Group Assessment of Resident Performance: Valuable for Program Director Judgment?

Marrigje E. Duitsman, MA, MD
Irene A. Slootweg, PhD
Imke C. van der Marel, MD
Marianne ten Kate-Booij, PhD, MD
Jacqueline de Graaf, MD, PhD
Cornelia Fluit, PhD
Debbie Jaarsma, PhD

ABSTRACT

**Background** Group discussion of resident performance is an emerging assessment approach in postgraduate medical education. However, groups do not necessarily make better decisions than individuals.

**Objective** This study examined how group meetings concerning the assessment of residents take place, what information is shared during the meetings, and how this influences program directors’ judgment of resident performance.

**Methods** In 2017, the researchers observed 10 faculty group meetings where resident performance was discussed and interviewed the program directors within a month after the meetings. We used a thematic framework analysis to identify themes from the transcribed meetings and interviews.

**Results** The information shared by group members during the meetings had 2 aims: (1) forming a judgment about the residents, and (2) faculty development. Most group members shared information without written notes, most discussions were not structured by the program director, the major focus of discussions was on residents with performance concerns, and there was a lack of a shared mental model of resident performance. The program directors who benefited most from the meetings were those who thought group members were engaged and summarized the information after every discussion.

**Conclusions** Unstructured discussions and a lack of a shared mental model among group members impede effective information sharing about resident performance with a developmental approach. Structured discussions with an equal amount of discussion time for every resident and creating a shared mental model about the purpose of the discussions and the assessment approach could enhance use of a developmental approach to assessing resident performance.

**Introduction**

Program directors are responsible for assessing resident performance. Until recently, they primarily accomplished this individually by interpreting multiple assessment data points, with or without consulting faculty. In recent years, the competency-based medical education (CBME) approach in medical education has emphasized group decision-making regarding resident performance as a new assessment approach in postgraduate medical education. Since 2013, in the United States, the Accreditation Council for Graduate Medical Education has required clinical competency committees (CCCs) to determine resident performance, and other countries are following this example.

Group decision-making related to resident performance is based on the concept that groups make better decisions than individuals if they discuss existing data and share and integrate new information, uniquely held by members. Group discussion can increase detection of resident problematic performance. Yet, the literature on group decision-making shows that reality often falls short of expectations. Biases, such as holding on to an initial opinion, believing things because others do, judgment influenced by overreliance on consensus or by emotions instead of objective data, and lack of discussion intensity can lead to ineffective information sharing and poor decision-making. This may jeopardize the validity of the judgments about resident performance.

Literature on CCCs includes guidelines to set up a meeting and reviews of literature on group decision-making that offer recommendations for maximizing the effectiveness of CCC processes. One study found most CCCs review resident performance by identifying problems, instead of using a developmental approach. The majority of studies on CCCs are conducted in the United States, where programs have acquired experience with creating and operating CCCs since 2013. It is unclear whether findings and recommendations apply to other nations and cultures. This information is relevant since many other nations are establishing CCCs.

DOI: http://dx.doi.org/10.4300/JGME-D-18-01069

Editor’s Note: The online version of this article contains a semistructured observation scheme and a semistructured interview guide.
Our study explored how group decision-making about resident performance takes place in the Netherlands, where CCCs are not yet a required component of assessment in postgraduate medical education. We also sought to understand how group discussions influence program directors’ judgment of resident performance, since the purpose of CCCs is to advise program directors about residents’ progress. We sought answers to the following questions: (1) What kind of information about residents is shared during a CCC meeting and how is this shared? (2) How does group information sharing influence program directors’ judgment about resident performance? The results provide insight into current practices and contribute to the understanding of effective group discussion about resident performance.

**Methods**

**Background and Participants**

In the Netherlands, most residents complete their postgraduate medical training in both general hospitals and university medical centers. All programs are competency-based. Residents are trained and supervised by all staff members, but program directors (PDs) are solely responsible for the assessment of residents’ progress. PDs are required to have an evaluation meeting with residents at least twice a year to provide feedback on their performance. Group meetings regarding resident assessment are not required, but most PDs organize meetings with faculty to discuss resident performance. The intent of these meetings is comparable to the purpose of CCCs in the United States: the group makes a decision about the level of performance of residents and advises the PD.

We purposefully sampled 10 group meetings in a range of medical specialties during which resident performance was discussed. We invited PDs for participation by e-mail. Based on ethnographic principles, we combined observations of program faculty meetings with interviews with the PD to answer the second research question.

**Data Collection and Analysis**

Data were collected from May to December 2017. CCC meetings were observed by 1 researcher (M.E.D. or I.A.S.), who recorded observations using a semistructured scheme based on Hauer and colleagues’ narrative review on group decision-making that scored whether and how often certain interaction occurred. Meetings were audio recorded and transcribed, and transcripts and field notes were analyzed by 3 researchers (M.E.D., I.A.S., I.C.M.), using a thematic framework analysis interaction schema. The researchers discussed findings using constant comparison until agreement was reached. Within a month of the group meeting we interviewed the PDs, asking them to reflect on the findings from our group observations, including the influence of the meetings on PD assessment of resident performance. Interviews were based on a semistructured interview guide, conducted by M.E.D. or I.A.S., and lasted about 60 minutes. They also were audio recorded, transcribed, and analyzed using a thematic framework analysis. We selected key themes by analyzing and discussing the data within the interview categories: process, content, and result. After 5 meetings and interviews and again after 10 meetings and interviews, themes were discussed in the whole research team. Saturation was reached when all data could be analyzed using the existing themes and no new themes were identified. Observation and interview protocols and tools are provided as online supplemental material.

The study was approved by the ethical board of the Netherlands Association for Medical Education. We obtained informed consent from all participants.

**Results**

Participants represented 10 medical specialties from 4 Dutch university medical centers (TABLE). Team size varied from 4 to 20 members, and all group members were program faculty. The number of residents discussed ranged from 2 to 32, and the duration of meetings ranged 22 to 96 minutes. All meetings were chaired by the PD.

The results for our research questions are presented below, and are supported by illustrative quotations from meetings or from PD interviews.
### Characteristics of Participants

<table>
<thead>
<tr>
<th>Medical Specialty</th>
<th>No. of Faculty Joining Meeting</th>
<th>No. of Residents Discussed</th>
<th>Duration of Meeting</th>
<th>Sequence in Which Residents Are Discussed</th>
<th>Time Per Resident</th>
<th>Frequency of Meetings</th>
<th>Form of Judgment Used in Meeting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obstetrics and gynecology</td>
<td>11</td>
<td>3</td>
<td>22 min</td>
<td>Random</td>
<td>“Problematic resident” more discussion time</td>
<td>3× per 10 months</td>
<td>CanMEDS competencies (general competencies)</td>
</tr>
<tr>
<td>Anesthesiology</td>
<td>10</td>
<td>32</td>
<td>96 min</td>
<td>Random</td>
<td>Youngest residents less discussion time</td>
<td>4× a year all residents</td>
<td>Assessment form developed by faculty (based on general CanMEDS competencies)</td>
</tr>
<tr>
<td>Surgery</td>
<td>8</td>
<td>15</td>
<td>65 min</td>
<td>Random</td>
<td>“Problematic resident” more discussion time</td>
<td>Once a month</td>
<td>Open questions (eg, What do you think about Peter’s performance?)</td>
</tr>
<tr>
<td>Medical microbiology</td>
<td>9</td>
<td>9</td>
<td>59 min</td>
<td>Eldest to youngest</td>
<td>“Problematic resident” more discussion time</td>
<td>Twice a year</td>
<td>Open questions (eg, What do you think about Peter’s performance?)</td>
</tr>
<tr>
<td>Pharmacology</td>
<td>9</td>
<td>4</td>
<td>65 min</td>
<td>Random</td>
<td>“Problematic resident” more discussion time</td>
<td>4× a year all residents</td>
<td>Assessment form developed by faculty (based on general CanMEDS competencies)</td>
</tr>
<tr>
<td>Geriatrics</td>
<td>6</td>
<td>4</td>
<td>30 min</td>
<td>Random</td>
<td>“Problematic resident” more discussion time</td>
<td>Once a month only residents who are problematic or have an evaluation meeting planned</td>
<td>Open questions (eg, What do you think about Peter’s performance?)</td>
</tr>
<tr>
<td>Cardiology</td>
<td>14</td>
<td>12</td>
<td>69 min</td>
<td>Random</td>
<td>“Problematic resident” more discussion time</td>
<td>4× a year all residents</td>
<td>Open questions (eg, What do you think about Peter’s performance?)</td>
</tr>
<tr>
<td>Intensive care</td>
<td>4</td>
<td>11</td>
<td>51 min</td>
<td>Random</td>
<td>“Problematic resident” more discussion time</td>
<td>6× a year</td>
<td>Open questions (eg, What do you think about Peter’s performance?)</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>8 (all faculty members)</td>
<td>3</td>
<td>49 min</td>
<td>Random</td>
<td>“Problematic resident” more discussion time</td>
<td>Every month residents who are scheduled for evaluation meeting</td>
<td>CanMEDS competencies (general competencies)</td>
</tr>
<tr>
<td>Dermatology</td>
<td>20</td>
<td>4</td>
<td>25 min</td>
<td>Random</td>
<td>“Problematic resident” more discussion time</td>
<td>Every month residents who are scheduled for evaluation meeting</td>
<td>Open questions (eg, What do you think about Peter’s performance?)</td>
</tr>
</tbody>
</table>
Sharing Information

Shared information during the group meetings had 2 aims: (1) forming a judgment and (2) faculty development.

**Forming a Judgment:** PDs were looking for divergent faculty opinions to form a broader picture of resident performance. One PD stated, “Without the faculty meeting I have no foundation [. . . ] I always imagine a picture of an elephant. I describe its tail, someone else describes its trunk, and a third person its feet. We need each other to paint the whole picture.” PDs also desired specific examples of performance to provide feedback to residents in semiannual evaluation meetings: “I need to gather ammunition with concrete examples of behavior.”

In many cases, faculty put a label on residents, uttering sentences like, “If it were my mother [who needed surgery], I would drive by really fast”; “She performs surgery like a rusty gate”; or “That is a good resident, because he does not bother me.”

**Faculty Development:** Part of the shared information was to help develop faculty by talking about the way residents should be supervised. As one PD indicated: “The meeting is a moment in which we can discuss training-related subjects.” It is also seen as an opportunity for faculty members to learn from each other: “One hears how colleagues tackle things and how they judge residents.” PDs used the meeting as an opportunity to instruct faculty. For example, we observed that when faculty mentioned an incident with a resident, PDs often insisted on completing a mini Clinical Evaluation Exercise or an Objective Structured Assessment of Technical Skills. We also observed several discussions about the assessment criteria used to judge resident performance.

**Interaction During Meetings**

Most meetings included jokes, and often there was a giggly atmosphere and a lot of laughter. Almost none of the faculty members brought written notes to the meetings, with most sharing information by heart and frequently repeating each other’s comments. Real discussions of the performance of residents were rare. One PD expressed disappointment: “They don’t listen to each other; there is no discussion at all.” Most of the time, PDs listened to the repetitive comments without interference. They also did not encourage faculty to share new information. When asked about this, a PD answered: “I don’t want to play bad cop. I don’t want to spoil the good atmosphere in the group.” Drawing conclusions or summarizing information was hardly ever done. The exception was one PD who repeatedly summarized information and drew conclusions: “With a summary I ask for consensus in the group. . . . this is what I have to feed back to the resident.”

**Influence on Judgment of Program Directors**

**Faculty Engagement:** During the interviews, PDs indicated that the group’s judgment of resident performance in meetings depended on the effort of the group members, and that members varied in how important they consider the meetings: “They check their agendas to find out what they can skip and this meeting is always the first thing to skip.” Low turnout was a problem for many PDs; “That is the flaw in forming a judgment.” PDs put a different value on the opinions of faculty: “There are faculty members I trust and others I don’t take seriously. I have more trust in colleagues who are dedicated to training residents.”

All PDs pointed out that the danger in meetings is that faculty have a tendency to echo each other’s opinions. PDs saw it as their obligation to filter opinions and only use information that seemed useful. One PD noted: “When mass hysteria arises I throw all that is said in the trash!” Another said some faculty members “constantly bring up the past, mostly a very distant and dirty past.” She explained she saw it as her obligation to ignore these negative stories.

According to the majority of PDs, meetings are an opportunity for faculty to unload their feelings about residents. As one PD illustrated, “They pour out all those opinions and that is that.” Another commented, “The meeting is a fixed moment of which faculty members know: Now I can finally tell something about that resident!” PDs noted that faculty often see the meeting as a brief time of leisure. “The meeting is a form of relaxation.” “It is about residents; it is not threatening.”

**Resident Performance:** All PDs found the meetings more influential when “problematic residents” are discussed. “We can make jokes about the good-functioning resident, but are serious about the problematic ones.” PDs explained that meetings are uncomplicated when residents function at the expected level: “If there are not too many problems, then it is easy, we don’t have to think about it. But, if there are problems, then we suddenly have to start dealing with that resident.”

This suggests residents with performance concerns received more discussion time than high-functioning residents. One PD saw this as a problem, arguing that all residents must be guided to develop to a higher level: “It is not fair, because we can also help good
residents to become excellent.” While residents with performance concerns received more discussion time, this rarely concluded with a plan of action. For residents without performance, groups did not make suggestions for how these residents could develop further.

**Program Directors’ Focus During the Meeting:** PDs indicated the main focus of the meetings was on creating a group feeling among faculty members, and that it is important to emphasize that faculty as a group were responsible for training the residents as a group. Several PDs reported they did not create much structure for the meetings due to concerns of harming the positive group feelings, adding that faculty needed an opportunity to blow off steam and express their opinions about the residents. PDs closed noting they wished the meetings were more valuable to them in forming a judgment of resident performance.

A few PDs stated the main focus of meetings was on evaluating resident performance, and that they asked faculty members to complete assessment forms prior to meetings. They found it important to create a broad picture of each resident by aggregating opinions from different faculty members and using these to provide feedback to the residents. They summarized assessment information at the end of each discussion, and reported it as useful for judging resident performance.

**Discussion**

In most meetings the assessment conversations were not structured by the PD and predominantly focused on residents with performance concerns. PDs who found the meetings useful were those who reported that faculty members were engaged, who summarized the information at the end of each discussion, and who formulated feedback to the resident.

Biases and low intensity discussions contributed to ineffective information sharing, similar to the findings of other studies. Unstructured discussions and lack of a shared mental model among group members were common attributes of CCC meetings that hinder effective information sharing and discussion of resident performance. When more information is shared, better decisions are possible, and meetings are more beneficial to PDs who have to make judgments about resident performance. To address PDs concerns structuring meetings, an option is to have the meeting chaired by another individual.

Meetings were more useful when faculty members were engaged, yet were often skipped by faculty. One reason could be that CCCs are not mandatory in the Netherlands. In contrast to some faculty members, PDs valued the meetings as important for their role in assessing residents and providing feedback.

We found that group members lacked a shared understanding of the purpose of the meeting. Group performance improves with shared mental models—a common understanding of the purpose of the group’s work, the task to be performed, and teamwork necessary to complete it. This makes it important to explain to group members the purpose of the discussions and the task they are expected to perform.

Group discussions largely focused on residents with performance concerns, similar to previous studies of CCCs. This is in keeping with a “dwell time” model of medical education, assuming that most residents will be competent at the completion of a prescribed number of years of training. In contrast, introduction of CBME emphasizes individualized learning plans and paths to competence. To fully embrace CBME, it is necessary to let go of the habit of just identifying problems in resident performance, use a developmental assessment model, and guide all residents on individual paths to becoming competent physicians. Highly structured discussions with an equal amount of time for every resident and clearly explaining the aims of assessment for all group members promote such a developmental approach.

Limitations of our study include a sample from 1 nation, and interview data limited to PDs, which may overemphasize the perspectives of this group. Future research should seek to gather data from multiple perspectives, including residents, and should assess the generalizability of our findings to other nations and cultures.

**Conclusion**

Unstructured discussions and lack of a shared mental model among group members were common attributes of CCC meetings that hinder effective information sharing and discussion of resident performance. This reduces the impact of this information on PDs’ judgments of resident performance and impedes a developmental assessment approach. Structuring discussions to ensure an equal amount of time for every resident and creating a shared mental model among
the group members will contribute to an effective developmental approach for assessing resident performance.

References


27. Hannah DR, Lautsch BA. Counting in qualitative research: why to conduct it, when to avoid it, and when to closet it. *J Manag Inf*. 2011;20(1):14–22.


**Marrigje E. Duitsman, MA, MD,** is Physician and a PhD Student, Health Academy, Radboud University Medical Center, Nijmegen, the Netherlands; **Irene A. Slootweg, PhD,** is Educationalist and PhD Researcher, Department of Public Health Education, Leiden University Medical Center, Leiden, the Netherlands; **Imke C. van der Marel, MD,** is Resident in Obstetrics and Gynecology, Department of Gynecology, Ikazia Hospital, Rotterdam, the Netherlands; **Marianne ten Kate-Booij, PhD, MD,** is Gynecologist and Director of the Residency Program, Department of Gynecologic Oncology, Erasmus University Medical Center, Rotterdam, the Netherlands; **Jacqueline de Graaf, MD, PhD,** is Professor of Internal Medicine and Director of Postgraduate Medical Education, Department of Internal Medicine, Radboud University Medical Center, Nijmegen, the Netherlands; **Cornelia Fluit, PhD,** is Associate Professor, Radboudumc Health Academy, Department of Research in Learning and Education, Radboud University Medical Center, Nijmegen, the Netherlands; and **Debbie Jaarsma, PhD,** is Professor of Innovation and Research in Medical Education, Center for Education Development and Research in Health Professions, University Medical Center, Groningen, the Netherlands.

**Funding:** The study was funded by the Dutch Federation of Medical Specialists.

**Conflict of interest:** The authors declare they have no competing interests.

The authors would like to thank the participants for their cooperation and sincere input.

Corresponding author: Marrigje E. Duitsman, MA, MD, Radboud Health Academy, Gerard van Swietenlaan 4, PO Box 9101, 6500 HB Nijmegen, the Netherlands, +31648495035, marloes.duitsman@radboudumc.nl

Received December 13, 2018; revision received April 17, 2019; accepted April 23, 2019.