaction with a real or simulated patient. Video-feedback is based on reviews of the recordings of real or simulated interactions. Video-feedback on real interactions required physicians to record some patients during their surgery hours and bring those recordings to the training sessions.[19–21,24,26,29] In the Faulkner study,[18] feedback on the video recording of the roleplay was not provided during, but some time after the training. After analysis of the tapes, all participants received a written feedback sheet together with their videotape.

A final, often-applied educational method is *discussion*, either individually with the trainer or in the group of participants. Discussion is a valuable technique by which to exchange theoretical knowledge of the course with the participants' practical experiences. Discussion is a distinct part of nearly all programmes – mostly group discussion (Table 1). In three studies it is not an explicitly mentioned element of the training.[22,26–28] However, discussion can be an implicit part of any of the other educational methods.

The training curriculum may be *prefixed* (*teacher-centred*), meaning that the teacher is responsible for setting objectives, or *learner-centred* implying that the training objectives are largely determined by the learners and the content of the training is matched to their needs.[32] This approach was found in three training programmes.[14,26,29] These programmes are generally conducted in individual or small group settings, allowing much space for individual learning needs and experiences.

4 Duration and location of training

All studies but one[16] provide some information about the duration of the training; four only provide the number of hours,[14,17,22,25] another four only provide the time period for which training sessions are scheduled[21,23,26,27,28] and the remaining five studies provide both types of information. Insofar as it was mentioned, the training programmes took from 4 to 96 h and the training periods ranged from two days to six months. These totals have to be treated with caution, since they are never stated directly, but constructed from different sources of information. Information about the locations of the training programme was rarely provided. The impression is that graduates are generally trained in their departments, whereas postgraduates are invited to come over to the training institutes.

5 Instruments

Training effects are measured on different levels. On the first level the physicians' *subjective evaluations* about training effects are measured. These evaluations generally focus on the physicians' knowledge, skills and attitudes. Training effects on this level are important but not sufficient determinants of actual behavioural changes.[33]

The next level focuses on independent *behavioural observations* of doctor–patient interactions. Behavioural observations can be regarded as the most important indicator of training effects, since the interventions' aim to improve communication behaviours is tested most directly.

The final level involves measurement of *outcome effects* of the improved interaction with the patient. These three levels will be discussed in the following sections.

Physician self-ratings

Self-rating questionnaires are used in eight studies (Table 1). In two studies, physicians were asked to assess their own communicative behaviour.[25,26] In six studies the self-ratings related to the recognition of psychosocial problems of patients. Four of these studies focused on measurement of the physicians' knowledge of and attitudes towards psychosocial medicine in general.[19,26,27,29] The other two studies focused on the physicians' detection of psychosocial problems of real patients.[20,24]

A methodological weakness of the physician self-ratings is their sensitivity for response bias. Self-ratings are generally reactive measures; the measurement itself may influence the outcome, since the physicians are not blind to their training condition. Post-training improvements on self-ratings may not only be the result of a training effect, but may also reflect the willingness of the respondents to show that the training efforts have not been useless, for reasons described in cognitive dissonance theory, regardless of whether they experienced actual improvements.[34]

Behavioural observations

Behavioural observations applied in 12 studies (Table 1). The behavioural observations may relate to

interactions with real patients or simulated (standardized) patients. Both methods have their specific advantages. Interactions with real patients provide the most realistic image of the physicians' general performance, whereas the simulation method is more suitable to measure competence.[35]

Observation of real encounters is carried out in five studies.[14,16,20,22,25] Since the doctor–patient interaction is not only controlled by the physician but also by the patient, uncontrolled patient factors may obscure the measurement of a training effect. One solution to control for this variability, is to record for each physician a number of patient encounters per measurement time. Generally, this number ranges from five to nine (Table 1).

Another way to keep patient factors under control is to use simulated patients. With simulated patients the type and difficulty of the problems the physician has to face and the way they are presented are standardized. A disadvantage is that they give a more artificial interaction. Four studies used simulated patients.[18,19,21,23] Roter *et al.*[24] used simulated patients, not to measure communication skills, but to measure clinical proficiency.

In three studies both real and simulated interactions were analysed.[24,27–29]

Doctor–patient interactions are generally evaluated by recording them on audio- or videotape and by rating the interactions afterwards on various observation categories. Video recordings provide a more complete picture of the interaction than audio recordings, which lack the registration of non-verbal behaviours. Nevertheless, audio recordings can be applied more easily and unobtrusively. Audiotapes were used in four studies and video recordings in eight (Table 1). Two studies used both audio and video recordings, each in different contexts of real and simulated patients.[27,29]

To code the interview-behaviours, standard as well as study-specific observation instruments can be applied. Standard instruments are applicable in a wide range of situations. They aim at a comprehensive classification of all speech acts that occur in an interaction.[36] Well known instruments are: Bales Process Analysis System,[37] Stiles Verbal Response Modes[38] and Roter's Interaction Analysis System.[39] Standard rating systems are used in four studies.[14,16,22,23] Here, the number of observation categories ranges from 7 to 43.

Study-specific rating scales are often more compact, focusing solely on the target communication skills of the evaluated training.[36] This type is used in eight studies (Table 1). The number of observation categories ranges from 6 to 15.

Further, observation systems can differ in the applied rating methodology. In most systems the *frequencies* of target behaviours are rated. This method is applied in 11 studies. Other systems have a rating scale format which provides *judgements* of the performance of target behaviours. These global judgements were used in the Smith *et al.* study[29] where 11-point rating scales were applied to assess key interviewing behaviours. Each scale refers to a mix of key behaviours which have a common intention in the interaction.

Patient-outcomes

Measurements of patient-outcomes of improved communication behaviours are applied in nine studies (Table 1), seven concerning real patients[17,20,22,24,25,27–29] and two concerning simulated patients.[21,23]

Patient-outcomes can be *proximal measures* which are related more directly to the physician behaviour in the observed consultation, or *distal measures* concerning the more general functioning of the patient. The proximal measures concern patient-ratings of the physicians' communication skills[25] or satisfaction with the consultation.[17,21,22,23,27–29] The distal outcome measures concern compliance,[22,25] anxiety[17] and general health.[22,24,29]

6 Research designs

The study-design is fundamental to the generalizability of the results. In the reviewed studies, three different research designs are found.

Only three studies used an adequate design, which was the *pre-test–post-test randomized control group design*, which allows for conclusions about training effects on communication behaviours.[14,27–29] In two studies, assignment to the groups was not randomized.[22,26]

In five studies a *pre-test–post-test design* was applied.[16,17,19–21] A limitation of this design is that differences in pre- and post-intervention levels of performance cannot be ascribed exclusively to the intervention. Uncontrolled, non-specific factors may increase or attenuate the measured effect (e.g. the Hawthorne effect).

Three studies used a *post-test control group design*[23–25] and in one of these the control group assignment was randomized.[24] This design provides no proof that possible post-test differences in behaviour between the intervention and the control group are a result of the intervention, since the assumption that both groups have equal baseline levels of performance is not tested.

Random assignment of the participants to the conditions increases the likelihood that the pre-intervention levels are the same in both groups. In one study, a pseudo *post-test design* was used of which the measurements were partly integrated into the training.[18]

Most studies do not provide clear information about measurement intervals. Presumably, post-tests are generally performed directly after the training.[19–21,26–28] Three studies used intervals of one-month[14,29] or two-months.[23]

Follow-up measurements of behavioural observations are never used. Follow-up measurements of patient outcomes are found in two studies, respectively, involving compliance[25] and general health.[24] Sideris *et al.*[25] measured compliance one to three months after the consultation. Roter *et al.*[24] performed follow-up measurements of general health at two weeks and three and six months after the consultation. Follow-up measurements of physician self-ratings are used in one study, concerning a 15-month follow-up measurement of psychosocial attitudes and skills.[26]

7 Training effects

Results on behavioural observations

In two of the 12 studies which include behavioural observations, no conclusion can be drawn about training effects on communication behaviour. In the 1995 Smith *et al.* study[27,28] the results of the observations are not reported in the papers reviewed. These results may be reported in the 1998 Smith paper,[29] but this is not explicitly mentioned. In the Faulkner study,[18] conclusions about the effectiveness of the training are not allowed because of the study design, which neither includes pre-measurements nor a control-group. The results of the remaining 10 studies will therefore be reviewed.

In each study, the observation categories have been divided in three clusters, corresponding with the distinction in training objectives: *receptive behaviours*, *information behaviours* and *interpersonal and affective behaviours*. The results are presented in Tables 2a and 2b. Displayed in these tables are the total number of observation categories in each cluster and the number of significant categories for each study.

Of the 10 studies with behavioural observations, nearly all report some training effects. Only the short (4·5 h) intervention in the Levinson study[14] did not result in detectable changes in communication behaviour. Generally, effects are found on half, or less, of the rated number of observation categories. In the six studies with study-specific observation instruments, the number of significantly improved categories range from 0 to 10 of the 7–15 rated categories.[19–21,24,25,29] In the four studies in which standard instruments are applied, effects are found on only 2–12 of the 7 to 43 observation categories.[14,16,22,23] The highest average proportion of significantly improved to measured observation categories is found on *interpersonal and affective behaviour* (56%), the lowest proportion is found among the *information behaviours* (40%). The average course effects on *receptive behaviours* are in between (46%).*

In the study which used both simulated and real patients to evaluate the course effects on the physicians' behaviour, relatively more effects were found in the simulated interactions.[29]

When taking the type of research design into account, it is striking that the studies with the best designs report the fewest training effects. In two studies, training effects on behavioural observations are reported based on a pre-test–post-test (randomized) control group design. One of these, the randomized study, reports the absence of any training effect.[14] In the other (non-randomized) study, effects were found on only two out of seven categories.[22] In the study of Roter *et al.*,[24] with a post-test randomized control group design, four out of seven communication behaviours were found to be significantly higher in the intervention groups. Although this is a more positive result, it cannot with certainty be regarded as a training effect, because of the lack of a pre-test.

Results on physician self-ratings

The physician self-ratings reveal quite positive results (Table 2). In the two studies where the physicians had to assess their own communicative behaviour, they perceived their performance as having improved after the training.[25,26]

All six studies which measured the physicians' recognition of psychosocial problems of patients, reported improved recognition.[19,20,24,26,27,29] Follow-up measures of physician self-ratings are applied by Smith *et al.*[26] They found that after 15 months the increased attitudes towards psychosocial medicine had remained consistent, while the physician's self-assessment of psycho-social skills had declined.

Results on patient-outcomes

As Table 2 shows, in four of the six studies in which results are reported on the *satisfaction* of the patients

* The study of Levinson[21] is not included in these percentages because in this study the total number of observation categories per cluster is not provided.

with the physician-interaction, some training effects were found.[17,21,23,28] Evans *et al.*[17] found a significant result on only one of the 17 items. Gask *et al.*[21] and Roter *et al.*[23] used satisfaction ratings of simulated patients. It is doubtful that these results represent the opinions of real patients, since both groups may use different criteria. Smith *et al.*[28] found effects on two of the five satisfaction scales, concerning 'confidence in the physician' and 'general satisfaction'. Putnam *et al.*[22] and Smith *et al.*[29] found no training effects on satisfaction, which the authors mainly ascribed to a lack of sensitivity of their instruments. Patients in these studies were generally highly satisfied, so the satisfaction scores showed little variation. In a seventh study, data on patient satisfaction were not reported because of lack of sensitivity of the measurements due to high baseline satisfaction levels of the patients.[26]

In two studies *compliance* was measured.[22,25] Putnam *et al.*[22] found no effect which might be a result of a limited training effect on the communication behaviour. Sideris *et al.*[25] measured compliance one to three months after the consultation by comparing the health behaviours of the patient (use of medicines, diet, lifestyle) to the instructions given by the physician. It is unclear whether these patients were seeing their physicians for the first time. Since the patients in the study were all patients with a chronic disease, who can be expected to consult their physician regularly, it is questionable whether their health behaviour can be related solely to one single observed consultation.

No course effects were found on *health status*, measured in two studies.[22,29]

Three studies which measured course effects on *psychosocial health* revealed ambiguous results. Roter *et al.*[24] found lower GHQ levels of emotional distress in the intervention groups compared to the control group. However, no pre-course GHQ measurements were performed in this study. Smith *et al.*[29] found no course effects on the patients' GHQ scores. In the third study, the patients' post-test state-anxiety was significantly lower, whereas their trait-anxiety was significantly higher compared to the pre-test.[17]

Discussion

In this review an overview is presented of evaluation studies of eight graduate and six postgraduate training programmes on communication skills, published since 1985. The studies were compared with regard to target group, training characteristics, evaluation methods and reported results. The results of the studies reviewed indicate that clinically experienced physicians can be trained in communication skills. Training effects are reported on ratings of communication behaviour, on physicians' self-ratings and on patient-outcomes. When comparing the results of these three types of measures, most training effects are found on the physicians' subjective evaluation of their knowledge, attitudes and skills. Results with regard to the patient-outcomes mainly point to enhancement of satisfaction. Mixed results were found with regard to compliance and psychosocial health, and no improvements could be demonstrated on health status.

Since the primary goal of the studies reviewed is to improve doctor–patient communication, the discussion will focus on the behavioural observations. The behavioural observations show that overall training effects are generally found on less than 50% of the observed behaviours. This may be regarded as a limited result, especially in studies where study-specific observation systems are applied since, in these studies, effects are reasonably expected on all observation categories. This result may illustrate the difficulty in formulating complete protocols of adequate communication behaviour. The same phenomenon is also observed in the domain of clinical proficiency where 'obligatory' behaviours are frequently not performed, without harming the quality of care.[3,40] Likewise, good communication has certain degrees of freedom in using different behaviours to establish the same quality. Moreover, not every behaviour is relevant in every encounter; this may reduce the chance of finding course effects on all observed behaviours.

Another explanation of limited course effects on communication behaviour may be related to problems physicians experience translating the learned skills into daily practice. According to the model of Francke *et al.*,[33] a course affects primarily the knowledge, attitudes and skills of the participants. Whether these primary effects result in actual behavioural changes depends on their intention to change. Indeed Moorhead *et al.*[11] showed that even when the attitudes and knowledge of the physicians about patient-centred communication have been improved, this hardly results in a more patient-centred style of consultation behaviour. A limited intention to change among physicians is recognized by some studies in this review.[14,22,24] According to Putnam *et al.*,[22] physicians are reluctant to discuss the psychosocial problems of patients because they feel that they should do something to solve them. Since the solution to such problems is often beyond the power of the physician, this urge is highly frustrating. Feeling responsible to offer a solution but at the same time fearing not to be able to provide a solution, may result in dissatisfaction with consultations where problems cannot be resolved rapidly.[11] Consequently,

physicians may suppress communication skills which invite patients to discuss their psychosocial concerns.

The physicians' intention to change is not measured in any of the studies reviewed although, according to Francke *et al.*,[33] this is an important determinant in the expression of the acquired knowledge, skills and attitudes into actual performance. Another remarkable observation is that in all studies it is assumed straight away that the initial level of performance of the participants is poor by definition and therefore can and should be improved. However, the probability of relatively high initial levels of performance should not be discarded. Since most studies use volunteer participants, the likelihood of self-selection bias in recruitment may be substantial.[24] This self-selection may result in an over-representation of participants who feel relatively safe in communicating with patients. They do not fear the confrontation with their own behaviour, nor the criticism of others during the training. People who have no specific interest in the field of doctor–patient communication may interact relatively less adequately with their patients, but at the same time may feel less urged to participate in a training. High initial performance levels reduce the chance of finding training effects. The finding in this review, that training of *interpersonal and affective behaviours* is more successful than training of *information behaviours*, may be explained by this phenomenon. At baseline, physicians are possibly more skilled in the information domain than in the affective domain.

Other factors may limit the chance of finding training effects on communication behaviour. First of all, this review is concerned with training clinically experienced physicians, who have already developed certain persistent routines in their communication behaviour. For them it may be more difficult to acquire new skills or to modify existing behaviours. A related factor is that the physicians know that they are being studied, which could have enhanced their performance (Hawthorne effect). This may reduce the chance of finding differences between pre- and post-course measurements and between experimental and control groups.

Time is also an important factor in measuring behavioural changes. It may take some time before the newly acquired or modified knowledge, skills and attitudes become integrated into the daily routine behaviour of the physicians. This is illustrated by a study of Bowman *et al.*[41] who performed follow-up measurements among postgraduates who had previously participated in Gask's graduate study.[20] Bowman *et al.*[41] found a continued improvement of the level of communication behaviour of the participants, showing progressive integration during their daily clinical practice.

This finding has implications for the intervals at which training effects are measured. Most of the studies reviewed included only one post-course measurement, which is generally performed directly after the course. Only two studies applied delayed post-course measurements.[14,23] Nevertheless, these studies did not reveal noticeably more behavioural improvements than the other studies. It is unknown how much time the integration process of the newly acquired behaviours may take. Incorporation of follow-up measurements in the study design may provide more insight in this matter.

A final comment concerns the generalizability of the results. A remarkable finding is that, in studies with the most adequate research designs, the fewest results are reported concerning improvements of communication behaviours. When focusing on the two methodologically sound studies, the effectiveness of communication skills training does not look very promising.[14,22] This may be explained by the limitations of randomization itself, which may reduce the chance of finding significant results.[42] For example, randomized trials cannot take into account preferences of the participants towards the intervention. Black[42] concludes that randomized trials provide an indication of the minimum effect of an intervention, whereas non-randomized studies offer an estimate of the maximum effect.

In conclusion, this review shows that the training effects on the physicians' communication behaviour are generally rather limited. The most complete picture of the effectiveness of training programmes on communication skills can be obtained when effects are measured with adequate research designs which include follow-up measurements. Measurements should be performed on different levels, including the participants knowledge and attitudes, their actual performance in daily practice and patient outcomes. Future research should pay more attention to factors, like the participants intention to change, which facilitate or inhibit the translations of learned skills into routine consultation behaviour.

Acknowledgements

This research was granted by the Dutch Cancer Society, the Nijbakker Morra Foundation and the Nijenburgh Foundation.

References

1 Cassell EJ. *Talking with patients: the theory of doctor–patient communication.* Cambridge (Mass.): The MIT Press, 1985.

2 Cassell EJ. *The Nature of Suffering and the Goals of Medicine.* New York: Oxford University Press, 1991.
3 Bensing JM. *Doctor–patient communication and the quality of care. An observation study into affective and instrumental behavior in general practice.* (dissertation Erasmus) University of Rotterdam, Utrecht: NIVEL, 1991.
4 Dimatteo M. The physician–patient relationship: effects on the quality of health care. *Clin Obs Gynae* 1994;**37**:149–61.
5 Hargie O, Dickson D, Boohan M, Hughes K. A survey of communication skills training in UK Schools of Medicine: present practices and prospective proposals. *Med Educ* 1998;**32**:25–34.
6 Whitehouse CR. The teaching of communication skills in United Kingdom medical schools. *Med Educ* 1991;**25**: 311–8.
7 Wolfhagen HAP. *Kwaliteit Van Klinisch Onderwijs [Quality of Clinical Education]*. (dissertation) Limburg University, Maastricht: Thesis 1993.
8 Bruggemans E. *De communicatie tussen arts en patiënt. Een inventarisatie-onderzoek in opdracht van de Nederlandse Hartstichting* [Doctor-patient communication. A review]. Den Haag: De Nederlandse Hartstichting, 1991.
9 Frederikson L, Bull P. An appraisal of the current status of communication skills training in British medical schools. *Social Sci Medicine* 1992;**34**:515–22.
10 Batelaan MA. *Inventarisatie onderwijs communicatievaardigheden faculteiten genees-kunde.* [Survey of communication skills training in the Dutch medical schools]. Amsterdam: Free University, Department of Medical Psychology, 1995.
11 Moorhead R, Winefield H. Teaching Counselling Skills to Fourth-Year Medical Students: a Dilemma Concerning Goals. *Family Prac* 1991;**8**:343–6.
12 Roter DL, Hall JA. *Doctors Talking with Patients, Patients Talking with Doctors*. Westport: Auburn House, 1992.
13 Orbell S, Abraham C. Behavioral sciences and the real world: Report of a community interview scheme for medical students. *Med Educ* 1993;**27**:218–28.
14 Levinson W, Roter DL. The effects of two continuing medical education programs on communication skills of practising primary care physicians. *J Gen Intern Med* 1993;**8** (6):318–24.
15 Boshuizen HPA, Schmidt HG. De ontwikkeling van medische expertise; implicaties voor het praktisch en theoretisch medisch onderwijs. [Development of medical skills; Implications for medical education]. In: Metz JCM, Scherpbier AJJA, Van Der Vleuten CPM. eds. *Medisch Onderwijs in de Praktijk [Medical Education in Proactice]*. Nijmegen: Van Gorcum, 1995.
16 Bensing JM, Sluijs EM. Evaluation of an interview course for general practitioners. *Social Sci Medicine* 1985;**20**:737–44.
17 Evans BJ, Kiellerup FD, Stanley RO, Burrows GD, Sweet B. A communication skills programme for increasing patients' satisfaction with general practice consultations. *Br J Med Psychol* 1987;**60**:373–8.
18 Faulkner A, Argent J, Jones A, O'Keeffe C. Improving the skills of doctors in giving distressing information. *Med Educ* 1995;**29**:303–7.
19 Gask L, McGrath G, Goldberg D, Millar T. Improving the psychiatric skills of established general practitioners: Evaluation of group teaching. *Med Educ* 1987;**21**:362–8.
20 Gask L, Goldberg D, Lesser AL, Millar T. Improving the psychiatric skills of the general practice trainee: an evaluation of a group training course. *Med Educ* 1988;**22**:132–8.
21 Gask L, Goldberg D, Porter R, Creed F. The treatment of somatization: Evaluation of a teaching package with general practice trainees. *J Psychosomatic Res* 1989;**33**:697–703.
22 Putnam SM, Stiles WB, Jacob MC, James SA. Teaching the medical interview: an intervention study. *J Gen Intern Med* 1988;**3**:38–47.
23 Roter DL, Cole KA, Kern DE. An evaluation of residency training in interviewing skills and the psychosocial domain of medical practice. *J Gen Intern Med* 1990;**5**:347–54.
24 Roter DL, Hall JA, Kern DE, Barker R, Cole KA, Roca RP. Improving physicians' interviewing skills and reducing patients' emotional distress. *Arch Intern Med* 1995;**155**:1877–84.
25 Sideris DA, Tsouna-Hadjis P, Toumanidis ST, Vardas PE, Moulopoulos SD. Attitudinal educational objectives at therapeutic consultation: Measures of performance, educational approach and evaluation. *Med Educ* 1986;**20**:307–13.
26 Smith RC, Osborn GG, Hope RB, Lyles JS, Van Egeren LF, Henry R *et al.* Efficacy of a one month training block in psychosocial medicine for residents: a controlled study. *J Gen Intern Med* 1991;**6**:535–4.
27 Smith RC, Mettler JA, Stoffelmayr BE, Lyles JS, Marshall AA, Van Egeren LF *et al.* Improving residents' confidence in using psychosocial skills. *J Gen Intern Med* 1995;**10** (6):315–20.
28 Smith RC, Lyles JS, Mettler JA, Marshall AA, Van Egeren LF, Stoffelmayr BE *et al.* A strategy for improving patient satisfaction by the intensive training of residents in psychosocial medicine, randomized study. *Acad Med* 1995;**70**:729–32.
29 Smith RC, Lyles JS, Mettler J, Stoffelmayr BE, Van Egeren LF, Marshall AA *et al.* The effectiveness of intensive training for residents in interviewing. A randomized, controlled study. *Ann Intern Med* 1998;**128**:118–26.
30 Bandura A. *Social Learning Theory*. Englewood-Cliffs (NJ): Prentice Hall, 1977.
31 McAvoy BR. Teaching Clinical Skills to Medical Students: The Use of Simulated Patients and Videotaping in General Practice. *Med Educ* 1988;**22**:193–9.
32 Smith RC, Marshall AA, Cohen-Cole SA. The Efficacy of Intensive Biopsychosocial Teaching Programs for Residents: a Review of the Literature and Guidelines for Teaching. *J Gen Intern Med* 1994;**9**:390–6.
33 Francke AL, Garssen B, Huijer Abu-Saad H. Determinants of changes in nurses' behaviour after continuing education: a literature review. *J Advanced Nursing* 1995;**21**:371–7.
34 Festinger L. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press, 1957.
35 Pieters HM. *De Utrechtse Consult Evaluatie Methode. Vaardigheden in consultvoering van huisartsen in opleiding getoetst.* [The Utrecht Consultation Assessment Method. The assessment of consultation skills of trainees in the vocational training]. (dissertation) Utrecht University. Lelystad: MEDITekst, 1991.

36 Bensing JM, Tan LHC, De Boer SA, Mol SSL. LACONTO in everyday general practice. A multisystem comparison. In: *Doctor-Patient Communication and the Quality of Care. Methods and Measures; Process and Outcome. Report of an Invitational Workshop.* Mol SSL, Bensing JM. eds. Utrecht: NIVEL, 1991.

37 Bales RF. *Interaction Process Analysis: a Method for the Study of Small Groups*. Massachusetts: Addison-Wesley, 1950.

38 Stiles WB. Verbal response modes and dimensions of interpersonal roles: a method of discourse analysis. *J Personal Social Psychol* 1978;**36**:693–703.

39 Roter DL. *The Roter Interaction Analysis System (RIAS) Coding Manual.* Baltimore (MA): School of Hygiene and Public Health, John Hopkins University, 1997.

40 Grol R, Thielens V, Mokkink H, Zwaard T. Ideaal of werkelijkheid? Problemen bij de ontwikkeling en invoering van standaarden. [Ideal or reality? Problems in the development and implementation of medical standards]. *Huisarts En Wetenschap* 1988;**31**:392–7.

41 Bowman FM, Goldberg DP, Millar T, Gask L, McGrath G. Improving the skills of established general practitioners: the long-term benefits of group teaching. *Med Educ* 1992;**26**:63–8.

42 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;**312**:1215–8.

Received 22 *July* 1998; *editorial comments to authors* 15 *March* 1999; *accepted for publication* 14 *April* 1999