Medical Education

Self-evaluation and peer-feedback of medical students’ communication skills using a web-based video annotation system. Exploring content and specificity

Robert L. Hulsman *, Jane van der Vloodt

Academic Medical Centre, Department of Medical Psychology, Amsterdam, The Netherlands

A R T I C L E  I N F O

Article history:
Received 21 October 2013
Received in revised form 16 October 2014
Accepted 11 November 2014

Keywords:
Self-evaluation
Peer-feedback
Communication
Assessment

A B S T R A C T

Objective: Self-evaluation and peer-feedback are important strategies within the reflective practice paradigm for the development and maintenance of professional competencies like medical communication. Characteristics of the self-evaluation and peer-feedback annotations of medical students’ video recorded communication skills were analyzed.

Method: Twenty-five year 4 medical students recorded history-taking consultations with a simulated patient, uploaded the video to a web-based platform, marked and annotated positive and negative events. Peers reviewed the video and self-evaluations and provided feedback. Analyzed were the number of marked positive and negative annotations and the amount of text entered. Topics and specificity of the annotations were coded and analyzed qualitatively.

Results: Students annotated on average more negative than positive events. Additional peer-feedback was more often positive. Topics most often related to structuring the consultation. Students were most critical about their biomedical topics. Negative annotations were more specific than positive annotations. Self-evaluations were more specific than peer-feedback and both show a significant correlation. Four response patterns were detected that negatively bias specificity assessment ratings.

Conclusion: Teaching students to be more specific in their self-evaluations may be effective for receiving more specific peer-feedback.

Practice implications: Videofragmentrating is a convenient tool to implement reflective practice activities like self-evaluation and peer-feedback to the classroom in the teaching of clinical skills.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Self-monitoring and self-regulation are important responsibilities of medical professionals in the development and maintenance of professional competencies [1,2]. Physicians are expected to be able to evaluate their own strengths, weaknesses and learning needs in order to maintain a level of competence in accordance with the professional norm [3]. The importance of concepts like reflection, self-assessment, peer-assessment, peer-feedback are intrinsically connected to the Professional and Scholar roles of the CanMeds model, which is nowadays the leading model underlying many medical curricula in the world [4,5].

Self-evaluation and peer-feedback are particularly relevant and have been proven to be effective in medical communication skills training because the problems to be solved in medical communication are ‘ill defined’ [6]. In ill-defined problems the given state, the goal state, and effective operations are not fully predefined and have several unique but equally correct solutions [7–9]. For example, Epstein et al. [10] found that physicians show different ‘solutions’ in responding to patients’ expressions of worry (from most to least frequent): acknowledgement, inquiry, explanation, reassurance, empathy. Which of these responses is most appropriate requires appraisal of the context factors [11].

Self-evaluation and peer-feedback fit in the modern educational paradigm of reflective practice, a concept introduced by Donald Schön in 1983, which has gained popularity in recent years [12–14]. Reflective practice aims to develop critical thinking, problem-solving, and self-directed and lifelong learning skills through gaining new understandings, new perspectives, and new alternatives for future performance [12,15].
Reflective practice does not come naturally to most students and requires formal educational arrangements [16]. Video-based learning using a web-based platform for sharing video recordings of clinical performance, which students and peers can review and comment on, can be helpful [17,18].

By reviewing their work and that of others, students develop their evaluative skills and acquire a better understanding of the performance criteria [14,19,20]. Combining internal information from self-evaluation with external information from (peer-)feedback is considered most effective to optimize clinical performance [21,22]. While some studies stress the importance of external feedback as input for informed self-assessment, others show that external feedback is more effective in response to student self-evaluations [21,22].

Assessment of the quality of self-evaluations and peer-feedback is often defined in terms of accuracy, consistency across assessors, or concordance with teacher feedback [23]. However, the concordance between student and expert evaluations is often vulnerable to bias for two reasons. First, for an individual it is difficult to make objective observations of his or her own performance due to unconscious biases [3]. Second, external evaluations among medical faculty often lack consensus by valuing different aspects of the performance [24,25].

A different approach defines quality of self-evaluations and peer-feedback in terms of content and/or style characteristics [23]. Students’ ability to focus on content and style characteristics is generic and transferable to other settings. Self-evaluations and peer-feedback need to be specific to be effective [22,26,27]. Specificity is defined as the level of information presented in feedback messages [28]. Some studies rated specificity of reflections and feedback on three component levels, while others use five [22,23]. In a previous study among second year students we developed a system for coding self-evaluations, which includes three retrospective categories describing the event (behaviour) and its antecedents (motive) or consequences (effect), and two prospective categories describing an alternative strategy and its goal [29].

The aims of the present mixed method study are to explore quantitatively and qualitatively:

1. the characteristics of the self-evaluations and peer-feedback annotations of medical students’ video recorded communication skills.
2. how the characteristics of the peer-feedback relate to the characteristics of the self-evaluations.
3. features that may bias the assessment of the specificity of the self-evaluations and peer-feedback.

2. Methods

2.1. Participants and procedure

In 2009–2010 the Bachelor-Master structure was introduced in the curriculum of the Academic Medical Centre of the University of Amsterdam, following the Bologna Process to harmonize the systems of higher education in Europe [30]. Every fourteen days a group of about 14 year 4 students is starting with their Master programme. In July and August 2012 two student groups were invited to use the VideoFragmentRating (VFR) system, embedded in a regular history-taking skills training programme preparing them for the clinical rotations in their clerkships [31].

All Master students were trained in the Bachelor programme on history-taking, the functions of nonverbal and verbal active listening skills, and the principles of effective feedback. History-taking skills were practiced in small group trainings with simulated patients: (1) relationship building; (2) gathering reliable biomedical information about the complaints; (3) asking about the patient’s ideas, concerns and expectations; (4) structuring the consultation [32,33].

In the Master programme these principles were refreshed briefly in a plenary instruction, including the review and analysis of a demonstration video. Next, each student recorded individually a history-taking consultation with one of five available simulated patients for formative assessment. Videos were recorded with standalone handycams and uploaded in the VFR system.

VFR is a dedicated web-based application for review and annotation of video recording of clinical skills. In VFR the video and all markings and annotations are presented in a single screen. It is developed at the Department of Surgery of the University Medical Centre Groningen, The Netherlands. Students can upload their own video recording to the password protected video server. Based on the strict security requirements, only the student and the invited peers or supervisor are able to access and review the video recording and annotations of the student.

Students were instructed to review their video individually and to mark and annotate on the timeline two green bullets for successful performances and two red bullets for poor performances of the student (Fig. 1). Next, each student provided peer-feedback to a student who recorded a history-taking consultation with the same simulated patient as their own. Feedback annotations to the self-evaluations automatically have the same valence as selected by the student. Peers could also provide additional feedback by marking and annotating new events on the timeline. The valence of the additional feedback is decided by the peer.

2.2. VFR data extraction

Anonymized data were extracted from the log files of the VFR system. These data include per student the number of self-evaluations, the number of peer-feedback annotations, and the number of additional feedback annotations. Further is extracted the valence and the content of each annotation. The number of characters entered was computed for each annotation with the Microsoft Excel length (field) function.

2.3. Data coding

Content and specificity of annotations were coded in an iterative process by JV. The coding of the content was initially based on the criteria derived from the History Taking Assessment Scale (HTAS) which is used for formative and summative assessment of students in the Bachelor programme [33]. The initial coding structure was organized around the HTAS behavioural (sub)categories: courteousness and respect; asking for medical information; asking for ideas, concerns, expectations of the patient; structuring the conversation. In weekly meetings the codings of JV were discussed with RH, expanded and modified if required. Throughout the analyses differences in interpretation were resolved through discussion and re-examination of the annotations and codings. The coding scheme was developed by using MaxQDA software which is designed for qualitative and mixed methods data analysis [34].

The coding of the specificity of the annotations was based on a system developed in an earlier study [29]. Three retrospective categories relate to ‘describing the key event’; two prospective categories relate to ‘finding new solutions’. Box 1 provides a description of each category. Annotations containing more of these categories are considered more specific. Hence, the specificity score can range between 0 and 5.
2.4. Data analyses

Paired t-tests were used for the statistical analysis of the number and valence of annotations per student. The amount of text used and specificity scores of the annotations were analyzed using mixed-linear modelling to take into account that these characteristics of multiple measures (annotations) within students are correlated [35]. Analyzed in the models were the significance of the main and interaction effects of type of annotation (self-evaluation, peer-feedback, or additional feedback) and its valence (positive, negative) on amount of text and specificity of each annotation.

The associations between peer-feedback and the self-evaluations in the amount of text used and specificity codings were tested with mixed linear models. Prior to analysis variables were translated into standardized z-scores, to be able to interpret β-coefficients as correlations. Quantitative data were analyzed with SPSS Version 20.

3. Results

3.1. Participants

Participants were 25 year 4 students (15 women, 60%) out of 28 members of the two teaching groups starting the Master programme. One student could not participate and had to leave early, two other students dropped out due to technical failures of the video recordings of their history-taking consultations. The length of the 25 video recordings was on average 14.0 (SD 6.2) min.

3.2. Number and valence of self-evaluation and peer-feedback

The average number of self-evaluations and peer-feedback is displayed in Fig. 2a. Students entered in VFR on average 5.64 (SD 1.98) self-evaluation annotations; 2.36 (SD 0.81) comments with a positive valence and 3.28 (SD 1.40) comments with a negative valence (t = 3.99; df = 24; p < .001). Students marked more critical events in the video than the requested two positive and two negative events (t = 4.14; df = 24; p < .001).

Students entered in VFR on average 4.96 (SD 1.51) peer-feedback annotations; 2.08 (SD 0.70) in response to self-evaluations with a positive and 2.88 (SD 1.09) in response to self-evaluations with a negative valence (t = 3.84; df = 24; p < .001).

The average number of additional feedback annotations per student was 2.08 (SD 2.63). Additional feedback has more often a positive (1.40; SD 1.66) than a negative (0.68; SD 1.35) valence (t = 2.42; df = 24; p < .05).

3.3. Amount of text entered by type of annotation

Data of 317 annotations was extracted from the VFR log files: 141 self-evaluations (82 red valence 58.2%), 124 peer-feedback (73 negative valence 58.9%) and 52 additional feedback (16 negative valence, 30.8%). This shows that peer-feedback is provided to 87.9% of the self-evaluations.

The overall average amount of text entered per annotation was 160.3 (SD 121.4) characters, which is equal to about 20 words. The average amount of text entered by type and valence of the annotations is displayed in Fig. 2b. More text is entered for the self-evaluations (176.6; SD 128.2) than for the peer-feedback.
(160.3; SD 113.5) and additional peer-feedback (115.7; SD 111.4). More text is entered for comments with a negative valence (194.9; SD 131.4) than for comments with a positive valence (119.7; SD 93.8). This difference is much stronger for the additional feedback (Fig. 3). Mixed-linear modelling showed a significant main effect of additional feedback ($F = 8.04; \text{df} = 299.7; p < .005$) and negative valence ($F = 13.33; \text{df} = 290.0; p < .001$) on the amount of text used. A significant interaction effect was found between additional feedback and valence ($F = 4.15; \text{df} = 292.7; p < .05$).

A significant association was found between amount of text used in peer-feedback and in the corresponding self-evaluations ($\beta$-coefficient = 0.32; $p < .001$; $n = 124$).

3.4. Topics of the annotations

Following the topics addressed in the history-taking assessment scale (HTAS) a coding scheme with 23 content categories was developed (Table 1). Most annotations were coded with a single topic, 28 of the 317 annotations (8.8%) were coded with two content categories: 8 of the 141 (5.7%) self-evaluations, 18 of the 124 (14.5%) feedback, and 2 of the 52 (4%) additional feedback annotations.

Most topics relate to structuring the conversation (36.2%), followed by the medical perspective (22.6%), courteousness and respect (18.8%) and finally the patient perspective (11.0%). This pattern applies to the self-evaluations and the peer-feedback annotations.

The additional feedback annotations have a slightly different topic structure. Here, the ‘Other’ category is most frequent. This category is dominated by codes marked as ‘compliments’.

Examples of the topics of the self-evaluations and the peer-feedback are displayed in Box 2.

Table 2 presents the valence of the main topics of the self-evaluations. This table shows that self-evaluations often have a negative valence when related to the medical perspective (76.9%), or courteousness and respect (64.3%). Self-evaluations have often a positive valence when related to the patient perspective (64.7%).

Most of the topics coded in the peer-feedback match the topics coded in the corresponding self-evaluation. A topic switch is observed in only 18 of the 142 (12.7%) peer-feedback annotations, 12 (66.7%) of these in response to self-evaluations with a green valence.

In this example the topic of the self-evaluation coded as ‘structuring’ and the peer-feedback as ‘patient-perspective’:

SE: I'm interrupting the patient politely, (although rather delayed) and ask about her concerns about her complaints, after which I can continue with focussed questions about her medical symptoms. (S-69)

PF: Very good to see how you are reflecting the way the patient is experiencing her complaints. It is obvious she appreciates it. (S-74)

3.5. Assessment of specificity of the annotations

The specificity of the annotations is displayed in Table 3, showing the distribution of the five categories (behaviour, motive, effect, suggestion and goal) in the self-evaluations, peer-feedback and additional feedback. The Behaviour category is coded most frequently (71.6%) in the annotations, followed by the Suggestions category (39.7%). The Effect category is coded the least: 20.0% in all annotations.

Self-evaluations are most specific and contain on average 2.23 (SD 1.10), feedback 1.80 (SD 0.95) and additional feedback 1.31 (SD 1.02) categories per annotation. Regardless of type, annotations with a negative valence are more specific (2.21; SD 1.12) than annotations with a positive valence (1.56; SD 0.93). These patterns were confirmed in the mixed-linear model analysis of the specificity, which showed a significant main effect of additional feedback ($F = 9.84; \text{df} = 310.8; p < .01$) and of negative valence ($F = 27.04; \text{df} = 294.7; p < .001$).

The specificity of the peer-feedback is positively associated with the specificity of self-evaluations ($\beta$-coefficient = 23; $p < .01$; $n = 124$).
Box 3 shows two examples. The specificity of the first example is rated high since it contains all retrospective and prospective categories. The second shows a typical example of containing only Behaviour and Suggestion categories.

3.6. Response patterns that may bias the specificity assessment

Coding the specificity of the annotations showed response patterns in how students phrased their self-evaluations that may violate the validity of the specificity assessment. In a number of instances certain specificity coding categories remain implicit.

a. Suggestions are not coded in 79 self-evaluations and 57 of these (72.2%) have a positive valence. When positive behaviour is pointed out, there is no need to suggest alternative behaviour. This is in line with our observation that annotations with a positive valence are less specific than annotations with a negative valence.

For example:

SE: I explain I will make some short notes so I won’t forget anything, and I’m asking if that is okay. So, I explain why I will be writing every now and then and the patient will trust me I’m not forgetting anything. (S-47)

b. In 22 self-evaluations with a negative valence the Suggestion category is lacking. Pointing out the absence of behaviour or mentioning negative behaviour contains an implicit Suggestion the behaviour should have been the opposite:

For example:

SE: I haven’t asked about her emotional concerns. She had to come up with it herself at the end of the consultation. (S-58)

c. The Behaviour category is not explicitly mentioned in 16 self-evaluations. This is observed when the self-evaluation is pointing out behaviour that should have occurred, regardless what happened in the consultation. By stating what should have happened, the student is implicitly stating Behaviour that did not occur.

For example:

SE: It should be more clear what exactly the patient meant. Did the pain move from right to left, or was it radiant pain? (S-68)

d. Sometimes it is redundant to mention a Goal. Often the Goal of the suggestion is identical to the Motive for the actual behaviour. This reduces the need to describe the goal once again.

For example:

SE: I’m announcing I want to go back to the patient’s physical complaints. I’m doing so because it’s strange to jump back from psychosocial topics to the physical complaints. So I’m announcing I will switch topics. Maybe I should have completed the symptoms review first, but I think I solved this adequately this way. (S-77)

The implicit message is that sudden topic changes may be unpleasant for the patient. The goal would be to prevent this.

4. Discussion and conclusion

4.1. Discussion

This mixed method study explored the characteristics and interrelatedness of the self-evaluations and peer-feedback annotations of medical students’ video recorded communication skills.
Box 2. Topic coding examples of self-evaluation and peer-feedback.

Courteousness and respect
SE: I'm saying 'okay' regularly to signal that I'm listening. Afterwards the patient pointed out to me 'okay' can be interpreted incorrectly by the patient as: 'this is something good, it's nothing serious'. It seems to me better to replace 'okay' by confirmative nodding, or 'humming', or saying 'yes', and 'I understand' to indicate I'm listening? (S-46)
Pf: I was also suffering from saying 'okay' frequently. My solution is also 'humming', saying 'yes', or repeating what the patient just have said. (S-51)

Patient perspective
SE: I consider it appropriate I'm asking for clarification just now about his concerns (patient perspective) because he could be afraid of a brain tumour or another hereditary disease (so I could explore this further). It turned out he was concerned about high blood pressure. (S-49)
Pf: Very good question indeed, he expressed being worried, so it is important to explore this further. This will make the patient feeling understood. (S-44)

Medical perspective
SE: I'm asking about the onset of the complaints but I'm forgetting to ask about the time course. As a result I've no idea about the acute or gradual onset of the complaints. I should follow the medical dimensions of complaints more strictly in order not to forget this? (S-49)
Pf: By following the dimensions of complaints you will less likely forget dimensions. However, I'm afraid it will become very artificial (you are using a checklist). Maybe you can ask after how long ago something started, ask if it has become more or less worse. This will provide a more systematic approach, without needing to follow strictly the medical dimensions. (S-44)

Structuring the conversation
SE: Next she is going to tell about sitting up and how annoying it is to sleep in this posture. She told me before, so I'm trying to skip this and to continue with other questions (like here weight). However, a sudden change in topic may appear blunt. It's difficult when a patient is talking so much. (S-50)
Pf: I understand your issue. You might consider to announce you are going to switch topics. In this case however, the patient didn't mind. (S-55)

Other
SE: I should not have said this (diagnosis), presumably, even if it is true in this case. (S-43)
Pf: I'm not sure. But you are absolutely right. I prefer wouldn't say it, unless I'm 100% confident. (S-48)

SE, self-evaluation; Pf, peer-feedback; between brackets the respondent code. (S-xx)

This is in line with observations of Lanning et al. [36] who found that students' self-ratings of their communication skills were more negative than peer-ratings. The authors suggest that students have a different insight into communication with patients than peers and may be overly critical of their own performance. Peers on the other hand may be biased in a positive direction guided by friendship marking, not wanting to be too critical in their peer-feedback to preserve the interpersonal relationships with their colleagues. In our study this is reflected in the observation that additional feedback is frequently coded as 'compliments'.

The observed skewed balance between positive and negative comments may depend on the order in which self-evaluation and peer-feedback are provided. While some researchers stress the importance of feedback as input for reflective activities of students [21,37], others show that learners perceive feedback on self-selected events as having more educational value than feedback on externally identified significant events in their performance [38]. In our study the feedback sequence is initiated by the student's self-evaluations. Self-evaluation gives learners more control over the peer-review process, gaining ownership of learning [19,39]. While it is the student who primarily decides about the topics that need to be discussed. It is also the student who is pointing out poor performance, making it easier for peers to respond, and reducing the risk of retaliation.

4.1.2. Topics of self-evaluations and peer-feedback

The topics combined with the valence of the self-evaluations provide insight in the medical students' own perceptions of their performance and learning needs. The participants starting the first year of their Master programme, were in a transition programme preparing for clinical rotations, where history taking with clinical patients is one of the main tasks. Students were commenting most frequently on structuring their history-taking consultation and the least on topics related to the patient perspective. Considering the valence of the self-evaluations, students were more negative about asking for medical information and more positive about exploring the patient's perspective. The latter is not always confirmed in the literature. Peters et al. [40] analyzed self-evaluations of communication skills of second year medical students, based on interactions with patients. Their most frequently mentioned learning needs were skills related to the emotional aspects of the interaction, and to structuring the consultation.

Comparison of the topic-codings of the peer-feedback with the codings of the self-evaluations revealed that topic-switches are rare and occur most frequently to positive self-evaluations.

4.1.3. Specificity of self-evaluations and peer-feedback

Analysis of the specificity of the annotations revealed that students use on average two or three components in the self-evaluations, and less than two in the feedback annotations. The annotations referred most often to behaviours, followed by suggestions. In self-evaluations also motives were frequently reported. The findings suggest that the level of specificity of the annotations can be improved. This is also observed in studies by Canavan et al. [41], who analyzed the quality of written formative feedback on professional behaviours of physicians in training, and Gielen et al. [23] who observed a low occurrence of the justification category in peer-feedback. Feedback that is lacking specificity and bearing actionable information is less useful to learners [22,27,41].

The finding in our study that the specificity of self-evaluations and peer-feedback are correlated significantly suggests that peers are encouraged to be more specific in their feedback when the self-evaluations they respond to are more specific. This phenomenon is also observed in a study by Pelgrim et al. [22] who found a positive relation between the specificity of reflection and

and factors that bias assessment of the specificity of the self-evaluations and peer-feedback.

4.1.1. Balance between positive and negative comments

One of the findings is that the students marked and annotated more self-evaluations than the four they were assigned to do and also provided additional feedback, beyond the peer-feedback to the self-evaluations. Further, students paid more attention to performance that needed improvement. This is reflected both in the number of marked events and in the amount of text entered in annotations with a negative valence. Additional peer-feedback, however, had more often a positive valence.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Valence of the topics of the self-evaluations.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Green valence</td>
</tr>
<tr>
<td>Courteousness</td>
<td>10 (35.7%)</td>
</tr>
<tr>
<td>Patient perspective</td>
<td>11 (64.7%)</td>
</tr>
<tr>
<td>Medical perspective</td>
<td>9 (23.1%)</td>
</tr>
<tr>
<td>Structuring</td>
<td>28 (48.3%)</td>
</tr>
<tr>
<td>Other</td>
<td>4 (57.1%)</td>
</tr>
<tr>
<td>Total</td>
<td>62 (41.6%)</td>
</tr>
</tbody>
</table>
feedback among general practice trainees and their trainers, after conducting a consultation with a real patient. Students do not only learn from each other how to perform effectively, but also how to provide specific evaluations of clinical performance.

An important finding in this study is the existence of four response patterns in annotations that may bias the quantitative assessment of their specificity. Suggestions are often omitted in annotations with a positive valence, and when negative or the absence of behaviour is observed. The latter contain implicit suggestions of opposite behaviour. Further, observed behaviour is not always explicitly referenced when suggestions are presented. Finally, goals may remain implicit when they overlap with described motives of displayed performance. Consequentially, specificity ratings are negatively biased. These response patterns violate the validity of the quantitative assessment of specificity. Further research is needed to solve this problem.

4.1.4. Implications for research

Reflective practice is considered pivotal in the development and maintenance of professional competence [13]. To be able to study this relationship, reliable and valid coding schemes of reflective activities are required. Until now there is quite a variety in coding systems used [20,22,37]. The field may benefit from more standardization of the assessment methods of reflective practice. Most systems have in common that there should be references to specific behaviour, a judgement (positive/negative), an explanation or justification, and a suggestion for improvement. Our system contains these key elements.

4.1.5. Implications for teaching

A fundamental issue in the application of self-evaluations and peer-feedback is: how does it contribute to the outcomes of learning? Nofziger et al. [42] found among medical students that peer assessment affected their self-reported personal awareness, attitudes, or behaviours. Change was more likely when feedback was specific and described an area for improvement.

Formal requirements for practitioners to provide evidence of reflective practice are becoming part of licensing procedures of bodies like the Canadian College of Family Physicians, the British General Medical Council and American Accreditation Council for Graduate Medical Education (ACGME) [13,43].

Self-evaluation and peer-feedback are considered essential attributes that need to be integrated in medical education. We applied the web based VFR system which offers a dedicated environment for reviewing the video recorded performance and adding self-evaluations an peer-feedback. Evaluation of performance requires a save educational setting. Videos in VFR are password protected and it is the student who is in control when inviting peers or supervisors for feedback.

4.1.6. Limitations

There are important limitations to this study. This study was exploratory and included only a limited number of students. Data was collected in the context of testing the feasibility of using the web-based (VFR) system in a regular communication skills teaching programme for first year Master students. The primary focus of the teaching programme was refreshing the students’ history-taking skills in transition to the clinical rotations and not the improvement of the specificity of self-evaluations and peer-feedback. Future studies may focus on interventions to improve the specificity of self-evaluation and peer-feedback, and on their contribution in the development of reflective practitioners.

A final limitation is the connectedness of the valence of the peer-feedback with the valence of the self-evaluations. The videofragmenting programme offers no options for peers to assign independently the valence to their feedback.

4.2. Conclusion

This study shows that important CanMedS based strategies like self-evaluation and peer-feedback, which are highly valued in the modern reflective practice paradigm, can be applied in the context of video-based communication skills teaching. Teaching students to be more specific in their self-evaluations may stimulate peers to be more specific in their feedback.

---

**Table 3**

Specificity of the annotations.

<table>
<thead>
<tr>
<th></th>
<th>Self-evaluations (n = 141)</th>
<th>Feedback (n = 124)</th>
<th>Additional feedback (n = 52)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Retrospective</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour</td>
<td>125 (88.65%)</td>
<td>75 (60.48%)</td>
<td>36 (69.23%)</td>
</tr>
<tr>
<td>Motive</td>
<td>66 (46.81%)</td>
<td>21 (16.94%)</td>
<td>7 (13.46%)</td>
</tr>
<tr>
<td>Effect</td>
<td>30 (21.28%)</td>
<td>29 (23.39%)</td>
<td>6 (11.54%)</td>
</tr>
<tr>
<td><strong>Prospective</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suggestion</td>
<td>62 (43.97%)</td>
<td>48 (38.71%)</td>
<td>12 (23.08%)</td>
</tr>
<tr>
<td>Goal</td>
<td>32 (22.70%)</td>
<td>40 (32.26%)</td>
<td>7 (13.46%)</td>
</tr>
<tr>
<td><strong>Average number of elements:</strong></td>
<td>2.23 (SD 1.10)</td>
<td>1.80 (SD 0.95)</td>
<td>1.31 (SD 1.02)</td>
</tr>
<tr>
<td>Green valence</td>
<td>1.76 (SD 0.77)</td>
<td>1.65 (SD 0.97)</td>
<td>1.09 (SD 0.95)</td>
</tr>
<tr>
<td>Red valence</td>
<td>2.57 (SD 1.18)</td>
<td>1.90 (SD 0.94)</td>
<td>1.76 (SD 1.03)</td>
</tr>
</tbody>
</table>

---

**Box 3.** Specificity coding examples of self-evaluation and peer-feedback

**Example 1 (structuring): specificity = 4**

SE: Suddenly I'm talking about the working life of the patient [B], because she is introducing the topic herself [M] and I'm not finishing the systems review [B]; Would I have received more or better information [G] if I would return to her social and work related situation after having finished the systems review [S]? (S-69)

PF: The patient is constantly interfering with your planning [M]. You are asking about her stools [B] and she responds with talking about her work [E]. With such a patient maybe it is better to keep in your mind the topics you are asking about [G]. At that moment you might say: 'I'd like to finish my questions about your stools and I come back to your working life afterwards' [S]. (S-74)

**Example 2 (patient perspective): specificity = 2**

SE: The gentleman tells that his ill mother has passed away by now and I don't respond to that at all. I don't ask about how he was coping and not expressing his condolences [B] Would it have been better to do so? [S]. (S-76)

PF: I think I would not have done that either [B]. But you could have made a brief reflection, 'how sad' [S]. (S-71)

4.3. Practice implications

Peer monitoring and evaluating each others work is becoming common in clinical practice to maintain and improve the quality of care. Peer-feedback from colleagues is an important element of multi-source feedback (MSF) which is now obligatory for all physicians in the United Kingdom to retain their practicing license [44]. Comparable procedures are now developed in other countries [38,45]. Teaching self-evaluation and peer-feedback to medical students early in the curriculum is shown to result in improved feedback skills and enhanced comfort with feedback [46].

Videofragmentating is a convenient tool to bring self-evaluation and peer-feedback to the classroom in the teaching of clinical skills.

Conflict of interest

The authors have no conflicts of interest to report.

Acknowledgements

This study was conducted with the financial support of the ICTO Fund of the University of Amsterdam. The aim of the fund is to support implementations and innovations in e-learning. Statistical advice on mixed-linear modelling was obtained from M.G.E. Verdam and from P.T. Nieuwkerk, Academic Medical Centre Amsterdam, Department of Medical Psychology, The Netherlands.

References