

How do trained raters take context factors into account when assessing GP trainee communication performance? An exploratory, qualitative study

Geurt Essers · Patrick Dielissen · Chris van Weel · Cees van der Vleuten · Sandra van Dulmen · Anneke Kramer

Received: 4 November 2013 / Accepted: 5 May 2014 / Published online: 24 May 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Communication assessment in real-life consultations is a complex task. Generic assessment instruments help but may also have disadvantages. The generic nature of the skills being assessed does not provide indications for context-specific behaviour required in practice situations; context influences are mostly taken into account implicitly. Our research questions are: 1. What factors do trained raters observe when rating workplace communication? 2. How do they take context factors into account when rating communication performance with a generic rating instrument? Nineteen general practitioners (GPs), trained in communication assessment with a generic rating instrument (the

G. Essers (✉)

Department of Public Health and Primary Care, Leiden University Medical Centre, Hippocratespad 21,
2333 ZP Leiden, The Netherlands
e-mail: G.T.J.M.Essers@lumc.nl

G. Essers · P. Dielissen · C. van Weel · C. van der Vleuten · S. van Dulmen · A. Kramer
Department of Primary and Community Care, Radboud University Medical Centre, Nijmegen,
The Netherlands

C. van Weel
Australian Primary Health Care Research Institute, Australian National University, Canberra, Australia

C. van der Vleuten
Department of Educational Development and Research, Maastricht University, Maastricht,
The Netherlands

C. van der Vleuten
Northumbria University, Newcastle upon Tyne, UK

C. van der Vleuten
Copenhagen University, Copenhagen, Denmark

S. van Dulmen
NIVEL (Netherlands Institute for Health Services Research), Utrecht, The Netherlands

S. van Dulmen
Department of Health Science, Buskerud University College, Drammen, Norway

MAAS-Global), participated in a think-aloud protocol reflecting concurrent thought processes while assessing videotaped real-life consultations. They were subsequently interviewed to answer questions explicitly asking them to comment on the influence of predefined contextual factors on the assessment process. Results from both data sources were analysed. We used a grounded theory approach to untangle the influence of context factors on GP communication and on communication assessment. Both from the think-aloud procedure and from the interviews we identified various context factors influencing communication, which were categorised into doctor-related (17), patient-related (13), consultation-related (18), and education-related factors (18). Participants had different views and practices on how to incorporate context factors into the GP(-trainee) communication assessment. Raters acknowledge that context factors may affect communication in GP consultations, but struggle with how to take contextual influences into account when assessing communication performance in an educational context. To assess practice situations, raters need extra guidance on how to handle specific contextual factors.

Keywords Communication · Context · Assessment · General practice

Background

Assessment of clinical performance in real-life, authentic consultations, i.e. judging ‘good clinical practice’, has been proven to be a complex task (Crossley et al. 2011; Tavares and Eva 2013; Williams et al. 2003; Yeates et al. 2012). Raters in workplace-based assessment are not merely passive instruments but active agents who may take—more or less explicitly—various factors into account when rating clinical performance, including context (Durning et al. 2011; Govaerts et al. 2011; Kogan et al. 2011; McGill et al. 2011; Pelgrim 2013; Yeates et al. 2013). Contextual influences in the doctor-patient encounters are not only apparent in clinical performance as a whole. In recent years, several researchers have argued that context factors also have a specific impact on doctor-patient communication in health care (Bensing et al. 2003; de Haes and Bensing 2009; Eva 2003; Hulsman 2009). In daily practice consultations, it is therefore impossible to say beforehand, what “good communication” is, as contextual influences will be different in each encounter (Winefield et al. 1995; Zoppi and Epstein 2002).

In many studies on communication assessment, generic communication skills instruments are used and performance is largely expressed in terms of those generic skills only (Brown and Bylund 2008; Chan et al. 2003; Hobma et al. 2004; Kramer et al. 2004; Ram et al. 1999; Reinders et al. 2011; Salmon et al. 2011). Assessment of communication performance of GP trainees in daily practice consultations in generic terms suggests a suboptimal increase in performance on communication skills over the course of training (Kramer et al. 2004), although other studies show that some improvement can be found (van Es et al. 2013; Reinders et al. 2011). In all studies, a generic instrument was used to assess communication performance (Hobma et al. 2004; Ram et al. 1999). In order to obtain a fair and informative assessment, context should be incorporated into the assessment (Hulsman 2009). The generic nature of the skills listed therein, however, does not provide indications for the context-specific behaviour that is required in various practice situations (Eva 2003; Salmon and Young 2011). This may present a problem in assessment

as the absence of criteria makes it difficult to value and incorporate the specific context in which the communication takes place.

In previous exploratory studies, we identified several context factors in general practitioners' (GP) communication in daily practice (Essers et al. 2011). In these studies, context factors were identified as observable factors relating to either the doctor, the patient, or the consultation. Subsequently, we found that there is a relevant and meaningful influence on communication performance scores if context factors are explicitly incorporated in the assessment of communication in the workplace (Essers et al. 2013a, b). These studies made us wonder how raters, who regularly assess GP trainees' communication in videotaped, daily workplace consultations, take context factors into account when assessing GP trainees' communication performance. Therefore, our research questions in this study were: (1) What factors do GPs, working as trained raters in the GP Specialty training, observe when rating workplace communication? and, (2) How do raters take context factors into account when rating communication performance using a generic rating instrument? In this exploratory study, a qualitative methodology was used to generate rich information on the inner processes of the raters as spontaneously as possible.

Methods

Participants and procedure

In the Dutch General Practitioner (GP) Specialty Training programme, GP trainee communication performance is assessed several times during the 3 years of training on the basis of videotaped, real-life consultations (Ram et al. 2011). Trained GPs involved in such communication assessment in the GP Specialty training programme of the Radboud University Medical Centre and the Leiden University Medical Centre were invited to participate in the study. All these GP teachers had received training in performance assessment with the MAAS-Global (Ram et al. 2003). We obtained e-mail addresses through mediation of the first and third author (GE and PD). Potential participants ($n = 40$) received an e-mail informing them about the study and asking whether they would be interested in participating (September 2012). Subsequently, if they reacted positively, raters were invited to participate (October–November 2012). We aimed to interview a maximum of 20 and a minimum of 12 GP teachers, as saturation can be expected after 12–15 interviews (Guest et al. 2006). The invitations were sent to each rater individually and up to 2 reminders were sent. Informed consent was obtained prior to participation.

Study design

In this exploratory study, a qualitative methodology was used. Data were gathered from two sources. Firstly, the GP teachers were first asked to participate in a think-aloud protocol in order to reveal their thought processes while assessing videotaped real-life consultations. Two real-life videotaped GP trainee–patient consultations, in which the trainee provided counseling to real patients in general practice, were selected that depicted several context factors previously identified as potentially impacting communication (Essers et al. 2011). Both consultations (of approximately 8 and 13 min in length) were taken from a database of 467 videotaped consultations, which were recorded as part of a video-observation study performed by AMC Amsterdam (van Es et al. 2013). The 2 consultations contained different contextual conditions often seen in general practice.

Consultation 1 (common cold in a young child) included presence of both children and parents in the encounter as well as a language barrier. Consultation 2 (follow-up consultation) included the issues of possible urgency, multiple patient complaints, patient being knowledgeable about physical examination and specific patient behaviour: trivialising fear. Each participant assessed the same two videos, but we alternated the order in which the consultations were seen to prevent an order effect. The participants observed and assessed the videotapes individually and did not have access to any other clinical information (e.g. written medical records) about the patients. The only information they received was that the GP trainees were at 9 months in their first of 3 years of training and at that time working in a general practice. Participants were asked to think-aloud while observing and assessing the GP trainee performance.

Secondly, immediately following the think-aloud protocol, participants were subjected to a semi-structured interview, in which we explicitly asked them to comment on the influence of context factors on assessment. In this interview, raters were questioned on the thoughts they had concerning the context factors that were present in the consultation. Subsequently, they answered an open ended question on how they would consider these context factors to influence their assessment of GP trainee communication (see “[Appendix 1: Think-aloud protocol and interview questions](#)” section). All sessions (think-aloud and interview) were recorded on audiotape, transcribed verbatim and anonymised.

A member check was done by sending a written summary of the interview to the participants. The document verbally summarised the assessment of both consultations as well as the statements the interviewee had made regarding context factors he or she took into consideration in assessing the GP trainee communication.

Communication assessment instrument

The MAAS-Global was used as the instrument to assess GP trainee communication 1991 (Kramer et al. 2004; Ram et al. 2003; van Thiel et al. 1991). The MAAS-Global consists of 13 generic communication items that can be rated from 0 (‘absent’) to 6 (‘excellent’). Each item has three or four sub-items referring to criterion behaviour (see “[Appendix 2: MAAS-Global rating list for doctor-patient communication skills](#)” section). The raters were instructed to assess the trainee performance as they normally would and to think-aloud while rating the GP trainee communication performance.

Data analysis

Because of the explorative nature of this study we performed a qualitative content analysis using the constant comparative method (Glaser 1965). Atlas-ti 7.1 software was used to process the data. As a first step, in order to achieve data reduction, three researchers (GE, PD, and AK) independently read and coded the first three transcripts. In this interpretive process the raw data are broken down into larger chunks by assigning conceptual labels to text fragments from the interview. In Atlas-ti, the text fragments were labelled to indicate their source (think-aloud or interview questions). Following the initial coding of these three interviews, the three investigators met to compare and discuss the codes, resolve differences and develop a code book to be used for further analysis. With the resulting codebook, they independently coded a fourth interview, in which new codes could be added or codes altered. In the subsequent comparison, no more new codes came forward, so we assumed that saturation had been reached. Based on the discussion, the codebook was fine-tuned and the remaining interviews were coded by PD and GE.

Four researchers (AK, SD, PD and GE) studied the coded text fragments to look for emergent themes. Codes referring to the same phenomenon were grouped into categories and categories were grouped into themes. Next to the open coding we used the previously described context factors as codes within the theme ‘Context factors’ (Essers et al. 2011). For this paper, we primarily focused on codes and text fragments relating to context factors.

Ethical regulations

The study was performed according to Dutch privacy legislation. The privacy regulation was approved by the Dutch Data Protection Authority. All participating GP registrars and patients signed an informed consent form before the recording of the consultation. All participating raters were informed beforehand on the procedure and the subject of the study, and signed an informed consent form before the recording of the think-aloud interview. Approval by the Dutch medical education ethics board of NVMO was given for this observational study (NVMO-ERB 173).

Results

Characteristics of participating GP raters

A total of 19 GP communication raters (10 from Leiden University Medical Centre, 9 from Radboud University Medical Centre) participated in this study (7 male, 12 female). Mean number of years experience as a GP was 17.4 (SD = 11.7) ranging from <1 to 38 years. On average, they had 6.2 years of experience in communication assessment (SD 5.4, range 1–21 years). All participants reacted to the member check and found the summarisations adequately representing their thoughts. We included all interviews in our analysis, while saturation was reached after 15 interviews.

What context factors are perceived to influence communication rating?

Both from the think-aloud procedure and from the interviews we found that raters perceive various context factors to be present in the consultations. During the think-aloud procedure, based on observing two consultations, 17 raters spoke of context factors changing the communication between GP and patient. These factors stemming from the context of the consultation could be categorised into doctor-related, patient-related and consultation-related factors. Nine raters named doctor-related factors, 7 named patient-related factors and 16 named consultation-related factors. For example, a statement like “Well, it’s part of what going on here: these people [GP trainee and patient] obviously have seen each other before, so this is part of their way of communicating” (THAL—2243) was considered a doctor-related context factor “Doctor knows the patient”. Moreover, 15 raters named education-related factors affecting their rating the GP trainee-patient communication. For instance, the text fragment “If it’s a bad or borderline performance, I observe the video again. I’ll put it aside for a while and then look at it again to see if it’s really that bad. Because I feel sorry for the trainee.” (THAL—1145) was labelled as the education-related factor “Borderline performance”.

In the subsequent interview, all raters made statements about context factors influencing communication and communication rating, in part concerning the context of the consultation (17 on doctor-related, 13 on patient-related, 18 on consultation-related factors) and in part referring to the context of rating (18 on education-related factors). In total, 27 different factors were named to influence communication rating. The raters named fewer context factors spontaneously in the think-aloud process than in the subsequent interview. In the latter, they were specifically confronted with the questions on context influences. Quotations on context factors are presented in Table 1 (see Table 1).

How do raters take context factors into account when rating communication performance?

The participants had different views how to incorporate context factors into the assessment of the GP (-trainee) communication performance. All raters strived to be careful in assessing the GP trainee but, in determining a mark, 14 raters said that context factors make a difference, whereas 4 raters explicitly stated that for them it makes no difference. If, on the other hand, information on the context of the consultation is lacking or if the structure in the consultation is flawed, e.g. due to the presentation of two or more problems, many raters felt that assessment became more difficult and strenuous.

In situations where there is tension in applying the instructions of the Manual, this was for the most part resolved by giving narrative feedback to the GP trainee in terms of “I scored low, but you did well”. Some raters tried to compensate a possible low item score by heightening other item scores or by looking for mitigating factors in the situation, in order to do justice to the GP trainee: “If the consultation is complex, I think I would tend to grade the trainee higher.” (THAL 1130). Other raters feared that their assessment will lack all objectivity if context or circumstances were taken into account. They expressed their wariness to diverge from standard procedures and guidelines, because, in their view, it might lead to arbitrariness in assessment: “In all fairness, however, I think that the MAAS-Global will lose its value if you start taking all kinds of circumstances into account” (INT 2237). Some, however, seemed to come to a standstill in the conflict; they do not finish their sentence, or give a vague answer, or evade the issue:

- “Ah, well, yes, what does it matter, with such a simple complaint? How important should I make it...?”—THAL 2231;
- “If you don’t hear things, you can’t score them. However, in the context of a consultation it could be logical that it’s not there, but if you want to be objective, then... maybe you should rate it ‘not sufficient’, although that would not be fair”—INT 1145.

Only a few raters were clear on the way they would like to handle or amend on the scoring rules:

- “In this case, because it’s quite an easy problem, I’ll give her a 2 instead of a 1, as there is no real necessity to inquire any further.”—THAL 2248.
- “If in a consultation the diagnosis is known, I would like to be able to rate that item ‘not applicable’ instead of ‘zero’”—INT 2237.

Lastly, a combination of context factors might affect the way that communication is rated. Especially in the case of a borderline performance, other context factors may influence the rating, such as the complexity of the consultation, the doctor not being a native speaker, or the presentation of multiple complaints: “There, in the end I would mark

Table 1 Quotations on context factors affecting communication from think-aloud (THAL) and interview (INT)

Doctor-related factors	
Doctor knows the patient and his social context	Ah yes, maybe this goes with the territory; these people obviously know each other and they have seen each other more often (2243-THAL) In general, the GP will explore less profoundly than a trainee, because he thinks he already knows the patients' context (1145-INT) I can imagine you're inclined to skip some things, because you know patients already for a long time and know a few things of them, so you can almost guess what they are thinking—at least that's what I do.(2237-INT)
Doctor knows patient's medical history	Well, it's the same thing, knowing it or, if you don't know, explicitly checking the patient's medical history. (2235-THAL) For instance, a woman who has had breast cancer presents with pain in her side. Patient and GP know each other very well through the breast cancer period, then the GP might say "You think it has to do with the breast cancer?". And he may not go deeply into emotions, because they have gone through that before. (2248-INT)
Doctor knows patient's way of communicating	Suppose it is a repeat or check-up for something, then maybe it [the consultation] is not structured explicitly, because patient and doctor are both already acquainted with its structure. (2244-THAL)
Doctor is experienced	I would be more surprised to see such hesitant behaviour in an experienced GP (2244-THAL) So in the onset I ask general, open questions, where people have to think about. That shortens my consultations. I think that many more experienced GPs do it like that. (2248-INT)
Doctor's cultural and language background	If a person is not aware of the fine nuances of the language, like when you're a native speaker, then it is more difficult to catch all verbal cues, so they are missed more easily. (1130-THAL) Many foreign doctors that may have had a career as a surgeon in their home country and have come as a refugee and had a Dutch doctor's degree here, but are used to a totally different culture (2231-INT) You may have a terrible accent and be able to communicate very well, but the other way around is also possible (1141-INT)
Patient-related factors	
Specific (non-)verbal behaviour	The mother does not speak any Dutch, so the man has to translate it all. (2246-THAL) With patients who are very long-winded, you have to keep their attention focused, don't ask too many open questions and structure your consultations very strictly. (1130-INT)
Problem is known to both patient and doctor	That will be in follow-up consultations then. Oh yes, I can imagine that has an influence. (1145-INT) In those cases, you will go back to that and use part of your information from it. (2235-INT)

Table 1 continued

Patient is familiar with physical examination (PE)	There's not much explanation necessary here (2248-THAL) Yes, that has an impact, because you see that the patient can do with less information. (1145-INT) Many people, certainly when they're a little older, have had their blood pressure taken a few times in their lives. It would be redundant to go and explain what you're about to do. (2246-INT)
Consultation-related factors	
Follow-up consultation	It is a follow-up consultation, so that maybe makes asking for other reasons to attend not very logical, although it would have been good too if he had made an inventory. (1145-THAL) So, I think he could have mentioned what had previously happened, but that is not always done. (2243-INT) Yes, in the sense that you would expect that in a follow-up consultation they have had prior contact and know what was agreed upon. So, I think, building up the contact will go more swiftly. (2235-INT) Well, also in a follow-up consultation, the patient may have new questions or another request for help. So, there's still the duty to explore that. (2237-INT)
Follow-up consultation based on protocol (initiative by doctor)	Well, you could say that the doctor's request should be made very clear in such a consultation. (1130-INT) So, in that case, someone comes on your request, so you know why the patient is there. However, if you want to do it perfectly, you could say "I've asked you to come here, but do you have a question of your own?". (2246-INT)
Consultation in preventive care (initiative doctor)	No, I wouldn't explore a request for help! A protocolised consultation is meant for what it's meant for. I could still ask if the patient has questions regarding the subject, but other questions I'd refer to another time. (2237-INT) I think, in that context, the consultation is built up in a totally different way (2235-INT)
Problem is easily solved	Because it's such a simple problem, there's no obvious need to explore this further (2248-THAL) I can imagine that, in very easily solved problems, you won't go into emotions deeply. (2237-INT) The simpler a complaint is, the more you can leave the management choice up to the patient. (1133-INT)
Problem urgently needs medical care	This makes all my alarms go off: this man has had something weird (1130-THAL) Oh, yes, sure! In that case, there won't be much exploring emotions, summarising, information giving, request for help, or maybe not at all. (1145-INT)
Diagnosed problem is mainly psychosocial	You'll need to look at the nature of the consultation. It is quite a different thing if someone presents with pain in his ear or with a burn-out. In that case, you'll talk more about emotions and such (2243-INT) With psychosocial problems, it is of the utmost importance that you know what the patient expects of you (2248-INT)

Table 1 continued

More than one person (patient) is present	This is good: she absolutely shows to have attention for both the parent and the child (1141-THAL) What is important, is to explore the ideas of both parents and check if they feel the same way or have different questions. (1144-INT)
More than one problem is presented	If more than one problem is presented, I call it a complex consultation. Often you'll see that one problem is worked out well, but the other is somewhat disregarded (1136-THAL) Well, if someone presents with various problems, that calls for more skills of the GP. (2232-INT)
Presence of children	We subject children to all kinds of frightening things. Children tend to yell and defend themselves, adults have learned to keep silent and accept what is going on. (...) Trying to prepare children is okay, but especially with small children that won't help a lot, for at the moment they will react emotionally anyway (2248-THAL) I tend to take it into account when children are present, for it always brings more turmoil (2243-INT)
Rare problem or question is presented	If you suspect something you're not acquainted with, that's like playing chess on two boards: keeping contact with the patient and at the same time searching your memory 'what can this be?'. (2231-INT) If a patient comes and asks you to keep silent on her mother's diagnosis, that confronts you with a lot of complications. (2248-INT) I once saw a consultation with a deaf patient. The trainee did wonderfully well, with gestures and non-verbal communicating, and they really came to the core of the problem. There was hardly any verbal communication. (1136-INT)
Education-related factors	
Trainee level	Certainly in their first year of traineeship, this is a complex consultation, although you cannot take that into consideration (1136-THAL) On the other hand, I can rate a first year trainee's performance negatively, because I feel that it may be good to keep them awake and alert (2237-INT)
Borderline performance	I marked the score there, because I want to look at it again, and also look into the manual to see how the criteria are exactly formulated (1149-THAL) In the case of a borderline performance, I picture the consultation in my mind once more and pose myself the question "Do I find this good enough?" (...) So, if it's essential for the score, the trainee has the right that I take an extra close look at the things that were scored low (1130-INT)
Awareness of risk of inducing personal bias	I recognize my own style, that's why I label it positively. That's not altogether objective, but it's the background (2235-THAL) Well, I didn't like the doctor that much, he annoyed me a little, so that may have influenced my assessment. On the other hand, I have rated a video of a doctor that I did like, but whose performance was not good, as insufficient. But I can imagine that it has an influence on the assessment if you like a person or not (1149-INT)

Table 1 continued

High or low stake assessment	Altogether, such an [high stake] assessment has an impact on trainees. So, if, in some way, I have the feeling that things will turn out good, although the current performance is insufficient, it makes it more difficult to score it as such (2249-THAL) The other thing is knowing what depends on it. If it is low stake and I already tend to score it as insufficient, then I actually will rate it that way, because, hopefully, the feedback will stimulate the trainee to a better performance (2243-INT)
Correctness of trainee's reflections	Then, firstly, I would look up what she writes on it in her reflections. If she is not aware that this is not good, then that would trouble me greatly. I would feed that back to her and say "this does not suffice, what you did here" (1136-THAL) Sometimes, their own reflections on the consultation already indicate "this is going to be all right" (2232-INT)
Incorrect or incomplete diagnosis or management	If I think 'to my mind, from a medical perspective, this is not a good consultation and I don't agree with what he's doing', I find it sometimes very hard to assess the communication rightly. (2243-INT)

Numbers refer to interviewees

it 'sufficient' mainly because, eventually, he does have a diagnosis and management proposal for the arm" (THAL-2235).

Discussion

The results of this study reconfirm the influence of context on clinical performance assessment (Durning et al. 2011; Govaerts et al. 2011; Kogan et al. 2011). Raters named doctor-related, patient-related, consultation-related factors, stemming from the context of the consultation, and education-related factors, referring to the context of rating, to influence GP-patient communication assessment. These findings confirm the validity of previously found context factors for communication (Essers et al. 2011). For the greater part, they resemble those context factors and can also be classified into the same categories. It seems that the specific questions in the interview triggered considerations that stayed below the consciousness threshold in the think-aloud process. Thinking aloud during an assessment of a real-life workplace consultation may have taken up too much cognitive load to leave room for reflection at the same time (Tavares and Eva 2013).

In the current study, four other factors in the context of the consultation were named that can influence communication: the cultural or language background of the doctor, the presence of children, more than one problem presented by the patient, and the presentation of a rare and complicated medical problem. The cultural and language background of the doctor was not identified previously, but it aligns with studies that found that communication performance is related to cultural background and health care setting (Claramita et al. 2011, 2013). Moreover, GP trainees with a different cultural or language background were found to have more difficulties in communication tasks (Malau-Aduli 2011), whereas native speakers perform better in communication skills (van Es 2012). The presence of children and the presentation of more than one complaint were also found to be relevant context factors in other studies (Crossley and Davies 2005). For these factors to be treated in the same way in communication

assessment as the previously identified context factors—in the sense that they may lead to “justifiably absent (sub-)items on the MAAS-Global”, more research is needed (Essers et al. 2011). These findings show that our previous findings may not have been exhaustive and that, even within the limited setting of a GP-patient consultation, more relevant context factors for communication may be identified. It also indicates that context-specificity is a truly multi-faceted concept (Eva 2003).

The think-aloud procedure and interviews also generated rich information on the process of how raters come to a conclusion regarding the rating of communication performance. First of all, raters evidently put much effort into making a careful and balanced judgment of the GP trainee. They all seem well aware of the impact and value of their assessments to GP trainees and they try hard to do their job well (Tavares and Eva 2013). Secondly, many raters expressed a wish to see more consultations of the GP trainee in order to validate their judgment. This is consistent with findings that raters lack confidence in their judgment if they only see a trainee on one occasion (Tweed and Ingham 2010). Thirdly, especially in borderline performances, some raters explicitly named the complexity of the consultation to mitigate their judgment. This seems to support previous findings that raters tend to assign marks just above the pass mark threshold (Tweed and Ingham 2010).

Regarding our second research question how raters take context factors into account when rating communication performance, we found conflicting results. All raters recognised and acknowledged the importance of context, and could identify several context factors in the consultation influencing GP-patient communication, but they had different views with regard to taking this context into account in communication assessment. When questioned upon the topic, many appear to get caught in a conflict between their perceptions of the practice situations (the ‘real world’) versus the guidelines in the MAAS-Global Manual. In the Manual, however, raters are given the freedom to interpret the quality of the observed performance against the criteria for ‘excellent’ (Ram et al. 2003). Furthermore, the Manual states that different contexts may need additional instructions (Ram et al. 2003). The different ways that the raters handle context factors may well account for differences in the way the GP trainees are rated. In the training programmes in which the interviewed raters participate (Leiden and Nijmegen), the MAAS-Global is used in a summative way for first year GP trainees. Therefore, we would argue that these workplace-based assessments be used primarily as formative assessments in educational contexts and would caution against its use as summative. Secondly, considering the way that the raters are able to express the relationship between context and communication, the narrative feedback on the performance deserves a more prominent place in education than the ratings. The content and richness of the narratives is much more important for learning than a pass or fail. This corroborates recent findings from other studies on assessment and feedback in workplace-based learning (Pelgrim and Kramer 2013; Pelgrim et al. 2012). It is also consistent with current views on assessment programmes in medical education (van der Vleuten and Schuwirth 2005; van der Vleuten et al. 2012). Thirdly, it is necessary to address the relationship between context and communication in rater training. In our view, differences in how individual raters handle context factors can lead to different assessment outcomes (Tavares and Eva 2013; Yeates et al. 2013). Explicating verifiable context factors will help raters make a valid, non-depriving assessment of GP trainee communication, for it will fill up the current void in the MAAS-Global Manual that is now implicitly filled in.

Working as a rater within a medical education context, the interviewees also named several education-related factors that influence their assessments. These factors refer to

trainee education, in which communication performance assessment is an integral part of the training programme. In other research, factors similar to the education-related factors influencing raters of clinical performance have been brought forward, such as being aware of their personal bias when judging a GP trainee, and making a difference for the level of trainees (Yeates et al. 2011). Crossley et al., for instance, found that the existing methods for workplace-based assessment may reflect not assessors' differing assessments of performance, but, rather, different interpretations of poorly aligned scales. Scales aligned to the expertise of clinician-assessors and the developing independence of trainees were found to reduce assessor disagreement and increase assessor discrimination (Crossley et al. 2011). Additionally, the RIME framework for assessment, that describes ascending levels of clinical competence, has proven to generate reliable results (Ander et al. 2012; Pangaro 1999; Pangaro and ten Cate 2013; Tolsgaard et al. 2012). The statements made by some of the raters in our study suggest that these education-related factors are implicitly incorporated in order to somehow balance their assessments in the light of the GP trainee level. This is also a phenomenon that should be addressed in rater training.

It seems apparent that a generic communication assessment instrument like the MAAS-Global is both a help and a hindrance. A generic instrument has the advantage of potentially being used for many different situations in which communication assessment is required. At the same time, its power also seems its weakness: such an instrument may look like a 'one size fits all'. If the instrument is to be used on a larger scale to assess practice situations, it needs fine-tuning to the specific aspects of different situations or contexts. In the case of the MAAS-Global, the Manual needs to be expanded in order to take different context factors into account that were identified for GP consultations (Essers et al. 2011). Raters can subsequently be trained on how to incorporate context factors. That way, it will offer a better guidance to assess context-specific communication performance in daily practice.

Strengths and limitations

We asked our raters to assess the GP trainees as they normally would. This instruction had the effect that the participants considered the consultations as if they were submitted in the context of a summative assessment procedure in the GP training programme. However, the presented consultations were never recorded for that purpose. Several raters made remarks concerning the level of communication they saw versus what they expected in a consultation that was handed in for assessment. The instruction that they were to assess the GP trainee performance as they would normally do may have elicited this effect. Moreover, having in mind that the consultations were handed in for assessment purposes, participants may have been inclined to demonstrate to the researcher how well they complied to the assessment manual and inadvertently feel impaired to freely express their thoughts on or be critical on the assessment procedure in their institution.

Although we found confirmation for context factors we had previously identified (Essers et al. 2011), the overlap needs to be viewed with some consideration due to the difference in the perspective with which context was approached. In our previous study, we were looking for factors that could explain low item scores for GPs on the MAAS-Global (Essers et al. 2011). In the current study, we asked raters, while assessing a GP trainee, their views on factors that affected item scores on the basis of two selected consultations. As a consequence, apart from confirming previous findings, raters named factors that pertain to their context, i.e. an educational context in which their task is to assess GP trainee performance against a predefined standard using a validated instrument.

Another limitation in comparing the results with our previous study originates from the fact that the consultations observed were very different. In our previous study, we studied 18 consultations of experienced GPs, whereas in this study, two GP trainee consultations were shown.

Conclusions

Raters acknowledge that context factors affect communication in GP consultations to some extent, but they struggle with the way contextual influences should be taken into account when assessing communication performance in an educational context. To assess practice situations, raters need extra guidance on how to handle specific contextual factors.

Acknowledgments We thank all GPs who participated in the study. Special thanks go to Vera Atema who interviewed the Leiden raters, and to Dragana Milic, Fleur van den Bogaard, Tim Butterbrod, Ine Smeets, and Maxime Essers, who transcribed the recordings.

Conflict of interests The study was funded by the SBOH Foundation, employer of Dutch General Practitioner Registrars and funder of the national specialty training for General Practitioners. There are no conflict of interests.

Appendix 1: Think-aloud protocol and interview questions

Introduction and think-aloud protocol

Assessing GP trainees in their workplace is one of your tasks. It is an important task, for its aim is that the trainees learn from your feedback. We know that assessing workplace performance is not an easy job, and that raters all have their idiosyncratic ways in doing this. The aim of this study is to get a more specific idea of the considerations and thoughts that raters have when assessing communication performance using the MAAS-Global scoring instrument. The results of this study may be relevant to rater training in the future

We will perform this study by showing you two videotaped consultations consecutively, which will be followed by a number of questions after each video. It should be clear to you that this is not a test of your assessment abilities, nor will you be judged as a result of participating in this study. There is no right or wrong in the answers you give or in the thoughts you express! Nor is it of any concern to us if your judgment is right or wrong. We are merely interested in your thoughts concerning the GP trainees performance in that situation

We will therefore ask you to think-aloud during your assessment and to express all thoughts and considerations you have. You may act as if you are alone in this room and are talking to yourself. In order to carefully reproduce what you have said, we will record this interview on audio tape. Is it clear what we ask of you?

Now, you are going to see a consultation of a first-year GP trainee at nine months of his/her training programme. Suppose this is a situation in which you are to assess his/her communication performance. You can do this as you are used to do when assessing a video in a real assessment. However, now you are invited to express your thoughts and feelings aloud while doing so. You may use the papers we provide to make notes and the MAAS-Global scoring form for your rating. Our main leading question is: "What thoughts and considerations (regarding the consultation, the GP trainee, the patient and/or the assessment itself) are relevant to the assessment of the GP trainees communication performance?"

I will start the video and also kindly invite you to express your first impression as soon as you have these

Interview questions

Appendix continued

1. Having seen this consultation on video, what feedback would you give the GP trainee concerning his/her communication performance based on this?
2. What feedback do you have regarding the assessment itself?
3. In this consultation, you may have noticed X (e.g. more than one person was attending, or the patient seemed to know the PE he was about to undergo, or the GP trainee and the patient seemed to know each other, or the consultation seemed to take place on the initiative of the GP, or the problem presented seemed easy to solve, or the patient did not seem to master the Dutch language). Does this (in general) make a difference in how you rate the GP trainee's communication performance?
4. What do you consider most important when assessing communication performance?
5. Are there any other factors (concerning the trainee, the patient or the consultation) that you tend to take into account when assessing GP trainee communication performance?
6. Did the way you now assessed the trainee's performance differ from the way you usually do this? If so, how is it different?

Appendix 2: MAAS-Global rating list for doctor-patient communication skills

Each item is scored on a scale ranging of 0-6. For the items 2 and 4, the rating 'not applicable' is an additional option.

Communication skills for each separate phase

1. Introduction
 - Giving the patient room to tell his story
 - General orientation on the reason for visit
 - Asking about other reasons for visit
2. Follow-up consultation
 - Naming previous complaints, requests for help and management plan
 - Asking about adherence to management plan
 - Asking about the course of the complaint
3. Request for help
 - Naming requests for help, wishes or expectations
 - Naming reasons that prompted the patient to come now
 - Completing exploring request for help
4. Physical examination
 - Instructions to the patient
 - Explanation of what is being done
 - Treating the patient with care and respect
5. Diagnosis
 - Naming findings and diagnosis/hypothesis
 - Naming causes or the relation between findings and diagnosis
 - Naming prognosis or expected course
 - Asking for the patient's response
6. Management
 - Shared decision making, discussing alternatives, risks and benefits
 - Discussing feasibility and adherence

Appendix continued

-
- Determining who will do what and when
 - Asking for patient's response
 - 7. Evaluation of consultation
 - General question
 - Responding to requests for help
 - Perspective for the time being
 - General communication skills
 - 8. Exploration
 - Exploring requests for help, wishes or expectations
 - Exploring patient's response to information given within patient's frame of reference
 - Responding to non-verbal behaviour and cues
 - 9. Emotions
 - Asking about/exploring feelings
 - Reflecting feelings (including nature and intensity)
 - Sufficiently throughout the entire consultation
 - 10. Information giving
 - Announcing, categorizing
 - In small quantities, concrete explanations
 - Understandable language
 - Asking whether the patient understands
 - 11. Summarisations
 - Content is correct, complete
 - Concise, rephrased
 - Checking
 - Sufficiently throughout the entire consultation
 - 12. Structuring
 - Logical sequence of phases
 - Balanced division of time
 - Announcing (history taking, examination, other phases)
 - 13. Empathy
 - Concerned, inviting and sincerely empathetic in intonation, gesture and eye contact
 - Expressing empathy in brief verbal responses
-

References

- Ander, D. S., Wallenstein, J., Abramson, J. L., Click, L., & Shayne, P. (2012). Reporter-interpreter-manager-educator (RIME) descriptive ratings as an evaluation tool in an emergency medicine clerkship. *The Journal of Emergency Medicine*, 43, 720–727.
- Bensing, J., van Dulmen, S., & Bates, K. (2003). Communication in context: New directions in communication research. *Patient Education and Counseling*, 50, 27–32.
- Brown, R. F., & Bylund, C. L. (2008). Communication skills training: Describing a new conceptual model. *Academic Medicine*, 83, 37–44.
- Chan, C. S., Wun, Y. T., Cheung, A., Dickinson, J. A., Chan, K. W., et al. (2003). Communication skill of general practitioners: Any room for improvement? How much can it be improved? *Medical Education*, 37, 514–526.

- Claramita, M., Dalen, J. V., & Van Der Vleuten, C. P. (2011). Doctors in a Southeast Asian country communicate sub-optimally regardless of patients' educational background. *Patient Education and Counseling*, 85, e169–e174.
- Claramita, M., Nugraheni, M. D. F., van Dalen, J., & van der Vleuten, C. (2013). Doctor-patient communication in Southeast Asia: A different culture? *Advances in Health Sciences Education*, 18, 15–31.
- Crossley, J., & Davies, H. (2005). Doctors' consultations with children and their parents: A model of competencies, outcomes and confounding influences. *Medical Education*, 39, 807–819.
- Crossley, J., Johnson, G., Booth, J., & Wade, W. (2011). Good questions, good answers: Construct alignment improves the performance of workplace-based assessment scales. *Medical Education*, 45, 560–569.
- de Haes, H., & Bensing, J. (2009). Endpoints in medical communication research, proposing a framework of functions and outcomes. *Patient Education and Counseling*, 74, 287–294.
- Durning, S., Artino, A. R., Jr, Pangaro, L., van der Vleuten, C. P., & Schuwirth, L. (2011). Context and clinical reasoning: Understanding the perspective of the expert's voice. *Medical Education*, 45, 927–938.
- Essers, G., Kramer, A., Andriess, B., van Weel, C., van der Vleuten, C. P. M., & van Dulmen, A. M. (2013a). Context factors in family physician communication—patient encounters and their impact on assessing communication skills—an exploratory study. *BMC Family Practice*, 14, 65.
- Essers, G., van Dulmen, S., van Es, J., van Weel, C., van der Vleuten, C., Kramer, A. (2013b). Context factors in consultations of general practitioner trainees and their impact on communication assessment in the authentic setting. *Patient Education and Counseling*. doi: 10.1016/j.pec.2013.08.024. [Epub ahead of print].
- Essers, G., van Dulmen, S., van Weel, C., van der Vleuten, C., & Kramer, A. (2011). Identifying context factors explaining physician's low performance in communication assessment: An explorative study in general practice. *BMC Family Practice*, 12, 138.
- Eva, K. W. (2003). On the generality of specificity. *Medical Education*, 37, 587–588.
- Glaser, B. G. (1965). The constant comparative method of qualitative-analysis. *Social Problems*, 12, 436–445.
- Govaerts, M. J. B., Schuwirth, L. W. T., Van der Vleuten, C. P. M., & Muijtjens, A. M. M. (2011). Workplace-based assessment: Effects of rater expertise. *Advances in Health Sciences Education*, 16, 151–165.
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18, 59–82.
- Hobma, S. O., Ram, P. M., Muijtjens, A. M. M., Grol, R., & van der Vleuten, C. P. M. (2004). Setting a standard for performance assessment of doctor-patient communication in general practice. *Medical Education*, 38, 1244–1252.
- Hulsman, R. L. (2009). Shifting goals in medical communication. Determinants of goal detection and response formation. *Patient Education and Counseling*, 74, 302–308.
- Kogan, J. R., Conforti, L., Bernabeo, E., Iobst, W., & Holmboe, E. (2011). Opening the black box of clinical skills assessment via observation: A conceptual model. *Medical Education*, 45, 1048–1060.
- Kramer, A. W. M., Dusman, H., Tan, L. H. C., Jansen, J. J. M., Grol, R., & van der Vleuten, C. P. M. (2004). Acquisition of communication skills in postgraduate training for general practice. *Medical Education*, 38, 158–167.
- Malau-Aduli, B. S. (2011). Exploring the experiences and coping strategies of international medical students. *BMC Medical Education*, 11, 40.
- McGill, D. A., van der Vleuten, C. P. M., & Clarke, M. J. (2011). Supervisor assessment of clinical and professional competence of medical trainees: A reliability study using workplace data and a focused analytical literature review. *Advances in Health Sciences Education*, 16, 405–425.
- Pangaro, L. (1999). A new vocabulary and other innovations for improving descriptive in-training evaluations. *Academic Medicine*, 74, 1203–1207.
- Pangaro, L., & ten Cate, O. (2013). Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Medical Teacher*, 35, E1197–E1210.
- Pelgrim EAM. (2013). *Clarifying observation and assessment feedback in workplace-based learning, chapter 7*. Dissertation, Radboud UMC, Nijmegen. http://www.eampelgrim.nl/phd_thesis_pelgrim.pdf.
- Pelgrim, E. A. M., & Kramer, A. W. M. (2013). How can medical education benefit from the evidence on learners seeking and using feedback? *Medical Education*, 47, 225–227.
- Pelgrim, E. A. M., Kramer, A. W. M., Mekkink, H. G. A., & van der Vleuten, C. P. M. (2012). The process of feedback in workplace-based assessment: Organisation, delivery, continuity. *Medical Education*, 46, 604–612.

- Ram, P., Grol, R., Rethans, J. J., Schouten, B., van der Vleuten, C., & Kester, A. (1999). Assessment of general practitioners by video observation of communicative and medical performance in daily practice: Issues of validity, reliability and feasibility. *Medical Education*, 33, 447–454.
- Ram, P., van Thiel, J., & van Dalen, J. (2003). *MAAS-global manual 2000*. Maastricht: Maastricht University.
- Ram, P., et al. (2011). *National assessment plan for the general practitioner specialty training*. Utrecht: General Practitioner Specialty Training in the Netherlands.
- Reinders, M. E., Blankenstein, A. H., van Marwijk, H. W. J., Knol, D. L., Ram, P., et al. (2011). Reliability of consultation skills assessments using standardised versus real patients. *Medical Education*, 45, 578–584.
- Salmon, P., Mendick, N., & Young, B. (2011). Integrative qualitative communication analysis of consultation and patient and practitioner perspectives: Towards a theory of authentic caring in clinical relationships. *Patient Education and Counseling*, 82, 448–454.
- Salmon, P., & Young, B. (2011). Creativity in clinical communication: From communication skills to skilled communication. *Medical Education*, 45, 217–226.
- Tavares, W., & Eva, K. W. (2013). Exploring the impact of mental workload on rater-based assessments. *Advances in Health Sciences Education*, 18, 291–303.
- Tolsgaard, M. G., Arendrup, H., Lindhardt, B. O., Hillingsø, J. G., Stoltenberg, M., & Ringsted, C. (2012). Construct validity of the reporter-interpreter-manager-educator structure for assessing students' patient encounter skills. *Academic Medicine*, 87, 799–806.
- Tweed, M., & Ingham, C. (2010). Observed consultation: Confidence and accuracy of assessors. *Advances in Health Sciences Education*, 15, 31–43.
- van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, 39, 309–317.
- van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., et al. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34, 205–214.
- van Es, J., De Waard-Wieringa, M., & Visser, M. (2013). Differential growth in doctor-patient communication skills in GP trainees. *Medical Education*, 47, 691–700.
- van Es, J. M., Schrijver, C. J. W., Oberink, R. H. H., & Visser, M. R. M. (2012). Two-dimensional structure of the MAAS-Global rating list for consultation skills of doctors. *Medical Teacher*, e1–e6. doi:[10.3109/0142159X.2012.709652](https://doi.org/10.3109/0142159X.2012.709652).
- van Thiel, J., Kraan, H. F., & Vleuten, C. (1991). Reliability and feasibility of measuring medical interviewing skills: The revised Maastricht History-Taking and Advice Checklist. *Medical Education*, 25(3), 224–229.
- Williams, R. G., Klamen, D. A., & McGaghie, W. C. (2003). Cognitive, social and environmental sources of bias in clinical performance ratings. *Teaching and Learning in Medicine*, 15, 270–292.
- Winefield, H. R., Murrell, T. G. C., Clifford, J. V., & Farmer, E. A. (1995). Process and outcomes in general practice consultations: Problems in defining high quality care. *Social Science & Medicine*, 41, 6.
- Yeates, P., O'Neill, P., & Mann, K. (2011). Examining the box's contents. *Medical Education*, 45, 970–972.
- Yeates, P., O'Neill, P., Mann, K., & Eva, K. (2012). Seeing the same thing differently: Mechanisms that contribute to assessor differences in directly-observed performance assessments. *Advances in Health Sciences*. doi: [10.1007/s10459-012-9372-1](https://doi.org/10.1007/s10459-012-9372-1). [Epub ahead of print].
- Yeates, P., O'Neill, P., Mann, K., & Eva, K. W. (2013). 'You're certainly relatively competent': Assessor bias due to recent experiences. *Medical Education*, 47, 910–922.
- Zoppi, K., & Epstein, R. M. (2002). Is communication a skill? Communication behaviors and being in relation. *Family Medicine*, 34, 319–324.