Direct Observation Reassessed

Kathryn M. Andolsek, MD, MPH, FAAFP (@GME_MD)
Deborah Simpson, PhD (@debsimpson3)

The Challenge

Direct observation is embedded in the clinical learning environment, as faculty physicians supervise trainees to ensure high-quality patient care and assess their readiness for progressive autonomy. Direct observation provides opportunities for the *assessment of learning*—what a resident does in real time with real patients. It also can be used to *assess for learning*, to identify gaps between what a resident does well and what he or she needs to improve. As part of direct observation, faculty physicians typically are asked to complete an assessment based on their observation of a resident's performance in a real-life clinical context; faculty "are" the measurement tool.[A]

Rip Out Action Items

Program directors should:

1. Educate faculty, residents, and staff on the importance and practice of direct observation.
2. Mentor faculty on their role as assessor and how to best observe, judge, and represent their judgment through narratives and rating scales.
3. Simplify the process of direct observation and associated rating forms.
4. Clarify and distinguish the formative role ("How is the resident doing this activity at this moment?") from a summative assessment that uses multiple ratings, along with other data, to make judgments about trainee progression in the program.

What Is Known

Direct observation is used in 90% of graduate medical education programs,[1,B] and the education community can improve its value. Addressing the following common issues[C–E] can reduce measurement errors[2] and enhance the accuracy of direct observation:

1. *Rating tools* are often too long, include ambiguous items, or do not assess what is intended.
2. *The rating process* is cumbersome, time consuming, includes too few raters, or is not integrated with the program's larger assessment system.
3. *Raters* may have bias about a resident's performance based on prior experiences,[3] may delay completing or fail to complete ratings, not understand the assessment process, or lack competence themselves in the skill being assessed.[F,G]
4. *Limited variance* in ratings may occur, with overall high ratings or identical scores across different items, due to a lack of rater training or because rating forms are too long.

Effective forms for direct observations can be quite simple. For example, the Zwisch scale, which has 2 questions and a measure of case complexity, correlates well with longer instruments,[4] and a single-item rating of operative performance is equivalent to longer scales.[5] Also, narrative descriptors outperform numeric scales.[6] Sampling the *right* number of faculty is key: using 7 to 10 different faculty members minimizes rater idiosyncrasies and is feasible.[7] In sum, it is important to align the direct observation process with how faculty form judgments to simplify the process and to gain as many assessments as possible.[8,H–M]

How You Can Start TODAY[9,N,O]

Partner with your faculty to enhance their use of direct observation. Help faculty refine their role of teacher to include performance assessor or judge. Provide intermittent educational boosters and reminders about key features of direct observation.

1. *Acknowledge that we all have biases* by anticipating the typical sources of bias to mitigate them.
2. *Ask faculty to assess the trainee's performance observed during a specific, recent encounter* (*not* the resident's performance over the preceding day/week/month/rotation or a past experience with the trainee).
3. *Emphasize frequent, brief assessments* using a few items and narrative comments (including strengths) to improve identification of performance issues.
4. *Seek assessments from multiple raters* for summative judgments in order to limit bias.
5. *Leverage* opportunities for direct observation when faculty are already directly supervising residents.[P]
6. *Complete ratings within 72 hours of an encounter* as after that time the rating accuracy decreases.[Q]
7. *Simplify* rating forms and associated processes.

What You Can Do LONG TERM[9]

1. *Continuously build a culture that values direct observation*. Make the process transparent to both

DOI: http://dx.doi.org/10.4300/JGME-D-17-00328.1

Editor's Note: Numbered references refer to citations in the reference list. Lettered references (A, B, C . . .) are available as online supplemental material.

residents and faculty. Start by identifying several committed faculty champions.

2. *Involve faculty in cooperative discussions* to identify opportunities for improving its use, since the faculty are the "instrument" of assessment. Do faculty differentiate between direct observation as formative versus summative assessment, and how does this influence their assessment?
3. *Use faculty champions* to assess the strengths and opportunities in a real-time process. Do raters assess what they actually observe? Do the residents and Clinical Competency Committee consistently receive direct observation information? If so, how useful is that information?
4. *Review direct observation assessment forms*. Do other programs in your specialty or other programs at your institution use a simpler form or process? Identify which information is most needed for resident progress decisions, and design the assessment form to capture that information while deleting other items. If 1 type of form is more useful, can you eliminate the others? Always leave space for free text, and ask faculty to describe trainees' strengths and provide suggestions for improvement. When available, incorporate items that correlate with clinical outcomes.
5. *Prioritize 1 or 2 direct observation improvements* such as to increase its use on a particular rotation, or to simplify a form to capture its assessment. Develop an intervention that outlines goal(s), specific action plan(s), time line(s), and how success will be measured. Monitor to ensure that the intervention meets its goals.
6. *Develop a reliable, feasible, simple system* to facilitate assessment form completion and to ensure that trainees receive feedback. Apps have increased the number of evaluations, improved trainee satisfaction, and helped faculty learn milestones.[R–V] Video may also be a beneficial tool.[W]
7. *Benchmark assessors' performance* and provide feedback on a regular basis.
8. *Consider use of direct observation for scholarly work* as a key outcome measure for intervention studies.

References

1. Williams RG, Kim MJ, Dunnington GL. Practice guidelines for operative performance assessments. *Ann Surg*. 2016;264(6):934–948.
2. Paisley AM, Baldwin PJ, Patterson-Brown S. Accuracy of medical staff assessment of trainees' operative performance. *Med Teach*. 2005;27(7):634–638.
3. Williams RG, Sanfey H, Chen XP, et al. A controlled study to determine measurement conditions necessary for a reliable and valid operative performance assessment: a controlled prospective observational study. *Ann Surg*. 2012;256(1):177–187.
4. George BC, Teitelbaum EN, Meyerson SL, et al. Reliability, validity, and feasibility of the Zwisch scale for the assessment of intraoperative performance. *J Surg Educ*. 2014;71(6):90–96.
5. Williams RG, Verhulst S. Mellinger JD, et al. Is a single-item operative performance rating sufficient? *J Surg Educ*. 2015;72(6):e212–e217.
6. Regehr G, Ginsburg S, Herold J, et al. Using "standardized narratives" to explore new ways to represent faculty opinions of resident performance. *Acad Med*. 2012;87(4):419–427.
7. Williams RG, Verhulst S, Colliver JA, et al. A template for reliable assessment of resident operative performance: assessment intervals, numbers of cases and raters. *Surgery*. 2012;152(4):517–524.
8. Williams RG, Mellinger JD, Dunnington GL. A problem-oriented approach to resident performance ratings. *Surgery*. 2016;160(4):936–945.
9. Hauer KE, Holmboe ES, Kogan JR. Twelve tips for implementing tools for direct observation of medical trainees' clinical skills during patient encounters. *Med Teach*. 2011;33(1):27–33.

Kathryn M. Andolsek, MD, MPH, FAAFP, is Professor of Community and Family Medicine and Assistant Dean, Duke University School of Medicine; and **Deborah Simpson, PhD,** is Director, Medical Education Programs, Aurora Health Care, and Professor (Clinical Adjunct), Family & Community Medicine Medical College of Wisconsin and University of Wisconsin School of Medicine and Public Health. Both authors are members of the *JGME* Editorial Board.

Corresponding author: Kathryn M. Andolsek, MD, MPH, FAAFP, Duke University School of Medicine, 0159 Orange Zone DS, 201 Trent Drive, Durham, NC 27710, 919.668.3883, kathryn.andolsek@duke.edu