Development of the GPSE: A Tool to Improve Feedback on Procedural Skills in Residency

Melissa Nothnagle, MD, MSc; Shmuel Reis, MD, MHPE; Roberta Goldman, PhD; Agnes Diemers, MD, MSc

Background and Objectives: Learning procedural skills requires supervised practice with feedback. Observation of performance provides an opportunity for feedback, but the quality of feedback is often poor. We developed an instrument to improve feedback during procedural skills training in residency.

Methods: The Global Procedural Skills Evaluation (GPSE) was drafted based on qualitative data from field notes and literature on procedural skills training. Interview and focus group data described the current state of feedback and guided revisions to the instrument. An iterative process incorporated evidence from education research and input from teachers and learners to optimize the credibility and dependability of the instrument.

Results: The final GPSE includes a self-assessment prompt, five criteria (items) assessing performance, a rating scale quantifying the degree of intervention by the teacher, a global assessment, a rating of case difficulty, and suggestions for improvement.

Discussion: The GPSE is designed to improve feedback during procedural skills training and encourage learner self-assessment. Its development process may be of interest to those who use or produce performance assessment tools.

(Fam Med 2010;42(7):507-13.)

Supervised practice with feedback and formative assessment is essential for learning procedural skills. Feedback and formative assessment refer to information about performance that is intended to guide learning. Effective feedback on procedural skills provides unambiguous information about discrepancies between actual and desired performance. However, feedback in practice is often vague and evaluative (e.g., “good suturing”). Providing explicit criteria and tools to structure teachers’ observations can increase the precision and quality of feedback and thereby improve procedural skills learning.

In addition to feedback, reflection is also essential for procedural skills learning. Reflection involves “returning to an experience to examine it, deliberately intending that what is learned may be a guide in future situations, and incorporating it into one’s existing knowledge.”

Research emphasizes the importance of guidance during reflection; tools promoting structured feedback after direct observation may create opportunities for guided reflection.

Numerous tools exist for procedural skills assessment. Procedure checklists have been widely used to assess performance but must be created for each procedure and updated as techniques evolve. An alternative to checklists is a global rating scale, which includes dimensions of performance common to all procedures. Used to assess a single observed performance, global ratings have demonstrated superior psychometric properties to checklists in assessing performance in the simulation laboratory.

One such scale, the Objective Structured Assessment of Technical Skills (OSATs), is designed to assess surgical trainees’ performance. The original OSATs includes a task-specific checklist and a seven-criteria global rating scale. The tool has acceptable psychometric properties in the simulation laboratory but has not been widely studied in work settings. A feasibility study in one residency reported low rates of faculty completion; further, because assessors did not complete the forms immediately, the OSATs did not facilitate timely feedback. The OSATs was designed for surgical settings and does not include criteria.
such as pain management and communication that are important to assessment of procedures involving conscious patients.

To address limitations of the OSATs, British educators developed the Direct Observation of Procedural Skills (DOPS). Its 11 criteria include skills such as informed consent, analgesia, and communication. Assessors rate each criterion on a 6-point scale, comparing observed performance with expected performance at a given level of training. A feasibility study found that although most users qualified the DOPS as practical, only 33% of trainees completed any forms; lack of time was a key barrier. Simpler assessment tools may help overcome these feasibility problems in busy clinical settings.

Family medicine residency involves learning to perform procedures on conscious patients. Supervision of procedures is required, but assessment and feedback are not standardized. Because of the aforementioned problems with existing instruments, we developed a new procedural skills feedback tool, using a qualitative approach to optimize its properties. Citing the weaknesses of the psychometric model when applied to clinical assessment tools, several authors have advocated using the qualitative concepts credibility (analogous to validity) and dependability (analogous to reliability) to evaluate performance assessment tools.

Credibility depends heavily on the development process. For example, checklists or rating scales should embody the essential information about the performance being assessed. Content experts can identify the essential aspects of the skill in question, assure that the criteria included are clear and meaningful, and determine whether the assessment includes all the relevant criteria. The dependability of such a tool refers to how consistently and appropriately the assessors rate a given performance; this depends on inherent factors, such as clarity of the criteria and rating scale, and aspects of implementation, such as assessor training and adequate sampling.

This paper presents findings from a study in which we used qualitative research methods to design a tool, the Global Procedural Skills Evaluation (GPSE). It was designed to improve feedback based on direct observation of residents.

Methods

We conducted a multi-method qualitative study, collecting and interpreting data from multiple sources using an iterative process to develop and refine an instrument for procedural skills assessment. This process of triangulation, synthesizing data from multiple sources or multiple methods of inquiry, facilitates verification of findings in qualitative research.

Setting

The study was conducted at a family medicine residency based at a northeast US urban community hospital. The residency program has 39 residents and 13 full-time and 17 part-time faculty members. The hospital’s human subjects committee provided ethical approval, and all participants provided informed consent.

Design

Figure 1 shows the steps in the data collection and the iterative development process of the instrument. Following a review of the medical education literature on procedural skills training, the primary researcher recorded detailed field notes while supervising residents performing procedures. These notes, based on 12 patient encounters, focused on current state of feedback during procedural skills training, key elements of procedural skills, and logistics of providing feedback in the clinical setting.
Based on the literature review and review of field notes, an initial draft instrument was created. This incorporated features of existing tools, with fewer criteria and a simpler rating scale, and included a prompt to elicit learner self-assessment.

The instrument was then iteratively refined using data from 10 interviews and three focus groups with teachers and learners. All focus groups and interviews were conducted by one researcher using a moderator’s guide. Interviews and focus groups were audiotaped and transcribed.

Data collection with each method continued until saturation was reached. To increase reflexivity during analysis, emergent findings were shared with colleagues and expert coauthors.

Interviews included five family medicine residents and five family medicine faculty who were approached after performing or supervising a procedure. All those approached agreed to participate in separate, brief, semi-structured interviews focusing on the feedback given or received, important components of feedback and criteria for assessing procedural performance. Interviewees reviewed the current draft of the GPSE for clarity and comprehensiveness and provided their critique and suggestions for revisions. Interview data were used to revise the instrument before the next iterative cycle.

Additional data were collected using three 1-hour focus groups. The first was conducted at a national conference on procedural skills training, where seven of the 17 attendees (experienced family medicine educators) volunteered to participate. Two other focus groups involved five residents and five faculty members from the residency, purposefully recruited to include residents from all 3 years and with a range of interests in procedures. Teachers also had a range of interest and experience in teaching procedures. Participants reviewed the current version of the GPSE and commented on its comprehensiveness, clarity and independence of the criteria, usefulness of the rating scale, and feasibility. Revisions to the instrument incorporated findings from each focus group, and we conducted ongoing review of the medical and educational literature relevant to skills training to assure that decisions made during the revision process were supported by theory and research.

**Data Analysis**

Before the final revision of the GPSE, a formal analysis of all transcripts was conducted to ensure that all data were considered in the final version and to reexamine decisions made in the development process. We used a variation of the framework approach described by Pope and colleagues, beginning with familiarization with the data and identification of key concepts and recurring themes. One researcher identified themes in the data and from the skills training literature and created a framework for analysis. A second individual with expertise in skills training and qualitative research analyzed the transcripts independently. Through discussion the two agreed on a final thematic framework (Table 1).

---

**Table 1**

<table>
<thead>
<tr>
<th>Current Feedback and Supervision During Procedure Training</th>
<th>Documentation: procedure note, appropriate billing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired feedback: higher quality and more structured</td>
<td>Additional Elements to Record:</td>
</tr>
<tr>
<td>Timing: before, during, and after procedure</td>
<td>Previous experience (number of times previously performed)</td>
</tr>
<tr>
<td>Include positive feedback and suggestions for</td>
<td>Learner self-assessment</td>
</tr>
<tr>
<td>improvement</td>
<td>Global assessment</td>
</tr>
<tr>
<td>Time constraints and contextual stressors reduce</td>
<td>Positive feedback</td>
</tr>
<tr>
<td>or delay feedback</td>
<td>Suggestions for improvement</td>
</tr>
<tr>
<td>Learner preparation needs improvement</td>
<td>Difficulty of case</td>
</tr>
<tr>
<td>Learner self-assessment is rarely encouraged</td>
<td>Instructor self-assessment/experience with procedure</td>
</tr>
<tr>
<td>Scaffolding: Instructor often performs part(s) of</td>
<td>Rating Scale</td>
</tr>
<tr>
<td>procedure, providing less support as trainee’s skills</td>
<td>Degree of readiness for independent performance</td>
</tr>
<tr>
<td>increase</td>
<td>Degree of intervention by the instructor</td>
</tr>
<tr>
<td></td>
<td>Include “unable to evaluate”</td>
</tr>
</tbody>
</table>

**Implementation**

Time required may or may not be excessive (no consensus)
Paper versus electronic form (no consensus)
To promote completion:  
- include fewest meaningful criteria/items  
- use only for complex or risky procedures  
- complete immediately after encounter  
- require for promotion or credentialing  
- written self assessment may not be feasible
Instructor training:  
- brief training needed, emphasize eliciting learner self-assessment  
- include instructions on the form

**Anticipated impact on teaching/learning:**

Improved feedback (timely, more comprehensive)
Improved supervision
Improved performance (learners cued to prepare, discuss risks)

**Potential added benefits**

Tracking of procedures performed
Assessing readiness for independent performance
Data were indexed according to this framework, and researchers critiqued the GPSE based on this analysis and reached consensus on final revisions.

**Results**

“Learners” refers to resident participants, and “teachers” refers to faculty participants. “Participants” and “respondents” refer to both residents and faculty.

**Current State of Feedback**

Participants reported that current feedback on procedural performance is sometimes inadequate, especially due to time constraints. Teachers want to give, and learners want to receive, more structured and comprehensive feedback, including what was done well and suggestions for improvement. Both learners and teachers felt that feedback before, during, and after the performance was useful.

Teachers reported that residents are sometimes unprepared, not knowing the anatomy or steps of the procedure. Participants noted that teachers often provide verbal guidance or hands-on assistance, sometimes even taking over the procedure to maintain patient safety and quality of care.

Participants discussed key aspects on which feedback is needed most, grouping them into the following categories: preparation, medical knowledge, technical skills, communication/counseling, pain management, and documentation.

**Suggestions for Assessment Tool**

Participants felt that a rating scale quantifying the degree of intervention by the teacher was a good indication of the learner’s readiness to perform independently and was relatively easy to measure. They felt that the rating scale should include “unable to evaluate” in case some portion of the performance was not observed. Some participants suggested adding the number of times the learner had performed the procedure previously; however, some teachers felt this could bias their ratings and might be difficult to accurately quantify.

Teachers recommended including a self-assessment by the learner, reminding them to elicit this before providing feedback. Participants suggested that a global assessment of the performance be included, providing information beyond the individual criterion ratings. Learners favored including the teacher’s rating of the difficulty of the case, expressing a sense that this would make the assessment fairer.

In addition, some participants suggested including a teacher self-assessment of confidence with the procedure. Learners noted that teacher competence affected teaching and feedback quality. However, both teachers and learners acknowledged that teachers might be reluctant to document lack of confidence. Moreover, teachers pointed out that they would not supervise a procedure they did not feel confident performing.

**Feasibility**

Most thought the instrument was short enough to be completed rapidly. A few individuals expressed concerns that due to the time pressures of clinical teaching, faculty compliance would be a problem. Experienced residency directors expressed opposing views on feasibility.

To increase feasibility, both groups recommended including the smallest number of meaningful items, requiring teachers to complete the form immediately after the encounter and requiring completed forms for promotion or credentialing. Because of the time needed to complete it, teachers recommended against using this tool for very simple procedures. Some were concerned that written self-assessments by learners would not be feasible; they suggested offering flexibility, such as using it as a prompt for a verbal self-assessment. There was disagreement in the groups about whether to implement the tool on paper or electronically; regardless of the medium, access to the form during supervision was essential, so that it could guide supervision and be completed immediately.

Teachers felt that only brief training would be needed to use this tool. They recommended emphasizing the learner self-assessment, as most teachers are not in the habit of eliciting this before giving feedback. They suggested including instructions on the form itself.

Both teachers and learners predicted that use of this tool would improve procedural skills teaching and learning. They expected that it would make feedback more timely and comprehensive and improve the level of supervision. Teachers noted that it may improve learner preparation by setting higher expectations. Participants mentioned potential added benefits of this tool, such as tracking procedures performed and determining readiness for independent performance through review of multiple assessments of the same procedure.

**Development of the Assessment Tool**

To minimize the number of criteria included, teachers argued that preparation and medical knowledge could be combined; if learners lacked knowledge related to the procedure, they had not prepared adequately. Procedure documentation was excluded as a criterion since it often occurs hours later, precluding immediate feedback. The final draft of the instrument includes a total of five criteria and a global rating of the performance (Figure 2).

The rating scale reflects the amount of intervention by the teacher, who must quantify the extent to which he/she assisted on each aspect of performance. If no intervention was needed, the learner was performing independently. Like our participants, experts on feedback advocate eliciting learners’ self-assessments before delivering feedback, to promote reflection on their performance and allow them to voice their perceived strengths and deficits; a self-assessment prompt was
Figure 2

Global Procedural Skills Evaluation Form

Learner ___________________________ Date ___________________________

Instructor ___________________________ Procedure ___________________________

**Learner self-assessment:**
What did you do well?

What could you improve?

**Instructor assessment**

<table>
<thead>
<tr>
<th>Significant guidance provided</th>
<th>Some guidance provided</th>
<th>Performed independently</th>
<th>Unable to evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Preparation &amp; Medical knowledge</td>
<td>Indications, complications, patient positioning, relevant anatomy, equipment, steps of procedure, follow-up plan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Technical skills</td>
<td>Instrument handling, aseptic technique, efficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Attention to patient comfort</td>
<td>Appropriate analgesia, response to patient discomfort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Communication</td>
<td>Informed consent, communication during procedure, post-procedure instructions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Self-awareness and seeking help</td>
<td>Recognizes limits of own skills, seeks help appropriately</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Global assessment of today’s performance of this procedure:**

- Hands-on or verbal guidance provided with most aspects
- Guidance provided with some aspects
- Minimal guidance provided
- Performed independently

**Difficulty of this particular case:**
- Average
- Unusually difficult

**Suggestions for improvement:**

______________________________
Instructor signature
Conclusions

The GPSE is designed to help teachers take advantage of often-missed opportunities to provide detailed feedback and promote reflection on performance, by eliciting learners’ self-assessments and providing specific structured feedback. The instrument reflects the judgment of experts and future users of the tool who were involved in its development, as well as the theoretical basis of procedural skills learning. By including input from learners and teachers, reducing the number of criteria, and simplifying the scale, we addressed potential causes of feasibility problems encountered with existing procedural skills assessments. Several unique features emerged during the GPSE development process, including a prompt for self-assessment, a novel rating scale, and the choice of criteria included.

In light of growing consensus that reflection is the means by which feedback is assimilated into existing knowledge structures, facilitating reflection may be as important as providing high-quality feedback. Eliciting learners’ self-assessments after performance is one way to foster reflection. A study of feedback using the Mini-CEX (which does not include a self-assessment prompt) found that teachers usually did not elicit learner self-assessments before giving feedback. We anticipate that including a specific prompt to encourage learner self-assessment will help promote this activity. Recent reformulations of the concept of self-assessment discuss the importance of reflecting on external feedback to improve performance. Assessors will need training to facilitate learners’ reflection on feedback provided using the GPSE.

Besides retrospective reflection on a completed performance (reflection on action), reflection in action is essential to procedural skills training; this type of reflection occurs during performance, in response to novel situations or problems. This ability to self-monitor, recognize the limits of one’s competency, and seek help appropriately is essential for safe independent performance, as there will be no supervisor to assist or intervene during a procedure by providing additional patient counseling, revising the equipment setup, advising a change in hand position, or taking over completely when the learner cannot safely complete the procedure. Scaffolding must fade over time to support transfer of skills to independent performance.

Repeated assessments of a learner’s performance using the GPSE could document fading of scaffolding over time as skills improve.

To increase the fairness of the assessment, learners suggested including the teacher’s rating of the case difficulty. In addition, case difficulty may be important if serial assessments are to be used to determine readiness for independent performance. It would be desirable to know that a learner has encountered difficult cases and managed them appropriately or recognized the need to seek help.

Limitations

This study has some limitations. Because of the specialized nature of the inquiry, one researcher conducted the interviews and focus groups and made initial revisions to the instrument. Potential consequences of this choice include less willingness by participants to criticize the GPSE and influence of investigator biases on the development of the GPSE. Strategies used to increase reflexivity included reviewing the developing analysis with an independent expert.

Transferability of these findings should also be considered when interpreting our study. In the United States, family physicians provide many office procedures, residents usually receive direct supervision when performing procedures, and the medico-legal environment limits independence on the part of learners. Whether these factors affect relevance of this instrument outside the United States will need to be assessed.

Another limitation is that we developed this tool to assess moderately complex office procedures. It may or may not be useful for simpler or more complex procedures or for procedures in inpatient settings.

Final Comments

Our approach in developing the GPSE is not specific to procedural skills assessment; a similar process could be used to develop other performance assessment tools. A process that incorporates education theory and input from content experts increases a tool’s credibility.
Questions and suggestions from teachers and learners during the development process increase the clarity of the criteria and the rating scale. To optimize dependability of performance assessment tools, faculty training is needed prior to implementation.

Using qualitative research methods, the GPSE was designed to improve procedural skills learning by improving feedback and encouraging learner self- assessment and reflection. The development process, which incorporated the theoretical basis of procedural skills learning with the judgment of expert teachers and input from users, can be applied to developing other tools for performance assessment.

Acknowledgments: This paper was presented at the annual conference of the Association for Medical Education in Europe, August 2009, Malaga, Spain.

Corresponding Author: Address correspondence to Dr Nothnagle, Brown University, Department of Family Medicine, 111 Brewster Street, Providence, RI 02860. 401-729-2236. Fax: 401-729-2923. melissa_nothnagle@brown.edu.

REFERENCES