'It depends': The complexity of allowing residents to fail from the perspective of clinical supervisors

MEDICAL TEACHER
https://doi.org/10.1080/0142159X.2021.1984408

'It depends': The complexity of allowing residents to fail from the perspective of clinical supervisors

Jennifer M. Klasen[a], Pim W. Teunissen[b,c], Erik W. Driessen[c] and Lorelei A. Lingard[d,e]

[a]Clarunis, Department of Visceral Surgery, University Centre for Gastrointestinal and Liver Diseases, University Hospital Basel, Basel, Switzerland; [b]Department of Obstetrics and Gynecology, Maastricht University Medical Centre, Maastricht, The Netherlands; [c]School of Health Professions Education (SHE), Faculty of Health Medicine and Life Sciences (FHML), Maastricht University, Maastricht, The Netherlands; [d]Centre for Education Research and Innovation, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; [e]Department of Medicine, University of Western Ontario, London, Ontario, Canada

ABSTRACT

Purpose: Clinical supervisors acknowledge that they sometimes allow trainees to fail for educational purposes. What remains unknown is how supervisors decide whether to allow failure in a specific instance. Given the high stakes nature of these decisions, such knowledge is necessary to inform conversations about this educationally powerful and clinically delicate phenomenon.

Materials and methods: 19 supervisors participated in semi-structured interviews to explore how they view their decision to allow failure in clinical training. Following constructivist grounded theory methodology, the iteratively collected data and analysis were informed by theoretical sampling.

Results: Recalling instances when they considered allowing residents to fail for educational purposes, supervisors characterized these as intuitive, in-the-moment decisions. In their *post hoc* reflections, they could articulate four factors that they believed influenced these decisions: patient, supervisor, trainee, and environmental factors. While patient factors were reported as primary, the factors appear to interact in dynamic and nonlinear ways, such that supervisory decisions about allowing failure may not be predictable from one situation to the next.

Conclusions: Clinical supervisors make many decisions in the moment, and allowing resident failure appears to be one of them. Upon reflection, supervisors understand their decisions to be shaped by recurring factors in the clinical training environment. The complex interplay among these factors renders predicting such decisions difficult, if not impossible. However, having a language for these dynamic factors can support clinical educators to have meaningful discussions about this high-stakes educational strategy.

KEYWORDS
Clinical supervision; decision-making; intuition; failure; trainee learning

Introduction

Clinical supervisors are trusted with challenges such as providing learning opportunities, instructing, giving feedback to support trainees and nudging them to the edge of their clinical competence so that they develop (Kilminster and Jolly 2000; Kilminster et al. 2007; Cate et al. 2007; Farnan et al. 2010, 2012; Walton and Barraclough 2013; Martin et al. 2014, 2017; Swanwick et al. 2018; Gilchrist et al. 2021). On a daily basis, supervisors have to engage in decisions on how to allow trainees to strive and develop their skills (Kilminster and Jolly 2000; Cottrell et al. 2002; Kilminster et al. 2007). This will depend on different factors and some of the complexities of these decisions have been studied in the medical education literature on clinical supervision (Kogan et al. 2009; Hauer et al. 2014, 2015; Holzhausen et al. 2017; Gilchrist et al. 2021). Part of the challenge is that supervisors will find themselves in situations where they know or sense that the trainee working at the edge of their competence might fail in ways that potentially impact patient care (Bump et al. 2015; Finn et al. 2018; Li et al. 2021). In some instances, such failures can offer powerful learning experiences, but they confront supervisors with an apparent trade-off between trainee learning and patient safety (Wu et al. 1991; Mazor et al. 2005; Fischer et al. 2006; Wong et al. 2010; Bradley et al. 2013; Quillin et al. 2013; Klasen and Lingard 2019; Klasen

Practice points

- Clinical supervisors allow trainees to fail clinical activities for educational purposes, intervening, if necessary, to avoid harming patients.
- Supervisory decisions to allow failure are intuitive, influenced by four main factors – patient, trainee, supervisor, environment – which interact in a complex, non-linear manner.
- These supervisory decisions are high stakes and may not be predictable from one situation to the next.
- The development of a definition and a language for this phenomenon supports clinical educators in having meaningful discussions about this educational strategy.

CONTACT Jennifer M. Klasen jennifer.klasen@clarunis.ch Department of Visceral Surgery, Clarunis University Center for Gastrointestinal and Liver, Diseases, St. Claraspital and University Hospital Basel, CH-4002 Basel, P.O. BOX CH-4002, Switzerland

et al. 2019; Wong and Lim 2019). Although, we know how supervisors navigate trust and entrustment decisions to grant autonomy at a designated level of supervision (Dijksterhuis et al. 2009; Kennedy et al. 2005, 2009; Sterkenburg et al. 2010; Halpern and Detsky 2014; Hauer et al. 2014, 2015; Cate et al. 2016; Hashimoto et al. 2016; Holzhausen et al. 2017; Crockett et al. 2019), we know little about how supervisors decide what to do in situations where they could allow failure for educational purposes (Ross et al. 2011; George et al. 2018; Klasen et al. 2019).

Our recent interview study found that clinical supervisors from a variety of postgraduate specialty contexts reported allowing failure for educational purposes (Klasen et al. 2019). Supervisors reported allowing a variety of erroneous actions in different clinical settings so that trainees could learn by failing and recovering from it. Supervisors expected both technical and emotional benefits from these allowed failures for trainees, while they endeavored to limit the consequences for patients. The first content analysis of these reported failures sets the stage to understand the concept of allowing failure and has yielded a preliminary definition of the phenomenon: 'While supervising a trainee's clinical performance, the supervisor detects an imminent trainee mistake, has the opportunity to intervene but deliberately chooses not to do so because the educational gain for the trainee is perceived to outweigh the (potential) consequences for the patient' (Klasen et al. 2019). This definition provides a useful starting point for conversations about the educational practice of allowing failure. However, such conversations must also be informed by insights into how this educational practice is enacted. It is this gap that the current paper addresses, by exploring the question: what is the judgment process that supervisors employ to decide whether or not to allow trainee failure during a clinical performance? The exploration of how supervisors decide to allow failure in one situation but not another will allow us to refine the preliminary definition of this phenomenon, develop a shared vocabulary for discussing it in our medical education community, and provide a framework for faculty development efforts to prepare clinical faculty for these high-stakes supervisory decisions.

Methods

Research design

We used constructivist grounded theory (CGT) to explore the complex social process of supervisors allowing failure in the clinical workplace (Charmaz 2006; Watling and Lingard 2012). With approval by Swiss (EKOS) and Canadian institutional research Ethics, we chose to conduct individual interviews voluntarily, calculating that participants would feel comfortable speaking candidly with a colleague (JK). The data collection and analysis were iterative, and we, as constructivist researchers, acknowledge that our interpretation of both was shaped by our experiences, which we shared and discussed in regular meetings.

We proposed the term 'allowing failure for educational purposes' as a consistent starting point for our research program. We chose the term 'failure' for two reasons: first, the definition of failure as 'lack of success in doing something,' (https://Dictionary.Cambridge.Org/de/Worterbuch/Englisch/Failure) matches better with our context of clinical learning than the definition of 'error'; 'to make a mistake or to do something wrong' (https://Dictionary.Cambridge.Org/de/Worterbuch/Englisch/Error). Second, as educational scholars, we adapted this term from other domains of education, where authors have used 'failure' rather than 'error' if educational benefits can thus be obtained (Kapur 2015, 2016; Mylopoulos et al. 2018; Varpio and Nagler 2018; Steenhof et al. 2019; Young 2019; Steenhof et al. 2020). We aim to develop a common language in the literature of medical education that can build on discussions of learning from failure in other educational contexts.

Sampling strategies, settings and data collection

All study participants held a faculty position in a hospital and identified themselves as clinical supervisors, with varying stages of experience. We recruited participants in Switzerland and Canada from a range of clinical specialties (Klasen et al. 2019). We tried to purposely select our sample to secure information-rich participants who are interested in the study question and showed willingness to participate after hearing about our study topic (Patton 1990). We acknowledge that such participants have a subjective position. Further, our sample is not representative and therefore our results cannot be generalized. However, we endeavoured to interview a range of participants to understand diverse perspectives on this phenomenon. The presence of two discrepant examples in the dataset suggest that, even though participants self-selected for their interest in the study, we captured some who did not represent themselves as using this teaching strategy. Having an interviewer (JK) from the surgical field, surgical supervisors in senior faculty positions were interviewed at the beginning of the study (P1, P2, P4, P5) as a rapport building strategy (Figure 1). We expanded the interview process from surgical supervisors with experiences as clinicians and supervisors to other specialties and colleagues with medical education background, formal training such as a master's or Ph.D. degree, or medical education research (P3, P6, P7, P8, P9, P10). We expected them to be informed about clinical supervision concepts and be able to reflect in-depth on different supervisory strategies. Further, we included less experienced supervisors from all specialties with medical education background (P11-P16). We chose the last three interview candidates for specific reasons: P17 was recommended by a colleague interested in the study topic, working as a pediatrician. P18 presented an informal conversation before the interview as a discrepant case, saying not using the strategy of allowing failure rather than the last interview participant, who confirmed the dataset in perceiving to use this strategy (P19) regularly.

Including participants from two different sites had the intention to broaden our perspective on trust as it may relate to institutional culture, while both sites were conveniently chosen due to the research team settlement. Seventeen interviews were conducted in English with native and non-native speakers, while two were conducted in German for the participants' convenience and translated professionally afterwards. The semi-structured, individual interviews, conducted by JK, lasted between 45 and 75 min. All interviews were audio-recorded and subsequently transcribed verbatim. We explored participants'

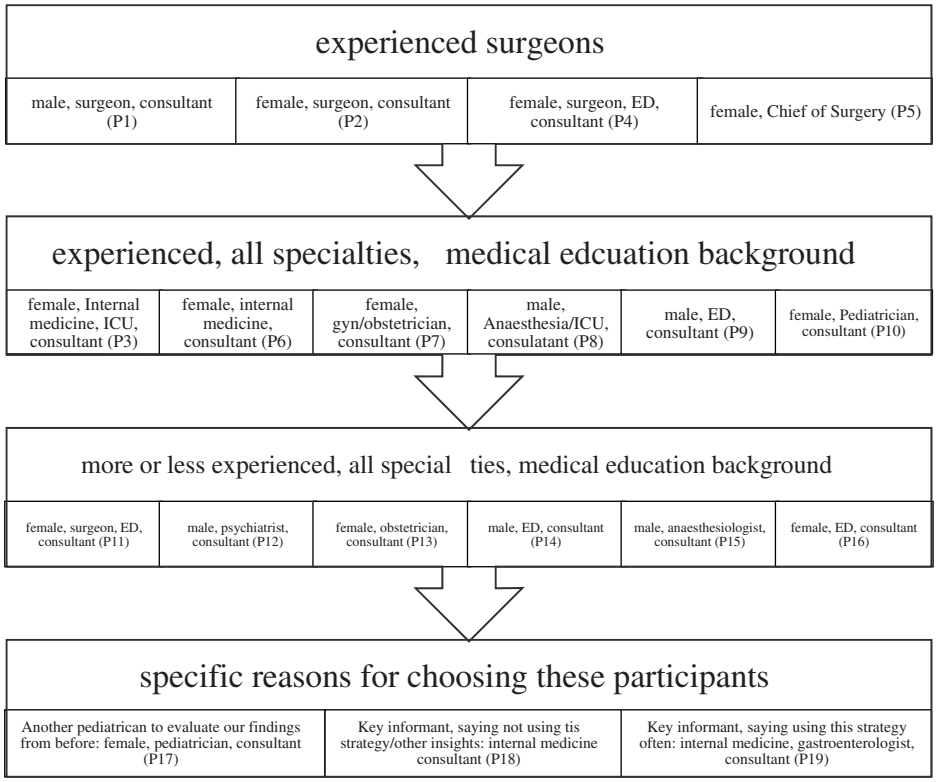

Figure 1. Step-by-step sampling strategy.

perceptions of learning from failure in general and allowing trainee failure in the clinical workplace. Initially, we used an 'easing-in' strategy, in case the idea of allowing failure might be perceived by participants as a sensitive topic. This strategy involved beginning the interview with a broad question about whether they had ever seen a trainee fail in clinical situations and whether they had anticipated that such a mistake might occur, and afterwards inquiring about the strategy of allowing trainees to fail for learning purposes. Because our first 4 participants did not exhibit inhibitions about discussing the strategy of allowing failure, we dropped this 'easing-in strategy' in later interviews in order to maximize the interview time. These interviews began with a brief introduction to the strategy of allowing failure in other educational settings and then asked about whether and how clinical supervisors use this strategy in clinical training. After 19 interviews of clinical supervisors, we perceived to reach the point of theoretical sufficiency, where no new codes or themes occurred (Morse 2000).

Data analysis and research team

In accordance with constructivist grounded theory (CGT), we collected and analyzed the data iteratively during a process of constant comparison of the data, refining the interview protocol and engaging in theoretical sampling to explore recurrent categories (Tavakol et al. 2006). Two kinds of analysis were conducted. During the analysis of the CGT data, we realized a richness in the narratives and decided to do a second qualitative content analysis of each reported instance of an allowed failure in the data which yielded a description of the key features of this educational phenomenon and a first definition of the phenomenon in a previously published paper (Klasen et al. 2019).

The constructivist grounded theory analysis reported in this paper differs from the already published content analysis in two main ways. First, it aims to explain how supervisors perceive making the decision to allow failure for educational purposes, and second, we used a different methodology (CGT) compared to the previously published one to answer our research questions (Klasen et al. 2019). Also, by providing two different research questions – how the phenomenon of allowing failure presents itself in the data and how supervisors decide to allow failure in different clinical situations, we used two distinct different analysis techniques to cover two different problems. As a first step in this analysis and following CGT procedures, we grouped related topics and defined these using gerunds, such as assessing and decision making, taking a risk, or calculating patient safety. We refined these recurring categories using the constant comparative method, in which new instances of a theme were compared to all existing instances, and the definition of each theme revised until it accounted for all dimensions of all instances (Tavakol et al. 2006). We used Quirkos as a qualitative software to support the analysis.

The international team consisted of three PhD-trained medical education researchers (LL, PWT, ED) and one PhD candidate (JK). JK and PWT are clinical supervisors in abdominal surgery and in obstetrics and gynecology, respectively. JK's surgical experience and interest shaped the direction of the results, and the sampling of the participants from surgical disciplines in being able to share experiences and rapport, especially in exploring trainees' failures during surgical procedures.

Results

Twelve women and seven men from 11 different Swiss and Canadian institutions were interviewed. The participants'

expertise as clinical supervisors ranged from 2.5 years to over 25 years. Thirteen of the 19 participants worked in nonsurgical environments such as emergency medicine, critical care, internal medicine or subspecialties, pediatrics, and psychiatry. The surgical specialties were represented by general surgeons and obstetricians/gynecologists.

Participants explained that they understood the decision to allow failure as an in-the-moment, intuitive phenomenon. We identified four main factors that they perceived as important influences on the decision to allow failure: patient-, trainee-, supervisor-, and environment factors. Supervisors portrayed these factors as interacting with one another and producing variability in their decisions.

Intuitive decisions

Although they often made the intuitive decision to allow failure in clinical situations, many participants declared that the interviews were the first time they had reflected about it outside the moment of allowing failure itself. The decision to allow trainees to fail was characterized as largely unconscious: 'I think I consciously don't do it [allowing failure] very often, but I think I unconsciously perhaps may be doing it more than I think' and 'I'm probably unconsciously allowing for it to happen without calling it that' (P12). As an emergency clinician reflected: 'And now, I am doing this interview, I really think of it. … I was forced to think consciously about things I'm doing unconsciously in my daily life, so it's very interesting.' (P16) In addition to characterizing the process as unconscious, participants also described it as intuitive, 'a point-to-point decision' (P9) arising 'in the moment' (P7). One supervisor described 'it was such a snap decision, too, it wasn't something that I hemmed and hawed over.' (P13)

Even as they characterized the decision to allow failure as unconscious and in-the-moment, supervisors tried to describe what these moments were like. In the interviews, we invited supervisors to reflect on these in-the-moment decision that they described as intuitive decisions. Some had already given this some consideration while others articulated their approach to such situations for the first time during the interview. Below, we illustrate how they tried to illuminate this decision, rationalizing feelings and different factors that they made explicit in the reflections during the interview.

Most interviews cited the role of 'feelings': 'The line between taking the risk (to allow failure) and taking over from the resident to make sure there is no further damage, or no damage at all, is only a feeling. This is very difficult to describe.' (P09) And every interview also presented logical explanations of factors that were weighed: 'But like the judgment if we don't want to talk about the environment, about the resident again, it's the patient as well. Like all these factors together it's like … I think it's a mixture of feelings and facts.' (P11) Another consultant confirmed:

> Yes, it's always feeling. It's always a combination. As I said before, a combination of the case, whether it's a stable situation, of the experience, whether you've experienced the same or a comparable situation before so it's always feeling which makes the decision. (P19)

Supervisors talked about their 'gut feeling' or 'intuition' being an important aspect of their decision to allow a failure for educational purposes. Describing this aspect of their decision-making, they acknowledged that 'It's not very scientific' (P2) and sometimes 'rather a feeling than a really conscious calculation' (P9), confirmed by another ER physician: 'I don't think I can calculate that, actually. I never consciously calculate before I do something.' (P16) Even the issue of potential risk for patients was described in this way: 'The line between taking the risk and taking over from the resident to make sure there is no further damage, or no damage at all, is only a feeling.' (P8)

As participants reflected on these in the moment decisions, they became more analytical.

In the next section, we illustrate in more detail the main factors that supervisors described weighing logically in deciding whether or not to allow failure.

Four main factors

Across all accounts of the decision to allow failure, four factors recurred consistently: patient, trainee, supervisor, and environment factors.

First, participants presented *patient factors* as the most important and predominant factor. Regardless of other factors, participants stated they would not allow failure if it posed unacceptable risk to the patient: 'I try to dose my interference with the severity of the sickness of the patients, but I think I never did, I just try to think when I saw, for example, in the resuscitation room, where they come in when they're really sick, I never let them do it by themselves.' (P3) Some supervisors declared other patient factors to consider if the patient was elderly, undergoing cancer treatment, or pediatric: 'If you have a young one, a very fragile person, for example, a child, of course, you don't tolerate anything, so I think it depends. I think you really look at your patients and think, well, here I have to be more careful, and you have to say more here.' (P2) or another consultant put it this way, drawing an example of an emergency colonoscopy:

> … Yes, as long as I see that the patient is stable. If the patient develops a tachycardia or a hypotension, of course I would not let the resident continue with the whole endoscopy. Then I would have to finish it but as long as a patient is stable it's okay. (P19)

Second, *trainee factors* were perceived as another main component influencing supervisors' decisions about allowing failure. The personality of young physicians was reported by many supervisors as a factor in their decisions. Confidence, both too much and too little, was recurrently discussed as part of what supervisors considered in relation to 'personality'. As one surgeon stated: 'if I had the feeling that the person is generally reasonable, then I would be way more open than if I have the impression the person is just Rambo and it's dangerous.' (P5)

As this quote illustrates, for this supervisor, an overconfident trainee would not be allowed to fail, while a trainee with 'reasonable' levels of confidence might be.

Knowledge of and comfort with trainees arose through the supervisor-trainee trusting relationship. Not uncommonly supervisors remarked that 'you need to have a certain relationship' (P1) to allow failure, explicitly connecting this to the idea of 'a foundation of trust' (P4) that must exist. As one participant explained, trainees '… have to

trust me. They have to think that I'm professional and that I don't judge them when they fail. I think this is the most important thing, that they're not afraid of me.' (P2) Such reflections suggest the relationship of trust might make supervisors more likely to allow failure. In the absence of such a relationship or with trainees they perceive as being afraid of them, they may be less likely to allow failure. Another participant explained how their knowledge of a trainee as 'a very careful person' or 'a bull in a china shop' (P1) informs their decision to allow failure.

Third, when considering allowing failure, participants also reflected on *supervisor factors* such as abilities and constraints. Many participants talked about their own comfort and confidence as a factor in their decision. As one supervisor stated: 'I think it's important that there be comfort with uncertainty, and you have to be of a predisposition that you don't need to have control over everything in order to allow failure to be part of how you teach.' (P12) All participants agreed that, in order to even contemplate allowing failure, supervisors need to '… be comfortable with the situation and the procedure and the problem itself' (P16). The data offer a striking range of examples where comfort level and confidence dictated the decision about allowing failure. For instance, one surgical supervisor acknowledged that she/he wouldn't tell the resident 'just cut it [the exposed anatomical structure] … if it's really you knew it was the vena porta, then I maybe can't fix it. And then it's game over for all of us' (P1), while another commented that 'if this mistake happens, I will be able to handle this afterwards' (P2), and a third reflected 'do I put them in situations where it's not unsafe for them to potentially fail? Yes, I give them those opportunities, that, for sure, but knowing that I can bail them out, I can back them up, I can make that work' (P18). While a general level of comfort and confidence was important, the decision to allow learner failure was also influenced by the supervisor's sense of their own current personal conditions. As this surgeon explained, 'sometimes, you want to risk more [so] that your student can learn, and on the other hand, you have days that you think, 'oh no, I have my own problems, so I don't want to add more'.' (P2) Fatigue was one such issue: as another supervisor explained, 'when you sleep three or four hours, you just don't have the nerves to go through this whole process' (P3) of allowing failure and attending to its technical and emotional consequences.

Last, in reflecting on such situations, participants noted how *environmental factors* influence their approach. For instance, in a case of an emergency endoscopy, the supervisor described how environmental factors such as having experienced nurses and sufficient time for the procedure can influence their decision:

> For example, once it is not like doing a mistake but in bleeding situations for example there are different methods to stop the bleeding. And sometimes I would prefer the one method and the resident for example would prefer another method, of which I am pretty sure it won't work but of course I let the resident try it, try to stop the bleeding. Most of it doesn't work and in such a situation I would prefer to have assistant endoscopy nurses who are experienced for example because the others might become very nervous because it's still bleeding. And then the whole situation would become nervous and that's not good, so experienced other staff of the ward and enough time. P19

Supervisors reported not to feel comfortable to allow failure in busy environments: 'When it's a stressful day, when I'm in a time rush or whatever, where there's time pressure, I'm not comfortable to do that kind of teaching method.' (P16). The lack of cognitive and emotional space in a busy environment was also a factor in dealing with the aftermath of being allowed to fail. Given their sense that 'it's me who has to afterwards clean up the mess again' (P3), supervisors wanted to know that the environment would allow for both this. They also recognized that 'the time to debrief' (P7) with the trainee was essential to 'make this a productive failure situation'. (P3)

'It depends': Interactions among the factors

The four main factors – patient, trainee, supervisor, environment – were not discrete; they interacted in a complex interplay to produce decisions about allowing failure. One ER physicians reflected on potential reasons, rationalizing:

> If there is a patient who is really annoyed and in pain and anxious and whatever then I wouldn't do it because it would make the situation just worse. If it's a child where you have really annoying or anxious parents, you wouldn't do it. If it's a patient who is patient [tolerant], then you can do it. Yeah, it depends on many things. But like the judgment if we don't want to talk about the environment, about the resident again, it's the patient as well. Like all these factors together … (P11)

In another elaborate example, these interactions were evident as participants reflected on how they approached particular situations. One obstetrician reported a representative scenario during a laparoscopic procedure 'where things had been kind of challenging throughout the case.' (P13) In trying to explain her 'snap decision' (P13), she began with her sense of how trainee factors and the supervisory-trainee relationship fed into the decision:

> I felt like he and I already have a pre-existing relationship and I know that he's capable. I also think that he feels I don't let him do enough, that I don't give him enough autonomy, so this seemed like a good opportunity to do so. … At one point, he asked me, well, how do you want to do this next step? And, I said, well, this is your case, so how would you like to do it? He made a suggestion and that's when I thought his suggestion is a) not a way I've ever seen it done before and b) I can think of how that would go wrong. But I said to him, okay, you have one chance. You get one try your way and if it doesn't work, we're doing it my way. (P13)

Layered on top of these trainee factors and supervisory relationship factors, however, she also perceived that their own state of 'being exhausted' factored into their decision, in this example articulated as:

> My guard was let down. …. I wonder if I was just so exhausted with having worried throughout this entire procedure that things were not going smoothly, that I kind of gave up. (P13)

Her recollection was further elaborated by their sense that environmental factors also played a role because

> … It was also the middle of the night, in the midst of a very busy call shift, where your brain is already thinking about 100 different other things, like, am I going to have to [perform a caesarean] section [on] that lady upstairs, will I have to do forceps for that woman, I hope the tracing has gotten better in that other room. (P13)

This multi-layered explanation is representative of how participants reflected at length on the various, intersecting factors that shaped their 'snap decisions'. However, there was no recurrent pattern discernable in such explanations regarding how the factors intersected, and, consequently, there was no strong sense of predictability from one decision-making instance to the next. Instead, our analysis suggests that the interplay of factors is dynamic and nonlinear, and whether a supervisor decides to allow failure or not 'depends on many things' (p11). The one exception was that supervisors consistently declared that the patient factor trumped the others, like the surgeon who said: 'it's always, yeah, it depends how dangerous it is for the patient' (P2) or the ER physician who explained: 'If it's a child where you have really annoying or anxious parents, you won't do it.' (P11)

Such detailed reflections suggest that supervisors are considering multiple factors, they are doing so in the moment, and the process is at least partly an intuitive one.

Discussion

This work provides insights into how clinical supervisors understand, after the fact and in what many acknowledged as their first conscious reflection on the process, their intuitive, in-the-moment, and sometimes high-stakes decisions to allow trainee failure. Clinical supervisors may find our results familiar from their own experience; however, our results provide the first empirical evidence of this familiar phenomenon. Supervisory decisions about allowing failure for educational purposes share a number of characteristics: they weigh different factors, prioritize the patient in the moment, and are intuitive. However, our results suggest key factors that influence these decisions, but the relationship among them is unstable, characterized by a recurring refrain of "it depends" in the data. Below we wrestle with this central finding and consider its implications for patients, supervisors and learners.

In conducting a CGT analysis, we set out to develop a model of how supervisors decide to allow trainee failure, with the hope that such a model might be used to inform, critique or even predict such supervisory decisions. However, while the four factors provide insight into how supervisors understand the risks and benefits of allowing failure for learning, they do not provide a model for understanding why failure might be allowed in one situation but not another. Supervisors' accounts demonstrated that, even when the factors appeared similar (e.g., a strong trainee, a trusting supervisory relationship, a stable patient, a familiar team), in one case a supervisor might allow failure and in another avert it. It just depends. And it depends on a set of factors that exist in dynamic, non-linear relation to one another. A factor such as an over-confident trainee, or the risk of patient bleeding, or an unfamiliar team member may get heavily weighted in one decision, but not the next.

Our data aren't the first to point to the complexity and non-linearity of supervisory judgement and decision-making in clinical settings. Holzhausen et al. described four factors influencing supervisors' entrustment decision-making – trainee characteristics, supervisor characteristics, characteristics of the task at hand and contextual factors (Holzhausen et al. 2017) – which resonate with those that we found in our study. They also acknowledged the complex interplay among these factors and the incomplete nature of their framework. Hauer et al. also explored the complexity of supervisory decisions in developing trust in their trainees while simultaneously caring for patients. A trusted supervisor-trainee relationship is influenced by different factors of the supervisor, trainee, the supervisor–-trainee relationship, task, and context, suggesting a similar non-linearity of supervisory decision-making (Hauer et al. 2015).

The intuitive nature of these decisions undoubtedly contributes to the non-linear, 'it depends' pattern of our results. What our participants describe as 'gut feeling,' drawing on intuition, resonates with the concept of implicit knowledge, composed of knowledge and practical experience acquired over several years, which is known to shape decisions and allow a quick reaction to a challenging situation (Norman et al. 2014; Norman et al. 2017; Sherbino et al. 2012).

Intuition is a complex construct, variously treated across different literatures but largely embraced as a necessary part of expert decision-making and recognized as a source of potential wisdom in the unconscious mind (Greenhalgh 2002; Edwards 2004; Gigerenzer 2007; Coombes 2010; Harteis and Frost 2012; Harteis and Billett 2013; Zander et al. 2016). Acknowledged as an essential feature of Type 1 reasoning in dual process theory, intuition is embraced as a necessary part of expert decision making and recognized as a source of potential wisdom in the unconscious mind (Sherbino et al. 2012; Norman et al. 2014; Norman et al. 2017). However, it is, as Kahneman and Klein put it, 'sometimes marvelous and sometimes flawed' (Kahneman and Klein 2009). Marvelous in that clinical supervisors are continuously recognizing and weighing multiple factors as they move through the day. But flawed in that they cannot afford everything equal attention, and they may not know why some factors influence them more than others. Here, it makes sense to see analogies to clinical or surgical decision-making. Physicians use the so-called System 1 thinking in recognizing patterns as rapid and intuitive decisions and System 2 thinking for analytical rationalizations of more complex situations (Flin et al. 2007; Sherbino et al. 2012; Norman et al. 2014; Norman et al. 2017). Our participants represent their decisions as intuitive, even unconscious, which suggests that we are dealing with a non-analytical phenomenon. And while interview data can offer insights into how supervisors may rationalize their decisions after the fact, we must take care not to assume that these data represent an analytical process in the moment. Therefore, in an effort to avoid the linear assumptions of most process models, we do not present a model of the four factors (Durning et al. 2015; Klasen and Lingard 2021). Instead, we offer a refinement of our preliminary definition of this educational phenomenon (Klasen et al. 2019). Italics signify added terms, while square brackets signify terms we have removed in this version: 'While supervising a trainee's clinical performance, the supervisor, *influenced by both intuition and a non-linear interplay of different factors*, detects an imminent trainee mistake, has the opportunity to intervene but [deliberately] chooses not to do so, because the educational gain for the trainee is

perceived to outweigh the (potential) consequences for the patient.' This definition includes the non-linear relationship of the four factors and the issue of intuition derived from this analysis; this definition also removes 'deliberately' because our data calls into question how often this decision is conscious and deliberate. Our refined understanding of the phenomenon of allowing failure for educational purposes has implications for patient safety, clinical supervision, and trainee learning.

Non-linear, 'it depends' decisions about allowing failure have implications for patient safety. Supervisors recognize that there is potential and uncertain impact on patient safety and, but they argue they would never allow it if they think it might harm a patient (Klasen et al. 2019). To take a critical standpoint on supervisors' analytical reflections on their decisions, of course they must assert that they won't let patients be harmed. If we take at face value their assertion that 'patient factors trump all', we cannot understand the nuanced reasoning they exert to judge degrees of acceptable harm in particular instances. In effect, our results show poignantly how supervisors' reason their way through the conundrum that keeping patients unaffected by trainee learning is a noble, necessary but ultimately unachievable goal. There is always a relationship between trainee development and patient risk, but it is complex one. For example, supervisors must consider not only the risk to today's patient, but also the risk to tomorrows. We might argue, a supervisory approach that absolutely safeguards today's patient increases the risk for future patients if the trainee hasn't been given the necessary autonomy to develop independence, while the relationship of both, trainees autonomy and patient safety is unknown but most likely nonlinear (George et al. 2018; Atkinson and Smink 2020). Thus, supervisors must constantly monitor not only the risk to the patient in front of them, but the hypothetical future risk associated with not allowing trainees to struggle, to fail and to learn.

As this suggests, non-linear, it-depends decisions about allowing failure also have implications for clinical supervision. Supervisors may struggle to make sense of their own decisions or, more problematically, may oversimplify them in posthoc reflections. As Gilchrist et al. have recently shown in their analysis of 10 supervisory dyads, trainee behavior is not a straightforward 'trigger' for a faculty supervisory response (Gilchrist et al. 2021). Rather, 'what appears to be a linear path towards an entrustment decision, may actually represent a complex interaction of factors' (Gilchrist et al. 2021, 6). Similarly, Hauer et al.'s study of how supervisors develop trust concludes that the process is a 'dynamically evolving' (Hauer et al. 2015, 793) and 'sometimes nebulous' one (Hauer et al. 2015, 792), and that supervisory entrustments 'involve a synthetic, holistic judgement that perhaps cannot be fragmented into milestones' (Hauer et al. 2015, 792). What are clinical supervisors to do with the knowledge that their decisions are complex and perhaps even ill-suited to the supervisory tools they're asked to use, like milestones and entrustment scales? First, we contend that supervisors need a language for talking about these nuances. That language likely can't be the linear language of the entrustment and assessment literature (Dijksterhuis et al. 2009; Kogan et al. 2009; Pangaro and ten Cate 2013; Ten Cate and Billett 2014; Cate et al. 2016); we worry that it may stifle rather than support such conversations. We encourage supervisors and postgraduate programs to use the four factors as a preliminary vocabulary for describing, justifying and debating their supervisory decisions about allowing failure for learning. The development of a language for these dynamic factors can support clinical educators to have meaningful discussions about this high-stakes educational strategy. We aim to provide a language and a framework for faculty development efforts to prepare clinical faculty for these high-stakes supervisory decisions. We do not intend this framework to translate immediately to scales and checklists, however, as this would minimize the richness and variability of the phenomenon.

We suggest that supervisors engage in a regular practice of reflecting on how different factors shape supervisory decisions on a regular basis, and focusing explicitly on recognizing variability and non-linearity by asking themselves, 'Why does today's decision differ from yesterday's'? We anticipate that such individual reflections and community discussions may also make more evident the role of intuition in these decisions and offer opportunities to reflect on its positive and problematic aspects by framing questions such as, 'When does intuition work, and when does it let us down?'

Last, there are implications for learners. If we view allowing failure as social interaction, intimately connected to the relationship between supervisor and trainee with a potential for a shared responsibility for such failures, then the nonlinear, it-depends nature of this phenomenon may be uncomfortable or confusing for trainees. If trainees are trying to gradually become more autonomous, they may experience a sense of randomness: sometimes I'm given autonomy, sometimes not. Sometimes that autonomy leads to failure, sometimes not (Mieczkowski et al. 2014; Biondi et al. 2015; Chen et al. 2017; Williams et al. 2017; Atkinson and Smink 2020). And while our participants agreed that the goal of allowing failure was to benefit trainee learning, the literature does not provide much evidence yet of this outcome. What do trainees learn from being allowed to fail? Beyond the educational impact, we also need to know the impact on trainee resilience and wellbeing of experiencing what may constitute sentinel emotional events (Lases et al. 2018; Bynum et al. 2019). We are currently engaged in research with trainees to explore their experiences of the educational strategy of allowing failure, including its impact on their learning, their well-being and their supervisory relationships.

Limitations

Choosing interviews might be seen as one of the limitations in revealing how supervisory decisions are made, especially if they are made navigated by intuition, limiting the insights individuals have into their cognitive processes afterwards (Zander et al. 2016). Based on our sampling strategy, we chose to recruit participants from different specialties from two sides, which might be problematic when defining failure and allowing it in a specific contextual background. What may be judged as a failure in one specialty (surgical vs. non-surgical) may not be judged as such in another. However, surgical supervisors are slightly

over-represented. On the one hand, this may be because the author and interviewer is a surgeon herself (Dwyer and Buckle 2009). On the other hand, instances of failure are more recognizable in the context of invasive procedures, and the surgical decision-making process may be easier to describe. Furthermore, what for one supervisor is a failure, another calls a mistake, and yet another defines it as an everyday learning process within and outside the zone of proximal development. This shows how relevant the use of language in the exploration of such phenomena is. Interviewing non-native speakers of English by a non-native speaker may also represent a shortcoming of the study regarding finding a common language to describe such phenomena (Lingard 2013; Helmich et al. 2017).

Conclusions

Clinical supervision requires balancing the demands of patient safety and trainee learning. A key part of this balance involves decisions about when and whether it is safe to allow a trainee to struggle and perhaps fail during a clinical situation. These decisions may be largely intuitive, and they appear to be dynamic and nonlinear. While some situations clearly dictate that the supervisor intervene to protect patient safety over training learning, in many other situations the decision is variable and dependent on the interplay among different factors. We suggest that these supervisory decisions should be the focus of faculty development in clinical training programs so that we can develop more explicit, shared thresholds for safely allowing trainee failure in the service of learning.

Acknowledgements

The authors thank the supervisors for their participation in this study.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Glossary

Allowing failure: 'While supervising a trainee's clinical performance, the supervisor, influenced by both intuition and a nonlinear interplay of different factors, detects an imminent trainee mistake, has the opportunity to intervene but chooses not to do so, because the educational gain for the trainee is perceived to outweigh the (potential) consequences for the patient.'

Funding

This study was supported by Department of Medicine POEM program at the Schulich School of Medicine & Dentistry, Western University.

Notes on contributors

Jennifer M. Klasen, MD, MME is a Ph.D. candidate of the School of Health Professions Education of the Maastricht University. She works as a surgical specialist in Switzerland.

Pim W. Teunissen, MD, PhD is director of the School of Health Professions Education at Maastricht University. He combines this role with clinical work as a gynecologist specialized in high-care obstetrics. He is also a professor of workplace learning in healthcare at Maastricht University.

Erik Driessen, PhD, is a Professor at the School of Health Professions Education at Maastricht University.

Lorelei A. Lingard, PhD, is a Professor in the Department of Medicine, and Senior Scientist in the Center for Education Research & Innovation, both at the Schulich School of Medicine & Dentistry at Western University.

ORCID

Jennifer M. Klasen http://orcid.org/0000-0001-8355-1606
Pim W. Teunissen http://orcid.org/0000-0002-0930-0048
Erik W. Driessen http://orcid.org/0000-0001-8115-261X
Lorelei A. Lingard http://orcid.org/0000-0002-4150-3355

References

Atkinson RB, Smink DS. 2020. Allowing failure so trainees can thrive: the importance of guided autonomy in medical education. BMJ Qual Saf. 29(9):704–705.

Biondi EA, Varade WS, Garfunkel LC, Lynn JF, Craig MS, Cellini MM, Shone LP, Harris JP, Baldwin CD. 2015. Discordance between resident and faculty perceptions of resident autonomy: can self-determination theory help interpret differences and guide strategies for bridging the divide? Acad Med. 90(4):462–471.

Bradley CK, Fischer MA, Walsh KE. 2013. Trends in medical error education: are we failing our residents? Acad Pediatr. 13(1):59–64.

Bump GM, Calabria J, Gosman G, Eckart C, Metro DG, Jasti H, McCausland JB, Itri JN, Patel RM, Buchert A. 2015. Evaluating the clinical learning environment: resident and fellow perceptions of patient safety culture. J Grad Med Educ. 7(1):109–112.

Bynum WE, Artino AR, Uijtdehaage S, Webb AMB, Varpio L. 2019. Sentinel emotional events: the nature, triggers, and effects of shame experiences in medical residents. Acad Med. 94(1):85–93.

Cate OT, Hart D, Ankel F, Busari J, Englander R, Glasgow N, Holmboe E, lobst W, Lovell E, Snell LS, et al. 2016. Entrustment decision making in clinical training. Acad Med. 91(2):191–198.

Cate O, Ten F., Scheele 2007. Viewpoint: competency-based postgraduate training: can we bridge the gap between theory and clinical practice? Acad Med. 82(6):542–547.

Charmaz K. 2006. Constructing grounded theory. London: Sage Publications.

Chen X(P), Sullivan AM, Alseidi A, Kwakye G, Smink DS. 2017. Assessing residents' readiness for or autonomy: a qualitative descriptive study of expert surgical teachers' best practices. J Surg Educ. 74(6):e15–e21.

Coombes R. 2010. Think rationally rather than intuitively to avoid diagnostic errors, doctors are told. BMJ. 341:c6705–c6705.

Cottrell D, Kilminster S, Jolly B, Grant J. 2002. What is effective supervision and how does it happen? A critical incident study. Med Educ. 36(11):1042–1049.

Crockett C, Joshi C, Rosenbaum M, Suneja M. 2019. Learning to drive: resident physicians' perceptions of how attending physicians promote and undermine autonomy. BMC Med Educ. 19(1):293.

Dijksterhuis MGK, Voorhuis M, Teunissen PW, Schuwirth LWT, ten Cate OTJ, Braat DDM, Scheele F. 2009. Assessment of competence and progressive independence in postgraduate clinical training. Med Educ. 43(12):1156–1165.

Durning SJ, Lubarsky S, Torre D, Dory V, Holmboe E. 2015. Considering "nonlinearity" across the continuum in medical education assessment: supporting theory, practice, and future research directions. J Contin Educ Health Prof. 35(3):232–243.

Dwyer SC, Buckle JL. 2009. The space between: on being an insider-outsider in qualitative research. Int J Qual Methods. 8(1):54–63.

Edwards IR. 2004. Intuition, evidence, and safety. Lancet. 364(9431):387.

Farnan JM, Johnson JK, Meltzer DO, Harris I, Humphrey HJ, Schwartz A, Arora VM. 2010. Strategies for effective on-call supervision for internal medicine residents: the superb/safety model. J Grad Med Educ. 2(1):46–52.

Farnan JM, Petty LA, Georgitis E, Martin S, Chiu E, Prochaska M, Arora VM. 2012. A systematic review: the effect of clinical supervision on patient and residency education outcomes. Acad Med. 87(4):428–442.

Finn KM, Metlay JP, Chang Y, Nagarur A, Yang S, Landrigan CP, Iyasere C. 2018. Effect of increased inpatient attending physician supervision on medical errors, patient safety, and resident education: a randomized clinical trial. JAMA Intern Med. 178(7):952–959.

Fischer MA, Mazor KM, Baril J, Alper E, DeMarco D, Pugnaire M. 2006. Learning from mistakes. Factors that influence how students and residents learn from medical errors. J Gen Intern Med. 21 (5):419–423.

Flin R, Youngson G, Yule S. 2007. How do surgeons make intraoperative decisions? Qual Saf Health Care. 16(3):235–239.

George BC, Dunnington GL, DaRosa DA. 2018. Trainee autonomy and patient safety. Ann Surg. 267(5):820–822.

Gigerenzer G. 2007. Gut feelings: the intelligence of the unconscious. New York: Viking.

Gilchrist T, Hatala R, Gingerich A. 2021. A collective case study of supervision and competence judgments on the inpatient internal medicine ward. Perspect Med Educ. 10(3):155–162.

Greenhalgh T. 2002. Intuition and evidence–uneasy bedfellows? The Br J General Practice. 52(478):395–400.

Halpern SD, Detsky AS. 2014. Graded autonomy in medical education — managing things that go bump in the night. N Engl J Med. 370(12):1086–1089..

Harteis C, Billett S. 2013. Intuitive expertise: theories and empirical evidence. Edu Res Rev. 9:145–157.

Harteis C, Frost F. 2012. Error orientation in the context of intuitive and competent behaviour: results of an exploratory study in the domain of emergency medicine. In: Bauer Johannes, Harteis Christian, editors. Human fallibility. Vol. 6. Dordrecht: Springer Netherlands. p. 141–153.

Hashimoto DA, Bynum WE, Lillemoe KD, Sachdeva AK. 2016. See more, do more, teach more: surgical resident autonomy and the transition to independent practice. Acad Med. 91(6):757–760.

Hauer KE, Oza SK, Kogan JR, Stankiewicz CA, Stenfors-Hayes T, Cate OT, Batt J, O'Sullivan PS. 2015. How clinical supervisors develop trust in their trainees: a qualitative study. Med Educ. 49(8):783–795..

Hauer KE, Ten Cate O, Boscardin C, Irby DM, Iobst W, O'Sullivan PS. 2014. Understanding trust as an essential element of trainee supervision and learning in the workplace. Adv Health Sci Educ Theory Pract. 19(3):435–456.

Helmich E, Cristancho S, Diachun L, Lingard L. 2017. How would you call this in English?': Being reflective about translations in international, cross-cultural qualitative research. Perspect Med Educ. 6(2):127–132.

Holzhausen Y, Maaz A, Cianciolo AT, Ten Cate O, Peters H. 2017. Applying occupational and organizational psychology theory to entrustment decision-making about trainees in health care: a conceptual model. Perspect Med Educ. 6(2):119–126.

Kahneman D, Klein G. 2009. Conditions for intuitive expertise: a failure to disagree. Am Psychol. 64(6):515–526.

Kapur M. 2015. Learning from productive failure. Learning: Res Pract. 1(1):51–65. .

Kapur M. 2016. Examining productive failure, productive success, unproductive failure, and unproductive success in learning. Edu Psychol 51(2):289–99.

Kennedy TJT, Regehr G, Baker GR, Lingard LA. 2005. "Progressive independence in clinical training: a tradition worth defending? Acad Med. 80(10):106–111.

Kennedy TJT, Regehr G, Baker GR, Lingard LA. 2009. 'It's a Cultural Expectation…' The Pressure on Medical Trainees to Work Independently in Clinical Practice. Med Edu 43(7):645–53.

Kilminster S, Cottrell D, Grant J, Jolly B. 2007. AMEE guide no. 27: effective educational and clinical supervision. Med Teach. 29(1): 2–19.

Kilminster SM, Jolly BC. 2000. Effective supervision in clinical practice settings: a literature review. Med Educ. 34(10):827–840.

Klasen JM, Driessen E, Teunissen PW, Lingard LA. 2019. Whatever you cut, i can fix it': clinical supervisors' interview accounts of allowing trainee failure while guarding patient safety. BMJ Qual Saf. DOI:10.1136/bmjqs-2019-009808.

Klasen JM, Lingard LA. 2019. Allowing failure for educational purposes in postgraduate clinical training: a narrative review. Med Teach. 41(11):1263–1269.

Klasen JM, Lingard LA. 2021. The butterfly effect in clinical supervision. Perspect Med Educ. 10(3):145–147.

Kogan JR, Holmboe ES, Hauer KE. 2009. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. JAMA. 302(12):1316–1326.

Lases SS, Slootweg IA, Pierik EGJM, Heineman E, Lombarts MJMH. 2018. Efforts, rewards and professional autonomy determine residents' experienced well-being. Adv Health Sci Educ Theory Pract. 23(5):977–993.

Li S-A, Acai A, Sherbino J, Chan TM. 2021. The teacher, the assessor, and the patient protector: a conceptual model describing how context interfaces with the supervisory roles of academic emergency physicians. AEM Educ Train. 5(1):52–62.

Lingard L. 2013. Language matters: towards an understanding of silence and humour in medical education: language matters. Med Educ. 47(1):40–48.

Martin P, Copley J, Tyack Z. 2014. Twelve tips for effective clinical supervision based on a narrative literature review and expert opinion. Med Teach. 36(3):201–207.

Martin P, Kumar S, Lizarondo L. 2017. When I say … clinical supervision. Med Educ. 51(9):890–891..

Mazor KM, Melissa a, Fischer H-L, Haley D, Hatem HJ, Rogers ME., Quirk 2005. Factors influencing preceptors' responses to medical errors: a factorial survey. Acad Med.. 80(10 Suppl):S88–S92.

Mieczkowski AE, Rubio D, Van Deusen R. 2014. Perceptions of internal medicine–pediatrics residents about autonomy during residency. J Grad Med Educ. 6(2):330–334.

Morse JM. 2000. Determining sample size. Qual Health Res. 10(1):3–5.

Mylopoulos M, Steenhof N, Kaushal A, Woods NN. 2018. Twelve tips for designing curricula that support the development of adaptive expertise. Med Teach. 40(8):850–854.

Norman GR, Monteiro SD, Sherbino J, Ilgen JS, Schmidt HG, Mamede S. 2017. The causes of errors in clinical reasoning: cognitive biases, knowledge deficits, and dual process thinking. Acad Med. 92(1): 23–30.

Norman G, Sherbino J, Dore K, Wood T, Young M, Gaissmaier W, Kreuger S, Monteiro S. 2014. The etiology of diagnostic errors: a controlled trial of system 1 versus system 2 reasoning. Acad Med. 89(2):277–284.

Pangaro L, Ten Cate O. 2013. Frameworks for learner assessment in medicine: AMEE guide no. 78. Med Teach. 35(6):e1197–1210.

Patton MQ. 1990. Qualitative evaluation and research methods. 2nd ed. Newbury Park, Calif: Sage Publications.

Quillin RC, Pritts TA, Hanseman DJ, Edwards MJ, Davis BR. 2013. How residents learn predicts success in surgical residency. J Surg Edu. 70(6):725–730.

Ross PT, McMyler ET, Anderson SG, Saran KA, Urteaga-Fuentes A, Boothman RC, Lypson ML. 2011. Trainees' perceptions of patient safety practices: recounting failures of supervision. Jt Comm J Qual Patient Saf. 37(2):88–95.

Sherbino J, Dore KL, Wood TJ, Young ME, Gaissmaier W, Kreuger S, Norman GR. 2012. The relationship between response time and diagnostic accuracy. Acad Med. 87(6):785–791.

Steenhof N, Woods NN, Mylopoulos M. 2020. Exploring why we learn from productive failure: insights from the cognitive and learning sciences. Adv Health Sci Educ Theory Pract. 25(5):1099–1106.

Steenhof N, Woods NN, Van Gerven PWM, Mylopoulos M. 2019. Productive failure as an instructional approach to promote future learning. Adv Health Sci Educ Theory Pract. 24(4):739–749.

Sterkenburg A, Barach P, Kalkman C, Gielen M, ten Cate O. 2010. When do supervising physicians decide to entrust residents with unsupervised tasks? Acad Med. 85(9):1408–1417.

Swanwick Tim, Forrest Kirsty, O'Brien Bridget C, editors. 2018. Understanding medical education: evidence, theory and practice. 3rd ed. Hoboken, NJ: Wiley-Blackwell.

Tavakol M, Torabi S, Akbar Zeinaloo A. 2006. Grounded theory in medical education research. Med Educ Online. 11(1):4607.

Ten Cate O, Billett S. 2014. Competency-based medical education: origins, perspectives and potentialities. Med Educ. 48(3):325–332.
Varpio L, Nagler A. 2018. The bumps and bruises from turning a blind eye: learning from our failures and surprises. Perspect Med Educ. 7(Suppl 1):1–3.
Walton M, Barraclough B. 2013. Clinical supervisors: are they the key to making care safer? BMJ Qual Saf. 22(8):609–612.
Watling CJ, Lingard L. 2012. Grounded theory in medical education research: AMEE guide no. 70. Med Teach. 34(10):850–861.
Williams RG, George BC, Meyerson SL, Bohnen JD, Dunnington GL, Schuller MC, Torbeck L, Mullen JT, Auyang E, Chipman JG, et al. 2017. What factors influence attending surgeon decisions about resident autonomy in the operating room? Surgery. 162(6):1314–1319.
Wong BM, Etchells EE, Kuper A, Levinson W, Shojania KG. 2010. Teaching quality improvement and patient safety to trainees: a systematic review. Acad Med. 85(9):1425–1439.
Wong SSH, Lim SWH. 2019. Prevention–permission–promotion: a review of approaches to errors in learning. Edu Psychol. 54(1): 1–19.
Wu AW, Folkman S, McPhee SJ, Lo B. 1991. Do house officers learn from their mistakes? JAMA. 265(16):2089–2094.
Young M. 2019. The utility of failure: a taxonomy for research and scholarship. Perspect Med Educ. 8(6):365–371.
Zander T, Öllinger M, Volz KG. 2016. Intuition and insight: two processes that build on each other or fundamentally differ? Front Psychol. 7:1395.