How Do Thresholds of Principle and Preference Influence Surgeon Assessments of Learner Performance?

Tavis Apramian, MA, MSc, PhD^{*}, Sayra Cristancho, PhD^{*,†}, Alp Sener, MD, PhD[‡], and Lorelei Lingard, PhD^{*,§}

^{*}Centre for Education Research and Innovation, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

[†]Department of Surgery, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

[‡]Division of Urology, Department of Surgery, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

[§]Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

Abstract

Objective—The present study asks whether intraoperative principles are shared among faculty in a single residency program and explores how surgeons' individual thresholds between principles and preferences might influence assessment.

Background—Surgical education continues to face significant challenges in the implementation of intraoperative assessment. Competency-based medical education assumes the possibility of a shared standard of competence, but intersurgeon variation is prevalent and, at times, valued in surgical education. Such procedural variation may pose problems for assessment.

Methods—An entire surgical division (n = 11) was recruited to participate in video-guided interviews. Each surgeon assessed intraoperative performance in 8 video clips from a single laparoscopic radical left nephrectomy performed by a senior learner (>PGY5). Interviews were audio recorded, transcribed, and analyzed using the constant comparative method of grounded theory.

Results—Surgeons' responses revealed 5 shared generic principles: choosing the right plane, knowing what comes next, recognizing normal and abnormal, making safe progress, and handling tools and tissues appropriately. The surgeons, however, disagreed both on whether a particular performance upheld a principle and on how the performance could improve. This variation subsequently shaped their reported assessment of the learner's performance.

Reprints: Tavis Apramian, Centre for Education Research and Innovation, Health Sciences Addition (Room 110), Schulich School of Medicine and Dentistry, Western University, London, ON, N6A 5C1, Canada. tavis.apramian@schulich.uwo.ca.

Disclosure: The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

Conclusions—The findings of the present study provide the first empirical evidence to suggest that surgeons' attitudes toward their own procedural variations may be an important influence on the subjectivity of intraoperative assessment in surgical education. Assessment based on intraoperative entrustment may harness such subjectivity for the purpose of implementing competency-based surgical education.

Keywords

competence; competency-based medical education; entrustable professional activities; procedural variations; surgical education

Competence-based assessment asserts that experts share an understanding of the criteria for a competent performance.^{1,2} Based on this premise, competence-based medical education initiatives are creating competencies,³ milestones,⁴ and entrustable professional activities (EPAs)⁵ to structure training and assessment procedures in postgraduate medical education. These communal standards are the building blocks by which programs will judge each resident's readiness to progress.⁶ But what if that premise is more fraught than it appears: what if experts do not have a shared understanding of what a competent performance in their specialty looks like?

Recent insights in surgical education make this more than a hypothetical question. Our previous research using observations and interviews with surgical trainees and staff showed that trainees must navigate the individual variations of their staff surgeons by differentiating their non-negotiable principles and their idiosyncratic preferences.⁷ Further complicating the picture, the threshold between the principles and preferences appears to shift from surgeon to surgeon. The threshold between which procedural variations constitute non-negotiable procedural principles and what variations constitute optional preferences can be meaningfully different between 2 surgeons in the same specialty. For example, one surgeon may believe that using a specific kind of grasper on bowel is non-negotiable and thus imply that the specific choice of grasper in a given procedure is a principle a learner must follow. Another surgeon may, however, allow a learner to use a grasper that is different from the one the surgeon would typically use and thus imply that choice of grasper for that step is a preference. Learners in the surgical workplace must sort through this complexity day after day.

We concluded from observing this phenomenon in previous intraoperative field research that the process of discerning a surgeon's threshold between what constitutes a principle and what constitutes a preference—a process we called “thresholding”—is a complex and constant form of daily work for learners.⁷ In follow-up studies, learners and surgeons alike stated that taking a deliberate approach to the process of thresholding was critical to learner success.⁸ And, perhaps most importantly, successfully reproducing a surgeon's threshold while under their supervision may mean retaining the primary operator role but failure often means relinquishing it and losing valuable intraoperative experience.⁹

The existence of principles and preferences and the fluidity of the threshold between them matters for competence-based assessment. The finding that what one surgeon might call a principle another might call a preference is embedded in the theory of thresholds. The theory

raises the troubling possibility that, although surgeons' competence judgments might attest to a learner's reproduction of core surgical principles, those judgments might equally attest to learners' replication of individual surgeon preferences. To discern to what degree either or both might be the case, more fundamental work needs to be done to understand the impact of thresholds on assessment. In particular, surgical educators need to know whether principles are shared across a surgical community, and, if they vary, we need a better understanding of how such variability factors into intraoperative assessment.

The consequences of not attending to procedural variation and thresholds may jeopardize the implementation of competence-based medical education in surgery. Competence-based medical education relies on effective workplace-based assessment tools. Unfortunately, none of the currently available assessment tools have proven adequate for licensure due to poor usability,¹⁰ low intersurgeon reliability,^{11,12} and the absence of evidence-based benchmarks of competence.^{13–15} More remarkably, assessment by expert surgeons may not differ significantly from crowd-sourced assessment by the lay public.^{16–20} Could it be that surgeons' and residents' perceptions and behaviors around procedural variation contributes to the low inter-rater reliability in intraoperative assessment? We designed the present study to use empirical intraoperative data to develop concrete hypotheses about the potential influence of procedural variation and thresholds on workplace-based assessment.

METHODS

Research exploring intraoperative assessment has traditionally taken quantitative statistical approaches.^{21–24} We decided to take a grounded theory approach to respond to calls in the literature for research that substantively examines the sociocultural forces that shape the apparent subjectivity of assessment in surgical education.^{25–27} To accomplish that goal, our study focused on a question around subjectivity in assessment emerging from our previous research^{7–9}: could it be that the surgeon herself contributes to the variability of thresholds of principle and preference? To answer this question we eliminated other sources of potential variability such as patient anatomy by using a video-guided interview technique²⁸ to elicit thick procedural descriptions.²⁹

We showed surgeons condensed video of a laparoscopic radical left nephrectomy performed by a senior learner (>PGY5) under surgeon supervision at a tertiary teaching hospital in Ontario, Canada. The learner had graduated from the program before the study. We selected 8 specific clips from the procedure which, once compiled, totaled 3 minutes 51 seconds. These selected clips were 12 to 80 seconds long (mean = 26 seconds). The clips were selected as potential intraoperative assessment moments depicting procedural steps in which multiple tools or approaches might be used and procedural steps ranging from low stakes (eg, dissecting colon from kidney) to high stakes (eg, dividing renal artery).

Data Collection

Surgeons were presented with the 8 “assessment moments” and asked to verbally assess the performance of the learner on that step of the procedure. We recruited the entire faculty from 1 surgical division to participate. All surgeons in the division shared the same specialty. We chose this “divisional sampling unit” because the goal of the present study was to explore

how variations may shape workplace-based assessment in competence-based surgical education, as postgraduate divisions may be asked to develop consensus on resident competence based on multisourced evidence of workplace performance.^{30,31} All surgeons in the division (n = 11) viewed the 8 video-recorded assessment moments for a total of 88 assessments. Interviews lasted 17 to 47 minutes (mean = 34 minutes). The interviewer invited the surgeon to comment on each assessment moment, pausing the video to give the surgeon time to respond.

Data Analysis

The interviews were audiorecorded and transcribed. The entire dataset (video clips, interview transcripts, reflective memos, and procedural documents) was compiled and analyzed with assistance from NVivo qualitative data analysis software. The statements of each surgeon about the assessment moments were compared with one another to explore emerging themes. The analysis was undertaken using the theory of *thresholds of principle and preference*⁷ as a sensitizing concept.³² The data from the present study were compared against previously identified principles and used to refine the theory of thresholds.³³ As expected in grounded theory research, data collection and analysis progressed simultaneously, and the results of the analysis informed our ongoing data collection by modifying follow-up probes in the interviews.³²

RESULTS

The findings of the present study support and expand on principles identified in previous research.⁷⁻⁹ The data, however, suggest that, although surgeons use similar generic language to describe principles, they disagree about what employing those principles in practice actually looks like. That is, the surgeons in our study held different thresholds between what they considered a principle that a learner was obliged to enact and what they considered a preference that a learner could understand as optional.

Describing Generic Principles

The notion that principles and preferences exist on a continuum from variations relevant to patient safety to variations irrelevant to the outcome of the procedure emerged again in this study. The surgeons distinguished between “the preferences and the principles”(URS01) and confirmed that the primary goal of surgical education “is to teach (residents) principles and (also) different ways to do things and then let them make their decisions on their own when they go into practice”(URS04). Although there were cases in our data of surgeons identifying some of the same moments in the video-recorded procedure as problematic, these points of agreement were relatively rare. The few that did emerge suggest that principles are clear cut in only a limited number of instances. For example, in one moment of the procedure the learner prepared to use a stapler to divide the renal artery. Many surgeons (n = 7, 64%) expressed concern about not being able to “see the end of the stapler”(URS08) so that, for the learner, “what’s behind (the renal artery) is blind...(and) he (or she) could be cutting through anything”(URS09) in an area in which “if they get bleeding or they don’t staple properly or the stapler misfires then it could be disastrous”(URS07). Some moments like the dissection of the renal artery suggested the possibility of consensus, but at no single

assessment moment were all the surgeons in the present study in consensus about the performance of the learner. Instead, when they did comment, their responses evoked themes that supported and refined the following generic principles espoused in previous research⁸:

- Choosing the right plane (Table 1)
- Knowing what comes next (Table 2)
- Recognizing normal and abnormal (Table 3)
- Making safe, efficient progress (Table 4)
- Handling tools and tissue appropriately (Table 5)

Applying Personal Principles

In no single assessment moment were all of the surgeons in the division in agreement about the performance depicted in the video. More importantly, even in moments in which many of the surgeons were united in critique about the procedural choices made in the video, their agreement about principles dissolved once asked to speak about the individual steps of the procedure and what they would do differently. The full depiction of this variation in all 8 clips has been included in Appendix 1, <http://links.lww.com/SLA/B225> with quotations from the data arranged from the most amount of disagreement to the least. In a condensed example from clip 4, we have described the disagreement of the surgeons in our sample as it pertained to a mid-procedure dissection performed between the kidney and the spleen (Table 6).

Surgeons Move Thresholds

The surgeons in our sample explained how they would approach teaching a particular variation and why their threshold of principle and preference might land differently than another surgeon's. Some described taking time to justify to residents why their own variation works because they “truly believe that memorization [of each surgeon's variations] isn't the way to go.” (URS06). Other surgeons felt strongly that the learner's job is to memorize and apply the surgeon's variations: “I don't always repeat (my variations to them). So I will tell them, ‘you're a big boy, you've got to learn and remember...’ You're in my OR, you do it my way” (URS03). Still other surgeons insisted that teaching should not increase the length of the procedure: “I would do something in about fifteen seconds that they would've taken two minutes to do or four minutes or five minutes or fifteen minutes to do. So I'd take (the tools) away, show them how it's done... If we don't get to the next part quickly your next patient gets cancelled... we're always battling the clock” (URS05). From detailed explanation to memorization to pacing expectations, the surgeons in our study approached teaching procedural variations to residents with broadly different attitudes. Our findings suggest that surgeons themselves place thresholds of principles and preferences differently and the surgeons' beliefs about the learner's role regarding variations influences where their thresholds fall for a given procedure.

Thresholding as Assessors—The surgeons in our sample disagreed about both how to apply principles and the consequences of failing to do so. Similarly, the surgeons in the sample were split on the value of the video clips on assessing competence. The video was

enough for some surgeons to determine competence, whereas others felt making any decision whatsoever about the competence of the learner from the video would be irresponsible. Based on what they had seen in the video, 4 of 11 surgeons (36%) assessed the learner as competent to perform the procedure independently. Another 4 (36%) stated that the learner was not yet competent, and the remaining 3 surgeons (27%) wanted more information about the learner before passing judgment. This variation persisted across surgeons who routinely perform this common procedure and those who do not.

The surgeons' attitudes toward variations played an important role in the decisions they made about the learner's competence. A surgeon who used different tools to isolate the renal artery than the tool initially chosen by the learner decided that the "management of the hilum" (URS03), and thus the procedure as a whole, had been incompetently performed. A different surgeon endorsed the learner as competent because "it's the same sort of technique I see myself using" (URS02). Another used a different approach to clips around the hilum but deemed the learner competent and the procedure to have been performed "very safely and very well" (URS07). This variation in threshold placement had implications for how the surgeons perceived they would proceed if this was their resident. What one surgeon would stop a resident from doing (thus indicating a principle) another would allow to proceed (thus indicating a preference) and yet another would praise as permissible mistake or even call an excellent performance.

DISCUSSION

The surgeons in our study agreed on the language of generic principles but appeared to disagree on how to apply them. This finding may help to explain some of the difficulties faced by intraoperative assessment in surgical education³⁴ particularly around poor reliability of surgeon assessors^{11,12} and a lack of evidence-based benchmarking.^{13–15}

Agreement on Generic Principles

The 5 generic principles endorsed by surgeons in the present study correlate with the generic principles proposed by current approaches to intraoperative assessment of technical skill. The Objective Structured Assessment of Technical Skill³⁵—the most frequently employed tool for intraoperative technical skills assessment in studies using global ratings¹⁵—asks surgeons to assess learners' respect for tissue; time and motion; instrument handling; knowledge of instruments; use of assistants; flow of operation and forward planning; and knowledge of the specific procedure on a 5-point Likert scale.³⁵ Other approaches to surgical technical skill assessment operate according to the same principles-based logic. The Ottawa Surgical Operating Room Score (O-SCORE)³⁶ prompts surgeons to assess learners' knowledge of procedural steps, and procedure-based assessment³⁷ instructs surgeons to assess learners' use of instruments appropriately and safely and the learner's ability to work at an appropriate pace with economy of motion. The language of principles remains remarkably consistent across the surgical assessment literature.

Disagreement on Applying Principles

Surgical principles may seem clear cut, but important questions about the application of these principles are beginning to emerge. Our previous research claimed that the way surgeons and residents speak of procedural variations begs the question: is a principle always a principle?⁷⁻⁹ The present study proposes an answer to that question. We found the same guiding principles were embedded within surgeons' statements about safety and competence, but how they applied those principles in practice appeared to vary. *What is a principle for one surgeon may well be a preference for another.*

Supporting evidence for the persistence of variations in surgical principles continues to emerge from the surgical practice literature. Birkmeyer et al³⁸ demonstrated that procedural approaches of practicing surgeons vary significantly and their rate of complications vary alongside them. Asch et al³⁹ showed that imprinting of procedural variations and their associated complication rates may follow surgeons from their postgraduate training programs to their jobs in independent practice. And, most convincingly, Davidson et al⁴⁰ used the example of pancreaticoduodenectomy to demonstrate that although surgical procedures may in theory have relatively few variations, in practice they are significantly more complex. Davidson et al⁴⁰ closely examined the pancreaticoduodenectomy of 5 surgeons and found they differed in "steps and techniques employed"^(p. 2) for 7 major areas of the procedure and 21 minor areas. They concluded that each of the 5 surgeons in the study essentially performed 5 different procedures but "all called (it) a Whipple's procedure"^(p. 2).⁴⁰ Although surgical procedures may be commonly thought to be as uniform as the textbook, close examination reveals meaningful heterogeneity in practice.

The surgical literature increasingly recognizes the variability in practice.^{41,42} But the literature on the educational impact of these variations remains scant. Our study attempts to close that gap by drawing attention to the implications of such procedural variation for intraoperative assessment in surgical education. The surgeons in our study offered divergent views on how to safely and proficiently perform 8 moments of a laparoscopic nephrectomy despite agreeing on the language of generic principles. Some of the surgeons pointed to the same moves and choices the learner made when they were identifying problems, but even in those instances the suggestions the surgeons made for fixing the problems appeared idiosyncratic. Consequently, surgeons themselves place thresholds differently and attitudes toward thresholds have significant implications for surgical education.

Proposing an Explanation for Challenges in Assessment

We found that surgeons' individual thresholds between what they considered a non-negotiable principle and an inconsequential preference were loosely coupled with their assessment of the learner's performance. By "loosely coupled" we mean that surgeons invoke their personal variations when assessing learners, but do not strictly apply them to the point of expecting mimicry. Instead, a surgeon's attitude toward what constitutes a principle versus a preference seems to partially impact their decision about the competence of a learner. *Ultimately, therefore, the degree to which a surgeon assesses a learner to be competent may be shaped by the surgeon's own threshold of principle and preference.*

The loose coupling of thresholds and assessment could help to explain some of the challenges encountered by researchers working on the intraoperative assessment of technical competence. Recent research suggests that assessment by expert surgeons may not differ significantly from crowd-sourced assessment by the lay public.^{16–20} And each of the 5 most recent reviews of technical skill assessment concluded that none of the currently available assessment tools are adequate for credentialing or licensure due to low intersurgeon reliability^{11,12} and the absence of evidence-based benchmarks of competence.^{13–15} Even when assessment tools have demonstrated impressive intersurgeon reliability they continue to find limited application due to poor usability¹⁰ and difficulty scaling up from the research setting to the clinical environment.^{34,43,44} Our findings can help explain why surgeons may look for their own variations and make competence judgments based on their individual thresholds of principle and preference.

Moving Forward

Investment in assessment tools that harness subjectivity of surgeons may prove crucially important for the future of intraoperative assessment. Competency-based reforms in surgical education have attempted, on the one hand, to seek out consensus between surgeons about standardized competence of learners^{34,45} and, on the other hand, to longitudinally aggregate individual surgeons' subjective intraoperative entrustment decisions.^{46,47} For example, current large-scale innovations in intraoperative assessment have focused on standardized assessment of technical skill using a high-stakes simulation-based practical examination⁴⁸ or on low-stakes intraoperative assessment using mobile applications such as SIMPL to record shifts in surgeons' subjective intraoperative entrustment over time.⁴⁷ Research on the validity of assessment using standardized surgical principles drawn from clinical evidence¹⁵ as in standardized assessment remains a laudable aspiration for an evidence-based surgical world. Until then, our findings suggest that surgical educators consider a “programmatic” approach to assessment using tools that collect large samples of multiple low-stakes observations.^{49,50}

Interest in entrustment-based assessment continues to grow in surgical education^{9,46,51} as undergraduate and postgraduate medical education shift toward using EPAs.^{52,53} Fortunately, entrustment-based intraoperative assessment tools such as SIMPL may account for the complexity of thresholds of principle and preference by harnessing surgeon subjectivity. New research on EPAs in surgical education, however, suggests simply choosing a new assessment tool may not address the cultural complexity caused by procedural variation. Field research has shown that postgraduate surgical programs using EPAs may be willing to officially entrust residents based on the *number of procedures the resident has watched* rather than on documented observations of their intraoperative performance.⁵⁴ This research indicates that even entrustment-based tools can be implemented in a way that ignores the importance of observational assessment and reconciling procedural variations in practice. Therefore, for entrustment-based assessment reforms to work in surgery, postgraduate programs and clinical competence committees³⁰ may benefit from using the theory of thresholds of principles and preference to guide explicit conversations about the complexities of variations in their home divisions. Procedural variations shape the way surgical educators think about competence. Considering

a cultural shift toward transparency around thresholds of principle and preferences may positively transform surgical education.

CONCLUSIONS

Implementation of intraoperative assessment for the purpose of licensure continues to evade surgical education. The findings of the present study *provide the first empirical evidence* to suggest that the surgeons' attitudes toward their own procedural variations may be the most important factor shaping the seemingly inescapable subjectivity in intraoperative assessment. Researchers and educators should consider the formative influence of surgeons' thresholds of principle and preference on assessment of learner competence when designing, testing, and implementing approaches to competency-based surgical education.

Limitations

We urge caution in interpreting our findings as some of the variation in assessment uncovered in our findings may be an artifact of the grounded theory design. We acknowledge that asking surgeons to make judgments based on decontextualized clips from procedures within their own specialty may seem to amplify the actual variations in practice.⁵⁵ To deal with this potential over-representation we negotiated closely over which moments of variation should be reported in the data. Our goal was to emergently build a dataset and theoretical framework that could inform further research in the area of procedural variations.

We chose to recruit an entire surgical division. Including 4 of the surgeons in the division who do not routinely perform laparoscopic nephrectomies may also over-represent procedural variations in the data; however, we chose this approach to sampling because it reflects the likely composition of clinical competence committees³⁰ based on surgical specialty rather than by subspecialty. Keeping this question of variability in mind, we encourage future researchers to use our qualitatively derived findings to deductively analyze how tightly coupled thresholds and assessment may be.

References

1. Holmboe ES, Sherbino J, Long DM, et al. The role of assessment in competency-based medical education. *Med Teach*. 2010; 32:676–682. [PubMed: 20662580]
2. Gruppen LD, Burkhardt JC, Fitzgerald JT, et al. Competency-based education: programme design and challenges to implementation. *Med Educ*. 2016; 50:532–539. [PubMed: 27072442]
3. Frank J, Mungroo R, Ahmad Y, et al. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach*. 2010; 32:631–637. [PubMed: 20662573]
4. Tekian A, Hodges BD, Roberts TE, et al. Assessing competencies using milestones along the way. *Med Teach*. 2015; 37:399–402. [PubMed: 25523010]
5. Ten Cate O. Trust, competence, and the supervisor's role in postgraduate training. *BMJ*. 2006; 333:748. [PubMed: 17023469]
6. Sklar DP. Competencies, milestones, and entrustable professional activities: what they are, what they could be. *Acad Med*. 2015; 90:395–397. [PubMed: 25803411]
7. Apramian T, Cristancho S, Watling C, et al. Thresholds of principle and preference: exploring procedural variation in postgraduate surgical education. *Acad Med*. 2015; 90:S70–S76. [PubMed: 26505105]

8. Apramian T, Cristancho SM, Watling CJ, et al. 'They have to adapt to learn': surgeons' perspectives on the role of procedural variations in surgical education. *J Surg Educ.* 2016; 73:339–347. [PubMed: 26705062]
9. Apramian T, Cristancho SM, Watling CJ, et al. 'Staying in the game': how procedural variations shapes competence judgments in surgical education. *Acad Med.* 2016; 91:S37–S43. [PubMed: 27779508]
10. Pereira EAC, Dean BJF. British surgeons' experiences of a mandatory online workplace based assessment portfolio resurveyed three years on. *J Surg Educ.* 2013; 70:59–67. [PubMed: 23337672]
11. Van Hove P, Tuijthof G, Verdaasdonk E, et al. Objective assessment of technical surgical skills. *Br J Surg.* 2010; 97:972–987. [PubMed: 20632260]
12. Ahmed K, Miskovic D, Darzi A, et al. Observational tools for assessment of procedural skills: a systematic review. *Am J Surg.* 2011; 202:469.e6–480.e6. [PubMed: 21798511]
13. Ghaderi I, Manji F, Park YS, et al. Technical skills assessment toolbox: a review using the unitary framework of validity. *Ann Surg.* 2015; 261:251–262. [PubMed: 24424150]
14. Shalhoub J, Vesey AT, Fitzgerald JE. What evidence is there for the use of workplace-based assessment in surgical training? *J Surg Educ.* 2014; 71:906–915. [PubMed: 24831439]
15. Szasz P, Louridas M, Harris KA, et al. Assessing technical competence in surgical trainees: a systematic review. *Ann Surg.* 2015; 261:1046–1055. [PubMed: 25119118]
16. Powers MK, Boonjindasup A, Pinsky M, et al. Crowdsourcing assessment of surgeon dissection of renal artery and vein during robotic partial nephrectomy: a novel approach for quantitative assessment of surgical performance. *J Endourol.* 2016; 30:447–452. [PubMed: 26597352]
17. Kowalewski TM, Comstock B, Sweet R, et al. Crowd-sourced assessment of technical skills for validation of basic laparoscopic urologic skills tasks. *J Urol.* 2016; 195:1859–1865. [PubMed: 26778711]
18. Deal SB, Lendvay TS, Haque MI, et al. Crowd-sourced assessment of technical skills: an opportunity for improvement in the assessment of laparoscopic surgical skills. *Am J Surg.* 2016; 211:398–404. [PubMed: 26709011]
19. Holst D, Kowalewski TM, White LW, et al. Crowd-sourced assessment of technical skills: differentiating animate surgical skill through the wisdom of crowds. *J Endourol.* 2015; 29:1183–1188. [PubMed: 25867006]
20. Ranard BL, Ha YP, Meisel ZF, et al. Crowdsourcing—harnessing the masses to advance health and medicine, a systematic review. *J Gen Intern Med.* 2014; 29:187–203. [PubMed: 23843021]
21. Bernstein M, Khu KJ. Is there too much variability in technical neurosurgery decision-making? Virtual Tumour Board of a challenging case. *Acta Neurochir.* 2009; 151:411–413. [PubMed: 19255714]
22. Kerver A, Leliveld M, Theeuwes H, et al. Inter-surgeon variation in skin incisions for tibial nailing in relation to the infrapatellar nerve. *Injury Extra.* 2010; 41:131–166.
23. Walter A, Buller J. Variability of reported techniques for performance of the pubovaginal sling procedure. *Int Urogynecol J.* 2002; 13:160–164.
24. Van der Vleuten C, Norman G, Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ.* 1991; 25:110–118. [PubMed: 2023552]
25. Govaerts M, Van der Vleuten CPM. Validity in work-based assessment: expanding our horizons. *Med Educ.* 2013; 47:1164–1174. [PubMed: 24206150]
26. Gingerich A, Van der Vleuten CPM, Eva KW, et al. More consensus than idiosyncrasy: categorizing social judgments to examine variability in mini-CEX ratings. *Acad Med.* 2014; 89:1510–1519. [PubMed: 25250753]
27. Van der Vleuten C, Schuwirth LW, Scheele F, et al. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol.* 2010; 24:703–719. [PubMed: 20510653]
28. Hogan T, Hinrichs U, Hornecker E. The elicitation interview technique: capturing people's experiences of data representations. *IEEE Trans Vis Comput Graph.* 2016; 22:2579–2593. [PubMed: 26731766]

29. Geertz, C. Thick description: toward an interpretive theory of culture. In: Martin, M., McIntyre, L., editors. *Readings in the Philosophy of Social Science*. Cambridge, MA: The MIT Press; 1994. p. 311-323.
30. French JC, Dannefer EF, Colbert CY. A systematic approach toward building a fully operational clinical competency committee. *J Surg Educ*. 2014; 71:e22–e27. [PubMed: 24878101]
31. Gardner AK, Scott DJ, Hebert JC, et al. Gearing up for milestones in surgery: will simulation play a role? *Surgery*. 2015; 158:1421–1427. [PubMed: 26013987]
32. Charmaz, K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. 2. Thousand Oaks, CA: SAGE Publications; 2014.
33. Glaser BG. The constant comparative method of qualitative analysis. *Soc Probl*. 1965; 12:436–445.
34. Arora S, Darzi A. Introducing technical skills assessment into certification: closing the implementation gap. *Ann Surg*. 2016; 264:7–9. [PubMed: 27272957]
35. Martin J, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997; 84:273–278. [PubMed: 9052454]
36. Gofton WT, Dudek NL, Wood TJ, et al. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med*. 2012; 87:1401–1407. [PubMed: 22914526]
37. Marriott J, Purdie H, Crossley J, et al. Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. *Br J Surg*. 2011; 98:450–457. [PubMed: 21254025]
38. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013; 369:1434–1442. [PubMed: 24106936]
39. Asch DA, Nicholson S, Srinivas SK, et al. How do you deliver a good obstetrician? Outcome-based evaluation of medical education. *Acad Med*. 2014; 89:24–26. [PubMed: 24280859]
40. Davidson SJ, Rojnica M, Matthews JB, et al. Variation and acquisition of complex techniques: pancreaticoduodenectomy. *Surg Innov*. 2016; 23:586–592. [PubMed: 27507575]
41. Reames BN, Shubeck SP, Birkmeyer JD. Strategies for reducing regional variation in the use of surgery: a systematic review. *Ann Surg*. 2014; 259:616–627. [PubMed: 24240626]
42. Urbach DR. Closing in on surgical practice variations. *Ann Surg*. 2014; 259:628–629. [PubMed: 24603297]
43. Lendvay TS, White L, Kowalewski T. Crowdsourcing to assess surgical skill. *JAMA Surg*. 2015; 150:1086–1087. [PubMed: 26421369]
44. Szasz P, Louridas M, Harris KA, et al. Strategies for increasing the feasibility of performance assessments during competency-based education: subjective and objective evaluations correlate in the operating room. *Am J Surg*. 2016 E-pub ahead of print.
45. Szasz P, Louridas M, De Montbrun S, et al. Consensus-based training and assessment model for general surgery. *Br J Surg*. 2016; 103:763–771. [PubMed: 27005690]
46. Teman NR, Gauger PG, Mullan PB, et al. Entrustment of general surgery residents in the operating room: factors contributing to provision of resident autonomy. *J Am Coll Surg*. 2014; 219:778–787. [PubMed: 25158911]
47. Bohnen JD, George BC, Williams RG, et al. The feasibility of real-time intraoperative performance assessment with SIMPL (System for Improving and Measuring Procedural Learning): early experience from a multi-institutional trial. *J Surg Educ*. 2016; 73:e118–e130. [PubMed: 27886971]
48. De Montbrun S, Roberts PL, Satterthwaite L, et al. Implementing and evaluating a national certification technical skills examination: the colorectal objective structured assessment of technical skill. *Ann Surg*. 2016; 264:1–6. [PubMed: 26764869]
49. Van der Vleuten C, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005; 39:309–317. [PubMed: 15733167]
50. Cook DA, Brydges R, Ginsburg S, et al. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015; 49:560–575. [PubMed: 25989405]
51. 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*. 2013; 71:e90–e96.

52. Englander R, Flynn T, Call S, et al. Toward defining the foundation of the MD degree: core entrustable professional activities for entering residency. *Acad Med.* 2016; 91:1352–1358. [PubMed: 27097053]
53. Warm EJ, Held JD, Hellmann M, et al. Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. *Acad Med.* 2016; 91:1398–1405. [PubMed: 27355780]
54. Van Loon KA, Teunissen PW, Driessen EW, et al. The role of generic competencies in the entrustment of professional activities: a nationwide competency-based curriculum assessed. *J Grad Med Educ.* 2016; 8:546–552. [PubMed: 27777665]
55. Gingerich A, Ramlo S, Van der Vleuten C, et al. Inter-rater variability as mutual disagreement: identifying raters' divergent points of view. *Adv Health Sci Educ.* 2016 E-pub ahead of print.

TABLE 1

Choosing the Right Plane

<p>The surgeons in our sample routinely declare the importance of “learning about tissue planes”^(URS08). But not only do they not always agree with the plane chosen by the learner in the video—“he [or she] is using the wrong plane to start with”^(URS05)—but they also do not suggest the same alternative plane they wish the learner to follow. They universally agree that learning to “follow anatomic planes... (is a) surgical principal”^(URS03) crucially important to becoming a surgeon, but enacting the principle changes when asked to assess an actual performance.</p>

TABLE 2

Knowing What Comes Next

Choosing the right tools and anticipating the next steps appears necessary for positive interactions between learners and surgeons. Surgeons report finding themselves “question(ing) the learner’s experience” ^(URS01) when the learner suggests tools, techniques, or steps that the surgeon themselves does not use. The surgeons report losing trust in learners when “they don’t seem to know what tool to ask for, or you ask them ‘OK, we’ve done this part of the procedure, what’s next?’ And if they don’t know it’s a concern to us...Definitely a red flag” ^(URS04) . These choices may be personal to surgeons: “one thing that drives us crazy, and you’ll see this over and over again, is the trainee using someone else’s technique on your patient ” ^(URS05) . Despite the personal nature of these variations, knowing what comes next in that surgeon’s approach can be a high stakes decision. Choosing the wrong tool, technique, or step can “blow your trust” ^(URS03) .
--

TABLE 3**Recognizing Normal and Abnormal**

The surgeons in our sample support the principle of recognizing the abnormal, but the importance of “recognizing the normal”^(URS05) also emerged in the present study. Performing a nephrectomy requires dividing the renal artery and renal vein. In speaking about these steps, the surgeons in our study often returned to the importance of being able to recognize procedural moments “when the stakes are high”^(URS01) like operating around the renal vessels. The surgeons stated that this ability primes residents to ask themselves as they dissect “am I going too far this way, am I going too far that way?”^(URS09), and to be ready to recognize abnormalities such as atypical anatomy, “neovascularization”^(URS07), or scar tissue. Attending to this principle allows residents to “be honest with their own level of skill when they know ‘hey, wait a minute, if I continue here I may put this patient at risk because...something looks just a little different’”^(URS04).

TABLE 4

Making Safe, Efficient Progress

Appearing to lack technical skill or knowledge of the procedure may lead surgeons to feel a resident has not sufficiently attended to the principle of making safe efficient progress. “Failure to progress”^(URS07) or lacking efficiency can lead surgeons to take over operating: “I’m very critical about this with the residents. If I take an extra minute to do every step... You doubled the time for the operation. You doubled the time that somebody has to be on the table”^(URS05). The surgeons in our sample **point out different unsafe or “safe maneuver(s)”**^(URS04) and highlight **different ways “movements were efficient”**^(URS01) **or inefficient**. Nevertheless, their comments about the learner’s performance share safety and efficiency as common concerns.

TABLE 5

Handling Tools and Tissue Appropriately

<p>The surgeons in our study often return to the importance of “learning to handle the instruments or learning to handle the tissue”^(URS08). Some surgeons report feeling that they could use the trajectory of the learner’s instruments to decide on whether the learner understands “the implications of the tool”^(URS01) and anatomy they’re working on. Surgeons watching the same video clips had different conclusions about which tissues could withstand which forces. Some surgeons “don’t like grabbing fat”^(URS07), whereas others are more cautious with the “thin tissue”^(URS02) of veins, but all of the surgeons in our sample spoke of the importance of learners’ understanding of tissue and tool usage.</p>
--

TABLE 6

Disagreement in Principle Application for Clip 4

<p>The learner extends the harmonic scalpel deeply between the spleen and the kidney. The surgeons significantly diverge in their perception of this maneuver: what some surgeons perceived as problem of technique—“burying your tools into the depths without truly knowing where you are”_(URS04)—others perceived as a problem of step sequence or plane selection. For example, one surgeon felt this clip was an example of “working on upper pole vessels that he [or she] doesn’t need right now”_(URS03). While another felt that the learner’s plane choice was problematic because the learner was “getting a little bit close to the capsule of the kidney”_(URS11). On the contrary, some of the surgeons had no qualms with the maneuvers used by the learner, seeing “nothing special”_(URS02) to comment on. The surgeons in the sample seem to agree on the generic names of principles, but their ideas about what constitutes a principle vary greatly.</p>
--