Communication study

Effects of mode of presentation on ratings of empathic communication in medical interviews

Jennifer Nicolai a,⁎, Ralf Demmel b, Karin Farsch b

a Department of Psychosomatic and General Internal Medicine, University of Heidelberg, Thibastr. 2, 69115 Heidelberg, Germany
b Department of Clinical Psychology, University of Münster, Fliednerstr. 21, 48149 Münster, Germany

ARTICLE INFO

Article history:
Received 31 May 2009
Received in revised form 21 October 2009
Accepted 24 October 2009

Keywords:
Physician–patient communication
Empathy
Nonverbal
Channel

ABSTRACT

Objective: To test the impact of nonverbal behaviour on the assessment of a clinician’s level of empathic communication.

Methods: One hundred volunteers were asked to assess a clinician’s level of empathic communication using the Rating Scales for the Assessment of Empathic Communication in Medical Interviews (REM). Participants were randomly assigned to three groups differing with regard to the level of nonverbal information made available to them. Participants either watched a simulated medical interview, listened to an audio-only version of this interview, or read a transcribed version of the interview.

Results: Compared to watching a video and listening to an audiotape, respectively, reading a transcribed version of the interview produced lower empathy ratings and interrater reliabilities.

Conclusions: The findings suggest that assessments of a clinician’s level of empathic communication may differ according to the level of nonverbal information made available to the raters.

Practice implications: Focusing on the verbal level of communication alone ignores the fact that empathy can be expressed through nonverbal means. Hence, nonverbal channels need to be taken into account in addition to the verbal channel when conducting research on empathic communication in health care.

1. Introduction

The efficacy of medical treatment depends on many factors beyond the somatic aspects of the disease and a physician’s clinical knowledge. Empathy plays a key role in establishing a trusting physician–patient relationship [1,2]. A trusting physician–patient relationship enhances the patient’s readiness to reveal emotional problems [3] and the motivation to participate in the process of decision-making [4], which in turn is associated with compliance [5], health outcomes [6], and general patient satisfaction [7].

Building on findings that empathy enhances communication [8] and that empathy is a teachable and learnable skill [9], it is not surprising that empathy is considered as one of the core skills in communication skills training for health care professionals [10]. However, training success can be assessed in various ways including self-ratings, (simulated) patient reports, or external observer ratings of physician' behaviours. There are two major sources of variability in observational ratings of social interactions [11], making it difficult to compare the results of different studies. The first major source of variability stems from individual differences between raters. This source of variability is well recognized and usually well documented. A second major source of variability arises from the fact that the same interaction may be assessed using different sources of data. Research on physician–patient interaction is usually based on observational data analyzing videotapes, audiotapes, or transcripts of medical encounters. Videotaped interactions convey most information with respect to communication skills [12], but videotape recording is rather expensive [13] and potentially intrusive [14]. It is easier to obtain audiotapes, but they do not capture visual nonverbal behaviours [15]. Finally, transcription-based analysis is often used to study the verbal content of interpersonal interactions [16]. One advantage that transcripts have over video or audio recordings is that transcripts provide static information consequently facilitating the coding process. There is also evidence that coders might perceive visual information rather as distracting than supportive [13]. However, the process to transcribe interactions is very time consuming and expensive. Moreover, nonverbal behaviours are usually lost in verbal transcripts.

The mode of presentation of interactions varies greatly across different studies. Some researchers use videotapes, whereas others use audiotapes and still others transcripts or combinations of these. The decision for a particular mode of presentation is often based on convenience, availability, or ease-of-use rather than on validity. The mode of presentation of the interaction is important because it influences the amount of information available to the rater. Given...
that different modes of presentation vary greatly in the amount and type of information that they convey, it is reasonable to assume that the mode of presentation affects ratings. Thus, it is necessary to determine what constitutes the most accurate and valid source of data. However, there is surprisingly little research addressing this issue. For example, Weiss et al. [17] compared ratings of videotaped and audiotaped psychotherapy sessions using the Therapist Action Checklist/Patient Action Checklist [18] and found that interrater reliabilities were higher for ratings from videotapes. Contrasting this finding, Tracey and Guinee [11] demonstrated significantly lower intrarater reliabilities on the Interpersonal Communication Rating Scale [19] when transcripts were evaluated, whereas reliabilities did not differ between audiotaped and videotaped counseling interactions. Similar results were obtained in a more recent study by Riddle et al. [20] in which they compared evaluations of physician–patient interactions on the Moffitt Accrual Analysis System [21] and the Roter Interaction Analysis System [22] using audiotapes and videotapes. Whereas interrater reliabilities did not differ as a function of the mode of presentation, mean ratings were affected by the mode of presentation such that relational communication behaviours were coded differently. This is not consistent with the research by Weingarten et al. [23] who compared videotape and audiotape assessments of physicians’ patient-centred behaviour in primary care consultations using the patient-centredness scale [24]. They found no significant difference between audio and video assessment of patient-centredness. The latter result was further supported in a study by Dent et al. [13]. Support for differences between ratings of transcripts and audiotapes comes from Harrigan et al. [25], who found that ratings of empathy were higher for audiotaped than for transcribed physician–patient interactions.

The purpose of the present study was to investigate the effects of mode of presentation on ratings of empathic communication. This study addresses two research questions: (1) is the interrater reliability affected by the mode of presentation? (2) does the mode of presentation influence participants' ratings of empathic communication? Based on the review of the literature, the following hypotheses were evaluated:

(1) Given that transcripts carry less information than video- and audio recordings, it is expected that participants in the transcript conditions will vary mostly in their evaluations, that is, interrater reliability will be lowest. Since videotapes closely mirror natural interactions and convey most information with respect to communication variables, it is expected that interrater reliability will be highest for participants rating videotaped interactions.

(2) Given the fact that the results in the literature are heterogeneous we have no directed hypothesis with regard to mean differences between various modes of presentation. Thus, we expect mean ratings of perceived empathic communication to differ as a function of mode of presentation.

2. Methods

2.1. Participants

A total of 100 volunteers (84 females) – mainly undergraduates (97%) from the University of Münster, Germany – participated in the study (mean age = 22.19; S.D. = 5.26; range 18–50). Participants were considered eligible if they were native speakers of German and at least 18-year old. Participants were recruited through notices posted in the department of psychology. Participants were randomly assigned to three groups differing with regard to the level of nonverbal information made available to them. Participants either watched a simulated medical interview (n = 30), listened to an audio-only version of this interview (n = 30), or read a transcribed version of the interview (n = 38). The groups did not differ with respect to any background variables. Participants in the transcript condition were tested in small groups. Participants in the audio- and videotape condition were tested individually in a quiet room. Participants were reimbursed course credit for their time. Written informed consent was obtained from all participants prior to their participation.

2.2. Stimulus materials

The material presented to the participants was drawn from a training video for general practitioners (Merck KGaA). The video sequence (duration: 6 min 20 s) used for the purpose of this study shows a male physician seeking to motivate a male patient to refrain from drinking. The video recording of the simulated encounter was transcribed using standard orthography and eye dialect. The audio-only version of the interview was generated from the videotape.

2.3. Measures

Participants were asked to assess physician’s level of empathy using the Rating Scales for the Assessment of Empathic Communication in Medical Interviews (REM) [26]. REM comprises six items related to empathy (α = .93) and three items related to confrontation (α = .87). Empathy in REM is defined as the physician’s cognitive ability to perceive and understand the patient’s perspective and the behavioural ability to communicate this understanding to the patient [2]. Since there is evidence that practitioners mingle empathic and confrontational behaviour and that the positive effects of empathy are neutralized when confrontational behaviour is not eliminated [9], the confrontational factor highlights the aspect of talking down the patient so that the physician can make his/her own point. Convergent validity was established [26] by correlating the REM with the Motivational Interviewing Treatment Integrity Code [27] and with the Behaviour Change Counseling Index [28]. Participants rate the extent to which a physician displays a specific behaviour on a seven-point Likert scale. The two endpoints are described in specific behavioural terms, such as indicating that the physician showed a lot of interest in the patient’s opinion (seven points) or showed no interest in the patient’s opinion (one point). Items assessing confrontation were turned to give all items the same direction. Thus, a higher value indicates greater empathy and less confrontation.

2.4. Procedure

After participants gave written informed consent, they completed a brief questionnaire on background variables. On completion of the questionnaire, participants read a brief instruction sheet and then were presented with the medical encounter in one of the following formats: (1) videotape, (2) audio-only, or (3) transcript. After examining the medical visit, participants assessed the physician’s level of empathic communication. Each participant completed his or her own questionnaire and no discussion among participants was permitted. At the end of the experiment, participants were debriefed and dismissed. Experimental sessions lasted approximately 15–20 min.

3. Results

3.1. Reliability of ratings across presentation formats

To test our first hypothesis, the intraclass correlation coefficient (ICC) was employed to examine the interrater reliability [29]. The ICC used in this study assumes that the same observers scored each
target. The analysis was based on a two-way random effect model with absolute agreement definition. According to Cicchetti [30], ICC values below .40 are considered poor; values from .40 to .59 are fair; values from .60 to .74 are considered good; and values above .74 are excellent. ICCs were highest when videotaped interactions were rated and ICCs were lowest for transcript ratings (Table 1).

3.2. Mean differences of different presentation formats

To test our second hypothesis, a one-way multivariate analysis of variance (MANOVA) was conducted to determine the effect of mode of presentation on the assessment of the physician’s level of empathic communication. The MANOVA revealed a significant multivariate effect for mode of presentation [Wilks’ $\lambda = .79$, $F(4, 192) = 5.99$, $p < .001$; partial $\eta^2 = .11$]. Subsequent univariate tests revealed a significant effect of mode of presentation on empathy [$F(2, 97) = 11.82$, $p < .001$; partial $\eta^2 = .20$] but not on confrontation [$F(2, 97) = 1.01$, ns]. Post hoc tests were conducted to further explore the main effect of mode of presentation on ratings of empathy. Ratings on empathy differed significantly between transcript ($M = 4.25$; S.D. = 1.37) and audio ratings ($M = 5.25$; S.D. = 1.08, $p < .001$) and between transcript and video ratings ($M = 5.44$; S.D. = 0.72, $p < .001$). The REM items and their means and standard deviations are shown in Table 2.

4. Discussion

The present study examined the effects of mode of presentation on untrained observers’ ratings of a physician’s level of empathic communication. In line with our hypotheses, the results clearly demonstrate that mode of presentation is an important variable in determining the level of empathic communication. Consistent with previous research [11], interrater reliabilities were lowest when empathy was evaluated based on transcripts; whereas interrater reliabilities did not differ between ratings from videotaped and audiotaped interactions (see also [20]). Thus, the results indicate that important cues for evaluating empathy are lost in transcripts, in turn leading to greater variation in ratings. Moreover, empathy was rated differently when transcripts were used such that empathy scores were lower (see also [25]). It is further noteworthy that the audiotape produced similar empathy ratings and similar interrater reliabilities compared with videotapes, although visual cues are lost in audio recordings. These results are in accordance with earlier findings [13,23]. With regard to empathy, the results support previous suggestions that vocal cues contribute more to the assessment of interpersonal sensitivity than visual cues [31,32].

However, the results of this study have also shown that the mode of presentation is not associated with different ratings of empathic communication per se. Ratings of confrontation did not vary significantly as a function of presentation mode. This particular result is not surprising; since the confrontational items include explicit verbal expressions of confrontation (e.g., “Did the physician admonish the patient?”). With regard to confrontation, the verbal content of the message seems to be the predominant source of information. However, it is difficult to explain why interrater reliabilities for both audiotape and transcript ratings were poor, whereas the interrater reliability for videotape ratings was excellent. The full repertoire of nonverbal cues may provide important additional information that will help the rater to understand the meaning of a verbal statement more quickly and easily; particularly when the verbal message is not clear.

Generally, the results of the present study demonstrate that interrater reliabilities were highest for those modes of presentation providing most information. The magnitude of the interrater reliability coefficients may be a function of the number of elements available to the rater (i.e., transcript = verbal cues only, audio = verbal cues and vocal cues, video = verbal cues, vocal cues, and visual cues). Consequently, videotaped interactions involving verbal, visual, and vocal cues led to the highest agreement across raters for both empathy and confrontation, whereas ratings based on transcripts providing only verbal cues resulted in lowest interrater reliabilities. It is important to note that high interrater reliabilities do not necessarily imply high validity. It is possible that a highly reliable measure is not valid at all. In contrast to this, low interrater reliabilities always preclude high validity, because the error variances associated with low reliabilities must be unrelated to variance in the criterion. Since the present study did not include a gold standard measure of empathic communication, the deviations of the mean scores of a mode of presentation from the “true” value cannot be determined. Consequently, from the present study one can only conclude that different modes of presentations lead to different ratings of empathic communication. However, given that videotapes closely mirror natural interactions and convey most information with respect to communication variables, it is reasonable to assume that ratings based on videotapes are not only more reliable, but also more valid than ratings from transcripts.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Intraclass correlation coefficients as a function of presentation mode.</th>
</tr>
</thead>
<tbody>
<tr>
<td>REM subscale</td>
<td>Transcript</td>
</tr>
<tr>
<td>Empathy</td>
<td>.68$^{a}$</td>
</tr>
<tr>
<td>Confrontation</td>
<td>.18$^{a}$</td>
</tr>
</tbody>
</table>

Notes: REM: Rating Scales for the Assessment of Empathic Communication in Medical Interviews. ICC $< .40$ = low; $.40 < ICC $< .60$ = moderate; $.60 < ICC $< .75$ = good; ICC $> .75$ = excellent. Coefficients within a row not sharing the same superscript differ significantly at $p < 0.05$.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Descriptive statistics for the items and subscales of the REM.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
<td>Transcrip, M (S.D.)</td>
</tr>
<tr>
<td>1</td>
<td>Provided the opportunity for the patient to give his/her opinion</td>
</tr>
<tr>
<td>2</td>
<td>Treated the patient as an equal partner</td>
</tr>
<tr>
<td>3</td>
<td>Showed understanding of the patient’s point of view</td>
</tr>
<tr>
<td>4</td>
<td>Tried to put him/herself in the position of the patient</td>
</tr>
<tr>
<td>5</td>
<td>Showed interest in the patient’s opinion</td>
</tr>
<tr>
<td>6</td>
<td>Was responsive to the patient</td>
</tr>
<tr>
<td>Total score empathy subscale</td>
<td>4.25 (1.37)</td>
</tr>
<tr>
<td>7</td>
<td>“Preached”</td>
</tr>
<tr>
<td>8</td>
<td>Admonished the patient</td>
</tr>
<tr>
<td>9</td>
<td>Put the patient under pressure</td>
</tr>
<tr>
<td>Total score confrontation subscale</td>
<td>4.61 (1.29)</td>
</tr>
</tbody>
</table>

Note: REM: Rating Scales for the Assessment of Empathic Communication in Medical Interviews; items 7–9 are reverse scored to give all items the same direction. A higher value on items 7–9 indicates greater empathy and less confrontation.

Please cite this article in press as: Nicolai J, et al. Effects of mode of presentation on ratings of empathic communication in medical interviews. Patient Educ Couns (2009), doi:10.1016/j.pec.2009.10.014
4.1. Limitations

There are several important limitations to this study. First, participants were young, female college students. The stimulus material used in this study depicted only males. However, there are well-known gender differences in empathic communication [33]. For example, female physicians are more likely to actually engage in empathic communication [34] and females have been shown to be more accurate in decoding and transmitting nonverbal and vocal cues [35]. Future studies should systematically vary both the gender of the rater and the gender of the persons being rated. Second, the generalizability of the results may be limited because only one medical visit was used. Thus, it would be important to replicate this study with different medical consultations to determine if there are similar mode of presentation effects. Third, a limitation in terms of ecological validity stems from the fact that our study was based on a simulated physician–patient interaction rather than on a real physician–patient interaction. Fourth, the present study did not include a visual only condition (silent video). In order to disentangle the effects of the different presentation modes it is also necessary to discern nonverbal from vocal cues. Additional research on the assessment of physicians’ responses to simulated scenarios versus assessment of real consultations is needed. Finally, the present study does not offer an answer to the question how various categories of information available to raters affect the accuracy of their ratings. Clearly, future investigations of physician–patient interaction should pay close attention to nonverbal channels of communication. Given these limitations, the current results should provide direction and considerations for future studies, rather than being considered definitive.

5. Conclusions and practice implications

Taken together, the results suggest that ratings of transcripts are not equivalent to ratings of audiotaped or videotaped physician–patient interactions. Given that many important aspects of communication, such as visual and vocal nonverbal cues, are lost in transcripts, one may conclude that ratings of transcribed interactions lack both, reliability and validity. Consequently, our results suggest that transcripts seem to be inadequate means of representing empathic communication. Focusing on the verbal level of communication alone ignores the fact that empathy can be expressed through nonverbal means. Verbal statements cannot be studied and categorized in isolation, without concern for nonverbal factors.

References