

Postprint Version	1.0
Journal website	<a href="http://www.pec-journal.com/article/S0738-3991(11)00358-2/abstract">http://www.pec-journal.com/article/S0738-3991(11)00358-2/abstract</a>
Pubmed link	<a href="http://www.ncbi.nlm.nih.gov/pubmed/21784600">http://www.ncbi.nlm.nih.gov/pubmed/21784600</a>
DOI	10.1016/j.pec.2011.06.010

This is a NIVEL certified Post Print, more info at <http://www.nivel.eu>

## How do lay people assess the quality of physicians' communicative responses to patients' emotional cues and concerns? An international multicentre study based on videotaped medical consultations

MARIA ANGELA MAZZI<sup>A</sup>, , JOZIEN BENSING<sup>B</sup>, MICHELA RIMONDINI<sup>A</sup>, IAN FLETCHER<sup>C</sup>, LIESBETH VAN VLIET<sup>B</sup>, CHRISTA ZIMMERMANN<sup>A</sup> AND MYRIAM DEVEUGELE<sup>D</sup>

<sup>a</sup> Department of Public Health and Community Medicine, University of Verona, Italy

<sup>b</sup> NIVEL, Netherlands Institute for Health Services Research/Utrecht University, The Netherlands

<sup>c</sup> Division of Clinical Psychology, Whelan Building, University of Liverpool, Liverpool, UK

<sup>d</sup> Department of General Practice and Primary Health Care, University of Ghent, Ghent, Belgium

### ABSTRACT

Objective: to establish which kind of physician communicative responses to patient cues and

concerns are appreciated by lay people.

Methods: A balanced sample (259 people) was recruited in public places to participate in a full day observation of four videotaped standardized medical consultations. In a two-step procedure participants gave their individual quality ratings of the whole consultations and then of a set of four fragments from each consultation. They contained a patient negative emotional expression and the subsequent physician response, according to the VR-CoDES.

Results: Higher quality ratings were given to physician responses which provided space to the patient to talk and to the explicit expressions of empathy. The explicit responses were favored above non-explicit responses. Participants' global evaluation of the whole consultation affected their quality assessments of the fragments (halo-effect). In a multivariate model, lay people's background characteristics appeared to be relevant: to be female, of lower educational level and living in Belgium or Italy predicted higher ratings.

Conclusions: Providing space to patients is appreciated by all participants, combined with the need for tailor made communication.

Practice implications: To teach physicians listening skills and how to show empathy with distressed patients should be a core element in medical education.

## INTRODUCTION

There is growing plea for incorporating the public's experiences in the quality assessment of health care [1]. Doctor-patient communication is a domain 'par excellence' to include the public's perspective in quality assessments, as (a) health care users often have different priorities from health care providers [2], (b) tend to stress the importance of good communication [3] and [4], (c) often report quality problems in this area [5], and (d) these communication problems contribute to many adverse patient outcomes, such as non-adherence [6], formal and informal complaints [7], medical lawsuits [8] and patient dissatisfaction [5].

In short: doctor-patient communication is an area which is under scrutiny of the general population and could benefit from patients' input when trying to make improvements.

A key concept in research on doctor-patient communication, which reflects this orientation, is 'patient-centered care' (PCC). In a Cochrane review, patient-centeredness was defined as 'a philosophy of care that encourages shared control of the consultation, decisions about management of the health problems with the patient, and/or a focus in the consultation on the patient as a whole person who has individual preferences situated within social contexts (in contrast to a focus in the consultation on a body part or disease)'. This

review also concluded that 'patient-centeredness' is hard to define, and that more research is needed [9].

The concept 'patient-centeredness' claims to cover divergent areas: exploring the experience and expectations of disease and illness, understanding the whole person, finding common ground (partnership), health promotion, enhancing the doctor-patient relationship, and the realistic use of time [3] and [10]. However, there is little empirical evidence from the patients' perspective to support the precise structure of the model or to identify the components most important to patients [3].

De Haes et al. suggested to deconstruct communication in a number of meaningful elements, by – theoretically – deriving specific communication behaviors from the different aims and functions of the medical encounter, and – empirically – link these behaviors to predefined endpoints [11]. The following core functions of the medical encounter were distinguished: fostering the doctor-patient relationship, gathering information, providing information, (shared) decision making, enabling the patient and stimulating self-management, and responding to emotions.

Inspired by this approach, we decided to undertake a study in which ONE core function of the medical consultation was selected, which is vital for patient centered care, i.e. '*responding to negative emotions*', and ONE relevant endpoint is defined, i.e. *lay people's views on the quality of physicians' responses*. We decided to focus on a particularly sensitive area: physician's response to patients' cues or concerns, because, from the literature, we know that patients are often reluctant to reveal their real problems directly, providing subtle cues or concerns, instead [12]. This means that dedicated effort is required to get patients' emotions in the open. However, the literature also shows that health care providers often ignore patient cues and concerns [12], [13] and [14], thus leaving potential important topics unspoken [15]. For doctors, this seems to be a delicate area to maneuver, where a 'faux pas' is easily made. The quality assessment of lay people, who do not have specialized or professional knowledge of the subjects, could be helpful to develop empirically based guidelines and targeted skills training.

Lay people's perspective on the quality of physicians' response to emotions is a relevant topic, because up to date little is known about which physicians' responses to patients'

cues or concerns are appreciated or disliked by those on the receiving end: potential patients. When patients or the public are approached for the evaluation of health care, usually only general assessments are tapped from the respondents, such as whether there was sufficient time or attention, but without a clear benchmark, which would make it possible to compare people's opinions on the quality of care, based on the same concrete examples of communication. Yet, knowledge about which kind of communication is approved or disapproved by the public in situations where a physician has to deal with sensitive issues, could be helpful for physicians in choosing a communication strategy for those situations. It is also important to know when physician's communication evokes different reactions, because that would mean that physicians need to tailor their communication to patient preferences and expectations in those situations.

For this study we presented to lay people an identical set of various examples of a specific doctor–patient interaction in which the physician replies to patient's expression of a negative emotion, cue or concern.

The aims were to explore how lay people with different backgrounds assess the quality of the physician responses and how universal their quality assessments are. Are some types of communication always favored over others, or do they evoke differential responses in different respondents? We are also interested to learn whether people assess *the doctor* or *his/her behavior*: in other words to what extent is people's assessment of concrete communicative behaviors directly related to the type of communication at the micro-level, or determined by their general image of the doctor (halo-effect). The added value of studying these aspects in an international study with a wide variety of participants is, that we get to know how universal lay people's assessments of physician responses to patient cues and concerns are, putting to test the assumption that 'patient-centered communication' has the same meaning for all people, regardless who they are, and where they live.

## 2. METHODS

### 2.1. Design

The study is an international observational study which draws its name (GULiVer) from the four centers involved: Ghent University (Belgium), Utrecht University/NIVEL (the Netherlands), Liverpool University (United Kingdom) and the University of Verona (Italy). The benchmark material consists of an identical set of  $2 \times 4 = 8$  videotaped OSCE's (Objective, Structured, Clinical Examinations), used to examine the quality of communication of medical students as part of their final exams. The videotapes cover two different scenarios (period pain and vaginal discharge) and were selected to represent a maximum variation in the quality of communication according to the examiners. For the present study these videos were subtitled (the Netherlands and Belgium) or dubbed (Italy), reflecting the common way international television programs are handled in the respective countries. In each country, the same videotaped OSCE consultations were shown to 8 or 9 lay panels, each consisting of 6–8 citizens. Each lay panel observed (in random order) four different examples of the same scenario. In all four countries the same procedures were followed according to a detailed protocol [16], approved by the local ethics committee.

### 2.2. Sample

Participants were recruited in public areas, via calls in free local newspapers and word of mouth. Inclusion criteria were: age over 18 years; at least one GP-visit over the last 12 months; speaking the country's language. In order to ensure a heterogeneous distribution of the sample, the selection of participants was

stratified by gender (separate male and female panels) and age (18–30; 31–49;  $\geq 50$ ). The overall sample consisted of 259 participants, equally distributed across the centers and the stratification variables, as established by the study design and confirmed by the Generalized Linear Model (GLM) analysis [17] (deviance = 21.2 (df = 40);  $p = 0.99$ ).

### 2.3. Physicians' responses to patient cues and concerns

For the selection of consultation fragments all videotaped consultations had been coded to identify patient cues and concerns as well as doctor responses, applying the Verona Coding Definitions of Emotional Sequences (VR-CoDES) [18], [19] and [20]. This system consists of two parts for respectively patients' cues and concerns (VR-CoDES-CC) and provider responses (VR-CoDES-P). Cues are defined as "verbal or nonverbal hints which suggest an underlying unpleasant emotion and that lack clarity". Concerns are defined as "clear and unambiguous expressions of an unpleasant current or recent emotion that are explicitly verbalized with or without a stated issue of importance". The VR-CoDES-CC has a satisfactory interrater-reliability: Cohen's kappa of 0.70 ( $\pm 0.03$ ), percentage agreement 81.46 [18]. The validity of the coding system was confirmed by Eide et al., who replayed videotaped medical consultations to the patients involved and invited them to comment their contributions [21]. The VR-CoDES-P has two main axes for classifying provider responses, corresponding to the explicitness of the response (yes/no) and the amount of space for the patient (yes/no). As in the original study [19], the interrater reliability in this study was good with a percentage agreement of 90.2%; kappa = 0.86

### 2.4. Quality assessments

The lay people gave their quality assessment in a two-step procedure. In the morning-session, each group of lay people observed four consultations from the same scenario without any group discussion. The participants then, individually, rated each consultation on a 1 ("not at all satisfying") to 10 ("very satisfying") Likert-scale for the overall quality of communication. In the afternoon session, participants, individually, assessed the quality of four very short preselected fragments from each of these consultations, in which the patient on the video offers a cue to an underlying worry, or expresses a concern. The participants were asked to observe how the student-doctors respond to these patient expressions and to rate the appropriateness of these responses on a 1–10 Likert scale (1 = "I do not like the answer at all"; 10 = "I really like the answer").

### 2.5. Background characteristics of participants

In this study the following background characteristics of the participants were considered: age, gender, nationality and educational level (low, medium and high).

### 2.6. Statistical analyses

As there were no differences between quality ratings of physicians' responses to cues and concerns (mean = 6.1, sd = 2.1 versus mean = 6.3, sd = 2.1 respectively;  $t(4048) = -1.69$ ,  $p = 0.09$ ), these were lumped together. Cronbach's alpha was used to determine the internal consistency. Partial and Pearson's correlations were calculated to determine the interdependency of fragments within the same consultation, and their relationship with the global quality rating of that particular consultation given during the morning session (see

[Table 2] and [Table 3]). The relative contribution of the presented stimulus (physician, type of response) and the lay people's background characteristics on the quality ratings was determined applying a series of multilevel regression analyses (see Table 4). In a stepwise procedure, four different models were used which estimate each fragment quality assessment, starting with the empty model, adding blocks of information step by step, according to different sources: physician response to cue/concern (model 1), participants' global assessment (model 2) and their socio-demographic characteristics (model 3). The models were compared (see goodness of fit row) in terms of their explained proportion of variance with respect to the first level and second level analysis [22]. The multilevel analysis recommends considering the specific variance component at the two levels, and not the overall variance as used in ordinary regression [23].

### 3. RESULTS

#### 3.1. What kind of physician responses is positively or negatively valued?

The quality assessments by the lay people were normally distributed, had good internal consistency (Cronbach's  $\alpha = 0.81$  and  $0.84$  for the two scenarios), and showed a large range in scores of the perceived quality of physicians' responses on patient cues and concerns. The highest quality rating (mean 7.6; sd 1.8) was for a student-doctor who expressed empathy and partnership to a patient who was struggling with telling her story (Patient: *'no, mmm, to be honest I've always been a little bit uncomfortable coming to talk about. I don't know why. I just am really'*; Doctor: *'well I know it might, it's always uncomfortable talking to people about your sex life and very personal issues but we are always here to talk to'*). The lowest rating (mean 3.7; sd 2.0) was for a student-doctor who shut the patient down when she expressed her worries (Patient: *'well they do know about it because they know the reason I I don't hide the reason I'm taking it off, but obviously for them if I'm taking a week off every month it's not very good is it?'* Doctor: *'no. OK. Are you married at the moment?'*).

Responses which provide space to the patients were generally more appreciated than responses which reduced space (mean 6.5; sd 1.0, versus mean 5.6; sd 1.2;  $p < .01$ ). Whether the responses were explicit or non-explicit made a smaller, but still significant difference in the quality rating of the respondents (mean 6.3; sd 1.1, versus mean 6.0; sd 1.2;  $p < 0.01$ ). Combining both axes, the average quality rating for explicit, providing space responses (EP) was 6.5 (sd 2.0; 95% CI: 6.4–6.6), which is comparable to the average quality rating for non-explicit providing space responses (NP): mean = 6.5 (sd 2.2; 95% CI: 6.3–6.7). The average quality rating for explicit reducing space responses (ER) was 5.7 (sd 2.1; 95% CI: 5.5–5.8), while non-explicit reducing space responses (NR) had the lowest average quality rating: 5.4 (sd 2.1; 95% CI: 5.2–5.6).

Table 1 presents an overview of the identified types of physician responses to patient cues and concerns. The types of responses are ordered from high to low quality assessments.

#### [TABLE 1]

The overall most positively valued type of physician responses is: EP-empathic response, with an average score of 7.3, followed by NP-backchannels (mean 6.8). Interestingly, non-explicit empathy, despite providing space to the patient, was not appreciated by the



participants, (mean 5.2), in particular when this was followed by switching the topic of the conversation (Patient: *'and I'm just a bit worried that I've caught something'*; Doctor: *'right'*; Patient: *'cause I've not had it before'*; Doctor: *'yeah, totally understand, right, mmmm, just going back to your history again; mmmm, so you have been with the same partner for three years and, mmmm, just gonna ask you a bit about your menstrual cycles: have they been regular?'*).

Ignoring the words of the patients, switching to another topic, giving unsolicited information and advice, or shutting the patient down have all quality ratings below 6, which can be considered as 'insufficient quality'.

### 3.2. Are the quality assessments independent?

Table 2 shows that the specific fragment assessments of each consultation are related to the global quality assessment of the morning session. Moreover, there is a moderate and positive relationship between pairs of assessments within each student-doctor (see Table 3), as shown by the Pearson's correlation (column 4) and the partial correlation between pairs of fragments, when accounting for the presence of the global assessment (column 5). The positive relationship is confirmed, when the correlation is used to measure the variability among fragment assessments within each doctor (ICC, see column 3), with the exception for doctors A and B.

#### [TABLE 2 AND 3]

Are quality assessments only linked to type of physician responses, or also in the eye of the beholder? Table 4 presents the results of the multilevel analyses in four models, following a hierarchical procedure.

#### [TABLE 4]

Model 1 shows that participants appreciate a doctor who provides space to the patient ( $p$ -value < 0.05), while they do not pay attention to whether a doctor's response to the patient is explicit or non-explicit ( $p$ -value > 0.10). They also distinguish among different student-doctors, who could generically be classified into a "low-valued communication style" (doctors B, D and G) versus a "high-valued communication style" (doctors A, C, E, F and H). This first informative block results in 11% of variance explained at the first level ( $R^2$  level one). In this model no additional variance is explained at the participant level.

The addition of the global assessment of each consultation (Model 2) improves the proportion of variance explained at the first level (17%) but unexpectedly also at the second level (29%), revealing an effect of participants individual characteristics on their global assessments. Therefore in Model 3, the participants' characteristics were introduced, which added about another 19% of explained variance at the second level of analysis. The estimated parameters of the second level show that being elderly, female, of lower educational level, and from Belgium and Italy predicted significantly higher quality ratings as compared to being male, younger, of better education and from the United Kingdom and the Netherlands. Table 4 shows that the fourth model obtains the highest proportion of variance explained (17% and 42% at levels 1 and 2, respectively) compared to the "empty model", also showing the best goodness of fit value (lowest AIC).

## 4. Discussion and conclusion

### 4.1. Discussion

In this study the quality of communication was measured for one of the six core functions of the medical encounter [11], i.e. responding to patient emotions. This function is seen as highly important by patients, but is at the same time most problematic for physicians [24]. The relevance of this function for patients was

one of the main reasons for choosing lay people's assessments of physicians' response to patients' cues and concerns as the primary outcome for this study. The study shows some interesting results.

The lay people were able to discriminate between different student-doctors in a consistent way. These findings support the validity of the VR-CoDES-system for measuring cue-response sequences, and give empirical support to the feasibility of including lay people in the quality assessment of medical communication.

An important finding is the congruence of the lay people's quality assessments with the theoretical literature on doctor-patient communication: 'providing space', and in particular physician's explicit showing of empathy, indeed seems to have an universal value as the most adequate response to patient's implicit (cue) or explicit (concern) expressions of emotional distress. The study participants were very appreciative when a doctor expressed empathy in such a situation. Empathic statements obtained the highest quality rating from all participants, independently from their background characteristic and nationality. This result provides a firm empirical base for including expressions of empathy by physicians as a must in clinical guidelines for all health problems which evoke emotions.

While 'providing space' seems to be universally valued by lay people over 'reducing space' types of communication, the other main dimension of the VR-CoDES-P system (explicit versus non-explicit) does not evoke such an unequivocal reaction from the participants in this study. There is a small significant effect in favor of explicit responses, but this seems to have hardly any clinical relevance. An in-depth analysis of the results from Table 1 suggests that this dimension mitigates the impact of the providing/reducing space dimension, both in a positive and a negative way. For instance, when student-doctors reduced patients' space by giving information or advice, this was higher appreciated when it happened in a non-explicit way. Inversely, an empathic statement was less appreciated when it occurred in a non-explicit manner, as if the space-providing character of this communication was not picked up by the participants.

Another important result was that the quality assessments were interrelated and also related to the global assessment of the quality of communication, suggesting a halo-effect: it seems that the general atmosphere of the consultations had a radiating effect on the assessments of student-doctors' responses to patients' emotions. This finding suggests that a physician can switch topics, without receiving a very low quality rating for this space reducing behavior, as long as changing the topic of the conversation is embedded in space-providing communication, such as backchannels or an acknowledgement of patients' emotions. This is reassuring, because it means that physicians can make a 'faux pas', without being blamed, as long as his/her overall performance is appreciated. It also means that shutting patients down, for instance when moving to the next phase (or to the end) of the consultation (which is something that needs to be done!) does not necessarily result in low quality ratings, as long as it is embedded in a room-providing atmosphere. Whether the specific quality assessments may influence global quality assessments cannot be answered

with this study. This would require a design, which systematically varies the order of assessments.

The last issue that needs reflection is the influence of the lay people's background characteristics on the quality assessments. While the type of physicians' response ('providing space') was the strongest explaining factor in the models, a substantial proportion of the variance was explained by the individual characteristics of the participants. The more critical appraisal from younger, higher educated people is consistent with the literature, which shows that older and lower educated patients are usually more satisfied with the care delivered [4]. A plausible, but not yet empirically founded explanation could lie in the different stage of development of the vocational training for General Practitioners in these countries, with the UK and the Netherlands having a longer and stronger tradition in this area, in particular with regard to communication skills training. This suggests that English and Dutch patients are used to General Practitioners who are more proficient in communication as compared to Belgian and Italian patients, but more research on this issue is needed.

A strength of this study is that all participants have been exposed to the same identical set of video-fragments of medical consultations. To our knowledge, this is the first time that an international study on this scale has been performed. 259 participants from four countries have observed the same sets of videotapes according to the same strict protocol. Another strength is that the participants observed four different consultations of the same scenario, which made it possible to compare the performance of different student-doctors under the same conditions, and to set their standards about communication examples that they preferred. The third strong point is that the type of physician communication was standardized by applying the VR-CoDES to identify patients' negative emotional expressions and physicians' responses, which made it easier to compare the student-doctors communication performance. Together, this means that the quality assessments were highly standardized, which gives support to the robustness of our results.

The study also has limitations. The first is that we used videotaped OSCE-consultations which were recorded to test the medical students' communication skills during their final exams. The type of physician communication shown does not necessarily represent the communication of regular doctors. However, we saw a wide variation in student-doctors' responses, and we believe that the advantages of the standardized situation in which the observers could compare various communication styles in the same type of medical consultation, outweigh the disadvantage of working with medical students instead of regular doctors. The second limitation is that lay people were used for giving quality assessments. Although these people had experienced at least one medical consultation in the last year, they were not personally involved in the observed interactions and therefore not emotionally engaged. At the same time the absence of a personal involvement is one of the strengths of the present study, since it reduces the risk of complacency effects

leading to an overestimation of positive and/or negatives behaviors. Although the use of lay persons is a widely used methodology [4], [25] and [26], the validity of these assessments as compared to the assessments by real patients still needs to be confirmed. Finally, only Western-European countries were involved in the study, which means that results cannot automatically be generalized to other cultures. Replication in other countries could be worthwhile.

#### **4.2. Conclusion**

Some types of physician responses are differently valued by people. This would require from physicians personal sensitivity and tailor-made communication when discussing sensitive issues with patients from



different backgrounds. Therefore, the overall conclusion of this study is: providing space to patients who show signs of distress is universally appreciated by all participants. This means that doctors have one important trump card in these situations: 'when in doubt, express empathy!'

#### **4.3. Practice implications**

Space providing listening skills and expressions of empathy, which appear to make people forgive other occurring communication shortcomings of the health provider, should be core clinical skills to be targeted in medical education programs.

#### **Conflict of interest**

The authors have no potential conflict of interest related to the subject of the paper.

#### **Funding**

This study was made possible through a grant of The Dutch Ministry of Health, Welfare and Sports (National Fund for Patient-Oriented Research).

#### **Acknowledgements**

The Clinical Skills Team at The Medical School in University of Liverpool for supporting the study and assisting the recruitment and videoing of the summative examinations.

The lay panels in Ghent, Utrecht, Liverpool and Verona for their committed participation in the study.

The Dutch Ministry of Health, Welfare and Sports (National Fund for Patient-Oriented Research) for their financial support of the study.

Francesca Moretti, Corinne Geurtz and Marjolein van Bogaert for their help in study preparation and data collection.

#### **REFERENCES**

- [1] A. Coulter, After Bristol: putting patients at the centre, *Qual Saf Health Care* 11 (2002), pp. 186–188.
- [2] D.N.M. Wright, J.L. Corner, J.B. Hopkinson and C.L. Foster, The case for user involvement in research: the research priorities of cancer patients, *Breast Cancer Res* 9 (2007), p. S3 Available from: [http://www.sciencedirect.com/science?\\_ob=RedirectURL&\\_method=externObjLink&\\_locator=url&\\_issn=07383991&\\_origin=article&\\_zone=art\\_page&\\_plusSign=%2B&\\_targetURL=http%253A%252F%252Fbreast-cancer-research.com%252Fcontent%252F9%252FS2%252FS3](http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_issn=07383991&_origin=article&_zone=art_page&_plusSign=%2B&_targetURL=http%253A%252F%252Fbreast-cancer-research.com%252Fcontent%252F9%252FS2%252FS3).
- [3] P. Little, H. Everitt, I. Williamson, G. Warner, M. Moore and C. Gould et al., Preferences of patients for patient centred approach to consultation in primary care: observational study, *Brit Med J* 322 (2001), pp.468–472.
- [4] J.A. Hall, D.L. Roter, D.C. Blanch and R.M. Frankel, Nonverbal sensitivity in medical students: implications for clinical interactions, *J Gen Intern Med* 24 (2009), pp. 1217–1222.
- [5] R.S. Beck, S. Daughtridge and P.D. Sloane, Physician–patient communication in the primary care office: a systematic review, *J Am Board Fam Pract* 15 (2002), pp. 25–38.
- [6] E. Vermeire, H. Hearnshaw, P. vvan Royen and J. Denekens, Patient adherence to treatment: three decades of research. A comprehensive review, *J Clin Pharm Ther* 26 (2001), pp. 331–342.
- [7] M.M. Wofford, J.L. Wofford, J. Bothra, S.B. Kendrick, A. Smith and P.R. Lichstein, Patient complaints about physician behaviors: a qualitative study, *Acad Med* 79 (2004), pp. 134–138.
- [8] G.B. Hickson, C.F. Federspiel, J.W. Pichert, C.S. Miller, J. Gauld-Jaeger and P. Bost, Patient complaints and malpractice risk, *J Amer Med Assoc* 287 (2002), pp. 2951–2957.

- [9] S. Lewin, Z. Skea, V.A. Entwistle, M. Zwarenstein and J. Dick, Interventions for providers to promote a patient-centred approach in clinical consultations, *Cochrane Database Syst Rev* 4 (2001), p. CD003267.
- [10] M. Stewart, Towards a global definition of patient centred care. The patient should be the judge of patient centred care, *Brit Med J* 322 (2001), pp. 444–445.
- [11] H. de Haes and J. Bensing, Endpoints in medical communication research, proposing a framework of functions and outcomes, *Patient Educ Couns* 74 (2009), pp. 287–294.
- [12] C. Zimmermann, L. Del Piccolo and A. Finset, Cues and concerns by patients in medical consultations: a literature review, *Psychol Bull* 133 (2007), pp. 438–463.
- [13] R. Uitterhoeve, J. de Leeuw, J. Bensing, C. Heaven, G. Borm and P. Demulder et al., Cue-responding behaviours of oncology nurses in video-simulated interviews, *J Adv Nurs* 61 (2008), pp. 71–80.
- [14] J. Jansen, J.C. van Weert, J. de Groot, S. van Dulmen, T.J. Heeren and J.M. Bensing, Emotional and informational patient cues: the impact of nurses' responses on recall, *Patient Educ Couns* 79 (2010), pp. 218–224.
- [15] C.A. Barry, C.P. Bradley, N. Britten, F.A. Stevenson and N. Barber, Patients' unvoiced agendas in general practice consultations: qualitative study, *Brit Med J* 320 (2000), pp. 1246–1250.
- [16] Moretti F, Fletcher I, Mazzi MA, DeVeugele M, Rimondini M, Geurts C, et al., Travelling into the heart of good doctor–patient communication from a patient perspective. Study protocol of an international multicentre study. *Eur J Public Health* 2011, doi:10.1093/eurpub/ckr071.
- [17] P. Mc Cullagh and J.A. Nelder, *Generalized linear models* (2nd ed.), Chapman & Hall, London (1989).
- [18] C. Zimmermann, L. Del Piccolo, J. Bensing, S. Bergvik, H. de Haes and H. Eide et al., Coding patient emotional cues and concerns in medical consultations: the Verona coding definitions of emotional sequences [VR-CoDES], *Patient Educ Couns* 82 (2011), pp. 141–148.
- [19] L. Del Piccolo, H. de Haes, C. Heaven, J. Jansen, W. Verheul and J. Bensing et al., Development of the Verona coding definitions of emotional sequences to code health providers' responses (VR-CoDES-P) to patient cues and concerns, *Patient Educ Couns* 82 (2011), pp. 149–155. [20] [http://www.sciencedirect.com/science?\\_ob=RedirectURL&\\_method=externObjLink&\\_locator=url&\\_issn=07383991&\\_origin=article&\\_zone=art\\_page&\\_plusSign=%2B&\\_targetURL=http%253A%252F%252Fwww.each.nl%252F](http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_issn=07383991&_origin=article&_zone=art_page&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.each.nl%252F) [accessed 19.05.11.].
- [21] H. Eide, T. Eide, T. Tone Rustøen and A. Finset, Patient validation of cues and concerns identified according to Verona coding definitions of emotional sequences (VR-CoDES): a video- and interview- based approach, *Patient Educ Couns* 82 (2011), pp. 156–162.
- [23] J.J. Hox, *Multilevel analysis: techniques and applications*, Lawrence Erlbaum Associates, Mahwah (2002).
- [24] W. Levinson, R. Gorawara-Bhat and J. Lamb, A study of patient clues and physician responses in primary care and surgical settings, *J Amer Med Assoc* 284 (2000), pp. 1021–1027.
- [25] M.S. Mast, J.A. Hall, C. Köckner and E. Choi, Physician gender affects how physician nonverbal behavior is related to patient satisfaction, *Med Care* 46 (2008), pp. 1212–1218.
- [26] L.A. Fogarty, B.A. Curbow, J.R. Wingard, K. McDonnell and M.R. Somerfield, Can 40 s of compassion reduce patient anxiety?, *J Clin Oncol* 17 (1999), pp. 371–379.

## TABLES

**Table 1**  
Quality assessment of specific types of student-doctor responses to patients' cues and concerns.

VR-CoDES-P	Example	No. of fragments	No. of participants' assessment	Mean	Std. dev.	95% CI
EP – Empathic response	P: no mmm to be honest I've always been a little bit uncomfortable coming to talk about I don't know why I just am really. <b>D: well I know it might its always uncomfortable talking to people about your sex life and very personal issues but we are always here to talk to</b>	5	200	7.3	1.7	7.1–7.6
NP – Back channel	P: no it's its' not been too bad mmm I mean everythings been ok from that's concerned it did kinda make I thought it made my period a little bit heavier. <b>D: right</b>	5	246	6.8	1.8	6.6–7.1
EP – Affective Acknowledgement	P: in the last few months I've had loads of time off half because of the pain and the other half because I'm not sleeping very good so I'm just knackered. <b>D: so always playing on your mind as well</b>	4	124	6.7	1.3	6.5–7.0
NP – Acknowledgement	P: I've never I mean I've had normal period pains like normal period cramps but this is like 10 times worse. <b>D: ok right I've got the list of what's happening</b>	1	131	6.5	1.9	6.2–6.9
EP – Content Exploration	P: is definitely getting heavier and the pain I've never had the pain like that before <b>D: ok can you control the pain with pain killers or does it not really make a difference?</b>	7	211	6.4	1.3	6.2–6.6
NR – Information advice	P: and it's just not very nice. <b>D: ok it's a really common problem; lots and lots of people come with this kind of thing; are you on any medication at the moment?</b>	1	122	6.3	1.6	6.0–6.6
EP – Content Acknowledgement	P: not particularly I mean I've kind of I've had you know I've taken sick days and I've used holidays and when it's been over a weekend I've only had to take a couple of days off either side but they have actually asked to see me next week and I think this is because of the amount of time I'm taking off. <b>D: right so you don't think they are taking your problem seriously or</b>	1	124	6.2	2.0	5.9–6.6
NR – Ignoring	P: and I'm a bit worried about what the outcome of the meeting is gonna be really. <b>D: all right so just to write it down you said you've had problems with periods and its heavy period that you got together with pain as well?</b>	5	253	6.1	1.4	5.9–6.3
ER – Switching	P: and I went like abroad and slept with someone and I'm just worried I caught something. <b>D: ok ok have you it's a general question have you ever contracted anything before sexually?</b>	5	253	5.9	1.5	5.7–6.1
EP – Affective Exploration	P: in a way not that we want any children anyway but it's just horrible. <b>D: yeah you feel inadequate sort of compared to anybody else?</b>	1	124	5.8	2.0	5.4–6.1
NP – Active Invitation	P: but it's having to change on such regular intervals it's just impossible to work or do any normal things. <b>D: yeah so what are you main concern regarding?</b>	2	246	5.5	2.3	5.2–5.7
NR – Shutting down	P: well they do know about it because they know the reason I I don't hide the reason I'm taking it off but obviously for them if I'm taking a week off every month it's not very good is it? <b>D: no ok are you married at the moment?</b>	2	124	5.4	1.4	5.2–5.7
ER – Information advice	P: I just feel a bit worried in case is something really serious you know. <b>D: hopefully like the doctor will come in and like should be able to like help allay your fears have a look at this discharge and do some tests to find out what's been going on and hopefully everything will be ok and he never need know</b>	2	122	5.3	1.6	5.0–5.6
NP – Implicit empathy	P: and I'm just a bit worried that I've caught something. <b>D: right. P: 'cos I've not had it before. D: yep I understand totally understand...</b>	2	122	5.2	1.7	4.9–5.5

EP=explicit providing space; NP=non-explicit providing space; ER=explicit reducing space; NR=non-explicit reducing space.

**Table 2**  
Correlations between the global quality assessments and the specific quality assessments of student-doctor responses to patient cues and concerns.

Specific assessment	Global assessment <sup>a</sup>			
	Doctor A mean = 7.4 sd = 1.8	Doctor B mean = 6.8 sd = 1.7	Doctor C mean = 6.8 sd = 1.8	Doctor D mean = 6.5 sd = 1.8
Scenario: period pain				
Doctor A, fragment 1	<b>0.41</b> <sup>**</sup>	0.08	0.26 <sup>**</sup>	0.03
Doctor A, fragment 2	<b>0.40</b> <sup>**</sup>	0.07	0.13	0.05
Doctor A, fragment 3	<b>0.25</b> <sup>**</sup>	0.26 <sup>**</sup>	0.02	0.31 <sup>**</sup>
Doctor A, fragment 4	<b>0.39</b> <sup>**</sup>	0.01	-0.01	-0.09
Doctor B, fragment 1	-0.01	<b>0.53</b> <sup>**</sup>	0.09	0.08
Doctor B, fragment 2	0.01	<b>0.36</b> <sup>**</sup>	0.11	0.05
Doctor B, fragment 3	0.12	<b>0.39</b> <sup>**</sup>	0.13	0.15
Doctor B, fragment 4	0.22 <sup>†</sup>	<b>0.30</b> <sup>**</sup>	0.21 <sup>†</sup>	0.14
Doctor C, fragment 1	0.11	0.08	<b>0.13</b>	0.14
Doctor C, fragment 2	0.31 <sup>**</sup>	0.03	<b>0.25</b> <sup>**</sup>	-0.13
Doctor C, fragment 3	0.10	0.13	<b>0.25</b> <sup>**</sup>	0.12
Doctor C, fragment 4	0.13	0.11	<b>0.29</b> <sup>**</sup>	-0.09
Doctor D, fragment 1	0.02	0.05	0.06	<b>0.47</b> <sup>**</sup>
Doctor D, fragment 2	-0.15	0.10	0.05	<b>0.47</b> <sup>**</sup>
Doctor D, fragment 3	-0.02	0.17	0.06	<b>0.15</b>
Doctor D, fragment 4	-0.20 <sup>†</sup>	0.23 <sup>†</sup>	-0.04	<b>0.55</b> <sup>**</sup>
Scenario: vaginal discharge				
Doctor E, fragment 1	<b>0.15</b>	0.02	0.08	-0.02
Doctor E, fragment 2	<b>0.16</b>	0.14	0.23 <sup>†</sup>	0.22 <sup>†</sup>
Doctor E, fragment 3	<b>0.16</b>	-0.01	0.06	0.12
Doctor E, fragment 4	<b>0.22</b> <sup>†</sup>	0.19 <sup>†</sup>	0.13 <sup>†</sup>	0.18
Doctor F, fragment 1	0.04	<b>0.26</b> <sup>**</sup>	-0.01	0.15
Doctor F, fragment 2	0.02	<b>0.41</b> <sup>**</sup>	0.16	0.35 <sup>**</sup>
Doctor F, fragment 3	0.12	<b>0.48</b> <sup>**</sup>	0.14	0.32 <sup>**</sup>
Doctor F, fragment 4	0.09	<b>0.39</b> <sup>**</sup>	0.24 <sup>**</sup>	0.31 <sup>**</sup>
Doctor G, fragment 1a	0.10	0.25 <sup>†</sup>	<b>0.26</b> <sup>†</sup>	0.07
Doctor G, fragment 1b	0.08	0.08	<b>0.48</b> <sup>**</sup>	0.23 <sup>†</sup>
Doctor G, fragment 2	0.04	0.04	<b>0.34</b> <sup>**</sup>	0.20 <sup>†</sup>
Doctor G, fragment 3	0.19 <sup>†</sup>	0.17	<b>0.25</b> <sup>**</sup>	0.20 <sup>†</sup>
Doctor G, fragment 4	0.24 <sup>**</sup>	0.06	<b>0.50</b> <sup>**</sup>	0.24 <sup>**</sup>
Doctor H, fragment 1	0.21 <sup>†</sup>	0.01	0.14	<b>0.27</b> <sup>**</sup>
Doctor H, fragment 2	0.10	0.02	-0.01	<b>0.30</b> <sup>**</sup>
Doctor H, fragment 3	0.12	0.16	0.17 <sup>†</sup>	<b>0.31</b> <sup>**</sup>
Doctor H, fragment 4	0.29 <sup>**</sup>	0.23 <sup>**</sup>	0.31 <sup>**</sup>	<b>0.42</b> <sup>**</sup>

Boldface number = possible halo-effect (fragment nested in consultation).

<sup>a</sup> Global assessment = "How do you rate the quality of the communication of this doctor?"

<sup>†</sup> 0.05 < p < 0.01.

<sup>\*\*</sup> p < 0.01.

**Table 3**  
Specific quality assessments of student-doctors' responses to cues and concerns.

	Mean	Standard deviation	Intraclass correlation (ICC)	Range of correlations within doctor	Range of partial <sup>a</sup> correlations
Scenario: period pain					
Doctor A	6.5	2.0	0.07	0.15-0.53	0.05-0.44
Doctor B	6.3	1.9	0.04	0.30-0.56	0.19-0.48
Doctor C	6.5	2.0	0.18	0.19-0.49	0.16-0.46
Doctor D	5.5	2.4	0.30	0.23-0.79	0.18-0.71
Scenario: vaginal discharge					
Doctor E	6.6	1.9	0.14	0.30-0.57	0.28-0.56
Doctor F	6.0	1.9	0.13	0.32-0.54	0.23-0.45
Doctor G	5.9	2.2	0.15	0.27-0.58	-0.01-0.45
Doctor H	6.2	2.2	0.25	0.28-0.52	0.19-0.46

<sup>a</sup> Partial correlation is obtained removing the global assessment effect.

**Table 4**

Multilevel linear regression models on specific quality assessments ( $n=3537$ ) of student-doctors' responses to cues and concerns.

	Intercept only Model (0)	+Fragments info Model (1)	+Halo effect Model (2)	+Responders info Model (3)
Fixed part	Coeff. (s.e.)	Coeff. (s.e.)	Coeff. (s.e.)	Coeff. (s.e.)
First level variables				
Explicit vs. non-explicit		-0.06 (.08)	-0.06 (.08)	-0.06 (.08)
Provide space vs. reduce space		1.36 (.08)*	1.36 (.08)*	1.36 (.08)*
Video				
Doctor A vs. Doctor E		-0.60 (.17)*	-0.85 (.15)*	-0.86 (.15)*
Doctor B vs. Doctor E		0.12 (.17)	0.08 (.15)	0.08 (.15)
Doctor C vs. Doctor E		0.29 (.16)	0.27 (.15)	0.26 (.14)
Doctor D vs. Doctor E		-1.09 (.16)*	-1.00 (.15)*	-1.01 (.14)*
Doctor F vs. Doctor E		0.22 (.14)	0.22 (.13)	0.22 (.13)
Doctor G vs. Doctor E		-0.67 (.11)*	-0.63 (.11)*	-0.61 (.11)*
Doctor H vs. Doctor E		-0.15 (.12)	-0.05 (.12)	-0.03 (.12)
Global assessment to consultation			0.34 (.02)*	0.34 (.02)*
Second level variables				
Gender (female vs. male)				0.22 (.10)*
Age in class				
31-50 vs. <30				-0.06 (.13)
>50 vs. <30				-0.30 (.13)*
Education				
Medium vs. lower				-0.41 (.18)*
Higher vs. lower				-0.64 (.18)*
Country				
IT vs. NL				0.40 (.15)*
UK vs. NL				0.01 (.15)
BE vs. NL				0.38 (.16)*
Constant	6.18 (.06)*	5.60 (.13)*	3.32 (.18)*	3.70 (.31)*
Random part				
ICC	0.17 (.02)	0.20 (.02)	0.16 (.02)	0.13 (.02)
Residual std deviation (within participants)	1.90 (.02)	1.79 (.02)	1.73 (.02)	1.73 (.02)
Std deviation of random intercept (between participants)	0.89 (.05)	0.89 (.05)	0.75 (.05)	0.68 (.04)
Goodness of fit				
AIC	17085	14602	14335	14263
(a) $R^2$ level-one (responses)		0.11	0.17	0.17
(a) $R^2$ level-two (participants)		0.00	0.29	0.42

Explained proportion of variance (by fixed effects) using the Snijders-Bosker approach ([24]; p. 99).

\* $p < 0.05$ .