Learning to teach: A novel method for assessing surgical trainees' teaching and operative knowledge

Leah Furman [a], [*], Eliza Beth Littleton [b], Christof Kaltenmeier [b], Giselle G. Hamad [b]

[a] *University of Pittsburgh School of Medicine, 3550 Terrace Street, Pittsburgh, PA, USA*
[b] *Department of Surgery, University of Pittsburgh School of Medicine, 200 Lothrop Street, Pittsburgh, PA, USA*

ARTICLE INFO

ABSTRACT

Background: The purpose of this study was to investigate the feasibility of a simulated teaching activity as an assessment of surgical knowledge and teaching competencies.
Methods: In this prospective observational study, 15 residents and 1 fellow in the Department of Surgery watched three video clips of laparoscopic cholecystectomies and provided feedback to a participant learner. Qualitative and statistical analysis identified differences in surgical knowledge and teaching strategies.
Results: As compared to senior trainees, junior trainees were more likely to speculate on the learner's actions (p = 0.033), identify which actions looked correct (p = 0.028), and speculate more on the learner's thoughts (p = 0.02). Senior trainees noted case difficulty more frequently (p = 0.028), identified more actions that looked incorrect (p = 0.004), and speculated more about the learner's emotions (p = 0.033).
Conclusions: A simulated teaching scenario successfully assessed operative and teaching competencies, suggesting a novel assessment method.

© 2020 Elsevier Inc. All rights reserved.

Introduction

A significant concern in the training of today's surgeons is the varying success of the transition from junior to senior resident and, ultimately, attending surgeon.[1—3] To track and support this transition, surgical educators need brief, objective measures of the wide range of intraoperative skills and knowledge needed to be an excellent surgeon. In addition, because teaching is a formal ACGME Surgery milestone, educators need methods to teach and assess teaching competencies. The specific direction from the ACGME is that the resident must perform as a "highly effective teacher with an interactive educational style [who] engages in constructive educational dialogue".[4] It is expected that senior trainees, who may be either residents more advanced in their training or fellows, must be able to lead junior learners through a case.

Even for excellent surgical educators, intraoperative teaching is a complex, intense interaction between a teacher and learner.[5] Educators assess the learner continuously during cases to gauge how much autonomy can be safely allowed at any given moment.

Meanwhile, they direct, discuss, question, gesture, intervene, correct, retract, support tissues, switch instruments, and sometimes teach without saying a word.[6,7] This often-subtle interplay is difficult to capture and describe, much less evaluate. Despite the challenge, the field must develop feasible methods to gauge trainees' progress toward becoming an interactive, constructive educator. A study from 2019 found that out of 105 general surgery residency programs, 27 had a "Residents as Teachers" program. However, only four used teaching assessments based on actual observations of trainees' teaching.[8] This highlights an opportunity to expand the available methods for objectively assessing trainees' teaching competencies.

A teacher knows not only teaching strategies, but also the subject. Therefore, assessing teaching competencies involves also assessing the trainee's intraoperative competencies. A 2008 meta-analysis examined what learners and teachers believe are the traits of excellent clinical teachers, and found that two of the most valued traits relate to the teacher's surgical knowledge: "medical/clinical knowledge" and "clinical and technical skills/competence, clinical reasoning".[9] Similarly, science education research emphasizes that teaching competencies must comprise both "pedagogical knowledge" (PK), or how to guide learners to perform a task, and "pedagogical content knowledge" (PCK), comprising factual

* Corresponding author.
E-mail address: lmf110@pitt.edu (L. Furman).

knowledge about a subject, which surgeon educators sometimes call "content knowledge".[10]

In surgical education, there has been little research in support of the notion that teaching must, by default, reveal trainees' surgical competencies.[11–13] However, outside of the field, this principle underlies a training-development process known as "knowledge elicitation" (KE).[14,15] When engineers create a training tool, they often begin by interviewing subject matter experts, so they know what information the proposed system needs to have. One key KE method is "teach back," where the expert teaches the interviewer about a subject.[16] The expert's responses give the interviewer the "expert answers" that will inform training materials and assessment tools. To this end, we designed a vignette-based assessment method influenced by the "teach back" interview, where we cast surgical trainees in the role of "educator" to elicit PCK and PK. In contrast to the typical "teach back" method, we were interested only in trainees', not experts', responses.

Our goals for this study were twofold. First, we sought to establish the feasibility of our method for eliciting teaching knowledge. For the assessment to work, the trainees' responses must encompass both PCK, operative content knowledge, and PK, teaching knowledge. Most of our participants' responses would have to show them thinking deeply about teaching as they answer, and there should be reliable patterns in the responses because this would suggest the method elicits rational behavior. To accomplish this goal, we identified and categorized common themes relating to PK and PCK expressed by participants throughout their interview. Second, we endeavored to demonstrate the utility of this method as an assessment tool for PCK and PK. This was achieved by looking for evidence of differences in trainees' responses; specifically, we compared themes expressed by trainees between groups of varying experience levels. Sometimes learners may not be able to easily articulate knowledge because it is based in doing and in a specific context. We believe the vignette aspect of the method can assess some competencies that trainees might not be able to summarize or talk about in conventional testing formats.

Methods

We conducted an observational interview study approved by the University of Pittsburgh Institutional Review Board. Participation was open to general surgery trainees (residents and fellows) in the Department of Surgery who were 18 years of age or older and employed by the University of Pittsburgh Medical Center.

Recruitment was completed via email and in-person. Authors GH (Department of Surgery faculty) and CK (Department of Surgery resident) contacted all general surgery residents with an email script using the University of Pittsburgh Medical Center email directory. These emails provided contact information for authors LF (study principal investigator and administrator) or EL (LF's faculty advisor and senior study administrator) to schedule an appointment to complete the study. General surgery trainees were also notified about the existence of the study via face-to-face convenience sampling by a study team member or an attending physician who knew about the study. These potential participants were given contact information for LF or EL. Potential participants who verbally expressed interest in participating were contacted directly by LF or EL. Participation in the study was completely voluntary. Participants could choose to complete the study at their convenience, either on non-clinical time or otherwise, and there was no incentive for participation. To protect participants, as educators, GH and CK were not told who participated in the study even if they referred an interested potential participant to LF or EL for scheduling. Participation was known only to LF and EL. Additionally, all data analysis was performed with de-identified transcripts.

Each study session was approximately 15—25 min in length and consisted of a semi-structured interview anchored with three video clips and standardized prompts. Interviews were conducted by LF or EL. Participants read or listened to an informed consent script, after which they gave verbal consent to participate and recording of the interview began.

Participants were asked to state their post-graduate year (PGY). Then, three different video clips of a laparoscopic cholecystectomy were shown to participants in randomized order. These clips were chosen to show learners with a range of skills completing different operative tasks. In Video 1, the learner was a medical student using the laparoscope. In Video 2, the learner was a resident using the laparoscopic Kittner dissector. In this video, the dissector was briefly unable to be visualized but continued to dissect blindly, a maneuver the authors deemed to be unsafe or risky. Additionally, the gallbladder in this video was inflamed and particularly difficult to dissect. In Video 3, the learner was a senior resident using hook electrocautery.

Each video clip was preceded by an introduction that oriented the participant to the last step performed during the procedure and identified the learner's instrument(s). Participants were instructed to pause the video by pressing the spacebar if they would "give feedback to the learner." We chose the word "feedback" intentionally. Although "feedback" often means post-hoc case analysis, surgical educators also use it to refer to in-the-moment guidance. We wanted participants to comment as they would when interacting with a learner substantively enough that the learner could benefit from the interaction. Participants were not informed of the educational level of the learner shown in each video.

If a participant chose to pause the video, we asked them to describe the feedback they would give to the learner, and to mention the method by which they would give the feedback (e.g., verbal only, moving their instrument, pointing). We asked why they paused the video and asked them to speculate on the learner's thoughts. Afterwards, the video clip was played to the end. We then asked participants whether watching the end of the video changed the feedback they gave, and why or why not. We asked them to describe any educational activity that they would want the learner to complete before assisting in another such procedure. If a participant did not choose to pause the video, we asked nearly identical follow-up questions at the end of the video. However, instead of asking why they paused the video, we asked why they did not, and we did not ask if the end of the video changed their feedback. We concluded our interviews after participants watched all three videos. LF then transcribed the recorded audio file.

We used the constant comparison method described in Boeije (2002) for analyzing verbal data in interview transcripts.[17] LF and EL read all transcripts and highlighted phrases that seemed semantically related to (1) teaching and to (2) how to perform a laparoscopic cholecystectomy. They annotated the highlighted phrases with brief descriptive labels of (1) and (2). They discussed and grouped descriptive labels into categories and separately sorted select phrases taken from participants' answers to the open-ended questions into categories. No phrase was sorted into more than one category. The process was repeated until all phrases could be reliably sorted.

A signal detection analysis called "d-prime" (d') was used to determine reliability of the first draft of 22 themes we developed. LF and EL developed the inter-rater task in which raters would categorize 60 transcript excerpts into each theme. Of these excerpts, 40 were examples of a theme, while 20 were "distractors" representing no theme. Two raters unfamiliar with the study, one resident and one medical student, volunteered to independently rate the excerpts. Our aim was to conservatively judge whether the themes were intelligible on their own. As in signal detection

analysis, raters' responses were scored as either a hit, correct rejection, miss, or false alarm. The *d'* statistic was computed to capture these four categorizations of the excerpts. The *d'* method was chosen rather than the traditional Cohen's kappa because our constant comparisons returned more categories than kappa can address. The *d'* analysis has been empirically tested to provide an equally interpretable statistic as Cohen's kappa.[18]

We conducted statistical analyses on the themes that emerged using IBM® SPSS® Statistics Version 25 (Mission Hills, CA). For all statistical tests, a significance level <0.05 was used. We used binary logistic regression to determine whether there was a significant difference in whether a theme had ever been expressed, comparing both junior trainees (PGY-1 and PGY-2) to senior trainees (PGY-3 and above) and interns (PGY-1) to all other trainees (PGY-2 and above). This analysis was performed by examining the data for each of the three videos separately, in addition to combining data for all three videos and examining differences in whether a participant had ever expressed a theme during their interview. We used the Mann-Whitney test to examine significant differences in the number of times each theme was expressed, examining each video individually and the totaled thematic data from all three videos, and comparing junior trainees to senior trainees and interns to all others.

In addition to using statistical analysis to further examine the themes identified, we used binary logistic regression to determine significant differences in whether a participant chose to stop a video to give feedback. Again, we compared junior trainees to senior trainees and interns to all others. We completed this analysis for each of the three videos in addition to examining totaled differences in whether a participant had ever stopped a video. For Video 2, we also used binary logistic regression to determine significant differences in whether a participant ever mentioned that the gallbladder shown was "difficult" or "complicated." We compared junior trainees to senior trainees and interns to all others.

Results

Cohort characteristics

We interviewed 15 of 68 residents (22%) and one PGY-6 fellow of approximately 38 fellowship positions (3%). All participants were between PGY-1 and PGY-7. Four (25%) were PGY-1, 12 (75%) were PGY-2 or above, eight (50%) were junior (PGY-1 and PGY-2) trainees, and eight (50%) were senior (PGY-3 and above) trainees.

Inter-rater analysis

The initial analysis of transcripts resulted in 22 themes. At this first pass, we conducted the *d'* analysis, and it showed positive but mixed reliability. Rater 1 discriminated among the correct categories with a positive, non-zero *d'* statistic of 0.91 (large effect). Rater 2 discriminated among the correct categories with a positive, non-zero *d'* statistic of 0.54 (medium effect). The large number of categories and small number of excerpts per category that raters were presented with may have resulted in the medium effect for Rater 2. To ensure we had developed a coherent set of themes, we decided to resolve all differences by revising categories to fit the data more consistently. This resulted in 21 final categories. For example, we discarded first-pass categories we had called "Shares Strategy" and "Command or Directive" and instead created the categories "Recommended Action" and "Analogy" to better categorize the data ascribed to the original two themes. We also created new categories to capture more of each transcript, as we found that many participants' observations were not categorized by our initial

coding scheme. These new categories were "Correct Action" "Incorrect Action" and "No Educational Strategy Needed."

Thematic analysis

Of our 21 final themes, we identified six themes which differed significantly between interns and all other trainees, and three themes which differed significantly between juniors and seniors. The themes that differed significantly are shown in Table 1. We must note that coding for the category "Incorrect Action" was not straightforward and relied primarily on context clues rather than keywords, unlike the other listed categories. No participants labeled a learner's actions as "incorrect" or "wrong." Only one participant used the word "wrong" at all, but not to disparage what they were seeing, as the participant asserted: "there was nothing really wrong with technique; " and, "nothing was necessarily wrong or dangerous." Three participants did describe the learner as "struggling." However, most phrases coded as "Incorrect Action" were counterpoints to something that the participant noted the learner had done well, i.e., "the learner did X well, but they still did Y."

Statistical analysis

Binary logistic regression revealed no significant thematic differences between trainee groups for Video 1 or Video 3. For Video 2, we did find some significant differences, with junior trainees more likely to express "Lean In," interns more likely to profess "No Concern," and trainees PGY-2 and above more likely to mention "Didactic Educational Method" (Table 2). When examining combined thematic data from all videos, we found junior trainees were more likely to note "Correct Action," interns were more likely to speculate on "Learner's Actions," and trainees PGY-2 and above were more likely to speculate on "Learner's Emotions" and recommend "Didactic Educational Method" (Table 2).

Mann-Whitney testing revealed no significant differences between trainee groups for Video 1. In Video 2, juniors provided more "Recommended Actions" to the learner. In Video 3, interns made more speculations on "Learner's Thoughts," while trainees PGY-2 and above noted more "Incorrect Actions" (Table 3). After combining data, we found that trainees PGY-2 and above made more "Didactic Educational Method" recommendations (Table 3).

We also examined two categories of data that were not captured by thematic analysis: whether a participant paused for feedback, and whether a participant noted if the gallbladder in Video 2 was "difficult" or "complicated." Binary logistic regression revealed no significant difference between interns and all other trainees or junior trainees and senior trainees for any of the videos when comparing whether a video was paused for feedback. This same type of analysis revealed that seniors were more likely to note that the case in Video 2 involved a "difficult" or "complicated" gallbladder (p = 0.028).

Discussion

The goal of this study was to explore the feasibility of a vignette-based simulated teaching scenario for both eliciting and assessing teaching knowledge (PK) and operative knowledge (PCK). In order for the method to be conceptually feasible, it would need to elicit responses thematically related to PK and PCK, display variability, and show reliable patterns; additionally, to succeed as an assessment tool, it should suggest evidence of differences between trainees of varying educational levels. We examined trainees' knowledge instead of teaching attendings' knowledge, as at this time developing a scaled set of "correct" responses fell beyond the

Table 1
The nine themes that significantly differed between trainee groups and their characteristics. PK = pedagogical knowledge, PCK = pedagogical content knowledge.

Theme	Definition	Keywords	Transcript Example	PK, PCK, or Both?
No Concern	The participant states that they are not concerned about the safety of the learner's actions.	"not worried" "not concerned" "safe" "no danger"	"I didn't think that he was, or she was, in any danger"	Both
Lean In	The participant expresses that it is better to quickly correct the learner.	"jump in" "correct" "intervene" "early" "first" "before the end" "on-the-spot"	"I'm the proponent of, as I see things, to make the corrections, instead of doing it all at the end"	PK
Learner's Thoughts	The participant speculates about the learner's thoughts.	"thinking" "thought"	"They are thinking about the steps of the operation"	PK
Learner's Emotions	The participant speculates about what the learner may be feeling.	"nervous" "worried" "confident" "comfortable" (any named emotion)	"They're probably nervous in this situation"	PK
Learner's Actions	The participant speculates about what the learner is trying to do.	"trying to" "attempting to"	"They're just trying to dissect the tissue off"	Both
Didactic Educational Method	The participant suggests an educational method such as a lecture, textbook, or video.	"video" "lecture" "book"	"Maybe watching a video or two might help"	PK
Recommended Action	The participant states what the learner should do.	"recommend" "should do" "should try" "would tell to" "better to" "OK to"	"I would recommend that the person on camera use two hands to stabilize it"	PCK
Correct Action	The participant notes that the learner completed the right specific action.	"right" "correct" "fine" "nicely" "well" "ok" + specific action	"They're moving anterior and posterior to the gallbladder really nicely"	PCK
Incorrect Action	The participant notes that the learner completed the wrong action.	"should not do" "don't think" "struggling" "too much" "[learner] did X well, BUT"	"Playing around with the camera a little too much"	PCK

scope of this feasibility study. Instead, we wished to investigate an assessment method for both PCK and PK that could be further instrumented with scoring if needed.

Pedagogical content knowledge

The themes "Recommended Action," "Correct Action," and "Incorrect Action" are all direct reflections of a trainee's PCK, as they demonstrate the trainee's knowledge of the procedural steps. The themes "No Concern" and "Learner's Actions" show both PCK and PK, as the trainee must employ their surgical knowledge in order to state whether an action is safe ("No Concern"), and must be familiar with the steps of the operation in order to predict what the trainee will do next ("Learner's Actions"). These themes demonstrate that we were able to elicit PCK from the trainees through their feedback to a learner.

Our results also suggested an evolution of knowledge about the procedure between interns and more experienced trainees. Interns were more likely to express "No Concern" about safety in Video 2, in which the dissecting instrument was briefly out of sight in a critical spot. However, they and junior trainees appeared to sense something amiss, as these groups gave more "Recommended Actions" to the learner in this video overall. We suggest these responses reflect nascent knowledge. They may be capable of recognizing something is wrong and recall the right way to dissect, but they lack the knowledge to interpret the out-of-sight instrument as unsafe.

Results from Video 3 may further show the evolution of content knowledge. Unlike Video 2, Video 3 did not contain instances where the authors felt that safety could be called into question. The learner in Video 3 appeared to be operating expertly to PGY-1 participants, but those above intern level could still pick out at least one "Incorrect Action." Interns made no "Incorrect Action" observations during Video 3. It could be that the operating skills and technique of a learner as advanced as one in Video 3 may be difficult to critique from the standpoint of an intern.

Pedagogical knowledge

The themes "Lean In," "Didactic Educational Method," "Learner's Thoughts," and "Learner's Emotions," are all direct reflections of trainees' PK, as the first two are examples of a teaching strategy, while the other two are examples of empathy, a skill critical to constructive teaching. "No Concern" and "Learner's Actions" demonstrate both PCK, as explained above, and PK, as the former requires the knowledge of a teacher knowing when to intervene with a learner, and the latter is also an example of empathy. These themes show that PK, in addition to PCK, was also easily elicited from trainee feedback to a simulated learner.

There were several statistically significant thematic differences between trainee groups pointing to an evolution of teaching abilities during training. When examining differences in what trainees believed the learner was thinking, doing, or feeling, we found that interns were more likely to speculate on the learner's actions, while their more experienced counterparts were more likely to speculate on the learner's emotions. Interns also more frequently speculated on the learner's thoughts, although this was specific to Video 3. Interns appeared to mainly summarize what the learner was doing, narrating a play-by-play of thoughts and actions. Senior trainees

Table 2
Significant results of binary logistic regression analysis of thematic data. Video # "Total" indicates that this difference was found when the data was combined for all videos.

	Theme	Video #	P-value	# Junior Yes (%)	# Senior Yes (%)
Junior vs. Senior	Lean In	2 (Kittner)	0.028	3 (37.5%)	0 (0%)
	Correct Action	Total	0.028	8 (100%)	5 (62.5%)
	Theme	**Video #**	**P-value**	**# Intern Yes (%)**	**# PGY2+ Yes (%)**
Intern vs. PGY2+	No Concern	2 (Kittner)	0.033	3 (75%)	2 (16.7%)
	Learner's Emotions	Total	0.033	0 (0%)	6 (50%)
	Learner's Actions	Total	0.033	4 (100%)	6 (50%)
	Didactic Educational Method	2 (Kittner)	0.018	0 (0%)	7 (58.3%)
	Didactic Educational Method	Total	0.001	0 (0%)	10 (83.3%)

Table 3
Significant results of Mann Whitney testing of thematic data. Video # "Total" indicates that this difference was found when the data was combined for all videos.

	Theme	Video #	P-value	Mean Frequency Junior	Mean Frequency Senior
Junior vs. Senior	Recommended Action	2 (Kittner)	0.038	4.625	1.625
Intern vs. PGY2+	Learner's Thoughts	3 (Cautery)	0.02	2.5	1.083
	Didactic Educational Method	Total	0.013	0	3
	Incorrect Action	3 (Cautery)	0.004	0	1.167

focused on what emotions could be driving or distracting the learner from correct actions.

Some differences for Video 2 were also striking. We found that junior trainees were more likely to express that they preferred a "Lean In" method of teaching during this video, meaning they believed it better to intervene and correct right away when they saw the learner beginning to err. This finding suggests that senior trainees may be more likely to allow a learner to self-correct errors, though why senior trainees would be less directive than juniors is unclear. One possibility is that senior trainees' increased surgical knowledge hinders their ability to explain basic surgical concepts that they have mastered in the past, a sign of the "expertise effect".[19] Alternatively, senior trainees may feel more confident identifying errors as "non-critical" and thereby feel more comfortable allowing learners to safely self-correct.

Finally, there were some differences in the educational method recommended by trainee groups. Compared to junior trainees, senior trainees were more likely to recommend a "didactic" educational method, such as a textbook, lecture, or video, as a follow-up to Video 2 and overall. Didactic instruction would be a good fit for a learner who is naïve of a disease or a procedure. Senior trainees' choice of didactic instruction may indicate they discerned a naiveté in the learner's performance about the anatomy or the procedure.

Teaching repertoire

Table 1 shows that despite the trainees' lack of formal education in how to teach, the themes about teaching that emerged from the transcripts were sophisticated. As a group, these themes echo major theories in education. For example, "Lean In" reflects the "scaffolding" technique, in which teachers adjust their level of help to the learner's needs.[20] Trainees' transcripts also showed that they had several ways to guide surgical decision-making: addressing correct and incorrect actions and providing recommendations. They sought to anticipate what learners might be thinking and feeling and could provide a detailed narration of what they saw learners doing. This reflects the ACGME definition of the teaching milestone that residents must be *interactive*.

Trainees gave praise and criticism, but scrupulously avoided blunt critiques. This reflects the ACGME definition of the teaching milestone that residents must be *constructive*. As noted in Results *Thematic analysis*, most of the themes that we identified were anchored by keywords, with the exception of the theme "Incorrect Action." Curiously, we found that none of the participants in our study labeled the learner's actions as "wrong" or "incorrect." Instead, they used the word "struggling" if they observed an action that they did not agree with. Many remarks that were "Incorrect Actions" were prefaced with a contrasting "Correct Action." We did not ask participants why they did not explicitly label these actions. It is possible that this may be a bias introduced by the simulation and observation of the study, and that a trainee would have been more direct in another setting. It is also possible that our participants did not want to sound overly critical when making a correction. This may be why many chose to first highlight what a participant did well before noting something amiss. This may also reflect an application of the commonly used "feedback sandwich"

technique, wherein a critique is preceded and followed by positive feedback.[21]

Implementation

We suggest educators in a surgery department could use this method by coming to consensus on two or three competencies they wanted to assess. Vignettes excerpted from filmed cases can represent a spectrum of performance in teaching or operative competencies. Faculty can create questions and answers. For example, like Video 2, a vignette can be excerpted from a filmed case showing the active instrument heading out of view, among other kinds of errors. The trainee would be asked to list all actions in the film they would have done differently, and why. A rubric would be developed to score trainees' answers. Alternatively, one could time how long it takes for the trainee to notice the instrument veering off course and hit a key on a keypad. In *Future directions* we discuss the ways in which this method could be deployed educationally.

Future directions

We have explored the brief vignette-based interview as a potential assessment, but it seems reasonable that it could be used to teach. An attending and resident could use video vignettes to discuss operative or teaching competencies. Like video debriefing, this would be in-person and one-on-one, but they could also be used as starting points for a rich discussion at a resident boot camp, or among faculty in a group meeting. They can be embedded in online tutorials about teaching or about operative decision making with questions for the resident to answer about the material. The approach gives surgical trainees, particularly interns and junior residents, faculty-approved operative training material. Accordingly, these vignettes could be incorporated into a Residents as Teachers program.

Our method depends on the "teach back" strategy. Research has provided evidence to the common observation that teaching someone else increases the teacher's knowledge of the subject. A 2018 study suggested that "teach back" compels one to practice remembering the content so one can explain to a learner, finding that teaching helped participants remember material when they had to teach from their own memory without notes.[22] Another reason why teaching improves knowledge is the self-explanation effect.[23] Teaching obliges one to explain, and when an explanation is not serving the learner well, the teacher must elaborate and find other ways to convey material. Also, researchers have found that if medical students can be prompted to elaborate and explain a topic, they will retain more of the material.[24]

For the vignette approach to work, faculty should first come to consensus about a few critical or difficult points they want to illustrate for learners, such as when a learner is struggling or, by comparison, is ready to increase responsibilities. Note that unlike video debriefing, our vignettes were short and deployed uniformly across the participants. The films were curated a priori, de-identified, and none of the participants were in the films.

Creating a vignette is not an onerous task. Most of the effort is having productive discussions about what the vignettes should

illustrate. Once we chose to use laparoscopic cholecystectomy cases, we each searched for and shared clips of films that we had. We sought to avoid "floor" effects, which occur when interns have no familiarity with a procedure, and "ceiling effects," in which a procedure is too easy for the senior trainee to demonstrate advanced competencies. In our case, the laparoscopic cholecystectomy was selected because it is a standard part of the first-year curriculum but requires more years of mastery. We envision that deploying this method could provide the opportunity for on-the-spot correction and didactics for surgical trainees.

Limitations

Our study had several limitations. First, our sample size of 16 was relatively small. Despite identifying significant thematic differences, the sample size may have contributed to Type II error regarding our non-significant themes. Second, our convenience sampling might have introduced selection and self-selection bias. The data might not be representative of all the trainees in our program, nor trainees from other programs. As with many qualitative analysis studies, our study may be subject to observer bias, although we did our best to mitigate this with an interview script and by validating our findings with inter-rater analysis and statistical tests to find significant differences. Finally, as we conducted face-to-face interviews, our study may have been impacted by the Hawthorne effect.

Conclusions

We simulated a teaching scenario anchored with three short video clips of laparoscopic cholecystectomies to elicit surgical and teaching knowledge from general surgery trainees. Analysis of the feedback trainees provided to simulated learners revealed some significant differences related to both surgical and teaching knowledge between trainee experience groups based on postgraduate year. We suggest that this simulated teaching scenario could be adapted for use as a uniform, rich assessment and teaching method for trainees as they progress through their training as surgeon-educators.

Funding

Research data

Deidentified transcripts available upon requests.

Acknowledgements

Abbreviations

PK   pedagogical knowledge
PCK   pedagogical content knowledge
PGY   postgraduate year
KE   knowledge elicitation

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amjsurg.2020.10.018.

References

1. Sachdeva AK, Flynn TC, Brigham TP, et al. Interventions to address challenges associated with the transition from residency training to independent surgical practice. *Surgery*. 2014;155:867–882. https://doi.org/10.1016/j.jsurg.2013.12.027.
2. Sandhu G, Magas CP, Robinson AB, et al. Progressive entrustment to achieve resident autonomy in the operating room: a national qualitative study with general surgery faculty and residents. *Ann Surg*. 2017;265:1134–1140. https://doi.org/10.1097/SLA.0000000000001782.
3. Bohnen JD, George BC, Williams RG, et al. The feasibility of real-time intra-operative performance assessment with SIMPL (System for Improving and Measuring Procedural Learning): early experience from a multi-institutional trial. *J Surg Educ*. 2016;73:118–130. https://doi.org/10.1016/j.jsurg.2016.08.010.
4. Accreditation council for graduate medical education. The general surgery milestone project. Accessed 1 March 2020 https://www.acgme.org/Portals/0/PDFs/Milestones/SurgeryMilestones.pdf?ver=2015-11-06-120519-653; 2015.
5. Sutkin G, Littleton EB, Kanter SL. Intelligent cooperation: a framework of pedagogic practice in the operating room. *Am J Surg*. 2018;215(4):535–541. https://doi.org/10.1016/j.amjsurg.2017.06.034.
6. Sutkin G, Littleton EB, Kanter SL. How surgical mentors teach: a classification of in vivo teaching behaviors part 1: verbal teaching guidance. *J Surg Educ*. 2015;72(2):243–250. https://doi.org/10.1016/j.jsurg.2014.10.003.
7. Sutkin G, Littleton EB, Kanter SL. How surgical mentors teach: a classification of in vivo teaching behaviors part 2: physical teaching guidance. *J Surg Educ*. 2015;72(2):251–257. https://doi.org/10.1016/j.jsurg.2014.10.004.
8. Geary AD, Hess DT, Pernar LI. Resident-as-teacher programs in general surgery residency—Context and characterization. *Am J Surg*. 2019;76:1205–1210. https://doi.org/10.1016/j.jsurg.2019.03.004.
9. Sutkin G, Wagner E, Harris I, Schiffer R. What makes a good clinical teacher in medicine? A review of the literature. *Acad Med*. 2008;83:452–466. https://doi.org/10.1097/ACM.0b013e31816bee61.
10. Brovelli D, Bölsterli K, Rehm M, Wilhelm M. Using vignette testing to measure student science teachers' professional competencies. *Am J Educ Res*. 2014;2:555–558. https://doi.org/10.12691/education-2-7-20.
11. Busari JO, Scherpbier AJ. Why residents should teach: a literature review. *J Postgrad Med*. 2004;50:205.
12. Minor S, Poenaru D. The in-house education of clinical clerks in surgery and the role of housestaff. *Am J Surg*. 2002;184:471–475. https://doi.org/10.1016/s0002-9610(02)01001-2.
13. Pelletier M, Belliveau P. Role of surgical residents in undergraduate surgical education. *Can J Surg*. 1999;42:451–456.
14. Cooke NJ. Varieties of knowledge elicitation techniques. *Int J Hum Comput Stud*. 1994;41(6):801–849.
15. Hudlicka E. *Summary of Knowledge Elicitation Techniques for Requirements Analysis. Course Material: Human Computer Interactions*. Worcester Polytechnic Institute; 1997.
16. Geiwitz J, Kornell J, McCloskey B. *An Expert System for the Selection of Knowledge Acquisition Techniques*. California: Anacapa Sciences; 1990. Technical Report 785-2, Contract No. DAAB07-89-C-A044.
17. Boeije H. A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Qual Quantity*. 2002;36:391–409.
18. Grant MJ, Button CM, Snook B. An evaluation of interrater reliability measures on binary tasks using d-prime. *Appl Psychol Meas*. 2017;41:264–276. https://doi.org/10.1177/0146621616684584.
19. Bromme R, Rambow R, Nückles M. Expertise and estimating what other people know: the influence of professional experience and type of knowledge. *J Exp Psychol App*. 2001;7:317. https://doi.org/10.1037/1076-898X.7.4.317.
20. Sanders D, Welk DS. Strategies to scaffold student learning: applying Vygotsky's zone of proximal development. *Nurse Educat*. 2005;30:203–207. https://doi.org/10.1097/00006223-200509000-00007.
21. Dohrenwend A. Serving up the feedback sandwich. *Fam Pract Manag*. 2002;9:43–46.
22. Koh AWL, Lee SC, Lim SWH. The learning benefits of teaching: a retrieval practice hypothesis. *Appl Cognitive Psych*. 2018;32(3):401–410. https://doi.org/10.1002/acp.3410.
23. VanLehn K, Jones RM, Chi MT. A model of the self-explanation effect. *J Learn Sci*. 1992;2(1):1–59. https://doi.org/10.1207/s15327809jls0201_1.
24. Peixoto JM, Mamede S, de Faria RMD, et al. The effect of self-explanation of pathophysiological mechanisms of diseases on medical students' diagnostic performance. *Adv Health Sci Educ*. 2017;22(5):1183–1197. https://doi.org/10.1007/s10459-017-9757-2.