Feedback is a vital component of the learning process. Because self-assessment is notoriously inaccurate,1 assessments comparing observed performance to an objective standard are increasingly being emphasized.2,3 Such objective standards have been codified in the Accreditation Council for Graduate Medical Education (ACGME) milestones project,4 which specifies summative clinical competencies through longitudinal evaluation using a developmental progression model. Milestones denote a significant point in the development of competency-based outcomes (e.g., knowledge and performance).5 The milestones can track a learner’s progress and provide feedback to the learner on the evaluated gap between observed performance and the expected standard, which may facilitate performance improvement.6 Although the ACGME mandates that resident learners receive feedback on their performance, there is little guidance on how to communicate the results of milestone evaluations to resident learners. Without clear guidelines, it is uncertain what the milestone feedback process currently is or if milestone feedback is even being provided to neurosurgery resident learners.

The goal of feedback is to hone deliberate practice to achieve expertise.7 Furthermore, feedback should be provided for learning rather than of learning.8 Thus, feedback seeks to communicate the performance progression of the learner in relation to a standard, provide the learner with information to refine and improve performance, and supply the learner with the tools to more accurately self-evaluate their performance in anticipation of future independent practice.8

While the importance of feedback has been generally accepted,9 the impact of feedback is variable. Feedback can improve or impede performance depending on its quality.9,10 Feedback is the result of a complex interaction between participants2 and is situation dependent.11 The effectiveness of feedback is influenced by the teacher, the learner, the message imparted, the character of delivery,
the learner’s receptivity to the message, and the specific situation in which feedback is delivered, yet these components remain poorly understood.

Despite its recognized importance, resident learners do not feel that they receive appropriate, timely, and useful feedback. The 2013 ACGME resident survey indicated that satisfaction with feedback was poor, although no feedback. The 2013 ACGME resident survey indicated that satisfaction with feedback was poor, although no feedback. The 2013 ACGME resident survey indicated that satisfaction with feedback was poor, although no feedback. The 2013 ACGME resident survey indicated that satisfaction with feedback was poor, although no feedback.

In this study, we document the current state of milestone feedback with Mayo Clinic neurosurgery resident learners. This study focuses on the trainee’s perception of feedback delivery and its effectiveness in improving learner performance. It is hoped that this descriptive study will serve as a baseline for future educational interventions designed to optimize the feedback process.

**Methods**

Following institutional review board approval, 10 resident learners were interviewed face-to-face by the Mayo Clinic director of program evaluation and education (K.F.), who had no previous contact with the neurosurgical trainees. All trainees who had completed at least two neurosurgical rotations and who were rotating at the primary hospital were eligible to participate in the study. Trainees were provided information regarding study goals, interview procedures, and personal metrics collected, and all eligible residents chose to participate.

Interviews were scheduled on 15-minute intervals. Each interview explored milestone feedback delivery by specifically examining the mechanics of delivery, the culture of feedback, and the evaluation of feedback through a series of questions, prompts, and follow-up investigations. (Table 1).

<table>
<thead>
<tr>
<th>Specific Aim</th>
<th>Rationale</th>
<th>Question</th>
<th>Prompts</th>
<th>Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanics of delivery</td>
<td>Please describe how milestone feedback is reported to individual trainees.</td>
<td>Who does the reporting? Where does this occur? What method of communication is used? Further feedback after initial encounter?</td>
<td>Do the trainees have suggestions for improvement in delivery method?</td>
<td></td>
</tr>
<tr>
<td>Culture of feedback</td>
<td>What is your attitude toward receiving feedback?</td>
<td>Describe engagement.</td>
<td>Do trainees feel their attitude makes a difference and how?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What do you feel is the mentor’s attitude toward providing feedback?</td>
<td>Describe attitudes. Describe willingness to give/receive. Describe intangibles (e.g., awkwardness).</td>
<td>Do trainees feel the mentor’s attitude makes a difference and how?</td>
<td></td>
</tr>
<tr>
<td>Evaluation of feedback</td>
<td>Is the feedback useful? How do you use feedback?</td>
<td>How do you assess if you make changes?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>What changes do you make in response to the feedback? Does feedback change your view of your performance/is it accurate?</td>
<td>If changes are made, how do you assess whether they were effective? More feedback? If none, why not? How lasting are the changes?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rationale</th>
<th>Question</th>
<th>Prompts</th>
<th>Follow-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Results**

In all a priori themes of definition of feedback, feedback delivery, and impact of feedback and in the de novo theme of gender bias during feedback delivery was recognized and explored.

**Definition of Feedback**

Residents defined feedback as a “critique of a performance” or an “interpretation” of performance. The majority of trainees made clear distinctions between formative
and summative feedback. Most residents considered this “informal” (formative) and “formal” (summative). All trainees valued the immediacy and specificity of formative feedback and noted that this led to conscious behavioral changes.

Two distinct forms of summative feedback were described: attending-driven and milestone-driven. Some attendings traditionally have a face-to-face assessment with individual trainees on a quarterly basis to assess overall performance, which was valued, but not to the extent as formative feedback. All attendings provided written summative comments that were appreciated but noted to be nonspecific, with little advice for practice change.

Milestones were understood to theoretically provide objective measures of progress, but utility for residents was universally panned. Many recognized that the goal of the milestone program was to provide benchmarked feedback on progression and a snapshot of their current state, although they did not feel that the milestones provided useful personal data. For instance, one noted, “I think [the milestones] are a little flawed as they assume a linear progression, which I don’t think is necessarily accurate.” However, the majority considered them “boxes that need to be checked” but did not “clearly correspond to how I am developing as a neurosurgeon.” Another interviewee noted:

My personal opinion is that the drive towards more objective rubrics for standardizing how we are progressing is not really imparting value to me as a learner … it’s more paperwork for someone to fill out; it produces data that … people like you can study.

Feedback Delivery

Mechanism of Feedback

Residents noted multiple feedback delivery methods, including formative and immediate feedback, face-to-face moderately structured summative feedback, and milestone feedback. Key distinctions in feedback type were made based on immediacy and formality, with immediately relevant, specific, and personal feedback being considered most valuable. It is notable that milestone feedback was provided online using MedHub; thus, while the feedback was accessible, it was not personally presented or interpreted for the resident.

Culture of Feedback

Receptivity of feedback was probed. All trainees expected to receive feedback and knew what mechanisms were used to deliver feedback. Many noted that positive feedback was helpful to reinforce specific activities when given formatively. Summatively, most noted that positive feedback occurred with comments such as “good job,” which were warmly received but of no learning value. Residents focused on receiving positive feedback and pearls of wisdom to improve. In many cases, despite questionable learning value, one resident noted that it was nice to receive positive feedback to realize the work performed was “noticed.” Only one resident commented on requesting feedback, but noted, “a lot of people say, ‘please give me feedback,’ and what they mean is don’t give me negative feedback.”

In contrast, the receptivity to negative feedback was cautious and varied. More senior residents were receptive and reflective when presented with negative feedback. However, unlike positive feedback, all residents noted that they considered the veracity of the feedback before internalizing it. Feedback given because of a known mistake was accepted, while feedback that was perceived as erroneous or stemming from staff preferences for different but potentially equivalent situations was harder to incorporate into the learners’ worldview. Residents commented that if validated, learning as a result of negative feedback was internalized more easily than learning from positive feedback; “That kind of feedback doesn’t require a secondary reminder … I learn from my mistakes.” Another noted, “It is very easy, even in the next case, to start trying to practice what you were told.” A different trainee noted that negative feedback was “like a lightbulb,” prompting major change.

Feedback was consistently described in unidirectional terms from staff to trainee. There were no comments regarding questioning or clarification. No trainee commented on feedback as a conversation. Positive feedback was accepted with little question, while negative feedback was viewed with greater skepticism.

Only one senior resident noted delivering feedback to another trainee. The trainee used a face-to-face approach and pointed out that a behavioral difference was noted and sustained in the mentee. No other resident commented on providing feedback to fellow trainees or staff.

Quality of Feedback

Trainees reported that the quality of feedback was variable and dependent on staff and delivery method. Immediate face-to-face feedback was considered the most valuable because of its specificity and proximity to the behavior discussed. Summative feedback presented face-to-face with specific staff was deemed especially valuable.

Summative feedback delivered in the form of short written statements was generally considered devoid of learning: “It’s two or three sentences per quarter.” Common examples of summative feedback included “a pleasure to work with,” “good job,” and “should continue reading.” Specific areas of improvement or examples of performance were rarely noted. Finally, most felt that the milestone feedback was not of good quality. Because of the “limited window into your abilities as a learner,” inconsistencies in reporting, and categories that did not mesh well with experiences, the data were viewed poorly. A resident noted, “I could rate myself very accurately with those checkboxes … more accurately than the [attendings].”

Barriers to Feedback

The most cited barrier to feedback was the veracity of information. Interrater variability was noted, in that some raters used this scale to represent progress through residency (as it had been intended), and others used it to rate the performance of the resident at their PGY level. For instance, some raters would give a well-performing PGY-2 resident a grade of 2 or 3, reflecting their current abilities in relation to the aspirational milestone grade expected at graduation, while others would give a grade of 5 as a
reflection of excellent performance by a PGY-2 resident on that rotation. Trainees consistently felt that attending perception was inaccurate or random. One resident noted that use of the milestone system may result in assessment using limited data, feeling that staff were “going to end up either checking not applicable or just checking a random number.”

This ties in to the second most common barrier, that the categories evaluated may not be relevant to the rotation experiences. Trainees found milestone information difficult to understand in the context of their exposure to cases and experience. Thus, they found milestone performance gaps difficult to interpret. Third, residents noted that feedback was not specific enough to provide meaningful improvement, especially summative and milestone evaluations. Too little time spent providing feedback, a trainee’s willingness to receive feedback, and cultural assumptions were also noted.

Feedback Impact on Performance

One of the trainees stated, “One of the hardest things for human beings to change is their behavior.” All residents felt that immediate, in-person feedback provided the greatest opportunity for learning and practice change. One declined to call it feedback, noting, “I call it teaching.” The majority noted that summative feedback was often too general to result in them changing their practice and simply benchmarked their performance against that of their peers. Noting an inability to recall a time when feedback changed their behavior, one trainee remarked, “The reason is, most of the feedback is from these forms where they just check numbers.” Generally, feedback was described as unidirectional and concrete. Several residents note that they document and review formative feedback, often coding pearls to specific operative cases, or have a mental checklist they can quickly review if such a surgery is done again.

An exception was senior residents who were supervised but relatively independent in practice. They emphasized active participation in feedback, including reflection and deliberate steps to achieve suggested practice changes. They noted that part of the reflection included assessing why actions were perceived in a specific fashion, contemplating the veracity of the feedback, and ultimately choosing whether or not to incorporate that feedback into practice. There was an understanding that various treatment paradigms would be presented by the breadth of attendings, and they would ultimately choose a practice path based on their own knowledge and experience.

Additionally, senior residents noted an emotional component to feedback, which was often the most impactful despite being trying at the time. One noted that feedback caused them to question their knowledge, skill, and decision to pursue neurosurgical training at times, but upon self-reflection, including documentation of suggested improvements, felt that it was ultimately rewarding and provided an ability to internalize change. One described a particularly rough episode:

My staff was clearly disappointed in my performance: after the case [the staff] immediately set [sic] me down and told me of his disappointment and how I could have done better. And of course, at the time ... you think, I’m going to get fired. But especially as the years have gone by and I’ve reflected on that moment, it was ... a critical moment in my training that was very effective for my learning, and that same staff has become one of my closest allies over the years. So, while that was hard to hear, it was very effective feedback.

In contrast, while all residents noted the disappointment of negative feedback, junior residents justified their response. A trainee noted, “one of the natural responses to feedback is to feel defensive, and if you feel anger, the anger is justified.” However, no junior resident noted a positive result from their emotional response.

Gender Bias

Open-ended interviews noted a persistent de novo theme of gender bias. All 3 female residents and none of the 7 male residents were told they “lacked confidence,” a statement the women felt was untrue and insulting. Furthermore, 2 of the 3 females and none of the male residents perceived that attendings avoided negative feedback because of the presumed reaction by the resident, with one noting that the attendings were fearful that “I’m going to get mad at them.” The trainee elaborated:

It’s like oh everything’s fine, like when you ask, you know everything’s fine; and then you hear the rumbles, like oh well they did this and that annoyed me, and this annoyed me, and you know it’s like well you need to say that.

No specific themes manifested differences in the perception of feedback based on PGY level or trainee age with regard to culture or bias in presenting feedback. No resident commented on racial bias, although this was not probed specifically.

Discussion

Neurosurgical training is increasingly turning toward competency-based medical education (CBME) paradigms. The goal of CBME is to focus on outcomes, emphasize abilities, promote learner-centeredness, and de-emphasize time-based training.21 The milestone evaluation system has outlined the core competencies for neurosurgical training as reflected in the Matrix Curriculum.22 Within each domain, a trainee’s abilities are reflected in a progression from novice to master on a 5-point scale. Milestone reporting therefore serves as summative feedback of trainee progression to competence.

Shifting training from a time-based to a competency-based system changes the role of summative feedback. In a time-based system, summative feedback serves as a “safety check,” but it is generally assumed that a trainee will be adequately competent at the conclusion of their training. In contrast, in a competency-based system, summative feedback serves as a gatekeeper. Until a prespecified level of competency is achieved, a trainee cannot move forward in their training. For a successful transition to competency-based education, learners must be confident that the summative evaluation used to determine their ability to progress is robust and fairly applied.

However, trainees do not perceive the milestone system as adequate for assessing progression in a CBME model or in their professional development.14 Milestone feedback was intended to serve as an objective method of bench-
marking and identifying performance gaps in multiple domains. However, this study confirmed that trainees perceived a lack of timeliness, vague categories misaligned with experiences, erroneous data, and rater variability, all of which provided limited opportunity for personal growth. Milestones were seen as a box-checking enterprise without educational significance. Trainees valued face-to-face summative feedback more than milestone feedback. Universally, trainees contrasted milestone data with immediate formative feedback. Formative feedback was felt to have specificity that could be acted on in future situations. As organized neurosurgery continuously refines the milestones, two important things must change. First, trainees must perceive the feedback to be accurate and fair to improve individual practice and be viewed as an accurate benchmark. Second, the content of the feedback must be explicit enough to provide the trainee with meaningful information to identify learning gaps and implement behavioral change.

To address the validity of feedback, suboptimal interrater reliability must be improved. Evidence suggests that increasing familiarity with the milestone system has resulted in more consistent feedback. Enriched narrative descriptors in milestone competence grading provide clarity for raters and trainees and counter the tendency to grade based on expected norms of a particular training stage. A complaint noted by the interviewed trainees and identified in previous studies of neurosurgical residency programs. In a study involving radiology, urology, and emergency medicine residents, 1st-year residents received lower milestone scores during progressive years without a decline in in-service examination scores. This may indicate that clinical competency committees became more comfortable appropriately assigning lower ratings where appropriate, improving the quality and consistency of feedback. To further improve the consistency of ratings, national-level resources could be considered to harmonize the approach to milestone feedback.

Clear milestones with enriched narrative descriptors provide meaningful information to both the trainee and rater. The interviewed residents found it difficult to align the milestone information with their clinical experience and felt that this similarly impacted the ability of the rater to assess their competency. Misalignment between clinical experiences and milestone progression targets has been previously identified as a major barrier to implementation of the milestone system in neurosurgical training programs. Indeed, despite the milestones’ intention to provide greater fidelity than an overall grade of general performance, one study noted that 97% of the variance in competency grades has been attributed to the evaluation of “overall resident competency” or the rater’s general gestalt of performance.

Consideration should be given by national neurosurgical education organizations to improve the feedback process. The developmental progression model of milestone feedback and use of the Matrix Curriculum should continue to be emphasized to improve interrater reliability. Further tools to improve assessment in a transparent fashion have demonstrated utility in feedback veracity, consistency, and trainee internalization. One such tool is the Zwisch scale to assess intraoperative performance. Additionally, feedback is not one-sided. Residents should also receive tools to better utilize feedback and participate in the feedback process. This includes bidirectional discussion of performance and the ability to provide attendings with feedback on their educational delivery.

Disappointingly, in this study all of the interviewed female trainees perceived gender-based deficiencies in feedback. All females noted that they were told they were “not confident,” while none of their male peers reported the same feedback. Indeed, it has been shown that women are generally evaluated more harshly on traits, such as confidence, that are deemed masculine. Furthermore, all of the women also felt that the assessed lack of confidence was erroneous. Discordant feedback involving stereotypically masculine traits such as confidence not only promotes gender bias but also undermines attending credibility and trainee acceptance of valid feedback. Additionally, emphasis on masculine traits may increase stereotype threat among female trainees, ultimately harming performance.

The female trainees interviewed also perceived that the feedback they received was milder than that received by their male counterparts. They felt this was because of the attending’s reluctance to hurt their feelings and perception that the women were more emotionally fragile than their male counterparts. Extrapolating, this perceived gender bias may indicate that female residents are receiving poorer-quality instruction and mentoring. This study reinforces previous research demonstrating that evaluators have difficulty assessing competency independent of gender, which ultimately impacts milestone attainment. Further improvements in assessment and feedback must occur to reduce gender disparities in neurosurgical education.

This study was limited in its scope. In studying feedback, it is important to consider the faculty perspective on feedback and the overall resident performance. This study focused only on resident perceptions of feedback; however, faculty perspectives are necessary to complete the picture, specifically, the faculty’s impression of the culture of feedback, the willingness of trainees to accept feedback and modify their behavior, and the faculty’s own willingness to model this behavior by receiving feedback regarding their own performance. Additionally, the perceptions of both faculty and trainees should be correlated with overall performance and educational trajectory. Despite reaching saturation within a single center, exploring the themes at a national level would aid in future milestone improvements and inform the construction of tools to improve the feedback process. Although a nondepartmental interviewer was used, all trainees interviewed were aware that data collected would be studied by a faculty member. Should this study be repeated at other institutions, it would be worthwhile to have outside reviewers, as this may increase trainee openness.

**Conclusions**

Trainee perception of feedback in a single-center neurosurgical residency program exposed multiple areas of improvement. Summative and milestone feedback were criticized as being vague, misaligned with practice, and
often perceived as erroneous. Barriers to implementation of summative feedback included high interrater variability and the inconstant adoption of a developmental progression model. Gender bias was noted in the degree of feedback provided and language used. This paper can serve as a baseline to study improvements in the milestone feedback process and optimize learning.

References


**Disclosures**

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

**Author Contributions**

Conception and design: both authors. Acquisition of data: both authors. Analysis and interpretation of data: both authors. Drafting the article: Clarke. Critically revising the article: both authors. Reviewed submitted version of manuscript: Clarke. Approved the final version of the manuscript on behalf of both authors: Clarke. Study supervision: Clarke.

**Supplemental Information**

**Previous Presentations**

Portions of this paper were presented orally at the 2021 Mayo Clinic–Karolinska Institutet Scientific Meeting, September 20–21, 2021, held virtually.

**Correspondence**

Michelle J. Clarke: Mayo Clinic, Rochester, MN. clarke.michelle@mayo.edu.