PLOS ONE

Variance components of ratings of physician-patient communication: A generalizability theory analysis

Nicole Röttele[1]*, Christian Schlett[2], Mirjam Körner[1], Erik Farin-Glattacker[2], Andrea C. Schöpf-Lazzarino[2], Sebastian Voigt-Radloff[3], Markus A. Wirtz[4]

1 Institute of Medical Psychology and Medical Sociology, Faculty of Medicine, University of Freiburg, Freiburg, Germany, 2 Section of Health Care Research and Rehabilitation Research, Medical Center–University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany, 3 Institute for Evidence in Medicine (for Cochrane Germany Foundation), Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany, 4 Department of Research Methods, Freiburg University of Education, Freiburg, Germany

* nicole.roettele@mps.uni-freiburg.de

Abstract

Background

The ratings of physician-patient communication are an important indicator of the quality of health care delivery and provide guidance for many important decisions in the health care setting and in health research. But there is no gold standard to assess physician-patient communication. Thus, depending on the specific measurement condition, multiple sources of variance may contribute to the total score variance of ratings of physician-patient communication. This may systematically impair the validity of conclusions drawn from rating data.

Objective

To examine the extent to which different measurement conditions and rater perspectives, respectively contribute to the variance of physician-patient communication ratings.

Methods

Variance components of ratings of physician-patient communication gained from 32 general practitioners and 252 patients from 25 family practices in Germany were analyzed using generalizability theory. The communication dimensions "shared decision making", "effective and open communication" and "satisfaction" were considered.

Results

Physician-patient communication ratings most substantially reflect unique rater-perspective and communication dimension combinations (32.7% interaction effect). The ratings also represented unique physician and rater-perspective combinations (16.3% interaction effect). However, physicians' communication behavior and the observed communication dimensions revealed only a low extent of score variance (1% physician effect; 3.7%

communication dimension effect). Approximately half of the variance remained unexplained
(46.2% three-way interaction, confounded with error).

Conclusion

The ratings of physician-patient communication minimally reflect physician communication
skills in general. Instead, these ratings exhibit primarily differences among physicians and
patients in their tendency to perceive shared decision making and effective and open com-
munication and to be satisfied with communication, regardless of the communication behav-
ior of physicians. Rater training and assessing low inferential ratings of physician-patient
communication dimensions should be considered when subjective aspects of rater perspec-
tives are not of interest.

Introduction

Ratings of physician-patient communication (PPC) provide guidance for many important
decisions in health research, health care delivery and health care professional trainings. In the
field of health research, these ratings serve to promote evidence-based medicine. In doing so,
ratings of PPC are used to investigate the effects of PPC on relevant clinical outcomes (e.g.,
patient quality of life, satisfaction with care and health status) [1–3]. The ratings are applied to
assess physicians' communication skills, to reveal needs for improvement and to derive appro-
priate approaches for interventions [1, 2]. Finally, they are used to evaluate such interventions
and to decide whether they are effective and worthy of further implementation [4, 5]. In regard
to health care delivery, these ratings serve as quality criteria for health facilities [6, 7]. In terms
of health care professional training programs, ratings of PPC are incorporated to evaluate the
communication competence of students, to assess learning needs and to justify modifications
of curricula [8]. Accordingly, a complete understanding of PPC ratings is vital to progress and
to defensible decision making in research, education and health care delivery.

However, there is no gold standard, accepted measure across settings which is used to assess
ratings of PPC. For example, the different measures of PPC vary in the dimensions of PPC
they assess (e.g. relationship building, promoting information exchange, facilitating patient
involvement or satisfaction with the medical encounter) [9, 10]. In addition, measures of PPC
also differ in the perspectives from which the ratings are made (e.g., physicians, patients, neu-
tral observers) [11]. Thus, depending on the measure used, the ratings reflect PCC under spe-
cific measurement conditions. However, usually the aim is not to evaluate the quality of PPC
under specific measurement conditions but to draw generalizable conclusions across different
situations and conditions. For example, a neutral observer may assess a physicians' communi-
cation competence using the OPTION scale [12] to evaluate an intervention program aiming
to improve PPC. Then, one may not be interested in physicians' communication competence
how it is perceived by this specific rater on the specific dimension the OPTION scale assesses.
Instead, one may be interested in the physicians' overall communication competence across all
possible measurement conditions (e.g. across all communication dimensions and all possible
raters). Thus, it is important to reveal to what degree the assessed ratings reflect the construct
of interest itself rather than its specific operationalized variant. This can be done by investigat-
ing the amount of score variance of the assessed ratings attributable to the specific measure-
ment conditions. The less score variance is attributable to a specific aspect of the measurement
situation, the more generalizable are the assessed ratings over this specific aspect.

Due to the possible variations in measurement conditions when investigation PPC, there are multiple sources of variance, which may contribute to the total score variance of PPC ratings. First, score variance may be due to the observer perspective [13, 14]. Physicians and patients may have different consistent tendencies to rate PPC quality as either high or low, which is also known as leniency or severity bias [15]. This bias may be specific to particular communication dimensions or particular physicians, leading to interaction effects of rater perspective and communication dimensions or rater perspective and physicians [14]. For example, in the case of the dimension satisfaction with the medical encounter, there is evidence, that patients tend to be more satisfied than their physicians when individual encounters have been rated from both perspectives [16, 17]. Furthermore, in a meta-analysis of six studies investigating shared decision making (SDM) physician and patient ratings proved to be weakly associated, while physicians rated SDM to be higher [18] (r_{polyc} = .15; 95% CI [0.08–0.21]). Two additional studies, consulting neutral observers and patients also indicate that there may be score variance in PPC ratings due to the observed perspective: observers and patients did not agree in their ratings of the same consultation. Thus, the ratings may depend more on the perspective of the assessing person than on the quality of PPC itself [19, 20].

Second, there may be score variance of PPC ratings due to an interaction effect of the analyzed perspective and the assessed communication dimensions. Particularly, physicians' or patients' impressions created in one area may influence their judgements in another area (halo effect). This phenomenon will lead to correlations on the different communication dimensions among ratings from a particular perspective [13]. For example, patient satisfaction with the medical encounter may also lead to higher ratings of information exchange and SDM but may not increase physicians' ratings. Third, the assessed communication dimensions as well as an interaction effect of communication dimensions and investigated physicians may explain the score variance of PPC ratings. For example, ratings on satisfaction exhibit ceiling effects [15, 16]. Hence, observing the communication dimension of satisfaction may indicate a higher quality of PPC than observing other dimensions of communication, such as information exchange or SDM.

Previous studies did not reveal which of these specific characteristics of the measurement situation contribute to the variance of PPC ratings. When investigating the accuracy of PPC ratings, these studies typically focus on one global measurement error [20–22]. In doing so, all variance components accountable to the investigated type of error are incorporated into one single error term. None of these studies examines variability due to physician characteristics, rater perspective, or investigated dimensions of PPC and potential interaction effects between these factors simultaneously.

Generalizability theory (g-theory) allows simultaneous examination of multiple sources of variance [23]. By applying a generalizability study (g-study), it can be revealed how much of the score variance of PPC ratings is due to physicians' communication behavior, raters' perspective, communication dimensions, and interactions. Thus, in the context of PPC, g-theory outperforms previous used methods to investigate generalizability of ratings as it allows to separate the variability of ratings scores in different variance components instead of incorporating them into one global, unresolved error term. This process allows decision makers to enhance their understanding of the information included in the ratings and to draw more valid conclusions concerning the generalizability of PPC ratings across different measurement conditions. Two previous studies applying g-theory in the context of PPC investigated students' communication competence in the field of health care professional training programs [24, 25]. Both studies investigated PPC ratings assessed from observers. G-theory proved to be effective as it allowed to decompose variance components differentially. However, to our knowledge, there is no study applying g-theory to investigate PPC ratings assessed from self-report measures.

In the presented study we investigated variance components of PPC ratings gained from physicians' and patients' perspectives. To represent the multidimensional characteristic of PPC, variance components of PPC ratings were analyzed on three communication dimensions representing two major aspects of PPC and one global indicator of the quality of PPC using g-theory. The observed communication dimensions are SDM, effective and open communication and satisfaction with the overall communication.

Methods

Setting and study population

Cross-sectional survey data measuring the quality of PPC from 48 general practitioners (GPs) and 302 back pain patients belonging to 33 family practices were collected in Northern Bavaria (Germany). Data were assessed from September 2018 to August 2020 as part of the "Well Informed Physicians and Patients (GAP)" trial [26]. The trial was funded by the German Innovation Fund (Federal Joint Committee; Grant: 01NVF17010) and is registered in the German Clinical Trials Register (DRKS00014279). The study was approved by the ethics committee of the Albert-Ludwigs-University Freiburg (No. 559–17). All participants provided written informed consent to participate.

The Institute of General Practice at the University Medical Center Erlangen recruited GPs who recruited patients in their family practices. All GPs in Bavaria were eligible and were contacted in a two-step process: 1. Invitation by postal/fax through the Bavarian Association of General Practitioners (BHÄV). 2. By personal mail or telephone contact of the GPs. Participating GPs received information concerning the study, regulations for enrolling in the intervention program, and participation documents for the GP practice and the patients. Eligible patients received brief study information at their first consultation and were invited to participate in the study. If interest was confirmed, patients were informed and enrolled in the study by physicians. To facilitate recruitment, participating physicians received an incentive of 34€ per enrolled patient, and participating patients received a book voucher for 25€. Eligible patients were aged 18 years and older, were insured with a company health insurance fund and consulted the participating GP due to back pain symptoms. Sufficient German language skills to complete the study questionnaire were also required for study inclusion.

The intervention consisted of an online portal for the GPs and patients containing evidence-based information on treatment options for back pain. Once the GPs were recruited they were randomized into either control or intervention group. The allocation ratio was 2:1 for intervention and control group. GPs in the intervention group used the online portal during the consultations with their patients included in the study. Furthermore, they gave their patients login data for the online portal. Thus, their patients also had access to the online portal after the consultation. GPs in the control group continued to perform their usual consultation routines and their patients did not get access to the online portal.

Physician and patient ratings on the three dimensions measuring the quality of PPC were assessed using online and paper-pencil questionnaires. In particular, for physicians in rural areas and insufficient internet access as well as for patients completing questionnaires at the physician's office, paper-pencil questionnaires were more practical. Physicians were invited to complete an online questionnaire directly after recruiting the last patient and to give an overall rating of the quality of PPC for all their consultations with their enrolled patients. Physicians who did not fill in the online questionnaires after a survey period of four weeks, were sent paper-pencil questionnaires. To fill in the paper-pencil questionnaires, physicians had an additional time of six weeks (to ensure enough time for postal mail). All patients completed the questionnaire directly after the consultation and evaluated the quality of the PPC of this single consultation using paper-pencil questionnaires.

Patient and physician instruments

The three dimensions indicating the quality of PPC were assessed using two dyadic instruments (consisting of a physician and patient version), namely, the SDM questionnaire (SDM-Q) [27–29], the short version of the satisfaction scale of the P.A. INT questionnaire [30–32], and the effective and open communication scale of the KOVA questionnaire for patients [33]. The latter was adapted for physicians given that a physician version is not available. Additionally, instructions and items for the physician's questionnaire were adapted to provide an overall rating for all consultations. To validate these adaptations, the physicians' questionnaire was pretested in a sample of 7 GPs. All items of the questionnaires are presented in the S1 File.

The SDM-Q assesses SDM and consists of 9 items with participants responding on a 6-point Likert scale (ranging from "1" = "completely disagree" to "6" = "completely agree"; e.g. "My doctor asked me which treatment option I prefer" /" I asked my patients which treatment option they prefer"). A raw total score can be calculated by adding up the scores of all items. This score can be standardized into a scale ranging from 0 to 100 with zero indicating the lowest level of perceived SDM and 100 indicating the highest level of perceived SDM. The psychometric properties of the SDM-Q were evaluated in two previous studies in consultations with 324 patients with chronic disease and 29 GPs and other medical specialists [27, 29]. Both the patient version (α = .94) [27] and the physician version (α = .88) [29] showed good internal consistency. The patient version also revealed good construct validity with all items loading on one factor, which accounted for 62.4% of the total variance. Furthermore, a former version of the scale showed moderate correlation with the subdimensions of the Perceived Involvement in Care Scale [27].

The short version of the satisfaction scale of the P.A. INT questionnaire is a two-item scale assessing global satisfaction with PPC in medical consultations, with participants responding on a 5-point Likert scale (ranging from "1" = "completely disagree" to "5" = "completely agree"; e.g. "I am satisfied with the result of the consultation" / "I am satisfied with the result of the consultations"). Higher scores reflect higher satisfaction. A previous study with 474 inpatient rehabilitants and 60 physicians investigated the reliability of the questionnaire. High internal consistency was observed for the physician (α = .96) and patient versions (α = .95) of the satisfaction scale [32].

The scale effective and open communication of the KOVA questionnaire is a 10-item scale with participants responding on a 6-point Likert scale (ranging from "1" = "strongly disagree" to "6" = "strongly agree"; e.g. "Your physician listened carefully when you wanted to say something" / "I listened carefully when my patients wanted to say something"). A raw total score can be calculated by averaging the single scores of all 10 items. This raw total score can be standardized by transformation into a scale ranging from 0 to 100. Higher values indicate more open and effective communication behavior of the physician. The scale showed good internal consistency (α = .90) in a previous study with 703 back pain patients [33].

Statistical methods

A missing data analysis was conducted. Surveys with greater than 30% scale items missing were excluded from analysis [34]. Missing data were imputed using the expectation–maximization algorithm [35, 36]. All scale scores were transformed from 0 to 100. Internal consistency of the patient and physicians' versions of the three scales was assessed by Cronbach's alpha coefficients. As patients were nested within physicians, the intraclass correlation coefficient (ICC) was calculated to investigate patient score variance attributable to physicians [37]. A higher value indicates high consistency among the patient ratings of the same physician, i.e., the

variation of patient scores nested within the same physicians is relatively smaller than the variation between patient scores of different physicians. Three paired-samples t-tests were calculated to detect differences from physicians' and patients' mean scores of the examined communication dimensions. A significance level of $p < .05$ was chosen.

To identify the variance components representing the contribution of the investigated measurement conditions (facets) and their interactions to the total score variance of communication ratings, a g-study was conducted with a crossed two-facet random effects measurement design (P x R x C) [23, 38]. The object of measurement (differentiation facet) was physicians (P). Instrumentation facets included perspectives from which the scores were obtained (R) and communication dimensions for which the ratings were assessed (C). All facets were considered random. As physicians provide an overall rating for communication with all their patients, patient ratings belonging to the same physician were also summarized by calculating means. This process allowed for a crossed design. The g-study was performed with EduG 6e software [39], and all other statistical analyses were performed with SPSS version 26 [40].

Results

A total of 252 patient questionnaires and 32 physician questionnaires from 25 family practices could be included in the analysis. Thirty-nine patient questionnaires and 16 physician questionnaires were excluded because there was no questionnaire from the second perspective. Another 11 patient questionnaires were excluded due to missing data. Table 1 shows the socio-demographic data of the included physicians and patients.

The range of patients nested within physicians was from 1 to 38 (M = 7.88; SD = 7.83). The ICCs of patients' ratings were not significant for SDM and were only small for satisfaction and effective and open communication (ICC = .08). Thus, the same variability is noted in the communication scores from one patient to another patient of the same physician as from one patient of one physician to another patient of another physician.

Table 1. Characteristics of patients and physicians included in the study.

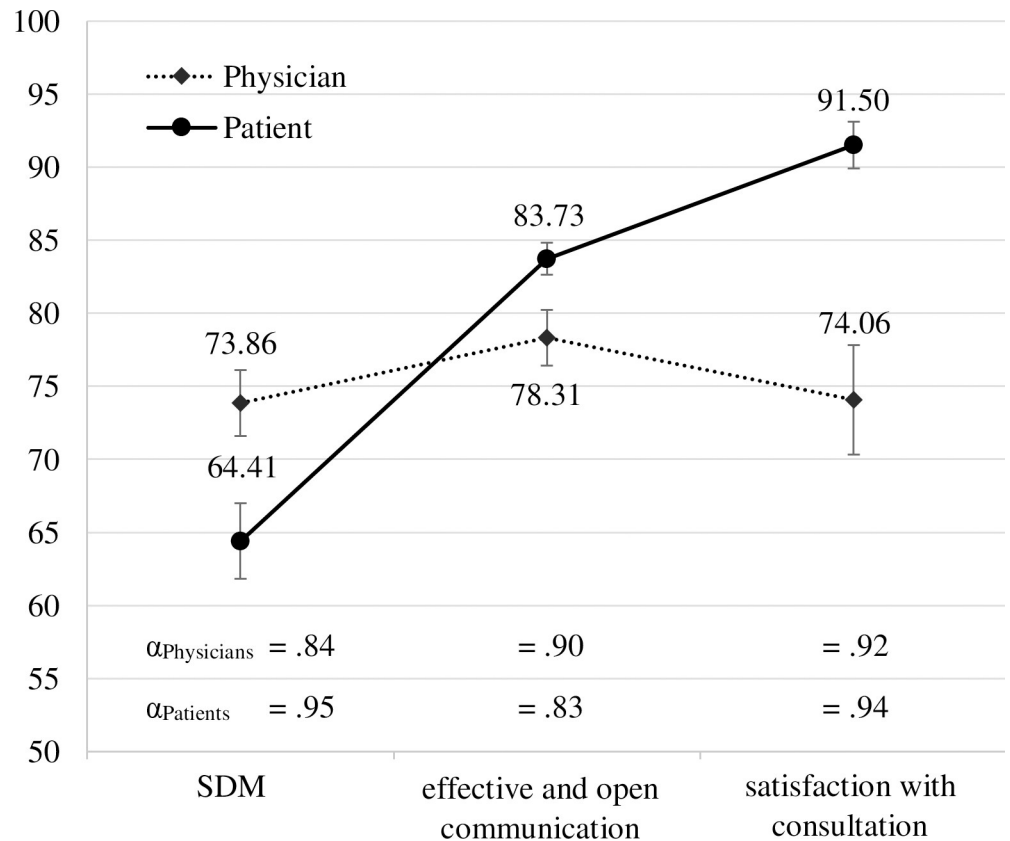|	Patients		Physicians	
	Mean	SD	Mean	SD
Age (years)	47.01	14.29	49.63	9.42
Work experience (years)	N/A	N/A	15.27	10.23
working hours per week	N/A	N/A	43.66	13.61
	N (total = 252)	%	N (total = 32)	%
Gender				
Male	101	40.1	20	62.5
Female	150	59.5	10	31.3
Type of practice	N/A	N/A		
Joint practice			20	62.5
Group practice			2	6.3
Single-partner practice			7	21.9
Other			1	3.1
German mother language	240	95.2	N/A	N/A
Educational level			N/A	N/A
Low	71	28.2		
Medium	98	38.9		
High	72	28.6		

Note. N/A not applicable.

Fig 1. Mean scores, standard errors and Cronbach's α for physician and patient rating scales.

Mean scores, standard errors and Cronbach's α for the ratings from physicians and patients' perspectives of the three communication dimensions are shown in Fig 1. T-test results indicate that physicians' and patients' rating levels differ significantly for all three dimensions, $t(31) = 3.11$, $p = .004$, $d = 0.55$ (SDM); $t(31) = -2.65$, $p = .012$, $d = 0.47$ (effective and open communication); $t(31) = -4.10$, $p < .001$, $d = 0.72$ (satisfaction). While physicians rate SDM to be higher, patients judge effective and open communication as well as satisfaction to be higher. All scales assessed revealed high internal consistency for both perspectives ($\alpha > .83$).

Table 2 provides variance components and G-coefficients for the ratings of PPC. The variance component attributable to physician effects indicates to what extent individual physicians differ in overall PPC. This variance component accounts for 1% of the total score variance.

The variance component for physician and perspective interaction (P x R) reflects variance attributable to rater perspectives and a particular physician. Depending on the particular physician, the congruence of rater perspectives varies systematically. This interaction accounts for approximately 16% of the total score variance. This interaction effect is shown in Fig 2. The difference in the overall PPC ratings between the two perspectives is for 12 physicians within the 95% CI with a mean difference of 0 indicating concordant ratings (diagonal). However, 13 physicians provided lower ratings for their own overall communication quality than their patients, whereas 7 physicians gave higher ratings.

The interaction effect of perspective and communication (R x C) becomes evident in Fig 1: leniency and severity biases of rater perspectives are communication dimension specific. The

Table 2. G-study estimates of variance components and G-coefficients.

Effect	VC[a]	Contribution to variance (%)	
P	2.791	1.0	
R	-25.533	0.0	
C	10.712	3.7	
P x R	46.808	16.3	
P x C	-5.498	0.0	
R x C	93.752	32.7	
P x R x C	132.337	46.2	
G-coefficient relative			0.06
G-coefficient absolute			0.04

Note.

[a] Estimated variance components. P = physician; R = rater-perspective; C = communication dimension; P x R = physician x rater-perspective; P x C = physician x communication dimension; R x C = rater-perspective x communication dimension; P x R x C = physician x rater-perspective x communication dimension.
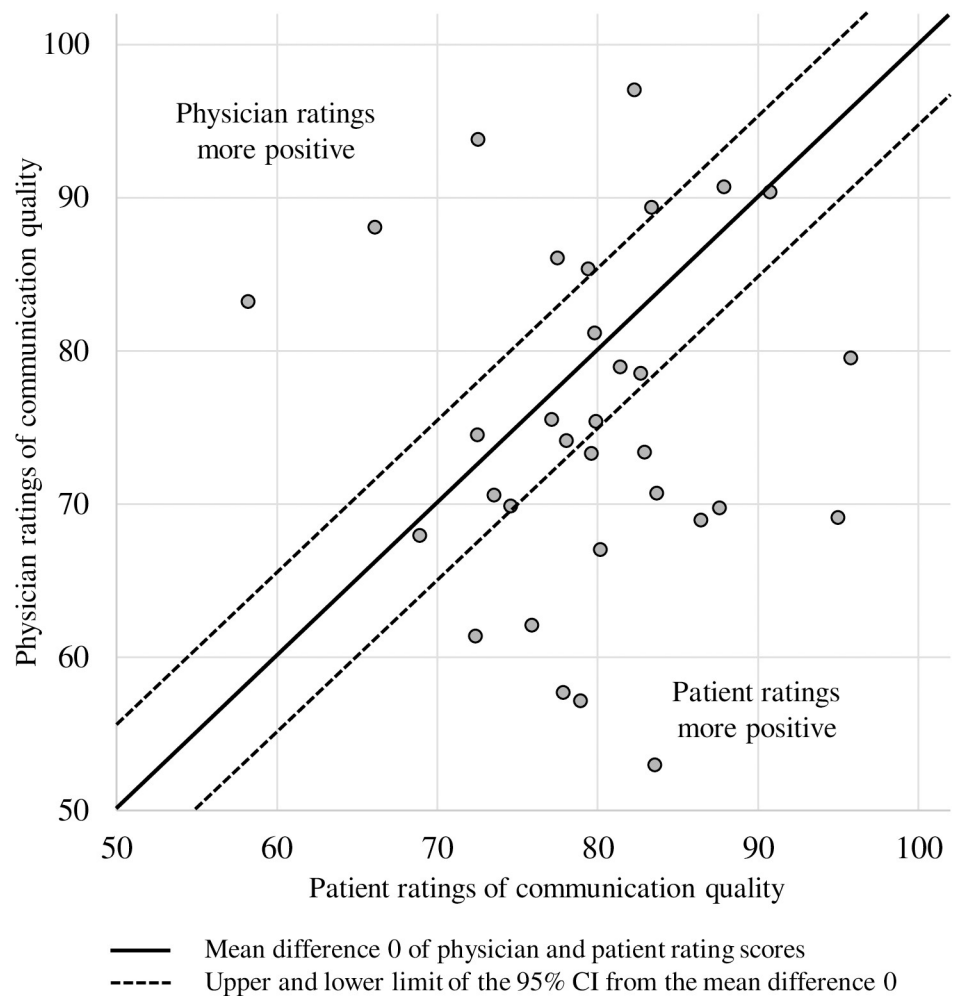
https://doi.org/10.1371/journal.pone.0252968.t002

Fig 2. Interaction effect of perspective facet and physician facet.

https://doi.org/10.1371/journal.pone.0252968.g002

g-study further reveals that this interaction effect explains 33% of the total variance in scores. Variance due to communication dimension indicates to what extent ratings on a specific dimension of PPC differ from ratings of other dimensions of PPC and contributes 4% to score variance. The residual variance component (P x R x C) accounts for approximately 46% of the score variance. This component reflects variance attributable to the three-way interaction and is confounded with error variance. The low g-coefficients indicate that physician and patient ratings have limitations in reliably assessing the overall communication of individual physicians.

Discussion

PPC ratings were minimally able to distinguish among the communication competence of individual physicians. Instead, the ratings more substantially reflect differences among physicians and patients in their tendency to perceive SDM, effective and open communication and to be satisfied with the communication. Thus, patients evaluate communication more effectively and openly and are more satisfied with communication in the consultation than their counterparts. In contrast, physicians perceive that they do better at SDM than their patients perceive. Furthermore, PPC ratings reflect physicians' and patients' leniency and severity bias unique to specific physicians. Thus, some physicians are rated in their communication more positively by their patients than indicated by their self-ratings. However, other physicians are rated more negative by their patients than indicated by their self-ratings. Finally, a high amount of score variance cannot be explained despite investigating the different variance facets P, R and C as well as their interactions.

The findings of this study are consistent with previous studies applying g-theory to examine ratings from neutral observers [24, 25] or in other contexts [14, 41] and are emphasized by studies investigating PPC ratings using other methods [18, 42]. Two studies examined ratings of PCC assessed from neutral observers to evaluate students communication competence in the context of health professional training programs [24, 25]. It was shown that students' communication competence could only be explained to a small extent (3% [24]; 11% [25]). This was also shown by a study examining the cultural competency of attending physicians using g-theory [41]. Physicians' cultural competency could only be explained to a small extent (3%). Thus, the physician effect was also very low. Instead, the ratings also reflected the observer's unique perceptions of specific physicians (47%; interaction effect of observer and physician) [41]. A meta-analysis investigating a broad variety of ratings in the context of psychological research showed similar results [14]. Approximately 37% of the score variance was attributable to raters' unique perceptions of particular targets and to raters' different tendencies to interpret the investigated rating scales. The latter is further emphasized by studies examining inter-rater agreement of physician and patient ratings [18, 42]. They indicate that physicians and patients tend to perceive single communication dimensions differently, whereas physicians rated the presence of SDM higher [43].

The different tendencies of the rater perspectives to rate the single communication dimensions of PPC may be due to several reasons. Different attributions of the meanings of the communication dimensions could provide different ratings from the perspectives [14]. Since patients are less familiar with the language of the items and instructions, their ratings may reflect broader aspects of communication than those from physicians [44]. Especially in the case of SDM ratings, physicians and patients may refer to distinctive content meanings given that patients' ratings do not only reflect the decision-making process. Instead, they may also reflect other aspects inherent in the patient's perspective, i.e., the alternative they focus on, their assessment of constraints on choices, and their evaluation of how good the best possible

solution is [45]. However, physicians may focus on their strengths and overlook their weaknesses when evaluating their own competence to involve patients in decision making. In doing so, they enhance their self-esteem and produce a self-serving bias. Furthermore, the importance of SDM is forced in health professionals' training and in evaluations of health care delivery quality. This phenomenon may lead physicians to reveal higher SDM ratings due to social desirability [46].

However, other sources of disagreement may be relevant regarding ratings of satisfaction with communication in the consultation. As subjective perception may be more important than the evaluation of physician performance, patients may be more satisfied with communication in consultations than physicians due to different expectations concerning medical encounters [47, 48]. Although physicians are educated about what aspects constitute good PPC, they may have expectations concerning medical encounters other than their patients, who may have formed their expectations due to former experiences. Thus, physician satisfaction ratings may more reflect on the degree to which medical standards are met. Patient ratings, however, may reflect to what extent their expectations were met based on former experiences. Another reason for the higher satisfaction ratings of patients could be a social desirability response bias. They may feel that positive comments are more likely to be accepted by survey administrators and therefore report greater satisfaction than they actually feel. Furthermore, being at the family practice when completing the questionnaires may have evoked the desire of patients to integrate themselves with medical staff, leading to more favorable ratings [47]. Finally, a halo effect may produce higher satisfaction ratings from patients, resulting in the contribution of unique impressions to the ratings [14].

The consequences of the high amount of variance in PPC ratings due to rater-perspective interaction effects should be considered separately for every single communication dimension as they underlie different conceptual frameworks of communication [11]. For example, satisfaction ratings are considered to be subjective judgments of an interaction between two individuals. Therefore, the individual perception and experience of the two parties is of interest. Thus, both variance due to rater perspective and variance due to the interaction of rater perspective and physician may reflect valid information. However, regarding ratings of effective and open communication behavior of the physician, communication is considered an observational behavior of one individual. It is not of interest how the behavior was interpreted by the perspectives but how the physicians acted. Hence, variance due to rater perspectives would be a source of error. In the case of SDM, both the observational behavior of the interaction and the perceived involvement could be of interest [49]. Thus, whether rater-perspective variance is a problem in SDM ratings depends on the research question.

To further facilitate the interpretation of PPC ratings, further research is needed. The development and evaluation of methods aiming to improve the reflection of study participants when responding to items of PPC may reduce rater biases, such as self-serving bias and halo effects. Additionally, investigating the decision-making process underlying when giving ratings of PPC would be valuable. Further examination of which characteristics lead to different ratings of the two perspectives for particular physicians will be performed. Furthermore, investigating further variance components that could not be explained in this study may provide a deeper understanding of the ratings. This information may also reveal additional indicators for interventions aiming to improve PPC. Additionally, conducting a g-study may be instructive taking into account not only self-assessments by physicians and patients, but also assessments by neutral observers. Finally, further research is necessary that goes beyond samples in the primary care setting with the indication of back pain.

Limitations

When interpreting the results, several limitations should be mentioned. First, physicians provided an overall rating of all patients included in the study. A study design where physicians give ratings for every single consultation instead of one overall rating may further have explained more score variance. However, as providing single ratings is an immensely greater effort for physicians, gathering overall ratings is a realistic scenario for studies observing physicians' ratings. Additionally, the overall rating of physicians may have produced systematic observer bias. For example, there may have been a halo effect leading one medical encounter with a good PPC to overall high ratings, or a recency effect could have resulted in patients who were treated most recently being more influential in physicians' ratings. Furthermore, this study showed that there is variability in the ratings of patients clustered within each physician. This might be due to individual patient perspectives and preferences [50] as well as due to varied performance of physicians with different patients [51]. However, the current study is not able to explore this due to physicians' generalised ratings of all encounters with their patients included in this study. Finally, five physicians completed a paper pencil questionnaire, whereas others completed an online questionnaire, which may have produced error variance but was not considered in this study.

Conclusion and practice implications

PPC ratings only minimally reflect physicians' communication. Furthermore, PCC ratings mostly reflect the interaction effects of rater perspective and communication dimension as well as the interaction effects of rater perspective and physician. This information has implications for settings where these types of rater variances are considered error sources. Consequently, rater training should be considered to reduce interrater disagreements [14]. In addition, measuring communication dimensions representing clear observational behavior requiring less rater inference would reveal more valid results. Additionally, researchers should provide detailed descriptions of the observed communication dimensions as well as pronounce which perspective was asked. This information will help decision makers and researchers draw conclusions about the quality of PPC. Finally, physicians tend to overestimate the extent of SDM they are practising. Hence, physicians should involve their patients more in SDM, even if they think they are already doing so sufficiently.

Supporting information

S1 File. Items of physician and patient questionnaire.
(PDF)

S2 File. Data set.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Nicole Röttele, Markus A. Wirtz.

Formal analysis: Nicole Röttele.

Investigation: Nicole Röttele, Christian Schlett, Mirjam Körner, Erik Farin-Glattacker, Andrea C. Schöpf-Lazzarino, Markus A. Wirtz.

Methodology: Nicole Röttele.

Project administration: Sebastian Voigt-Radloff.

Supervision: Markus A. Wirtz.

Validation: Nicole Röttele, Christian Schlett, Mirjam Körner, Erik Farin-Glattacker, Andrea C. Schöpf-Lazzarino, Markus A. Wirtz.

Visualization: Nicole Röttele.

Writing – original draft: Nicole Röttele.

Writing – review & editing: Nicole Röttele, Christian Schlett, Mirjam Körner, Erik Farin-Glattacker, Andrea C. Schöpf-Lazzarino, Sebastian Voigt-Radloff, Markus A. Wirtz.

References

1. Riedl D, Schüßler G. The Influence of Doctor-Patient Communication on Health Outcomes: A Systematic Review. Zeitschrift für Psychosomatische Medizin und Psychotherapie. 2017; 63:131–50. https://doi.org/10.13109/zptm.2017.63.2.131 PMID: 28585507

2. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet. 2001; 357:757–62. https://doi.org/10.1016/s0140-6736(00)04169-6 PMID: 11253970

3. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One. 2014; 9:e94207. https://doi.org/10.1371/journal.pone.0094207 PMID: 24718585

4. Légaré F, Adekpedjou R, Stacey D, Turcotte S, Kryworuchko J, Graham ID, et al. Interventions for increasing the use of shared decision making by healthcare professionals. Cochrane database of systematic reviews. 2018; 7:CD006732. https://doi.org/10.1002/14651858.CD006732.pub4 PMID: 30025154

5. Dwamena F, Holmes-Rovner M, Gaulden CM, Jorgenson S, Sadigh G, Sikorskii A, et al. Interventions for providers to promote a patient-centred approach in clinical consultations. Cochrane database of systematic reviews. 2012; 12:CD003267. https://doi.org/10.1002/14651858.CD003267.pub2 PMID: 23235595

6. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med. 2011; 9:100–3. https://doi.org/10.1370/afm.1239 PMID: 21403134

7. National Academies Press (US). Crossing the Quality Chasm: A New Health System for the 21st Century. Washington (DC); 2001.

8. Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R, Buffone N, et al. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. Academic medicine. 2004; 79:495–507. https://doi.org/10.1097/00001888-200406000-00002 PMID: 15165967

9. Epstein RM, Franks P, Fiscella K, Shields CG, Meldrum SC, Kravitz RL, et al. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues. Soc Sci Med. 2005; 61:1516–28. https://doi.org/10.1016/j.socscimed.2005.02.001 PMID: 16005784

10. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. Academic medicine. 2001; 76:390–3. https://doi.org/10.1097/00001888-200104000-00021 PMID: 11299158

11. Street RL. The many "Disguises" of patient-centered communication: Problems of conceptualization and measurement. Patient Education and Counseling. 2017; 100:2131–4. https://doi.org/10.1016/j.pec.2017.05.008 PMID: 28619273

12. Elwyn G, Hutchings H, Edwards A, Rapport F, Wensing M, Cheung W-Y, et al. The OPTION scale: measuring the extent that clinicians involve patients in decision-making tasks. Health Expect. 2005; 8:34–42. https://doi.org/10.1111/j.1369-7625.2004.00311.x PMID: 15713169

13. Hoyt WT. Rater bias in psychological research: When is it a problem and what can we do about it? Psychological Methods. 2000; 5:64–86. https://doi.org/10.1037/1082-989x.5.1.64 PMID: 10937323

14. Hoyt WT, Kerns M-D. Magnitude and Moderators of Bias in Observer Ratings: A Meta-Analysis. Psychological Methods. 1999; 4:403–24. https://doi.org/10.1037/1082-989X.4.4.403

15. Hoyt WT. Bias in participant ratings of psychotherapy process: An initial generalizability study. Journal of Counseling Psychology. 2002; 49:35–46. https://doi.org/10.1037//0022-0167.49.1.35

16. Zandbelt LC, Smets EMA, Oort FJ, Godfried MH, Haes HCJM de. Satisfaction with the outpatient encounter: A comparison of patients' and physicians' views. J Gen Intern Med. 2004; 19:1088–95. https://doi.org/10.1111/j.1525-1497.2004.30420.x PMID: 15566437

17. Hirsch O, Keller H, Albohn-Kühne C, Krones T, Donner-Banzhoff N. Pitfalls in the statistical examination and interpretation of the correspondence between physician and patient satisfaction ratings and their relevance for shared decision making research. BMC Med Res Methodol. 2011; 11:71. https://doi.org/10.1186/1471-2288-11-71 PMID: 21592337

18. Röttele N, Schöpf-Lazzarino AC, Becker S, Körner M, Boeker M, Wirtz MA. Agreement of physician and patient ratings of communication in medical encounters: A systematic review and meta-analysis of interrater agreement. Patient Education and Counseling. 2020; 103:1873–82. https://doi.org/10.1016/j.pec.2020.04.002 PMID: 32376141

19. Martin LR, Jahng KH, Golin CE, DiMatteo MR. Physician facilitation of patient involvement in care: correspondence between patient and observer reports. Behavioral medicine. 2003; 28:159–64. https://doi.org/10.1080/08964280309596054 PMID: 14663923

20. Kasper J, Heesen C, Köpke S, Fulcher G, Geiger F. Patients' and observers' perceptions of involvement differ. Validation study on inter-relating measures for shared decision making. PLoS One. 2011; 6: e26255. https://doi.org/10.1371/journal.pone.0026255 PMID: 22043310

21. Davis DA, Mazmanian PE, Fordis M, van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. Journal of the American Medical Association. 2006; 296:1094–102. https://doi.org/10.1001/jama.296.9.1094 PMID: 16954489

22. Brouwers M, Rasenberg E, van Weel C, Laan R, van Weel-Baumgarten E. Assessing patient-centred communication in teaching: a systematic review of instruments. Med Educ. 2017; 51:1103–17. https://doi.org/10.1111/medu.13375 PMID: 28762538

23. Brennan RL. Generalizability Theory. New York: Springer; 2001.

24. Uzun NB, Aktaş M, Aşiret S, Yorulmaz S. Using Generalizability Theory to Assess the Score Reliability of Communication Skills of Dentistry Students. Asian Journal of Education and Training. 2018; 4:85–90. https://doi.org/10.20448/journal.522.2018.42.85.90

25. Karabilgin OS, Vatansever K, Caliskan SA, Durak Hİ. Assessing medical student competency in communication in the pre-clinical phase: objective structured video exam and SP exam. Patient Education and Counseling. 2012; 87:293–9. https://doi.org/10.1016/j.pec.2011.10.008 PMID: 22169634

26. Voigt-Radloff S, Schöpf AC, Boeker M, Frank L, Farin E, Kaier K, et al. Well informed physician-patient communication in consultations on back pain—study protocol of the cluster randomized GAP trial. BMC Fam Pract. 2019; 20:33. https://doi.org/10.1186/s12875-019-0925-8 PMID: 30803433

27. Scholl I, Kristion L, Härter M. Fragebogen zur Partizipativen Entscheidungsfindung (revidierte 9-Item Fassung). Klinische Diagnostik und Evaluation. 2011; 4(1):46–9.

28. Scholl I, Kriston L, Dirmaier J, Buchholz A, Härter M. PEF-FB-Doc. Fragebogen zur partizipativen Enscheidungsfindung (Arztversion) 2010.

29. Scholl I, Kriston L, Dirmaier J, Buchholz A, Härter M. Development and psychometric properties of the Shared Decision Making Questionnaire—physician version (SDM-Q-Doc). Patient Education and Counseling. 2012; 88:284–90. https://doi.org/10.1016/j.pec.2012.03.005 PMID: 22480628

30. Dibbelt S, Schaidhammer M, Fleischer C, Greitemann B. Patient-doctor interaction in rehabilitation: the relationship between perceived interaction quality and long-term treatment results. Patient Education and Counseling. 2009; 76:328–35. https://doi.org/10.1016/j.pec.2009.07.031 PMID: 19683407

31. Dibbelt S, Schaidhammer M, Fleischer C, Greitemann B. Patient-Arzt-Interaktion in der Rehabilitation: Gibt es einen Zusammenhang zwischen wahrgenommener Interaktionsqualität und langfristigen Behandlungsergebnissen? [Patient-doctor interaction in rehabilitation: is there a relationship between perceived interaction quality and long term treatment results?]. Rehabilitation (Stuttg). 2010; 49:315–25. https://doi.org/10.1055/s-0030-1263119 PMID: 20963673

32. Fleischer C. Arzt-Patienten-Interaktion als Beitrag zur Qualitätssicherung in Reha-Kliniken.: Diplomarbeit im Fachbereich Humanwissenschaften, Fach Psychologie. Osnabrück: Universität Osnabrück 2006.

33. Farin E, Gramm L, Schmidt E. Taking into account patients' communication preferences: instrument development and results in chronic back pain patients. Patient Education and Counseling. 2012; 86:41–8. https://doi.org/10.1016/j.pec.2011.04.012 PMID: 21570795

34. Wirtz M. Über das Problem fehlender Werte: Wie der Einfluss fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann. [On the problem of missing data: How to identify and reduce the impact of missing data on findings of data analysis]. Rehabilitation (Stuttg). 2004; 43:109–15. https://doi.org/10.1055/s-2003-814839 PMID: 15100920

35. Little RJA, Rubin DB. Statistical Analysis with Missing Data. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2002.

36. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods. 2002; 7:147–77. https://doi.org/10.1037/1082-989X.7.2.147 PMID: 12090408

37. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods. 1996; 1:30–46. https://doi.org/10.1037/1082-989X.1.1.30

38. Cronbach LJ, Gleser GC, Nanda H, Rajaratnam N. The Dependability of Behavioral Measurements. Theory of Generalizability for Scores and Profiles. New York: Wiley; 1972.

39. Swiss Society for Research in Education Working Group. EDUG user guide. 2006.

40. IBM Corp. IMB SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp 2019.

41. Lucas T, Lakey B, Arnetz J, Arnetz B. Do ratings of African-American cultural competency reflect characteristics of providers or perceivers? Initial demonstration of a generalizability theory approach. Psychol Health Med. 2010; 15:445–53. https://doi.org/10.1080/13548506.2010.482141 PMID: 20677082

42. Kenny DA, Veldhuijzen W, van der Weijden T, Leblanc A, Lockyer J, Légaré F, et al. Interpersonal perception in the context of doctor-patient relationships: a dyadic analysis of doctor-patient communication. Soc Sci Med. 2010; 70:763–8. https://doi.org/10.1016/j.socscimed.2009.10.065 PMID: 20005618

43. Driever EM, Stiggelbout AM, Brand PLP. Shared decision making: Physicians' preferred role, usual role and their perception of its key components. Patient Education and Counseling. 2020; 103:77–82. https://doi.org/10.1016/j.pec.2019.08.004 PMID: 31431308

44. Chesser A, Reyes J, Woods NK, Williams K, Kraft R. Reliability in patient-centered observations of family physicians. Fam Med. 2013; 45:428–32. PMID: 23743944

45. Entwistle VA, Watt IS, Gilhooly K, Bugge C, Haites N, Walker AE. Assessing patients' participation and quality of decision-making: insights from a study of routine practice in diverse settings. Patient Education and Counseling. 2004; 55:105–13. https://doi.org/10.1016/j.pec.2003.08.005 PMID: 15476997

46. Worthington RL, Mobley M, Franks RP, Tan JA. Multicultural counseling competencies: Verbal content, counselor attributions, and social desirability. Journal of Counseling Psychology. 2000; 47:460–8. https://doi.org/10.1037/0022-0167.47.4.460

47. Sitzia J, Wood N. Patient satisfaction: A review of issues and concepts. Soc Sci Med. 1997; 45:1829–43. https://doi.org/10.1016/s0277-9536(97)00128-7 PMID: 9447632

48. Ahmed F, Burt J, Roland M. Measuring patient experience: concepts and methods. Patient. 2014; 7:235–41. https://doi.org/10.1007/s40271-014-0060-5 PMID: 24831941

49. Street RL, Mazor KM. Clinician-patient communication measures: Drilling down into assumptions, approaches, and analyses. Patient Education and Counseling. 2017; 100:1612–8. https://doi.org/10.1016/j.pec.2017.03.021 PMID: 28359660

50. Lindsay SE, Alokozai A, Eppler SL, Fox P, Curtin C, Gardner M, et al. Patient Preferences for Shared Decision Making: Not All Decisions Should Be Shared. J Am Acad Orthop Surg. 2020; 28:419–26. https://doi.org/10.5435/JAAOS-D-19-00146 PMID: 31567900

51. Aelbrecht K, Hanssens L, Detollenaere J, Willems S, Deveugele M, Pype P. Determinants of physician-patient communication: The role of language, education and ethnicity. Patient Education and Counseling. 2019; 102:776–81. https://doi.org/10.1016/j.pec.2018.11.006 PMID: 30527850