

## ASSESSMENT

# Seeing but not believing: Insights into the intractability of failure to fail

Andrea Gingerich<sup>1</sup>  | Stefanie S. Sebok-Syer<sup>2</sup>  | Roseann Larstone<sup>1</sup> | Christopher J. Watling<sup>3</sup>  | Lorelei Lingard<sup>4</sup>

<sup>1</sup>Northern Medical Program, University of Northern British Columbia, Prince George, British Columbia, Canada

<sup>2</sup>Emergency Medicine, Stanford Medicine, Stanford University, Stanford, California, USA

<sup>3</sup>Department of Clinical Neurological Sciences, Centre for Education Research and Innovation, Schulich School of Medicine and Dentistry, London, Ontario, Canada

<sup>4</sup>Department of Medicine, Centre for Education Research and Innovation, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

### Correspondence

Andrea Gingerich, Northern Medical Program, University of Northern British Columbia, 3333 University Way, Prince George, British Columbia V2N 4Z9, Canada. Email: Andrea.Gingerich@unbc.ca

### Funding information

This study was fully funded by The Medical Council of Canada Research in Clinical Assessment grant programme.

### Abstract

**Context:** Inadequate documentation of observed trainee incompetence persists despite research-informed solutions targeting this failure to fail phenomenon. Documentation could be impeded if assessment language is misaligned with how supervisors conceptualise incompetence. Because frameworks tend to itemise competence as well as being vague about incompetence, assessment design may be improved by better understanding and describing of how supervisors experience being confronted with a potentially incompetent trainee.

**Methods:** Following constructivist grounded theory methodology, analysis using a constant comparison approach was iterative and informed data collection. We interviewed 22 physicians about their experiences supervising trainees who demonstrate incompetence; we quickly found that they bristled at the term 'incompetence,' so we began to use 'underperformance' in its place.

**Results:** Physicians began with a belief and an expectation: all trainees should be capable of learning and progressing by applying what they learn to subsequent clinical experiences. Underperformance was therefore unexpected and evoked disbelief in supervisors, who sought alternate explanations for the surprising evidence. Supervisors conceptualised underperformance as: an inability to engage with learning due to illness, a life event or learning disorders, so that progression was stalled, or an unwillingness to engage with learning due to lack of interest, insight or humility.

**Conclusion:** Physicians conceptualise underperformance as problematic progression due to insufficient engagement with learning that is unresponsive to intensified supervision. Although failure to fail tends to be framed as a reluctance to document underperformance, the prior phase of disbelief prevents confident documentation of performance and delays identification of underperformance. The findings offer further insight and possible new solutions to address under-documentation of underperformance.

## 1 | INTRODUCTION

Inadequate documentation of observed trainee incompetence persists. Since Dudek and colleagues' seminal 'failure to fail' article<sup>1</sup> in 2005, a steady stream of findings have shown that assessors can often resist providing below-average ratings and writing comments that might be interpreted negatively, even when unsatisfactory clinical performance has been recognised.<sup>2-11</sup> Assessors are concerned that low ratings and constructive comments will be perceived as confrontational,<sup>12</sup> a type of harassment,<sup>9</sup> a form of punishment<sup>5</sup> or negatively overwhelming to the trainee.<sup>13</sup> Grade inflation,<sup>14</sup> delayed submission of assessments<sup>6</sup> and use of informal backchannels to circumvent formal assessment documentation<sup>15</sup> continue amidst the implementation of counterstrategies.<sup>16</sup> Inadequate documentation is particularly problematic for competency-based assessment approaches because both misclassification of incompetent trainees as competent and insufficiently documented evidence of clinical performance lead to inaccurate assessments of trainees.<sup>17</sup> The persistence of the failure to fail phenomenon, despite targeted strategies to combat it, suggests that documenting observed incompetence may be a singularly challenging task for assessors.

Perhaps incompetence is inadequately documented because it is inadequately defined. Although medical education has elaborate frameworks<sup>18-21</sup> to meticulously describe competence, incompetence tends to be referred to indirectly as 'not competent.' This definition is a good example of how incompetence is described in direct comparison to competence: 'Less than fully competent physicians or trainees fail to maintain acceptable standards in one or more areas of professional physician practice, whereas incompetent physicians lack the abilities (cognitive, noncognitive, and communicative) and qualities needed to perform effectively within the scope of professional physician practice.'<sup>22</sup> This definition also differentiates physicians who need improvement in certain areas to be fully competent from those who are deemed incompetent because they lack what is required to be competent. Our research focuses on how the latter is conceptualised and recognised by physicians when supervising trainees.

As illustrated in the above definition, incompetent physicians can be described in terms of what they are not doing competently and their skills or qualities are referred to using descriptors such as improper, inefficient, inappropriate and unacceptable.<sup>23-25</sup> Similarly, competence is often conceptualised as an amount of something, and not having enough of 'it' is by default categorised as incompetence. Accordingly, incompetent physicians are described as having deficits<sup>11</sup> and deficiencies<sup>22</sup> using adjectives such as limited, little and incomplete<sup>26</sup> to describe their knowledge and skills. By contrast, there are examples where competence has been conceptualised as developmental levels<sup>18</sup>: a series of milestones<sup>27</sup> or stages, such as in the RIME (Reporter-Interpreter-Manager-Educator) model of assessment.<sup>18</sup> The descriptors for each stage focus on what learners are doing. For example, they are collecting and reporting facts during the Reporter stage. In this model incompetence is the performance of basic, simple and rudimentary skills at a stage of development

where the performance of more sophisticated, complex skills is expected. Therefore, there exist in the medical education literature multiple conceptualisations of incompetence.

It has been said that the assessment of incompetence both reflects and reproduces its conceptualisation.<sup>28</sup> The varied conceptualisations of incompetence, as outlined above, each require assessors to enact and document their assessment judgements in different ways. A misalignment between the construct of incompetence imposed by the assessment tool and the conceptualisation used by assessors may impede documentation. Moreover, a lack of adequate language to articulate judgements of incompetence may complicate documentation. Although assessment designers have implemented alternate and seemingly more polite labels, such as struggling, in difficulty, at risk, problem or impaired,<sup>25,29</sup> such labels are still perceived as pejorative or disparaging.<sup>30</sup> This may be because in addition to cognitive considerations, explorations of assessors' resistance to fail underperforming trainees reveal a multitude of contributing social factors, including emotional reactions, relationship concerns with the trainee, personal repercussions, and implications for the trainee within the clinical learning system.<sup>1,5,7,9,12,13,31</sup> These findings suggest that communicating about incompetence, by assigning low ratings and providing negative comments, is not merely a cognitive activity but also a social act with social consequences.

Although much research has focused on understanding competence, there has been less exploration of how assessors conceptualise, communicate and navigate incompetence when supervising trainees. We need to identify the social rules regarding what can and cannot be said about incompetence when communicating with others. Understanding the social impact is imperative because the identification and clear communication of performance as incompetent is surely the foundation of strategies to correct and to build trainees' skills. We also need to gain insight into the social process of describing incompetence in order to align assessment approaches in ways that support and encourage assessors to share their judgements and comments. Inadequately documented assessment information from multiple supervisors in multiple contexts across time prevents programmes from effectively identifying trainees who need remediation. Therefore, to inform further improvement to our assessments, we examine physicians' experiences with supervising trainees who may be demonstrating incompetence to address the question: How do physician supervisors conceptualise and recognise incompetence in trainees?

## 2 | METHODS

Our research aimed to build an understanding of how physician supervisors conceptualise and recognise trainee incompetence when confronted by it in the clinical workplace. As such, we used constructivist grounded theory methodology<sup>32</sup> to guide data collection and analysis because it allows for a focus on the social processes involved when physicians encounter incompetence when interacting within a supervisor-trainee relationship. We scheduled semi-structured interviews and analysed the data in iterative

cycles such that the analysis of one cycle informed data collection in a subsequent cycle. More specifically, the analysis informed revisions to the interview guide and facilitated theoretical sampling of participants. This study received approval (E2018.0613.065) through the University of Northern British Columbia's research ethics review board.

## 2.1 | Participants and recruitment

Recruitment began with asking key informants, such as current and former programme directors, to forward email invitations to physician colleagues with clinical supervision and trainee assessment experience at the University of British Columbia (UBC) Faculty of Medicine. The UBC Medicine has numerous sites distributed across British Columbia, Canada, and includes both undergraduate and postgraduate programmes. The trainees complete university studies, undergo a highly competitive admissions process into 4 years of medical school, and then undergo a separate competitive selection process into a residency programme of at least 2 years' duration. Initial recruitment focused on purposive sampling of physicians from all specialties, clinical settings and geographies, and with diverse demographic characteristics. We then used snowball sampling to extend invitations to physicians with experience supervising trainees demonstrating incompetence, as identified by previous participants. Analysis of interview data informed theoretical sampling through the snowball method to recruit physicians with potentially confirmatory and discrepant perspectives and experiences to elaborate our understanding of the identified concepts and themes. To elaborate how 'fixable' incompetence was perceived to be, for example, we sought physicians who were respected for their involvement with remediation and those who were known for their involvement with trainee dismissals. Between September 2018 and March 2019, we recruited 22 physicians ( $n = 17$  female) from family practice ( $n = 13$ ) and non-surgical specialties ( $n = 9$ ). Participants practised in rural, northern and/or remote communities ( $n = 11$ ) or in large metropolitan cities ( $n = 11$ ). A total of 13 physicians had 10-20 years of clinical supervision experience (range, 4-25+ years).

## 2.2 | Data collection

We conducted interviews by telephone and in person. Interviews lasted an average of 50 minutes. Two researchers interviewed each physician: a research team member (RL) oversaw the logistics of the interview and asked questions from the interview guide, and the primary investigator (AG) asked prompting questions informed by iterative analysis of previous interviews and sensitising concepts, such as descriptors of unprofessional behaviour<sup>33</sup> and remediation as support,<sup>34</sup> from the medical education literature. The interviewers recorded notes during the interviews and debriefed after every interview to discuss their interpretations of the responses and efficacy of the interview questions.

Following our constructivist grounded theory methodology,<sup>32</sup> we took an iterative approach to the interview process. Interview questions invited physicians to share their experiences supervising medical students and residents who demonstrated 'incompetence.' After early participants told us that 'incompetent' and 'incompetence' were the wrong terms to use for trainees, we expanded our terminology to include multiple variations for underperformance, such as struggling, in difficulty, problematic, least competent, below expectations, barely competent, etc. We use the terms underperformance and underperforming as replacements because their definition 'requiring intervention beyond the normal level of supervisor-trainee interaction'<sup>2</sup> best aligns with our participants' descriptions of their trainees and their experiences supervising them. From this point, we continued to ask for their: reactions to our use of the terms 'incompetence' and 'incompetent'; descriptions of underperforming trainees, and understandings of competence, incompetence and underperformance. We also asked for elaboration on how underperformance was recognised and then managed. Appendix S1 is a representative interview guide compiled from all five iterations.

Typically, three or more different trainees were described within an interview. Often physicians spontaneously offered a set of trainees as representative examples and then contrasted their characteristics or performances within the descriptions. We used prompting questions to probe physicians' memories if a diverse selection of trainees was not offered. In a handful of interviews, the focus concentrated on one or two exceptionally memorable trainees and we used questions to assist physicians in describing as many aspects of their experience as possible. Interviews were audiorecorded and transcribed by a professional transcriptionist. Data were de-identified prior to analysis.

## 2.3 | Data analysis

Three members of the research team (AG, SSS-S and RL) completed and compared line-by-line coding on two early transcripts with diverse content. This facilitated a thorough review of the interview responses to identify what was being described and how it was described, so that the key fragments of the data could be studied in relation to others. The interviewers (AG and RL) read, re-read and discussed transcripts to inform changes to coding and the interview guide. Sensitising concepts from the failure to fail literature, such as concerns with documenting evidence of underperformance, were used to identify gaps in the data and inform revision of the interview guide and coding structure. The primary investigator (AG) analysed all transcripts as each cycle of three to five interviews was completed 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 focused codes through constant comparison.

Constructivist grounded theory recognises that meaning is constructed through the differing perspectives of the researchers and through their interaction with the participants and the

data.<sup>32</sup> Our team included researchers who study assessment by focusing on raters (AG, SSS-S) and have experience as a physician supervisor of trainees and postgraduate dean (CJW), along with constructivist grounded theory expertise (CJW and LL) and experience (AG, RL and SSS-S). Constructivist grounded theory's focus on social processes is aligned with the primary investigator's (AG) programme of research that has used social cognition as a lens to study physicians as they supervise and assess trainees. We used notes to track 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. Analysis informed modifications to the coding framework until it sufficiently summarised the incoming interview data. This occurred after 16 interviews. Analysis was stopped after 22 interviews when no new concepts were being coded within the framework despite additional interview data collected using modified interview guides. We then mapped and discussed the links and patterns between the focused codes to identify theoretical codes. The theoretical codes form the basis of our results by representing an interpretation that is abstracted from, and yet still grounded in, the data. We used NVivo Version 11 (QSR International Pty Ltd., Doncaster, Vic, Australia) for data management and coding.

### 3 | RESULTS

Examining incompetence through analysis of supervisors' experiences with trainees who had demonstrated *incompetence* was more difficult than anticipated, because most participants reacted strongly and negatively to our use of these terms for trainees. The terms were described as *pejorative* and *derogatory*, *too binary*, *too absolute*, and not applicable to trainees who had never been deemed competent. Supervisors also expressed concern that such labels could be potentially harmful for trainees who were learning more slowly and needed more support. Here, a supervisor offers an analogy that summarises the reactions to our use of the terms:

*... if you use the word incompetent, it's often—for me at least—there is an expected level of competence that wasn't met. So for this medical student, there wasn't an expected level of competence that wasn't met, it was 'huh, these are the quirks of the training system where not everybody can be considered competent at any given time.' But I'm trying to think of an analogy to compare it to ... like just because somebody's not clever, you don't call them stupid.*

(P12)

Our participants perceived an apparent social rule: that it is wrong to refer to trainees as incompetent. When physicians did give examples of trainees who may have met criteria to be characterised as incompetent, they tended to describe trainees from years past, with

the exception of a few physicians who described trainees who were currently on prolonged remediation contracts. Despite majority agreement on 'incompetent' and 'incompetence' being inappropriate terms, there was no consensus on preferred alternatives. Because their descriptions best align with the terms 'underperforming' and 'underperformance',<sup>2</sup> we use those terms as replacements.

In the following we detail how supervisors' expectations for medical training underpin three related concepts: (a) a conceptualisation of competence as progression made through expected engagement with learning; (b) a phase of disbelief when those expectations are unmet, and (c) a conceptualisation of underperformance as insufficient engagement with learning resulting in problematic progression for the trainee that is unresponsive to intensified supervisory efforts.

#### 3.1 | Progressing towards competence

When sharing their experiences as a supervisor, physicians described notable trainees by comparing their performance to the expectations they had for all trainees. Supervisors expressed a belief that '*our trainees are all smart*' (P9) due to the stringent admissions and selection processes: '*I really do believe our system does select for those who are able to acquire knowledge quite well and usually rapidly also*' (P7). Trainees were expected to show up on time, to be prepared, to participate in clinical experiences, to interact with teaching and feedback, and to put in a good effort as well as striving to do better. They were expected to be humble, to recognise their limitations and to ask for help when it was needed, '*You make mistakes—that's how you learn and that's expected*' (P4), and it was expected that trainees would acknowledge and be remorseful for those mistakes.

Supervisors believed that trainees would be good learners and expected them to be ready, willing and able to engage with learning. A commitment to learning was essential because physicians described medical training and the pursuit of competence as a progression: '*I think of them as progressing along the continuum and then eventually they get to that point where they're ready to step on to the next phase*' (P9). Language that indicated progression included expressions like 'move forward,' 'change as they progress through,' 'it happens gradually,' 'progress towards competence,' 'made that leap' and being 'on' or 'off-track.' Trainees were on a progression through the different courses, years of the programme, '*multiple rotations and multiple different preceptors and a million different attempts*' (P6).

The core concept of progression was engagement in repeated cycles of learning. As this supervisor begins to describe, the learning cycles require participating in an experience, acquiring feedback from the experience, and then applying that new learning in the next similar experience: '*I picture someone's journey towards competence as a student or a trainee who is open to feedback and successfully incorporates feedback into clinical practice in order to reach competence*' (P11). This supervisor further clarified what progression entails and how engaging in cycles of learning is the foundation of learning in the clinical workplace:

*I would say, 'These could've been the potential ramifications of what you were trying to do or the way in which you presented the patient and I think you should review x, y and z.' And the student would come back within the next two or three days and we would have a similar presentation and would've taken that feedback and would've applied it and sort of read up on things and made sure that they could actually change in the way in which I gave feedback previously.*

(P11)

All trainees were expected to demonstrate more rudimentary skills before acquiring more sophisticated clinical skills over time. The 'million different attempts' enabled trainees to progress from basic skills to increasingly more sophisticated skills and from straightforward cases to more complicated cases. Supervisors anticipated that trainees would have different strengths and weaknesses and interests but they still expected all trainees to make use of clinical experiences as learning opportunities. They described customising their supervision for trainees by discerning learning needs and accommodating them through the provision of clinical experiences, teaching and feedback:

*I just really see each learner very much as an individual and on their own individual path in medicine [...] you really have to spend the time with the person, ask them what they're working at and then see if what they're working at is congruent with what you think they should be working at. [...] they learn often the same things in different order and some people have more trouble with some stuff because of their particular personalities or the training that they've already had [...] to realize that that has nothing to do with what they might be capable of but with what training they've had already and what they are comfortable and confident in and that's why it's not that they're better, but that they're just in a different place because of their past experience.*

(P15)

Supervisors expected the majority of trainees to be stellar, with a few trainees taking longer to learn. When trainees were engaging in cycles of learning but required more attempts to implement teaching and feedback in similar subsequent scenarios, they were considered to be progressing slowly but not underperforming. Their slower progression could be accommodated with supervision and required more teaching and feedback based on a greater number of clinical experiences and even prolonged time in the programme:

*I felt that some people would take a really long time to achieve some sort of basic level of competence [...] at one particular moment in time, they may not be performing to the level that's expected of them, so they're not*

*competent in that sense but that doesn't mean they're not going to become competent with more time and mentorship.*

(P8)

In summary, when competence is conceptualised as an ongoing progression that requires engagement with learning, trainees are on track so long as they are making progression by engaging with learning.

### 3.2 | Disbelieving the unexpected

As supervisors recalled how they first noticed that a trainee might be underperforming, it was common for the descriptions to contain uncertainty in what they were seeing. They seemed perplexed by the situation and expressed difficulty in understanding what was happening:

*... despite having gone over with her prior to the interview all of the elements that needed to be included in say a complete history and then directly observing her doing it, she still wasn't able to include all of the required elements. I guess that had happened multiple times and they were having difficulty understanding why-when they had just gone through it with her-she couldn't then perform what they'd just asked her to do when she was observed to do it.*

(P22)

Supervisors used words like 'unbelievably,' 'surprising,' 'shocking,' 'odd' or 'weird' to label their earliest observations of worrisome behaviours. 'It shatters your whole perception' (P19) of a trainee when his or her actions reveal that he or she has not learned a basic concept. For example, he or she wants to prescribe a drug to raise something that was already too high and doing so would likely result in a 'dire outcome' for the patient. Phrases conveying how unexpected, and even bewildering, the experience was were used: 'I felt it was strange because I've never really seen that before' (P16).

Examples were shared where it was so unexpected to see signs that a trainee was not engaging with learning that it was almost unbelievable. The disbelief was accompanied by supervisors' expressions of doubt in their observations and second-guessing their interpretations. Accordingly, as supervisors were noticing concerning patterns of performance, they described giving the trainee 'the benefit of doubt' (P1) and 'a few more days just to see and make sure I've interpreted everything correctly' (P14) to prevent relying too much on a 'one-off clinical experience' (P4). They also described consciously challenging their initial impressions and searching for bias. However, we did not see expressions of denial or avoidance in acknowledging that a trainee might be underperforming. Rather, we noted an eagerness to better understand and make sense of the trainee's performance. Here a supervisor rapidly lists the questions

that come to mind to understand why the trainee is behaving in an unexpected way:

*... definitely it's an uncomfortable feeling. It's uncomfortable because I think—well first you definitely wonder why: why are they struggling or incompetent? And is it because of lack of adequate training; is it because they're struggling from an emotional point; is it lack of competence; is it just that they've never seen something before and that's where it's coming? So I think it does raise a lot of questions and [laughs] and uncertainty.*

(P18)

Put simply, it was expected that trainees would make mistakes and need guidance but it was not expected that a trainee could have extensive and ongoing difficulty engaging with learning. Believing that trainees are good learners and expecting them to engage with learning is the basis for how physicians conceptualise medical training and competence. This foreshadows the conceptualisation of underperformance as insufficient engagement, resulting in problematic progression that frustrates supervisory responses.

### 3.3 | Conceptualising underperformance

Supervisors described how underperforming trainees had a pattern of performance that became obvious over time: 'They keep asking the same questions. They keep repeating their mistakes. You keep correcting them on it and regardless of what you say or do, they just don't have the ability to retain that information and move forward with it and learn the skills' (P17). Accordingly, these trainees were described as performing at a noticeably more rudimentary stage than expected and as compared to their peers. There were several examples of residents described as performing 'like a medical student' and some physicians lamented how 'it was disheartening that somebody like this could've gotten this far along in the system' (P12). Their progression had stalled.

When a trainee's progression stalled it was ascribed to his or her inability to adequately engage with cycles of learning. Underperforming trainees could not apply learning from one situation to the next: 'they're not able to take that structured learning about "this is a CHF [congestive heart failure] problem, here's the approach to CHF that we just talked about with this patient, take it to the next one," and they're still not able to quite apply it' (P10). Underperforming trainees were sometimes referred to as 'very, very time-consuming and very, very frustrating' (P22) because they increased the supervisor's workload. Supervisors described increasing teaching, increasing feedback and being more vigilant, especially after concerning incidents.

With stalled progression signalling underperformance, supervisors described looking for causes of why the trainee who was trying to engage with learning was unsuccessful. The common causes they described fell into three categories: (a) a current temporary situational factor that was distracting their attention away from learning; (b) a medical or psychiatric condition that was interfering with

learning, or (c) a learning disability that was hampering learning. By identifying a cause of the stalled progression, supervisors expressed more confidence in directing the trainee towards appropriate support. Those working in remediation described needing to determine exactly what was impeding learning to restart progression:

*... when I think about [my remediation students], they've all made progress. We've reached a point at some point, or something flipped in them and we had a breakthrough, and then it was not smooth sailing but it was definitely upward sailing, right. There were several months of just grumbling along, nothing changing, being frustrated and suddenly, most of the time, there's like a switch that occurs and off they go.*

(P19)

When it became clear that a trainee was trying but 'spinning their wheels in our program' (P7) and repeatedly unable to apply learning that should have been gained from one experience to a subsequent similar experience, supervisors expressed concern that the trainee could not progress and should be flagged:

*... but there's still that hard line at some point that if you can't pick it up or you can't reach those milestones [...] you're given the chances, you're given the extra resources [...] there's been remediation. There's been time off [...] you're usually with one preceptor. It didn't work with that clinic so they were sent to another place. Didn't work with that place [...] at some point there is that time to say 'you're incompetent' or 'this is not for you' or 'something' like that.*

(P4)

Physicians also described experiences where trainees appeared to avoid engaging with learning experiences. As with the trainees who were unable to engage, they looked for explanations and causes for why trainees did not engage in cycles of learning. A common, and often first suspicion, was that the trainee lacked interest in the rotation or discipline:

*There's a lot of prenatal and pregnancy care in my [family practice] office. If I'm matched with a resident that's not interested in that, then it's not going to be on the forefront of their mind to improve their skillset and they're not going to be as motivated to do that. Whereas they might be if they were in a different clinic which focused on things that they were more interested in.*

(P17)

Supervisors expressed disappointment when they realised trainees were not interested in their area of medicine and adapted by trying to customise teaching to the trainee's area of interest or by focusing only on the fundamentals that needed to be learned. It was expected that these trainees would fully engage with rotations or topics that

were of greater interest to them. Disinterest in the topic was unacceptable if it adversely impacted patient care or team function. Other explanations included attributing the lack of engagement to an attitude or motivation problem.

Supervisors expressed more concern when they thought trainees were not engaging with learning because they were oblivious to the learning they required. Not recognising the need for improvement, even after it had been clearly pointed out by supervisors, tended to be attributed to lack of insight:

*I've also seen scenarios where it seems like there isn't even curiosity or they don't even seem to recognize or care that maybe they're falling short on an element or on a scale or an area of knowledge they should have. I'm thinking of one student who was, ya, almost like they were kind of just very shallow in their approach—their gathering of the history, minimal physical exam, not a very broad differential—but they also didn't seem to really care (laughs) much about it either. You know, because I think it's different if someone says 'No I'm sorry this is something I've never seen before but I'm going to do reading around it' or 'I'll look it up later' versus not even having that insight.*

(P18)

Insight was seen as something innate and lack of insight as something that could not be adequately addressed by the supervisor or learned by the trainee. Some programme directors and residency committee members even declared that it was the most important factor in residency, as expressed by this participant:

*We often say when we're choosing residents that insight may be the single most important thing in residency, more than how much knowledge do you have in your head or anything else because insight is what allows you to improve yourself all the time and that's what training and being a good doctor is about—is continually improving your knowledge, your skills, your experience, your communication, etc—and if you can't look at yourself critically and say, 'Yes, I lack in this and therefore I have to do this better.' If you can't do that then that's a big hurdle to being a good doctor.*

(P5)

Physicians seemed even more frustrated when they felt that trainees had decided that they had nothing to learn from the clinical experiences or supervisors and then refused or ignored feedback and teaching. Such trainees were seen as uncoachable and unresponsive to teaching: *'when given feedback [they] would get immediately defensive and give excuses as to why things had gone the way they had or justifications as to why they actually were performing well. They weren't willing to hear the feedback'* (P22). The choice to disengage from learning because learning was considered unnecessary tended to be attributed to

arrogance, narcissism, overconfidence or lack of humility. Supervisors were concerned that these trainees would not recognise their limitations and not ask for help when they were beyond their scope. They referred to such trainees as 'potentially dangerous':

*... every year I would say I get at least one student with an attitude that kind of scares me. Somebody who knows it all already or knows it better than you and is correcting you to patients, or who is so sure that they've told you the right diagnosis—so much more sure than I am and they've just started third year [medical school]—and that attitude concerns me for safety reasons, for the patient, and it concerns me about the learner's attitude towards learning and their humility in this very scary job.*

(P15)

## 4 | DISCUSSION

We set out to examine physicians' experiences with supervising trainees who may be demonstrating incompetence to better understand their conceptualisation of incompetence. We found, however, that supervisors bristled at the term 'incompetence' and did not agree on a preferred alternative. Because the definition of an underperforming trainee<sup>2</sup> best aligned with their descriptions, we used that term instead. Our analysis of physicians' experiences supervising underperforming trainees offers two related conceptualisations: (a) competence as progression through developmental phases that is dependent on engagement in experiential learning, and (b) underperformance as problematic progression due to inability or unwillingness to engage with learning that is unresponsive to intensified supervision. Although the failure to fail phenomenon was not our intended focus of study, our findings offer further insight into its intractability.

Expectations anchor both the conceptualisation of progression towards competence and the conceptualisation of underperformance as problematic progression. Supervisors believe that trainees can succeed. They expect them to arrive at their clinical workplace able and willing to engage in the cycles of learning that are needed to progress through developmental levels towards competence. Although incompetence tends to be conceptualised as those who 'fail to demonstrate competence,'<sup>22</sup> the supervisors' emphasis on progression aligns best with the conceptualisation of competence as a series of stages with increasingly more difficult skills, as exemplified by developmental frameworks<sup>18</sup> such as the RIME assessment method<sup>18</sup> and milestones.<sup>27</sup> Supervisors may be more inclined to document the specifics of what the trainee is doing rather than to describe what he or she is not doing well enough.

Supervisors expect that a small percentage of trainees will require more supervision, increased feedback and a greater number of learning cycles to progress more slowly through the programme. But, because they demonstrate the capacity to engage with learning, and because their progression can be supported through intensified

supervision, slowly progressing trainees are not considered to be underperforming. It is, therefore, the process of learning that is emphasised in these conceptualisations. Expecting trainees to try, practise and engage with ongoing cycles of learning aligns well with a growth mindset of learning.<sup>35</sup> The expectation for learning to occur through interactions within a chain of experiences in the real world fits well with experiential learning theories that do not limit learning to a purely cognitive activity.<sup>36,37</sup> These findings point to a potential incongruence between supervisors' inclination to focus on process and policymakers' demands for assessment to focus on outcomes, which is worthy of further investigation.

Although supervisors appear quite prepared to accept that different trainees will arrive with different levels of skill and knowledge, they expect all to show up and make an effort. Underperforming trainees may upend these expectations; like a first encounter with a black swan,<sup>38</sup> supervisors may pause when they encounter a trainee who repeatedly cannot or will not engage with the learning process. When peoples' expectations are violated they can experience something - a feeling, a mood, an arousal state - that we have labelled disbelief, although it has been variously labelled as uncertainty, disequilibrium and a feeling of the absurd elsewhere.<sup>39</sup> Regardless of the specific terms used, when peoples' experiences are inconsistent with their expectations they can have an uncomfortable feeling that motivates resolution of the inconsistency by somehow making sense of the situation.<sup>40</sup> Consequently, we have identified a relevant phase of cognition that precedes the classic failure to fail phenomenon in some cases: a phase we call disbelief.

Recognition of a disbelief phase that precedes the failure to fail phenomenon may help to explain why solutions have not eradicated the failure to fail phenomenon. Those solutions target the lack of support in reporting, unwillingness to report and limited ability to report underperformance that occurs after supervisors have identified underperformance.<sup>41</sup> During the disbelief phase, supervisors are unsure of what they are seeing; they have not yet made meaning of it and it is challenging to determine what evidence is noteworthy and attributable to the trainee. It may require time to work through this phase, and thus identification of underperforming trainees could be delayed. Although caution in assessment is typically a virtue, and supervisors should be supported to gather sufficient information prior to documenting high-stakes assessment judgements, trainees may move on before their supervisor has graduated from disbelief to belief to action. As has been described in the failure to fail literature,<sup>3,42</sup> if the duration of the relationship is too short and does not allow for a sufficient number of interactions, then the supervisor may be unable to identify an unexpected case of underperformance. Increasing the duration of rotations<sup>4</sup> has been used to target the failure to fail phenomenon and may help to accommodate the disbelief phase. Other failure to fail solutions that may also target this phase include encouraging discussion between supervisors, such as strategic learner handover<sup>43</sup> and mentorship training sessions,<sup>44</sup> to help supervisors process their experiences and understand unexpected trainee performance. Disbelief might also be addressed in faculty development materials by acknowledging that not all trainees should

graduate<sup>3</sup> and demystifying underperformance by clearly describing trajectories and outcomes for underperforming trainees.<sup>45</sup>

Although slow learning was expected, inability and unwillingness to engage with learning were not. The disbelief phase may represent a form of cognitive dissonance called either disconfirmed expectancy or failed prophecy.<sup>46,47</sup> Although cognitive dissonance tends to be framed as a shortcoming, recent theories suggest that it could be an asset in situations that involve difficult decisions.<sup>48</sup> In such theories, the discomfort of dissonance signals that there is a conflict between cognition and corresponding actions that are incompatible; this cognitive inconsistency needs to be resolved so that an action can occur.<sup>49</sup> In the context of clinical supervision, the unexpectedness of encountering insufficient engagement with learning may signal a contradictory supervisory action and evoke dissonance. For example, intensifying supervision is such a common action to support underperforming trainees that it can define underperformance,<sup>2</sup> so we would anticipate dissonance as a supervisor realises intensification efforts are futile and should cease because they are ineffective for supporting an underperforming trainee's progression. Future research using this lens may help us differently understand peculiar assessor reactions, such as the teacher's dilemma of feeling failure as a teacher when assigning a negative assessment to a student.<sup>50,51</sup> Solutions could target the identified cognitive inconsistencies to promote improved assessment documentation.

It is useful to consider how the language used for instructions, prompts and scales on our assessment forms could be revised in light of these findings. Supervisors resisted identifying trainees whose progress required additional learning experiences, teaching and feedback as underperforming, and yet these trainees could benefit from recognition of their increased needs. Intensifying supervision is such a common response that documentation of how the supervisor is supporting the trainee's progression could potentially serve as an early marker of slowed progression and underperformance. Level of supervision scales<sup>52</sup> could serve as a prototype for targeted development. Similarly, the notion of problematic progression that cannot be fully accommodated through supervision is a defining element for recognising underperformance. Designs that help supervisors to document when their supervisory efforts are not supporting trainee progression could be an alternative way to flag trainees who need formal remediation. Such approaches would expand the assessment focus from describing the trainee and his or her performance to describing the interactions between the supervisor and trainee.

Supervisors discussed trainees' engagement with learning as a key feature of underperformance. Assessment forms that document trainees' engagement with learning cycles would allow supervisors to acknowledge the effort made by the trainee. Documentation of the process of learning might help to identify slowly progressing trainees and to differentiate them from those who cannot or will not engage in learning. It could include descriptions of trainees' response to the supervisor's feedback, evidence of them incorporating feedback into subsequent similar situations, indications of insight, descriptions of work ethic and invested effort, and other language used by supervisors<sup>53</sup> that captures the

process of learning. Such language could help to describe the interactions between trainee, supervisor, patient and workplace in addition to cognitive capabilities of the trainee. Assessors do write about engagement with the learning process in their comments,<sup>54,55</sup> although such comments can be interpreted negatively by others.<sup>56</sup> Therefore, strategies to counter associated stigma may be needed to implement assessments specifically designed to document learning process engagement. Documenting engagement with experiential learning could complement the documentation of learning outcomes.<sup>57</sup>

#### 4.1 | Limitations

It is important to reflect on the limitations of this research. The belief that all trainees should be able and willing to engage with learning may be especially strong in Canada where societal needs determine the number of trainees granted access to highly competitive medical programmes. Further investigation is needed to determine if the disbelief phase is unique to specific conditions. Our use of snowball sampling relied on participants' networks and may have contributed to a greater number of women agreeing to participate than men, no participation from surgical specialties and limited participation from procedure-dominant specialties. Because there may be differences in the beliefs and expectations for trainees when there is a stronger focus on psychomotor skills, physical demands and artistry, further study is needed.

### 5 | CONCLUSIONS

Physicians conceptualise competence as trainees progressing through developmental levels by engaging in experiential learning when supported by supervisory efforts. Underperformance depends on expectations, however, and our findings show that a phase of disbelief may precede the failure to fail phenomenon. This phase of apparent cognitive dissonance may prevent confident documentation of performance and delay identification of underperformance. Although supervisors are uncomfortable with the term incompetence, they can recognise underperformance, characterising it around three related elements: (a) inability or unwillingness to engage with learning; (b) stalled or problematic progression, and (c) unresponsiveness to intensified supervisory efforts. The elements could be used individually or in combination for assessment purposes and may allow for identification of a broader range of trainees, from those who are progressing slowly with additional supervision and learning opportunities to those who despite intensified supervision are no longer progressing or no longer engaging effectively with learning. Further research is needed to determine how best to align the supervisors' focus on the process of learning through social interactions in the clinical workplace with contemporary education frameworks and assessment practices.

#### AUTHOR CONTRIBUTIONS

AG, SSS-S, CJW and LL contributed to the design of the study. AG and RL collected data. All authors (AG, SSS-S, RL, CJW and LL) contributed to data analysis and manuscript preparation, 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 Lindsay Mathews and Kevala Van Volkenberg. We thank the physicians who participated and everyone who invited others to participate.

#### CONFLICTS OF INTEREST

None.

#### ETHICAL APPROVAL

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

1. Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med*. 2005;80(Suppl 10):S84-S87.
2. Barrett A, Galvin R, Steinert Y, et al. A BEME (best evidence in medical education) review of the use of workplace-based assessment in identifying and remediating underperformance among postgraduate medical trainees: BEME Guide No. 43. *Med Teach*. 2016;38(12):1188-1198.
3. 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.
4. Daelmans HEM, Mak-van der Vossen MC, Croiset G, Kusrkar RA. What difficulties do faculty members face when conducting workplace-based assessments in undergraduate clerkships? *Int J Med Educ*. 2016;7:19-24.
5. Mak-van der Vossen M, Peerdeman S, van Mook W, Croiset G, Kusrkar R. Assessing professional behaviour: overcoming teachers' reluctance to fail students. *BMC Res Notes*. 2014;7(1):368.
6. Christensen MK, O'Neill L, Hansen DH, Norberg K, Mortensen LS, Charles P. Residents in difficulty: a mixed methods study on the prevalence, characteristics, and sociocultural challenges from the perspective of residency program directors. *BMC Med Educ*. 2016;16(1):69.
7. 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.
8. Bandiera G, Lendrum D. Daily encounter cards facilitate competency-based feedback while leniency bias persists. *Can J Emerg Med*. 2008;10(1):44-50.
9. McQueen SA, Petrisor 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.
10. Guerrasio J, Furfari KA, Rosenthal LD, Nogar CL, Wray KW, Aagaard EM. Failure to fail: the institutional perspective. *Med Teach*. 2014;36(9):799-803.
11. Guerrasio J, Garrity MJ, Aagaard EM. Learner deficits and academic outcomes of medical students, residents, fellows, and attending

- physicians referred to a remediation program, 2006–2012. *Acad Med*. 2014;89(2):352–358.
12. Watling CJ, Kenyon CF, Schulz V, Goldszmidt MA, Zibrowski E, Lingard L. An exploration of faculty perspectives on the in-training evaluation of residents. *Acad Med*. 2010;85(7):1157–1162.
  13. Kogan JR, Conforti LN, Bernabeo EC, Durning SJ, Hauer KE, Holmboe ES. Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Med Educ*. 2012;46(2):201–215.
  14. Alexander EK, Osman NY, Walling JL, Mitchell VG. Variation and imprecision of clerkship grading in us medical schools. *Acad Med*. 2012;87(8):1070–1076.
  15. Adams KE, Emmons S, Romm J. How resident unprofessional behavior is identified and managed: a program director survey. *Am J Obstet Gynecol*. 2008;198(6):692. e1–e5.
  16. Nixon LJ, Gladding SP, Duffy BL. Describing failure in a clinical clerkship: implications for identification, assessment and remediation for struggling learners. *J Gen Intern Med*. 2016;31(10):1172–1179.
  17. McGill D, van der Vleuten C, Clarke M. Supervisor assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review. *Adv Health Sci Educ*. 2011;16(3):405–425.
  18. Pangaro L, ten Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Med Teach*. 2013;35(6):e1197–e1210.
  19. General Medical Council. Good medical practice. [GMC website]. <https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/good-medical-practice>. Accessed March 8, 2020.
  20. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach*. 2007;29(7):642–647.
  21. Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach*. 2007;29(7):648–654.
  22. Hauer KE, Ciccone A, Henzel TR, et al. Remediation of the deficiencies of physicians across the continuum from medical school to practice: a thematic review of the literature. *Acad Med*. 2009;84(12):1822–1832.
  23. Caulford PG, Lamb SB, Kaigas TB, Hanna E, Norman GR, Davis DA. Physician incompetence: specific problems and predictors. *Acad Med*. 1994;69(Suppl 10):S16–S18.
  24. Yao DC, Wright SM. National survey of internal medicine residency program directors regarding problem residents. *JAMA*. 2000;284(9):1099–1104.
  25. Dupras DM, Edson RS, Halvorsen AJ, Hopkins RH, McDonald FS. 'Problem residents': prevalence, problems and remediation in the era of core competencies. *Am J Med*. 2012;125(4):421–425.
  26. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Med Educ*. 2011;45(6):560–569.
  27. Holmboe ES, Edgar L, Hamstra S. *The Milestones Guidebook*. Chicago, IL: Accreditation Council for Graduate Medical Education; 2016.
  28. Hodges B. Medical education and the maintenance of incompetence. *Med Teach*. 2006;28(8):690–696.
  29. Cleland J, Leggett H, Sandars J, Costa MJ, Patel R, Moffat M. The remediation challenge: theoretical and methodological insights from a systematic review. *Med Educ*. 2013;47(3):242–251.
  30. Yao DC, Wright SM. The challenge of problem residents. *J Gen Intern Med*. 2001;16(7):486–492.
  31. 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.
  32. Charmaz K. *Constructing Grounded Theory*. London, UK: SAGE Publications Ltd.; 2014:1–416.
  33. Mak-van der Vossen M, van Mook W, van der Burgt S, et al. Descriptors for unprofessional behaviours of medical students: a systematic review and categorisation. *BMC Med Educ*. 2017;17(1):164.
  34. Bourgeois-Law G, Varpio L, Regehr G, Teunissen PW. Education or regulation? Exploring our underlying conceptualisations of remediation for practising physicians. *Med Educ*. 2019;53(3):276–284.
  35. Dweck CS, Yeager DS. Mindsets: a view from two eras. *Perspect Psychol Sci*. 2019;14(3):481–496.
  36. Dreon R. Framing cognition: Dewey's potential contributions to some enactivist issues. *Synthese*. 2019;1–22. <https://doi.org/10.1007/s11229-019-02212-x>
  37. Elkjaer B. *Pragmatism: a learning theory for the future*. In: Illeris K, ed. *Contemporary Theories of Learning*. Philadelphia, PA: Routledge; 2009:82–97.
  38. Taleb NN. *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*, 2nd edn. New York, NY: Thomson/Texere; 2004.
  39. Proulx T. The feeling of the absurd: towards an integrative theory of sense-making. *Psychol Inq*. 2009;20(4):230–234.
  40. Proulx T, Inzlicht M, Harmon-Jones E. Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends Cogn Sci*. 2012;16(5):285–291.
  41. 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.
  42. 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.
  43. Humphrey-Murto S, LeBlanc A, Touchie C, et al. The influence of prior performance information on ratings of current performance and implications for learner handover: a scoping review. *Acad Med*. 2019;94(7):1050–1057.
  44. 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.
  45. Fazio SB, Papp KK, Torre DM, DeFer TM. Grade inflation in the internal medicine clerkship: a national survey. *Teach Learn Med*. 2013;25(1):71–76.
  46. Harmon-Jones E, Harmon-Jones C. Cognitive dissonance theory after 50 years of development. *Z Sozialpsychol*. 2007;38(1):7–16.
  47. Festinger L, Riecken HW, Schachter S. *When Prophecy Fails: A Social and Psychological Study of a Modern Group that Predicted the Destruction of the World*. New York, NY: Harper Torchbooks; 1956.
  48. Brannon SM, Gawronski B. Cognitive consistency in social cognition. In: Hogg M, ed. *The Oxford Encyclopedia of Social Psychology*. New York, NY: Oxford University Press; 2018.
  49. Harmon-Jones E, Harmon-Jones C, Levy N. An action-based model of cognitive-dissonance processes. *Curr Dir Psychol Sci*. 2015;24(3):184–189.
  50. Mak-van der Vossen M. 'Failure to fail': the teacher's dilemma revisited. *Med Educ*. 2019;53(2):108–110.
  51. Lankshear A. Failure to fail: the teacher's dilemma. *Nurs Stand*. 1990;4(20):35–37.
  52. 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.
  53. Ginsburg S, McIlroy J, Oulanova O, Eva KW, Regehr G. Toward authentic clinical evaluation: pitfalls in the pursuit of competency. *Acad Med*. 2010;85(5):780–786.
  54. Ginsburg S, Gold W, Cavalcanti RB, Kurabi B, McDonald-Blumer H. Competencies, 'plus': the nature of written comments on internal medicine residents' evaluation forms. *Acad Med*. 2011;86(Suppl 10):S30–S34.
  55. White JS, Sharma N. 'Who writes what?' using written comments in team-based assessment to better understand medical student performance: a mixed-methods study. *BMC Med Educ*. 2012;12(1):123.
  56. Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: faculty interpretations of narrative evaluation comments. *Med Educ*. 2015;49(3):296–306.

57. Bacchus M, Ward DR, de Groot J, Lemaire JB. How evidence from observing attending physicians links to a competency-based framework. *Med Educ*. 2017;51(6):633-644.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** 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;54:1148-1158. <https://doi.org/10.1111/medu.14271>