




RESEARCH ARTICLE

The shift from disbelieving underperformance to recognising failure: A tipping point model

Andrea Gingerich¹  | Stefanie S. Sebok-Syer²  | Lorelei Lingard³ | Christopher J. Watling³ 

¹Division of Medical Sciences, University of Northern British Columbia, Prince George, British Columbia, Canada

²Department of Emergency Medicine, Stanford University, Stanford, California, USA

³Schulich School of Medicine & Dentistry, Centre for Education Research & Innovation, Western University, London, Ontario, Canada

Correspondence

Andrea Gingerich, Division of Medical Sciences, University of Northern British Columbia, Prince George, BC, Canada.
Email: andrea.gingerich@unbc.ca

Funding information

Medical Council of Canada, Grant/Award Number: Research in Clinical Assessment grant

Abstract

Context: Coming face to face with a trainee who needs to be failed is a stern test for many supervisors. In response, supervisors have been encouraged to report evidence of failure through numerous assessment redesigns. And yet, there are lingering signs that some remain reluctant to engage in assessment processes that could alter a trainee's progression in the programme. Failure is highly consequential for all involved and, although rare, requires explicit study. Recent work identified a phase of disbelief that preceded identification of underperformance. What remains unknown is how supervisors come to recognise that a trainee needs to be failed.

Methods: Following constructivist grounded theory methodology, 42 physicians and surgeons in British Columbia, Canada shared their experiences supervising trainees who profoundly underperformed, required extensive remediation or were dismissed from the programme. We identified recurring themes using an iterative, constant comparative process.

Results: The shift from disbelieving underperformance to recognising failure involves three patterns: accumulation of significant incidents, discovery of an egregious error after negligible deficits or illumination of an overlooked deficit when pointed out by someone else. Recognising failure was accompanied by anger, certainty and a sense of duty to prevent harm.

Conclusion: Coming to the point of recognising that a trainee needs to fail is akin to the psychological process of a tipping point where people first realise that noise is signal and cross a threshold where the pattern is no longer an anomaly. The co-occurrence of anger raises the possibility for emotions to be a driver of, and not only a barrier to, recognising failure. This warrants caution because tipping points, and anger, can impede detection of improvement. Our findings point towards possibilities for supporting earlier identification of underperformance and overcoming reluctance to report failure along with countermeasures to compensate for difficulties in detecting improvement once failure has been verified.

1 | INTRODUCTION

Coming face to face with a trainee who needs to be failed is a stern test for many supervisors. How supervisors respond to the test is highly consequential, and potentially catastrophic, for the trainee. It can also be consequential for the supervisor, many of whom report workload and emotional repercussions that interfere with assessment demands.^{1–6} When negative repercussions are prioritised, not reporting a trainee's failing performance becomes framed as *reluctance*^{1,6–19} (or even unwillingness)^{20,21} to engage in the assessment process. In response, countermeasures to reluctance have been incorporated into assessment designs with changes that range from revised forms through to reconfigured systems. Countermeasures on assessment forms include having supervisors record the close supervision they provided for a trainee during a task to avoid the more daunting task of recording negative evaluative judgements.²² For example, supervisors may be less reluctant to select the 'talk them through' rating on an O-SCORE (Ottawa Surgical Competency Operating Room Evaluation) scale²³ than to assign an 'unsatisfactory' rating on a Mini-CEX (Mini Clinical Evaluation eExercise) scale.²⁴ Countermeasures in assessment systems include the programmatic assessment approach of separating assessment moments from decision moments to relieve the pressure of simultaneously functioning as the assessment data provider and the progression decision maker.²⁵ For example, supervisors may be less reluctant to note domains requiring further improvement on a low stakes assessment that is combined with many other assessments and used by other stakeholders to make progress decisions,²⁶ than to stop a trainee's progression by submitting a single 'unsatisfactory' in-training evaluation report.²⁷ However, these countermeasures still rely on individual supervisors to provide critical comments as assessment data and committees composed of similar individuals to make high stakes decisions to mandate remediation, alter progression, or remove a trainee from a programme. Although early evidence from the implementation of various countermeasures is promising,^{28–30} findings of bureaucratic tick-boxing and token comments suggest that important assessment information may not always be provided.^{28,30–32}

Previous work exploring supervisors' reluctance to fail trainees helpfully invites solutions for gleaning assessment information withheld by supervisors.^{1,2,16,19,20,33,34} Less attention has been directed to earlier phases of process—those critical moments where failure is noticed, characterised and confirmed. In recent work, we sought to better understand how supervisors recognise trainees who are underperforming and we identified a phase of disbelief as supervisors encountered unexpected signs of underperformance.³⁵ This disbelief was related to surprise or bewilderment at encountering an underperforming trainee and was strikingly similar to the unsettling and uncomfortable gut feelings experienced by nursing instructors while interacting with underperforming nursing students.³⁶ Although supervisors tended to give the trainee the benefit of the doubt, as is common,^{6,8,12,16,37,38} they were also inclined to gather additional observations to make sense of what they were seeing.^{35,36} The shift from disbelieving to recognising *underperformance* hinged on

perceiving how responsive the trainees (and the deficits) were to teaching and coincided with a realisation that continued, intensified supervisory efforts would likely be futile.³⁵ However, the shift from disbelieving underperformance to recognising *failure* has been underexplored.

Although this shift is a specific aspect within the larger process that eventually results in decisions about reporting evidence of failure, it is a critical aspect to understand in order to improve engagement in assessment processes. While failure is admittedly rare, it is so consequential when it does occur—and perhaps even more so when it should have occurred but does not—that it demands careful study. Therefore, in this study we ask the question: How do supervisors recognise failure? In particular, we invited supervisors to describe their experiences working with the least competent trainees to examine the shift from disbelieving early signs of underperformance to recognising failure.

2 | METHODS

2.1 | Overview

This study is situated within a programme of research investigating how supervisors assess trainees who demonstrate incompetence. Our earlier study was not focused specifically on recognising failure; however, we identified some themes related to recognising profound underperformance during prolonged remediation and dismissals from programmes. We subsequently investigated these underexplored themes with additional recruitment of physician supervisors using purposive and theoretical sampling. As such, this study combines data collected previously to explore how supervisors recognised underperformance with new data collected to better explore how they recognised that an underperforming trainee should not be allowed to progress and needed to be failed. Our focus on the shift from disbelieving signs of underperformance to recognising failure involves studying how supervisors think while interacting with others. We used constructivist grounded theory³⁹ methodology (CGT) to guide data collection and analysis to focus on the experiences of individual supervisors as they interact with others according to the rules and expectations of social contexts. This project received approval (E2018.0613.065) through the University of Northern British Columbia's research ethics review board.

2.2 | Participants and recruitment

In 2018–2019, we interviewed 22 physicians from non-surgical specialties in British Columbia, Canada, and characterised underperformance as stalled progression due to the *inability* to engage in learning or precarious progression due to *unwillingness* to engage in learning.³⁵ As noted previously, we decided to deepen our understanding of how supervisors experience working with markedly underperforming trainees and how they come to recognise when a

trainee's performance merits failure, by engaging in additional data collection. Since the majority of the original participants were female and all were from non-surgical specialties, we concentrated new recruitment efforts on male participants and those from surgical and procedurally dominant specialties. We used key informants and snowball sampling to identify and invite physicians with experience failing or remediating trainees in the clinical setting, in either the 4-year medical school programme or in one of the postgraduate residency programmes. We sought participants with varying levels of experience with supervision from diverse specialties, clinical settings and numerous University of British Columbia (UBC) Faculty of Medicine sites distributed across the province to maximise variation in perspectives. Between May 2019 and January 2020, we recruited and interviewed an additional 20 physicians (7 male) of whom 13 were from surgical or procedure-dominant specialties. The final participant sample thus consisted of 42 participants (11 male) representing 17 specialties/subspecialties. A total of 28 participants were from medical specialties (e.g. family medicine, internal medicine, psychiatry, medical oncology), and 14 participants were from surgical or procedure-dominant specialties (e.g. orthopaedics, obstetrics and gynaecology and emergency medicine). Participants practised in rural, northern and/or remote communities (20); or in large metropolitan cities (22) in the province of British Columbia, Canada. All shared experiences from a clinical supervisor role; 17 participants also shared experiences from a leadership role that involved contributing to high stakes progression decisions. The majority named their leadership role 'programme director', and to protect their identity, we use this term to refer to all of the leadership roles.

2.3 | Data collection

Following CGT methodology, we took an iterative approach to the interview process. We began by inviting experiences with supervising underperforming trainees in the clinical workplace in the first 22 interviews. Due to negative reactions to our use of the terms 'incompetent' and 'incompetence',³⁵ we tried various terms and prompts to elicit conversation on underperforming trainees. As we had sufficient data describing struggling and underperforming trainees, the latter 20 interviews focused on the more extreme cases of underperformance that had the most potential to be indicative of situations where failure should be considered. Interview questions invited participants to share their experiences supervising medical students and residents who profoundly underperformed, required extensive remediation, invoked disciplinary actions or were asked (or should have been asked) to leave the programme. Analysis informed data collection through iterative cycles. As depicted in Appendix 1, the four iterations of the semi-structured interview guide evolved from a focus on eliciting rich descriptions of profoundly underperforming trainees, to emphasising conversation about the substantial impacts on the supervisor, to encouraging supervisors to put their experiences with failing trainees into words and through to inviting descriptions of failing, or not failing, underperforming trainees.

Interviews averaged 51 min in duration (ranging from 37–68 min) were conducted by telephone or in-person, audio-recorded, transcribed by a professional transcriptionist and de-identified prior to analysis.

2.4 | Data analysis

The preliminary codes identified in the first 22 interviews that pertained to failing trainees were used to inform data collection for the subsequent interviews. Analysis of the subsequent 20 interviews continued iteratively with each cycle of 3–6 interviews analysed to identify focused codes that labelled the major themes and concepts. The full research team met after each cycle to discuss the interview data, identify themes and modify the evolving codes that labelled supervisors' recognition of failure and responses to recognised failure. As per CGT, we used the codes to define what was happening in the data by labelling our understanding of the participants' perspectives that was developed by attending to the content that they shared, the language they used and how they shared that information.⁴⁰ Although we did not complete a detailed linguistic analysis, we did code for salient linguistic features⁴¹ in the transcripts and listened to the recordings to code for volume, hesitation and emotion in the spoken words. We used constant comparison across the entire set of 42 transcripts to revisit, expand and refine the codes. The coding framework sufficiently summarised the data after 38 interviews with the last 4 interviews reinforcing, elaborating and clarifying the identified concepts. We then discussed the links and patterns represented by the focused codes to identify theoretical codes. The theoretical codes represent an interpretation that extends beyond, and yet is still grounded in, the data.⁴⁰

In CGT, meaning is constructed through the differing perspectives of the researchers and through their interaction with the participants and the data.⁴⁰ As such, our team engaged with analysis drawing on their expertise with studying supervision and clinical learning contexts (AG, CW, LL), assessment and raters (AG, SS-S), social cognition (AG), linguistics (LL) and lived experience supervising trainees and overseeing supervisors as postgraduate dean (CW). Discussions included reflexivity on how each researcher's perspective contributed to interpretation of the data. For example, one member of team has experienced medical training first as a learner, then as a supervisor and then as a leader. While he did not experience failure or remediation as a learner, he did as a supervisor and even more so as a leader with responsibility for oversight of training programmes. He shared his experiences with the research team, and we reflected regularly on the tension between the educational responsibility to support learners and the social responsibility to ensure that learners only progress if it is safe for them to do so. The non-clinicians on the team reflected on their expectations as patients and as family members of patients, as well as on their experiences as educators, to further contextualise this tension. Reflexivity also included scrutiny of the enacted research methods. For example, we reviewed transcripts to verify that the spoken interview questions did not lead participants to use particular

language or elicit specific reactions. We used notes to record decisions regarding data collection and to serve as an audit trail. We used memos to record our evolving understanding of the data during the analytic process. We used NVivo software version 11 (QSR International Pty Ltd, Chadstone, Victoria, Australia) for data management and coding.

Participants are referred to by a letter to indicate their specialty as 'M' for medical or 'S' for surgical or procedure-dominant followed by a number to indicate their interview sequence.

3 | RESULTS

According to supervisors' experiences, recognising failure involves overcoming the 'benefit of the doubt' by detecting patterns of failing performance in the workplace that are 'beyond a doubt' indicative of failure. We present the findings for three interrelated aspects of our inquiry: First, we present *what* is recognised as failure by describing the *patterns* of trainee performance indicative of failure. Then, we present how supervisors *express emotion* while recounting their experiences of working with underperforming trainees and recognising failure. Finally, we present how recognition of failure extends beyond the supervisor and supervisor-trainee relationship to involve verifying and *building consensus* with colleagues.

3.1 | Identifying patterns

Supervisors across medical and surgical specialties described a phase of disbelief when first encountering underperforming trainees: 'it's just baffling to me' when upper year residents cannot do 'some bare minimum things' that are 'basic' skills (S35). Programme directors shared that supervisors called them "on day two of the rotation and they will say 'I just can't believe this'" (M26) when they discovered, for example, that a trainee who was considered to be intelligent and dedicated to the residency programme was underperforming: 'I sat down and listened to her explanation to the patients. I was surprised with the inaccuracy of information. I was surprised with the way the information was delivered' (S30). They gave the trainee the benefit of doubt during this phase and resisted documenting observations as they were trying to make sense of what they were seeing:

'There were some times early on where I kind of put it up to, as I say, him going through that transition phase and kind of settling in, if you will, because I just felt like it didn't quite meet the bar that I needed to trigger a formal assessment. We always talked about things afterwards but not to necessarily write it down on paper with those interactions.' (S42).

The shift from disbelieving signs of underperformance to recognising failure necessarily includes the need for supervisors to identify underperformance that is representative of failure. They characterised failure by quantifying its frequency, consistency and magnitude across incidents: '*repeatedly* not doing something the way that I tried to show them *numerous* times how to do it' (S35).

Programme directors similarly characterised failure by quantifying it through reports from multiple supervisors, often in different contexts, describing different incidents, over time, despite receiving feedback: 'repeated rotations having difficulty with him' (S28) and 'I compiled all the verbal complaints that people had [reported to me as program director] and wrote him this 10 page letter' (M29). Therefore, recognising failure hinged on finding 'a pattern, it didn't just happen once' (S25) along with a determination that the trainee is 'for some reason unteachable' (S30), 'unsupervisable' (M36) and 'completely unfixable' (M6) because 'they just couldn't do it' (S34) and 'just couldn't learn' (M10) 'so they're not going to learn it' (M13).

We identified three patterns within the descriptions of how failure was recognised: overwhelming evidence, egregious error and outsider's insight. The overwhelming evidence pattern features 'persistent difficulty' and 'repeated demonstration of inadequate' skills (S24) that the trainee 'should've learned by then as a bread and butter item' (M10). The pattern could be restricted to particular domains or be global: 'it's the whole picture, it's not being prepared, it's not being engaged, it's not having any technical skills' (S35). This was a pattern of overwhelming evidence compelling action from the supervisor and the programme: 'there was a multifaceted failure on his part to really do the job to the point where I and everyone was very concerned about him and his abilities [... he was] the one we failed.' (S34).

In the egregious error pattern, recognising failure was triggered by the discovery of a disconcerting error or omission. A lone incident of great magnitude that involved a core skill could be sufficient in indicating failure, even for a trainee who had not been previously classified as underperforming: 'surgically, they were actually not a bad operator, and they were actually a very nice person, but missing something so obvious and egregious to me was a huge red flag and came out of nowhere' (S37). Similarly, a single egregious blunder that followed several smaller and excusable mistakes signalled a failure, because it too revealed a substantial gap in their knowledge, skills or understanding that risked patient safety. Specific examples cannot be included here because the 'career damaging' details (S37) could identify the trainees, but this supervisor went on to explain that 'essentially it's the same as building an airplane without wings' without realising that it does not have wings when your speciality is all about rebuilding airplanes. The necessity to document failure was increased when the trainee seemed not to recognise the significance of their actions: "'remember when you got into [iatrogenic] bleeding when you were trying to do the operations' and he would be like 'oh ya, but that's normal'" (S39). Similarly, supervisors described being more inclined to engage in the process of failing a trainee when they did not show appropriate regret or remorse for the mistake: "it was this flippant nature to getting the feedback [on a never-event] so it wasn't like 'oh my goodness I can't believe I did that, I'm so sorry'" (S25). However, lack of recognition and lack of remorse were both described as rare and unexpected.

The third pattern involved recognition sparked by having an outsider, like a colleague from another department or a trainee,

divulge a troublesome experience to the supervisor that illuminated a deficit that they had been 'a little bit blind to' (S42). The outsider's insight drew attention to a particular difficulty the resident was grappling with: '[the omission] was so concerning to the attendings—it wasn't me [an obstetrics attending] specifically—it was a family doctor that said 'you know what—this guy is unsafe' because of his behaviour' (S38). The off-the-cuff comment from someone with first-hand experience seemed to cause the previously overlooked pattern to click into place and to confirm that action needed to be taken: 'it really helped show us a little bit of the spectrum and recognise that we really did have an area that he needed increased feedback and support' (S42).

3.2 | Expressing emotion

When we pressed supervisors to go back to the moment when they first realised that what they were noticing was failure, they could not pinpoint it and could not put it into words: 'I can't totally remember. One moment ... but, you know, I can't totally remember one moment, ummm ... and, you know, I mean ...' (M36). There were examples that contained hints of a shift towards thinking that something that they were noticing now may become problematic and need to be dealt with in the future:

'... with her I would worry that it's going to be something ongoing because it seemed more of a personality source of the problem than just an off-day or a sporadic one-time kind of thing. So, I don't know. She might be able to compensate and learn from these type of things as they come up and adjust in her way through. Or she might, you know, for lack of a better term basically piss people off all the way through and eventually gets enough negative feedback and comments that something—some bigger measure—is taken by the program.'(M13)

The expletive in the above quote was not an isolated usage. Mild swearing accentuated the descriptions of profoundly underperforming trainees: 'an example of somebody who has screwed up because of this' (M15) as in 'really screwing up left and right' (S30). It was also used in expressing their reaction to supervising a failing trainee including: 'wondering why the hell am I doing this anymore' (M27) because of 'how much that pissed me off' (M41). Some abruptly stopped midsentence to apologise for their tone and choice of words: '... he had this very loquacious diarrhoea that would just—sorry, I'm sounding really awful, aren't I—but, he would just like not shut up ...' (M29). In addition, participants raised their voices while describing their experiences with failing trainees. For example, while sharing how they determined that a trainee was being dishonest: 'most of the time they just keep lying—and this is the hard thing—and I find this so hard myself because I'm like I'M RIGHT HERE, DO YOU THINK I DON'T KNOW?' (S39). They raised their voices when describing deception and manipulation: 'I can think of 2, maybe 3 trainees, where they're—I don't know, pathological, like they're just poison—the personality is just like 'OH MY GOD, HOW DID YOU GET INTO MEDICINE' kind of thing' (M14). They also used vivid imagery including allusions of

being 'persecuted' (M41) to convey the impact that profoundly underperforming trainees could have on them:

'I'll tell you what I'm sentenced to—a six month period of time where in my workplace I'm watching someone, cringing the whole time, feeling as though someone is bullying my patients and if I talk to them they get angry with me—this is my worst fear [...] and I would've probably opened a vein if I had to teach him for six months.' (M36)

The emotion that was most commonly named while sharing specific experiences of recognising failure was frustration: 'I don't begrudge the trainee for making the wrong decision—I think clinical judgement, that comes with time—I think I was frustrated that they didn't have the ability to see the error in their ways and to call for help for a sick patient' (M7). When we asked directly what it was like to supervise trainees who were failing, frustration again appeared as a common response: 'it's frustrating, it's kind of eye-opening' (M18). Our interpretation was that the combination of mild swearing and raised voices used while recalling those experiences was indicative of anger.

Another aspect of how participants shared their experiences was a distinct change in the language that they used. Supervisors tended to describe failing trainees from a current state of certainty using harsh, unequivocal language: 'It was a disaster' (M2) 'he's a disaster' (M26) 'he was dangerous' (S28) and 'entirely inappropriate' (S30). Our analysis identified that their choice of phrasing frequently included the use of boosters. According to Hyland, boosters are a communicative strategy used to express conviction, confidence and commitment to a statement and may be used to persuasively solicit support for particular assumptions or conclusions.⁴¹ We have underlined examples of the boosters used within the quotes presented in this paragraph. Supervisors grappled with 'fearing that I'm going to get sued because of this [omission by the resident ...] was a really big miss and I was really upset by it' (M20). In rare cases, the actions of a trainee 'prompted a huge lawsuit' (M29) that involved the supervisor in legal proceedings. However, they emphasised that determining a trainee was 'not meeting a standard of patient safety—it's actually causing harm' (S25) necessitated documentation of failure: 'I can't pass someone if they're downtight dangerous to the patient' (M14). It then became their obligation to report failure to prevent harm to patients and the profession: 'that is my duty. That is, I have to. That's—otherwise I am letting down the student and I am letting down the public' (M16) because 'it's obviously our responsibility in terms of the safety of our patients' (M41). After making the shift, the only appropriate option was 'this one had to be flagged' (M31), even though it was a rare and difficult option: 'in my career I think I've failed one person and they really needed to be failed' (M36). Such conviction may have been needed because reporting failure, and even underperformance, could come at a cost, such as having the trainee file 'a complaint' (M17) against them. In contrast to the benefit of the doubt that filled their descriptions of first discovering that a trainee might be underperforming, the descriptions of having recognised that a trainee should not progress in the programme were filled with certainty, firm language, intense emotion and a sense of duty to prevent harm that held the potential to mobilise action despite foreseeable costs.

4 | BUILDING CONSENSUS

Up until this point we have focused on the supervisor and presented recognition of failure as something that occurs mostly in the minds of individuals but recognising failure also included social interactions with colleagues. In particular, recognising failure involved supervisors discussing what they were seeing with others to make sense of it and to verify their interpretations. After that, having failure become *recognised*, so that a trainee's progression to the next phase was altered, required consensus across multiple people within a group of supervisors, department or committee. We now turn our focus to the collective experience of recognising failure. Programme directors described departments and committees reaching a point where they had decided 'we're not going to tolerate it—period—full stop' (S34) because when it is 'pathological enough it's going to be a serious issue then it usually is actually better served for everybody, including the trainee, to remove them than to graduate them' (S37). However, it required agreement and documentation from multiple supervisors:

'For the truly treacherous trainee that I think would provide dangerous patient care, we all have that responsibility to stand forward and we often do that united. Meaning we look for consensus among our physician group so it's not just one preceptor saying 'this one did this.' It's a 'did you have something like that in your week?' 'Yes, I had something too and so did one of my other colleagues', okay, so the three of us if we put it all down together, it's a strength in numbers, we can't really be questioned. If it's just a one-off of something, then it could be really hard to pin down.' (S40)

This collective recognition of failure was helpful in bringing attention to a trainee who needed closer observation. However, when it was the case that 'everybody was noticing' (M41), it was acknowledged that the trainee's reputation could affect how they were supervised and assessed. Underperforming trainees were discussed among supervisors in order to provide increased supervision. This was done informally when a colleague initiated 'discussing with the preceptors and making sure the other preceptors know this isn't just a one-off or that this person is potentially dangerous' (M12). It also happened formally within the programme in response to a serious incident: 'had someone not picked up that error and stopped it, something bad could've happened to the patient and so it became a big discussion within our training committee and with the attendings' (M5). Supervisors discussed times when the extra attention was deserved: 'once you have a microscope on you it's very hard to shake the microscope—I think sometimes it's justified' (M14). There were examples where it had become clear that the trainee perceived the extra attention as 'being mistreated' (S22), but the supervisors and programme interpreted it as necessary supervision and appropriate feedback. This difference of opinion commonly involved the perception that the trainee was being resistant to feedback. For example, when in a discussion with the programme director, a medical student would 'explain why actually what he had done was the right thing to do and that [preceptors] had misinterpreted or they didn't understand or that they were wrong and he was right' (S22). This behaviour

tended to be attributed to 'lack of insight [... with thinking] everybody is out to get me [...] instead of actually recognising, no, there are some actual competency issues' (S25).

The collective attention could also create situations where the trainee could not readily demonstrate improvement. Supervisors acknowledged that 'once you've made a mistake, you're under a microscope and suddenly everything you do is examined that much closer' (M19), and this extra attention caused stress for the trainee and interfered with their learning. They shared examples where they felt the extra attention was not justified, such as when there was a 'personality conflict between the preceptor and the resident and that poor resident can't do anything right—and everything's documented—but this person can't emerge from the sinkhole that they've gotten into' (M2). Residency programme directors discussed how difficult it was to manage residents with reputations for being involved with 'very scary events' (S39) within their small learning communities. It could require strategies like sending them 'out of province to get a fair assessment of their operative skills from people that didn't know their history' (S39).

Despite widespread recognition of profound underperformance warranting failure, it was reported that some of the trainees did not fail the rotation or were not dismissed from the programme. When insufficient documentation or impotent assessment processes allowed such trainees to graduate, programme directors would experience 'moral distress over [...] this resident] passing and being released to the public' (M29). But it happened because 'even when there's irrefutable evidence that a resident is a serious, serious problem, getting them out of the program entirely is very, very difficult to do' (S37). Experiences with failing trainees profoundly impacted programmes through targeted and substantial changes to policies, curricula or assessment.

5 | DISCUSSION

Failing a trainee is not easy. While the literature on the 'failure to fail' phenomenon tends to emphasise supervisors' reluctance to fulfil the expectations of their roles, our work adds considerable depth of understanding to a collective process that is complex and often arduous. We found that the shift from disbelieving underperformance to recognising failure involves a process of verifying failure by quantifying its magnitude or pervasiveness through patterns of persistent or notorious and non-rectifiable underperformance. The identified patterns include discovering that a trainee had committed an 'egregious' blunder that reveals a substantial learning gap, receiving a report from an impartial colleague that verifies a trainee's behaviour is inappropriate and can no longer be ignored and repeatedly seeing that a trainee cannot do what was expected of them and/or continues to do what was not expected of them. All patterns signalled that the trainee was at risk of providing dangerous patient care, even with supervision. At that point, supervisors and programme directors felt compelled to engage in the process to fail a trainee to uphold their duty to prevent harm to patients and the

profession.^{5,6,9,10,12,16,33} Additional insights can be gained by further elaborating the shift from not believing early signs of underperformance to recognising that the trainee should be failed.

5.1 | A tipping point model of recognising failure

A key contribution of our work is examining how patterns of failure are recognised as credible signal rather than anomalous noise. The shift from disbelieving underperformance to recognising failure seems akin to a threshold being crossed and is reminiscent of the psychological process of a *tipping point*.^{42,43} Research on tipping points focuses on how small changes from new social information compound into impression updating that shifts the corresponding person from one category into another.⁴³ Tipping points can mark the shift from attributing someone's behaviour to situational factors to attributing it to their preferences, personality, morality and other characteristics of the person.⁴⁴ It is an active process where people *diagnose* when a change has been made rather than just *note* that a change has occurred.⁴³ By conceptualising the process of recognising a failing trainee as a tipping point, we can draw on that literature to further explore the process.

Tipping points imply a shift towards becoming more likely to act in response to identifying change.^{42,43} The action inducing aspects of a tipping point resonate with our participants' recollections that recognising failure can mobilise the programme to put an underperforming trainee under a much needed microscope. However, participants were also concerned, once the microscope was focused on a trainee, that they could have difficulty shaking it, even when it was no longer deserved. The asymmetry of tipping points helps to explain this. Tipping point research shows that people more quickly diagnose a trend as declining than the same rate of change as improving and have a tendency to dismiss possible signs of improvement as lucky flukes.^{42,43} Furthermore, tipping points happen faster and with less evidence than people estimate,⁴⁵ which implies that those involved with assessing trainees may overestimate how diligent they are in determining that failure is warranted. A similar asymmetry in moral tipping points shows that moral decline (i.e. committing bad behaviours) is more willingly punished than identical moral improvement is rewarded (i.e. committing good behaviours).⁴⁴ Although efforts to address the failure to fail phenomenon have focused on alleviating reluctance to document underperformance, tipping point research highlights how the process can go awry with a disinclination to see improvement after underperformance has been verified. We may want to reframe supervisors' initial disbelief and time spent searching for patterns of behaviour as helpful in fending off premature labelling of a trainee as failing. Our findings open a new conversation on helping supervisors to appropriately fail underperforming trainees.

Research into tipping points has tended to focus on its cognitive aspects, without much attention given to the influence of emotions. Our work identifies anger as co-occurring with the shift towards recognising failure. Emotions and emotional repercussions for

supervisors have been identified as fuelling reluctance to fail trainees.^{3,7,8,11,13,16,46} We suggest that it may also play another role. Anger may be functioning as a catalyst in, or perhaps is a by-product of, recognising failure. Our methodology does not allow its precise role to be determined, but it did allow anger to be expressed and described in interviews. It struck us as remarkable, since angry reactions (including profanity) have rarely been reported in failure to fail research.⁴ But it is consistent with psychology research where demonstrations of incompetence have elicited contempt.⁴⁷ When people are angry, they have been found to be more likely to perform an evaluative task that is expected to involve a negative evaluation.⁴⁸ Anger is a moral emotion⁴⁹ that involves appraising another person's responsibility for wrongdoing with the goal to correct the resulting perceived wrong.⁵⁰ It is an emotion that people experience when their values and moral convictions have been threatened.^{50,51} It is experienced when social norms regarding hierarchical and communal obligations have been violated⁵² through the breaking of social rules that guide interactions like reciprocity, honesty, fairness and concern for others.^{50,53} Although anger tends to be seen as a negative emotion that needs to be controlled, 'justifiable anger can help mobilise people to defend their interests and values'.⁵⁴ It may be that the action tendencies of anger help to prompt supervisors to overcome disbelief in order to recognise failure.

Anger, however, may sometimes be a dangerous catalyst for action. Anger can reduce perspective-taking, such as recognising differences and making inferences about someone else's point of view.⁵⁵ It can impede consideration of new information that might prompt a reconsideration of currently held views.⁵⁴ It can fuel a biased search for information to attack the violator⁵⁶ along with confirming evidence to bolsters one's views while downplaying opposing evidence and arguments.⁵⁴ Anger is an *other condemning* emotion⁴⁹ associated with feelings of being antagonistic and punitive.^{57,58} It lessens chances for reconciliation.⁴⁷ Importantly, it can be antecedent to harassment, bullying or mobbing in the clinical workplace.^{59,60} The mobilising aspects of anger may help supervisors to overcome reluctance in failing a trainee, but there is the possibility for those aspects to distort subsequent information seeking and decision making. Our assessment systems may be able to benefit from strategically capitalising on supervisors' emotional responses to underperformance, but they must ensure that countermeasures are incorporated to offset any associated distortions.

In summary, we outline a tipping point model for recognising failure. Upon being confronted by signs of underperformance, supervisors may experience disbelief and give the trainee the 'benefit of the doubt' as they seek additional evidence through increased supervision to make sense of what they are observing. When the trainee's progression remains stalled and unresponsive to intensified supervision, underperformance is identified. If an egregious error is discovered, or an outsider's insight substantiates overlooked signs, or a pattern of overwhelming evidence of continued underperformance follows, failure becomes recognisable. Together with anger and a sense of duty to prevent the trainee from causing harm, the threshold of the tipping point is crossed with recognition 'beyond a doubt' that

the trainee should fail. Having tipped into the certainty of recognition, reporting of failure through engagement in assessment processes may be mobilised but discerning signs of restored progression and improvement may be obscured.

5.2 | Implications for assessment

Although our study did not investigate potential impacts on assessment designs, we will use the tipping point model and our improved understanding of how supervisors recognise failure to pose questions for further reflection. A tipping point model offers a pre- and post-tipping point phase. In the pre-tipping point phase, there may be disbelief and uncertainty in what is being observed and therefore reluctance to document those incomplete understandings. Are our workplace-based assessments based on a sufficient period of time working with trainees to allow supervisors to move through any disbelief and report substantiated information? If this is not possible, do our assessments support supervisors in documenting incompletely understood observations and experiences in ways that the system can effectively use as assessment information? Assessment processes could encourage supervisors to record their doubts along with their observations and feedback if programme directors and competence committees could appropriately situate the assessments within the longitudinal and multifaceted array of assessment data that they have access to but the supervisor does not. Competency committees could then focus on the early identification of patterns of underperformance that individual supervisors with limited interaction time cannot.

In the post-tipping point phase, there may be anger, certainty and a sense of duty that mobilises recognition of failure. If supervisors were encouraged to use feelings of frustration as a signal to document low stakes assessments of trainee performance, could documentation of constructive assessment comments be increased? On the other hand, do our systems have countermeasures in place to combat the possibility that information provided after failure has been recognised may be overly focussed on decline and to ensure that evidence of improvement is accessible to decision-makers? Supervisors could be informed that it is more difficult to see signs of improvement once they have recognised a trainee is failing and be encouraged to note aspects that the trainee is doing well (or better) along with the aspects that they continue to struggle with. Similarly, programme directors and competence committees could prioritise specifying expected patterns of improvement in advance⁶¹ with their noted absence indicative of failure.

5.3 | Limitations and future research

Since our methodology relied on participants recalling past supervisory experiences, we cannot verify that the trainees they described were underperforming nor determine the appropriateness of wanting to stop their progression in the programme. Although our methods

included coding linguistic elements, expressions of speech and changes in spoken volume along with discussing our interpretations within the research team, our interpretation of emotions remains subjective. We also infer that the anger expressed and reported while recalling experiences during interviews is indicative of anger experienced while recognising failure. We needed to use various questioning techniques during interviews to encourage discussion on this sensitive topic. Our lead-in questions using prompts like 'least competent trainees' or 'trainees who caused the most worry' may have generated discussion on extreme cases at the expense of less obvious cases of failure. We did invite contrasting examples to be shared and allowed time for participants to discuss additional examples as ways to address this limitation. Our analysis focused on the shift from disbelieving early signs of underperformance to recognising that a trainee should be failed. By closely examining this link in the process, we were able to draw connections to tipping point research and use it to gain understanding of the process. However, the tipping point concept was not a sensitising concept that directed our analysis, as it was introduced during analysis that followed the completion of data collection. Its use as a sensitising concept in future studies could shape the content and format of interviews to centre the concept of the tipping point. In addition, we did not focus our analysis on the steps following recognition of failure. Further research is needed to continue examining how supervisors engage with subsequent steps in the process such as documenting failure, removing a trainee from a programme and responding to appeals within competency-based education and programmatic assessment systems. We encourage investigation into how our assessment systems can be refined to support earlier identification of underperformance and increased reporting of failure while also correcting for decreased recognition of improvement and other imperfect assessment evidence.

6 | CONCLUSION

We have shown that recognising failure is no easy feat, and deciding how to respond to patterns of failure is equally complex. Our model of the shift from disbelieved to recognised failure as a tipping point and our identification that anger co-occurs with the tipping point raises the possibility for emotions to be a driver of, and not only a barrier to, recognising failure. The tipping point model offers a reframing to support supervisors in reporting suspected underperformance before the tipping point has been reached so that early support for the trainee, and the supervisor, can be initiated by decision-makers. But it also warns that crossing the threshold of the tipping point, with accompanied anger and a sense of duty to prevent harm, risks ossifying the sense of certainty that a trainee needs to fail. Thus, it offers the impetus to ensure countermeasures within our assessment systems compensate for difficulties in detecting improvement once failure has been verified. Failure is a rare event and likely to remain distressing for many involved but enhanced understanding of how it is recognised, documented and monitored can inform continued improvements to assessment systems.

AUTHORS' CONTRIBUTION

All authors contributed to design of the study. AG collected data and led analysis and manuscript preparation with input from all other authors. All authors approved the final manuscript for publication and agree to be accountable for all aspects of the work by ensuring that all questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

ACKNOWLEDGEMENTS

We are grateful for the research support provided by Roseann Larstone, Lindsay Mathews and Kevala Van Volkenberg. We thank the physicians and surgeons who participated and everyone who invited others to participate.

FUNDING INFORMATION

We are grateful for funding awarded by the Medical Council of Canada's Research in Clinical Assessment grant program.

CONFLICT OF INTEREST

None.

ETHICS STATEMENT

This study received approval (E2018.0613.065) through the University of Northern British Columbia's research ethics review board.

ORCID

Andrea Gingerich  <https://orcid.org/0000-0001-5765-3975>

Stefanie S. Sebok-Syer  <https://orcid.org/0000-0002-3572-5971>

Christopher J. Watling  <https://orcid.org/0000-0001-9686-795X>

REFERENCES

- Cleland JA, Knight LV, Rees CE, Tracey S, Bond CM. Is it me or is it them? Factors that influence the passing of underperforming students. *Med Educ*. 2008;42(8):800-809.
- Hunt LA, McGee P, Gutteridge R, Hughes M. Failing securely: The processes and support which underpin english nurse mentors' assessment decisions regarding under-performing students. *Nurs Educ Today*. 2016;39:79-86.
- Dennis AA, Foy MJ, Monrouxe LV, Rees CE. Exploring trainer and trainee emotional talk in narratives about workplace-based feedback processes. *Adv Health Sci Educ Theory Pract*. 2018;23(1):75-93.
- Finch J, Taylor I. Failure to fail? Practice educators' emotional experiences of assessing failing social work students. *Soc Work Educ*. 2013;32(2):244-258.
- Hughes LJ, Mitchell M, Johnston AN. Failure to fail'in nursing—A catch phrase or a real issue? A systematic integrative literature review. *Nurse Educ Pract*. 2016;20:54-63.
- Hughes LJ, Mitchell ML, Johnston AN. Moving forward: Barriers and enablers to failure to fail—A mixed methods meta-integration. *Nurs Educ Today*. 2021;98:1-7.
- Scarff CE, Bearman M, Chiavaroli N, Trumble S. Keeping MUM in clinical supervision: Private thoughts and public judgements. *Med Educ*. 2019;53(2):133-142.
- Scarff CE. Assessor grades and comments: Private thoughts and public judgements. *Doctoral Thesis*. 2020.
- Luhanga F, Yonge OJ, Myrick F. "Failure to assign failing grades": Issues with grading the unsafe student. *Int J Nurs Educ Scholar*. 2008;5(1):8-14.
- Brewer M, Duncanson K, Gribble N, Reubenson A, Hart A. An intervention to enhance the supervision of health science students who struggle during work placements. *Int J Work-Integrat Learn*. 2021;22(2):149-166.
- Mak-van der Vossen M, Peerdeman S, Van Mook W, Croiset G, Kusurkar R. Assessing professional behaviour: overcoming teachers' reluctance to fail students. *BMC Res Notes*. 2014;7(1):368.
- Nugent O, Lydon C, Part S, et al. Who is failing who? A survey exploration of the barriers & enablers to accurate decision making when nursing students' competence is below required standards. *Nurs Educ Pract*. 2020;45:1-8.
- McQueen SA, Petrison B, Bhandari M, Fahim C, McKinnon V, Sonnadara RR. Examining the barriers to meaningful assessment and feedback in medical training. *Am J Surg*. 2016;211(2):464-475.
- Bush H, Schreiber R, Oliver S. Failing to fail: Clinicians' experience of assessing underperforming dental students. *Eur J Dental Educ*. 2013;17(4):198-207.
- Rees CE, Knight LV, Cleland JA. Medical educators' metaphoric talk about their assessment relationships with students: 'You don't want to sort of be the one who sticks the knife in them'. *Assess Eval High Educ*. 2009;34(4):455-467.
- Yepes-Rios M, Dudek N, Duboyce R, Curtis J, Allard RJ, Varpio L. The failure to fail underperforming trainees in health professions education: a BEME systematic review: BEME Guide No. 42. *Med Teach*. 2016;38(11):1092-1099.
- Speer AJ, Solomon DJ, Fincher R-ME. Grade inflation in internal medicine clerkships: Results of a national survey. *Teach Learn Med*. 2000;12(3):112-116.
- Cohen GS, Blumberg P, Ryan NC, Sullivan PL. Do final grades reflect written qualitative evaluations of student performance? *Teach Learn Med*. 1993;5(1):10-15.
- Jervis A, Tilki M. Why are nurse mentors failing to student nurses who do not meet clinical performance standards? *Brit J Nurs*. 2011;20(9):582-587.
- Dudek NL, Marks MB, Regehr G. Failure to fail: The perspectives of clinical supervisors. *Acad Med*. 2005;80(Suppl 10):S84-S87.
- Friedman Z, Bould MD, Matava C, Alam F. Investigating faculty assessment of anesthesia trainees and the failing-to-fail phenomenon: A randomized controlled trial. *Can J Anesth*. 2021;68(7):1-8.
- Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: Outlining their usefulness for competency-based clinical assessment. *Acad Med*. 2016;91(2):186-190.
- Rekman J, Hamstra SJ, Dudek N, Wood T, Seabrook C, Gofton W. A new instrument for assessing resident competence in surgical clinic: The Ottawa clinic assessment tool. *J Surg Educ*. 2016;73(4):575-582.
- Norcini JJ, Blank LL, Duffy FD, Fortna GS. The mini-cex: A method for assessing clinical skills. *Ann Intern Med*. 2003;138(6):476-481.
- Van der Vleuten CP, Schuwirth L, Driessen E, Govaerts M, Heeneman S. Twelve tips for programmatic assessment. *Med Teach*. 2015;37(7):641-646.
- van der Vleuten C, Schuwirth L, Driessen E, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205-214.
- Ginsburg S, Eva K, Regehr G. Do in-training evaluation reports deserve their bad reputations? A study of the reliability and predictive ability of iter scores and narrative comments. *Acad Med*. 2013;88(10):1539-1544.
- Schut S, Maggio LA, Heeneman S, Van Tartwijk J, Van der Vleuten C, Driessen E. Where the rubber meets the road—An integrative review

- of programmatic assessment in health care professions education. *Perspect Med Educ.* 2020;10(1):1-8.
29. Griffiths J, Dalgarno N, Schultz K, Han H, Van Melle E. Competency-based medical education implementation: Are we transforming the culture of assessment? *Med Teach.* 2019;41(7):811-818.
 30. Prentice S, Benson J, Kirkpatrick E, Schuwirth L. Workplace-based assessments in postgraduate medical education: A hermeneutic review. *Med Educ.* 2020;54(11):981-992.
 31. Melvin L, Rassos J, Stroud L, Ginsburg S. Tensions in assessment: The realities of entrustment in internal medicine. *Acad Med.* 2020;95(4):609-615.
 32. Sherbino J, Regehr G, Dore K, Ginsburg S. Tensions in describing competency-based medical education: A study of canadian key opinion leaders. *Adv Health Sci Educ Theory Pract.* 2021;26(4):1-13.
 33. Black S, Curzio J, Terry L. Failing a student nurse: A new horizon of moral courage. *Nurs Ethics.* 2014;21(2):224-238.
 34. Carr J, Heggarty H, Carr M, et al. Reflect for success: Recommendations for mentors managing failing students. *Brit J Commun Nurs.* 2010;15(12):594-596.
 35. Gingerich A, Sebok-Syer SS, Larstone R, Watling CJ, Lingard L. Seeing but not believing: Insights into the intractability of failure to fail. *Med Educ.* 2020;54(12):1148-1158.
 36. El Hussein MT, Fast O. Gut feeling: A grounded theory study to identify clinical educators' reasoning processes in putting students on a learning contract. *J Clin Nurs.* 2020;29(1-2):75-84.
 37. Hughes LJ, Mitchell ML, Johnston AN. Just how bad does it have to be? Industry and academic assessors' experiences of failing to fail—A descriptive study. *Nurse Educ Today.* 2019;76:206-215.
 38. Deegan J, Burton T, Rebeiro G. Clinical assessment and the benefit of the doubt: What is the doubt? *Aust J Adv Nurs.* 2012;30(1):42-48.
 39. Charmaz K. *Constructing Grounded Theory.* London, UK: SAGE Publications Ltd; 2014:1-416.
 40. Charmaz K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis.* London, UK: Sage Publications Ltd; 2006:1-208.
 41. Hyland K. Boosting, hedging and the negotiation of academic knowledge. *Text Talk.* 1998;18(3):349-382.
 42. O'Brien E, Klein N. The tipping point of perceived change: Asymmetric thresholds in diagnosing improvement versus decline. *J Pers Soc Psychol.* 2017;112(2):161-185.
 43. O'Brien E. When small signs of change add up: The psychology of tipping points. *Curr Dir Psychol Sci.* 2020;29(1):55-62.
 44. Klein N, O'Brien E. The tipping point of moral change: When do good and bad acts make good and bad actors? *Soc Cognition.* 2016;34(2):149-166.
 45. Klein N, O'Brien E. People use less information than they think to make up their minds. *P Natl Acad Sci.* 2018;115(52):13222-13227.
 46. Castanelli DJ, Weller JM, Molloy E, Bearman M. Shadow systems in assessment: How supervisors make progress decisions in practice. *Adv Health Sci Educ Theory Pract.* 2020;25(1):131-147.
 47. Hutcherson CA, Gross JJ. The moral emotions: A social-functional account of anger, disgust, and contempt. *J Pers Soc Psycho.* 2011;100(4):719-737.
 48. Wiltermuth SS, Tiedens LZ. Incidental anger and the desire to evaluate. *Organ Behav Hum Decis.* 2011;116(1):55-65.
 49. Greenbaum R, Bonner J, Gray T, Mawritz M. Moral emotions: A review and research agenda for management scholarship. *J Organ Behav.* 2020;41(2):95-114.
 50. Gibson DE, Callister RR. Anger in organizations: Review and integration. *J Manage.* 2010;36(1):66-93.
 51. Mullen E, Skitka LJ. Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, group differentiation, or anger? *J Pers Soc Psychol.* 2006;90(4):629-643.
 52. Rozin P, Lowery L, Imada S, Haidt J. The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *J Pers Soc Psychol.* 1999;76(4):574-586.
 53. Leary MR, Diebels KJ, Jongman-Sereno KP, Fernandez XD. Why seemingly trivial events sometimes evoke strong emotional reactions: The role of social exchange rule violations. *J Soc Psychol.* 2015;155(6):559-575.
 54. Suhay E, Erisen C. The role of anger in the biased assimilation of political information. *Polit Psychol.* 2018;39(4):793-810.
 55. Yip JA, Schweitzer ME. Losing your temper and your perspective: Anger reduces perspective-taking. *Organ Behav Hum.* Dec. 2019;150:28-45.
 56. Young MJ, Tiedens LZ, Jung H, Tsai M-H. Mad enough to see the other side: Anger and the search for disconfirming information. *Cognition Emotion.* 2011;25(1):10-21.
 57. Goldberg JH, Lerner JS, Tetlock PE. Rage and reason: The psychology of the intuitive prosecutor. *Eur J Soc Psychol.* 1999;29(5-6):781-795.
 58. Jung H, Young MJ. Motivated to confront: How experiencing anger affects anchoring bias. *J Behav Decis Making.* 2019;32(4):493-503.
 59. Samsudin EZ, Isahak M, Rampal S. The prevalence, risk factors and outcomes of workplace bullying among junior doctors: A systematic review. *Eur J Work Organ Psy.* 2018;27(6):700-718.
 60. Kalaichandran A. *In America, Becoming a Doctor Can Prove Fatal.* Boston Globe; 2019.
 61. Wilkinson TJ, Tweed MJ, Egan TG, et al. Joining the dots: Conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress. *BMC Med Educ.* 2011;11(1):1-9.

How to cite this article: Gingerich A, Sebok-Syer SS, Lingard L, Watling CJ. The shift from disbelieving underperformance to recognising failure: A tipping point model. *Med Educ.* 2021; 1-12. doi:10.1111/medu.14681

APPENDIX A: Key questions from four iterations of semi-structured interview guides

Goal of question	Questions and prompts that elicited answers for analysis
<p>Opening question: <i>Obtain demographic information and start to build rapport</i></p> <p>First iteration: <i>Invite memories of trainees who are most likely to be those with the potential to fail or to be failed. Use prompts to elicit descriptions of specific experiences. Try to find any descriptors that are used in reference to the trainees.</i></p>	<p>Please tell us a little bit about yourself? Such as where you are located, the type of work you do and your experience working with trainees?</p> <p>We'd like you to think back and reflect on the least competent trainees you have ever worked with. How many trainees come to mind? How many of them did you fail or flag in some way? Please tell us about one you failed and one you did not fail. Were they demonstrating incompetence? Please tell us a little bit more about realising that this trainee did not have [knowledge/insight/confidence/trustworthiness/ability]. How did it become apparent to you? How problematic was this? Did it create issues for patients? How did you change your supervision? How were they at implementing your feedback? What information did you include in your assessment of them? Did you alert anyone about this? Formally or informally? Did they receive remediation? Did they overcome it? What words did you use to describe this trainee? How did others describe this trainee? What words come to mind to describe this trainee?</p>
<p>Second iteration: <i>Invite stories of how working with these trainees impacts the supervisor and their supervision. Prompt for descriptions of what is noticed and recognised.</i></p>	<p>We'd like you to think back and reflect on the least competent trainees you have ever worked with. Please tell us about a trainee who required much more supervision than you had expected. What happened? Why did you increase supervision? How did you increase supervision? What would it take for you to then decrease supervision? Tell me more about that? [overconfidence, lack of confidence, poor insight, poor communication, lack of knowledge] How did it impact you? What feedback did you give to the trainee? What information did you include in your assessment of them (e.g. Did you fail or flag them in some way)? Thinking back on all of the least competent trainees you have supervised, which one caused you the most worry? Please tell us about them. Describe a typical encounter with them What was problematic about their performance? What was your response? (e.g. How did you react? Feel? What did you do?) Please tell us about any other trainees you flagged (or considered flagging or should have flagged). Or please tell us about any trainees you supervised who should not have graduated from your program but did.</p>
<p>Third iteration: <i>Use more encouraging language to acknowledge difficulty in describing their experiences. Prompt for contrasting examples (e.g. if they start with professionalism example then ask for technical skills/medical expertise example).</i></p>	<p>We know some of the questions are challenging to answer so please take your time. We'd like you to describe what it's like as you are discovering that a trainee is not learning and progressing as they should, even if it feels awkward and difficult to put into words. Please tell me about a trainee who could not do what you expected them to do—when it was problematic enough that you had to deal with it in some way. We'd really like to understand what it's like to be supervising a trainee like that. When you are supervising them, what are some of the earliest things you noticed? How did you become aware that the trainee was in difficulty? (What captured your attention? What was going through your mind?) Tell me more about the [identified issue]</p>

(Continues)

Goal of question	Questions and prompts that elicited answers for analysis
	<p>Could you give us some concrete examples of what that behaviour looked like?</p> <p>At what point would it be/was it a concern?</p> <p>How does a trainee like that impact you?</p> <p>How did you handle it?</p> <p>How fixable do you think this is?</p> <p>How did the conversation go when you were discussing [feedback] with the trainee?</p> <p>Okay, and then did you end up flagging this resident?</p> <p>Why did you (not) flag this trainee or how did you go about assessing/flagging them?</p> <p>What about difficulty with the [ask about a contrasting example]? It may be difficult to put it into words but I want you to pick an example of [this contrasting example] and kind of go back in time and walk us through when you begin to realise that this trainee lacks [example]. Go through the steps.</p> <p>Okay, up until that point in [contrasting example], were you getting any hints, anything drawing your attention about this trainee?</p> <p>So when you intervened in [that example], how did the trainee react?</p> <p>How typical is that for your specialty?</p> <p>And did you work with that resident again? What was that like?</p> <p>And then how do you describe a situation like that in their assessment?</p>
<p>Fourth iteration:</p> <p><i>Add more emphasis on eliciting examples relevant to failing a trainee or failing to fail a trainee. Prompt for details on flagging along with descriptions of experience.</i></p>	<p>Please tell us about any trainees you supervised who should not have graduated from your program but did.</p> <p>Why should not they have graduated? Why did they?</p> <p>What was it like supervising them?</p> <p>How do you describe a trainee like that to your colleagues?</p> <p>How did other people describe this trainee?</p> <p>How did you assess that trainee?</p> <p>Please tell us about any other trainees you flagged (or considered flagging or should have flagged).</p> <p>Please describe the trainee and the situation in as much detail as possible.</p> <p>What was it like supervising them?</p> <p>How did you give feedback on this? How did the feedback conversation go with them?</p> <p>Did you fail or flag them? Why or why not?</p>
<p>Closing questions:</p> <p><i>Bring closure to the interview. Allow time for missed concepts or examples to be brought forward. Pick the question that has not been touched on yet.</i></p>	<p>What are your thoughts on the words ‘competent’ and ‘incompetent’?</p> <p>What has been most difficult put into words?</p> <p>What are your thoughts on Failure-to-Fail?</p>