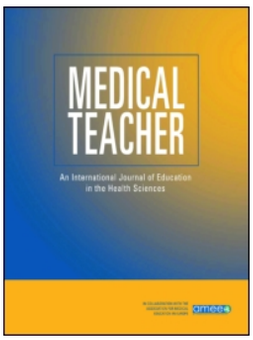

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: http://www.tandfonline.com/loi/imte20

The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42

Monica Yepes-Rios, Nancy Dudek, Rita Duboyce, Jerri Curtis, Rhonda J. Allard & Lara Varpio

To cite this article: Monica Yepes-Rios, Nancy Dudek, Rita Duboyce, Jerri Curtis, Rhonda J. Allard & Lara Varpio (2016): The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42, Medical Teacher

To link to this article: http://dx.doi.org/10.1080/0142159X.2016.1215414

MEDICAL TEACHER, 2016
http://dx.doi.org/10.1080/0142159X.2016.1215414

MEDICAL TEACHER
Taylor & Francis
Taylor & Francis Group

BEME GUIDE

The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42

Monica Yepes-Rios[a], Nancy Dudek[b], Rita Duboyce[a], Jerri Curtis[a], Rhonda J. Allard[a] and Lara Varpio[a]

[a]Uniformed Services University of the Health Sciences, Bethesda, MD, USA; [b]Ottawa Hospital Rehabilitation Centre, University of Ottawa, Ottawa, ON, Canada

ABSTRACT

Background: Many clinical educators feel unprepared and/or unwilling to report unsatisfactory trainee performance. This systematic review consolidates knowledge from medical, nursing, and dental literature on the experiences and perceptions of evaluators or assessors with this failure to fail phenomenon.

Methods: We searched the English language literature in CINAHL, EMBASE, and MEDLINE from January 2005 to January 2015. Qualitative and quantitative studies were included. Following our review protocol, registered with BEME, reviewers worked in pairs to identify relevant articles. The investigators participated in thematic analysis of the qualitative data reported in these studies. Through several cycles of analysis, discussion and reflection, the team identified the barriers and enablers to failing a trainee.

Results: From 5330 articles, we included 28 publications in the review. The barriers identified were (1) assessor's professional considerations, (2) assessor's personal considerations, (3) trainee related considerations, (4) unsatisfactory evaluator development and evaluation tools, (5) institutional culture and (6) consideration of available remediation for the trainee. The enablers identified were: (1) duty to patients, to society, and to the profession, (2) institutional support such as backing a failing evaluation, support from colleagues, evaluator development, and strong assessment systems, and (3) opportunities for students after failing.

Discussion/conclusions: The inhibiting and enabling factors to failing an underperforming trainee were common across the professions included in this study, across the 10 years of data, and across the educational continuum. We suggest that these results can inform efforts aimed at addressing the failure to fail problem.

Background

A cornerstone of successful medical education is the honest and accurate assessment of trainee performances. Medical educators and researchers have long been developing and improving their assessment strategies (e.g. objective structured clinical examinations, multiple choice question exams, etc.) and conceptual frameworks for performance expectations (e.g. competency based medical education, entrustable professional activities, etc.) to ensure that graduates are highly skilled medical professionals. Despite these efforts, the assessment of trainees remains fraught with challenges. One notable challenge is the assessment and reporting of trainees who are failing. A growing body of evidence indicates that clinical teachers who evaluate or assess trainees (hereby referred to as assessors) feel unprepared and/or unwilling to report a trainee's failing performance (Tonesk & Buchanan 1987; Cohen et al. 1990, 1993; Speer et al. 1996; Hatala & Norman 1999; Fitzgerald et al. 2010; Guerrasio et al. 2014; Luhanga et al. 2010; Brown et al. 2012; Cleland et al. 2013; Guerrasio et al. 2014). Dudek et al. (2005) described faculty members' perspectives on failing to fail underperforming trainees and identified four major barriers to reporting poor performance to include: (1) lack of documentation, (2) lack of knowledge of what to document, (3) anticipation of an appeal process, and (4) lack of remediation options (Dudek et al. 2005). Over the past decade scattered reports from the medical (Horvath 2010), nursing (Larocque & Luhanga 2013) and dental (Bush et al. 2013) literature have raised similar concerns about failing to fail underperforming trainees. Some researchers have proposed solutions to aspects of this complex issue (Duffy 2006; Carr et al. 2010; Luhanga et al. 2010; Jervis & Tilki 2011; Earle-Foley et al. 2012; Bush et al. 2013; Fazio et al. 2013; Larocque & Luhanga 2013; Pratt et al. 2013), but the improvements demonstrated are often small in scale [see Box 1].

Practice points

- We identified six barriers to failing underperforming trainees: (1) evaluator's professional considerations, (2) evaluator's personal considerations, (3) trainee related considerations, (4) unsatisfactory evaluator development and evaluation tools, (5) institutional culture and (6) consideration of available remediation for the trainee.
- We identified three enablers supporting assessors' willingness to fail a failing trainee: (1) duty to patients, to society, and to the profession, (2) institutional support such as backing a failing evaluation, support from colleagues, evaluator development, and strong assessment systems, and (3) opportunities for students after failing.

CONTACT Monica Yepes-Rios, MD, FACP ana.yepes-rios@usuhs.edu Department of Medicine, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd, Bethesda, MD 20814, USA

Supplemental data for this article can be accessed here.

Box 1. Proposed interventions to address failure to fail.

Proposed intervention	Reference
Base interventions on education theory	Cleland et al. 2008
Faculty development for assessors to be aware of passing underperformance	Monrouxe et al. 2011
Faculty development and support	Fazio et al. 2013
Culture shift to affect change	
Transparency in reporting grade distribution	
Faculty development, increase skill and comfort with failing	Vezeau & McAllister 2009
Mentorship preparation, coaching, early support	Black et al. 2014
Deliberate reflective practices	Debrew & Lewallen 2014
Faculty development for assessors to articulate their concerns	Dudek 2005
Consistent assessments	
Evaluation of grading process systems and outcomes	Heaslip & Scammell 2012
Examine students' experience	Larocque & Luhanga 2013
Policies and procedures to support assessors	
Education preparation for ethical competence	Gopee 2008

To date, no comprehensive literature review has consolidated the knowledge on assessor experiences and perceptions of the failure to fail phenomenon. This review aims to systematically consolidate and analyze a decade of knowledge from medical, dental and nursing literature relating to assessor's ability and willingness to report poor clinical academic and professional performance, and thereby help advance effective interventions. We hope it assists health professions educators to develop effective solutions to this multifaceted and important issue.

Methods

This study's methodology was reviewed and approved by the Best Evidence in Medical Education (BEME) Collaboration, (http://www.bemecollaboration.org), and approved by the lead institution's ethics board. We report the methodology and results in accordance with PRISMA guidelines (Moher et al. 2009).

A systematic review of the literature was conducted, relating to the failure to fail phenomenon across three healthcare disciplines (i.e. medicine, nursing, and dentistry), to capture the experiences and perceptions of assessors when presented with underperforming or failing trainees at any level of the health professions education continuum (e.g. medical students and residents). Search strategies and terms, listed in Appendix 1 (available online on the Journal website as Supplemental Material) and online at BEME Reviews (http://www.bemecollaboration.org), were applied to the English language literature in CINAHL, EMBASE, and MEDLINE via Ovid (In-Process & Other Non-Indexed Citations and MEDLINE 1946 to Present) from January 2005 to January 2015.

In May 2014, running the search strategies generated a library of 4625 publications. In January 2015, an update brought the total to 5330 publications. The team developed a set of inclusion and exclusion criteria (described in Appendix 1, available online on the Journal website as Supplemental Material). To ensure a broad scope of inclusion, we included all articles related to evaluator experiences of and/or perceptions of the failure to fail phenomenon in relation to trainee education in a clinical setting. We excluded studies related to the development or vetting of assessment tools, and to the development or assessment of curriculum content, of educational interventions, of peer mentoring, and of tools for assessing specific trainee skills, competencies, and/or knowledge.

The research team examined the titles of a random sample of 150 citations to test the inclusion and exclusion criteria. Satisfied that these criteria captured the literature related to the failure to fail phenomenon, the library was divided in two, with two or three reviewers assigned to each half of the library (reviewer team 1 = ND, JC; reviewer team 2 = LV, RD, MY). Each team reviewed the titles of the articles, applying the inclusion and exclusion criteria to half the library (2665 titles). If one team could not reach consensus, the other team of reviewers examined the citation and came to a consensus decision. Through this process, the library was reduced to 266 titles. Each team reviewed 133 abstracts to determine if inclusion and exclusion criteria applied. Any abstract deemed relevant by one reviewer moved on to full article review, yielding a library of 134 articles for full review. Finally, each article was analyzed against the inclusion and exclusion criteria (half by each reviewer team) yielding a library of 24 articles. We hand-searched the reference lists of the 24 articles to identify any omitted articles. This search added 4 citations for a final library of 28 articles. (Figure 1 outlines the selection process and Table 1 presents the selected articles – these are available online on the Journal website as Supplemental Material). DistillerSR® Systematic Review Software was used to support analysis across all steps of the literature review process.

Data were abstracted and synthesized for demographics (i.e. publication year, country, trainee group (i.e. medical, dental, or nursing)), evaluator demographics, setting (i.e. academic institution or other clinical site), type of study methodology (i.e. qualitative, quantitative, mixed, review), type of data analysis method (descriptive, constant comparison) and theory used to inform the research. Study quality was assessed with the Critical Appraisal Skills Program – UK tool checklist (CASP UK 2013). Since the focus of this review is to capture the assessors' personal insights and perceptions of failure to fail, it is not surprising that the review yielded mostly qualitative studies of focus groups or individual interviews ($N = 15$) and review papers ($N = 7$). The three quantitative studies reported frequencies of responses to specific survey questions on failure to fail. For the qualitative and review papers, the authors' engaged in a thematic analysis and developed a descriptive, theme-based classification system through several cycles of analysis, discussion and reflection by the whole research team. The survey questions in the quantitative studies were aligned with the theme-based classifications and therefore, their results were incorporated into the thematic analysis.

Results

Demographic

A descriptive summary of the demographics, setting, and type of study methodology of the searched papers is presented in Table 1 (available online on the Journal website as Supplemental Material). The majority of the failure to fail literature comes from Nursing (19/28) and Medicine (6/28). Very little demographic information describing the assessors is included in the publications. All qualitative studies were rated of very good quality since they met 8 or 9 of the 10 quality criteria of the CASP-UK checklist.

Thematic

We constructed six themes relating to the barriers assessors face when failing a poorly performing trainee. These were: (1) assessor's professional considerations, (2) assessor's personal considerations, (3) trainee related considerations, (4) unsatisfactory assessor development and assessment tools, (5) institutional culture and (6) consideration of available remediation for the trainee.

We constructed three themes relating to the factors enabling assessors to take on the challenge of failing underperforming students. These facilitators were: (1) duty to patients, to society, and to the profession, (2) institutional support such as backing a failing evaluation, support from colleagues, assessor development, and strong assessment systems, and (3) opportunities for students after failing.

We found that there were no differences in the themes reported across the medical, nursing, and dentistry literatures.

Barrier 1: assessor's professional considerations

One barrier commonly described across the reviewed papers involved assessors being mindful of the impact that failing a student had on them professionally. These professional considerations were so significant for the assessors that it was simply easier to pass underperforming trainees than to fail them. Failing a student involved extraordinary amounts of extra work and time, amounting to a significant deterrent to failing underperforming trainees (Dudek et al. 2005; Rutkowski 2007; Cleland et al. 2008; Luhanga et al. 2008; Carr et al. 2010; Luhanga et al. 2010; Watling et al. 2010; Earle-Foley et al. 2012; Fazio et al. 2013; Larocque & Luhanga 2013; Guerrasio et al. 2014). The time-consuming processes of documentation, discussion, and planned intervention moved assessors away from other important responsibilities:

> Registered nurses are required to prioritize clinical tasks to deliver care, comfort relatives and provide information for other professionals. As a consequence, a student's assessment may be deemed low priority. (Rutkowski 2007)

Assessors reflected that limited exposure to individual students combined with the low priority given to assessment work, translated into having insufficient information to properly assess the trainee's competence (Rutkowski 2007; Cleland et al. 2008; Gopee 2008; Luhanga et al. 2008, 2010; Deegan et al. 2012):

> The lack of (staff) continuity . . .resulted in a poorly performing student being graded as "satisfactory" for her first attempt at a procedure, and because she was evaluated by different members of staff, she continued to be graded as such. Thus her inability was not picked up. (Bush et al. 2013)

Further compounding this professional consideration was a fear of litigation (Williams et al. 2005; Cleland et al. 2008; Watling et al. 2010; Earle-Foley et al. 2012; Guerrasio et al. 2012; Bush et al. 2013) and unease with the appeal process (Dudek et al. 2005; Williams et al. 2005; Cleland et al. 2008; Jervis & Tilki 2011; Earle-Foley et al. 2012; Guerrasio et al. 2014; Larocque & Luhanga 2013). One paper suggested that fear of litigation was so great that assessors intentionally opted to delay assigning a failing grade for fear of repercussions (Duffy 2006). Indeed, the fear of legal consequences was a significant contributor to the failure to fail phenomenon:

> We feel that the "burden of proof" rests squarely with the program. . . .[We] are very hesitant to consider dismissal for fear that we will be dragged into court. (Guerrasio et al. 2014)

Assessors also reported concerns that failing a student would harm their professional standing. Assessors worried that they would be blamed for creating an uncomfortable learning environment, or would be labeled as a poor teacher (Rutkowski 2007). Furthermore, assessors were mindful that failing trainees could result in lower evaluations of their own performance (Rutkowski 2007; Jervis & Tilki 2011; Larocque & Luhanga 2013; Pratt et al. 2013), which could negatively impact their reappointment or tenure aspirations (Pratt et al. 2013):

> Before you know it, instead of the resident being judged, you are being judged. (Watling et al. 2010)

Lastly, assessors' professional considerations were not only self-oriented; they shared concerns that failing a student negatively affected the overall work environment (Duffy & Hardicre 2007; Rutkowski 2007).

Barrier 2: assessor's personal considerations

The reviewed articles reported that assessors felt a sense of personal failure and guilt when failing a trainee (Duffy 2006; Rutkowski 2007; Cleland et al. 2008; Luhanga et al. 2008; Earle-Foley et al. 2012; Larocque & Luhanga 2013; Pratt et al. 2013; Black et al. 2014). This self-blame was a barrier that was most apparent in less experienced assessors (Cleland et al. 2008):

> I think part of it is looking at yourself too because you are supposed to be getting this young nurse ready to step out into the professional world, and if she fails, maybe it's something you didn't do right. (Luhanga et al. 2008)

Assessors also struggled with the personal strain of conflicting responsibilities—that is, the responsibilities of supporting trainees, but also accurately assessing them (Cleland et al. 2008; Vezeau & McAllister 2009). As members of a profession dedicated to caring for others, assessors were reluctant to assign failing grades because that act could be viewed as uncaring behavior. (Earle-Foley et al. 2012; Pratt et al. 2013):

> Reporting underperformance was not seen as part of the broad, supportive culture, referred to as being "caring" by clinical tutors and "educating" by non-clinical tutors. (Carr et al. 2010)

Another aspect of assessors' personal considerations involved the close relationship that often develops between the trainee and the assessor (Cleland et al. 2008; Luhanga et al. 2008, 2010; Jervis & Tilki 2011; Bush et al. 2013; Guerrasio et al. 2014). Assessors genuinely liked the underperforming trainee (Guerrasio et al. 2014), wanted to be liked by trainees and colleagues (Cleland et al. 2008; Luhanga et al. 2008; Bush et al. 2013), or feared having to socialize with failed trainees in future situations (New South Wales Nurses Association 2010). Assessors also worried about being unfair in their assessments if they disliked an underperforming trainee (Cleland et al. 2008).

An additional barrier was the emotional toll involved in failing a trainee. Failing a trainee was described as an unpleasant, emotionally fraught experience (Duffy 2006; Duffy & Hardicre 2007; Gopee 2008; Cleland et al. 2008; Carr et al. 2010; Watling et al. 2010; Jervis & Tilki 2011; Bush et al. 2013; Larocque & Luhanga 2013; DeBrew & Lewallen 2014):

> You go home and go over it time and time again. Is this the right decision? Have I missed something? Is it a personality clash? It gives you sleepless nights! I'm not exaggerating! (Jervis & Tilki 2011)

Assessors were reluctant to engage in a conflict with the failing trainee (Cleland et al. 2008; Luhanga et al. 2008; Brown et al. 2012; Larocque & Luhanga 2013), or be challenged by the trainee (Cleland et al. 2008). They preferred to avoid dealing with the unhappy, angry trainees they would be failing (Fazio et al. 2013):

> The students cry, are upset and telling you X, Y, Z. It is like emotional blackmail. They put pressure on mentors to bend, not to refer, turn a blind eye and stuff like that. (Luhanga et al. 2008)

Barrier 3: trainee related considerations

The articles reviewed also reported assessor considerations related to the underperforming trainees themselves. Assessors weighed the decision to fail the trainee against the effect the failure would have on the trainee (Luhanga et al. 2008; Carr et al. 2010; Earle-Foley et al. 2012; Larocque & Luhanga 2013; Pratt et al. 2013; Black et al. 2014; Guerrasio et al. 2014). For example, they considered the impact of a failing grade on the trainee's career goals (Earle-Foley et al. 2012), and financial security (Cleland et al. 2008; Luhanga et al. 2008; Carr et al. 2010; Larocque & Luhanga 2013; Pratt et al. 2013; Guerrasio et al. 2014):

> For a student, failing a placement can be a significant loss experience: loss of self-esteem, time, education money, certification, and career. "I don't think any of us want to see somebody throw four years of their life out the window." (Larocque & Luhanga 2013)

The timing of the failure in relation to the trainee's career progression was another consideration. One paper described assessors wanting to allow the trainee to "exit medicine with dignity" (Carr et al. 2010). Assessors were reluctant to fail trainees early in their training believing that the trainee had time to improve; but they were equally reluctant to fail those whose training was advanced:

> There is reluctance because it is a third year and they have come that far. (Jervis & Tilki 2011)

Other trainee related considerations included concerns about trainees' emotional reaction and distress, (Gopee 2008; Brown et al. 2012; Bush et al. 2013; Larocque & Luhanga 2013; Pratt et al. 2013; Black et al. 2014; Guerrasio et al. 2014; Larocque & Luhanga 2013; Pratt et al. 2013), personal well-being and confidence (Bush et al. 2013), and self-esteem (Fazio et al. 2013; Larocque & Luhanga 2013). Assessors noted a desire to protect trainees (Bush et al. 2013), feeling that failure could be "very destructive" (Watling et al. 2010) and "stigmatizing" (Carr et al. 2010):

> She clearly so desperately wanted it ... her life depended on her passing this placement and she would do anything she could to pass it ... you can't help but be affected by those issues! (Black et al. 2014)

Assessors were less likely to report underperformance when the trainee was perceived as being aware of their difficulties, committed to the health profession, and/or actively trying to improve (Carr et al. 2010; Larocque & Luhanga 2013).

Barrier 4: – unsatisfactory assessor development and/or evaluation tools

Assessors reported feeling unprepared, lacking experience and lacking confidence in their evaluation role (Cleland et al. 2008; Luhanga et al. 2008; Carr et al. 2010; Luhanga et al. 2010; Brown et al. 2012; Deegan et al. 2012; Heaslip & Scammell 2012; Earle-Foley et al. 2012; Bush et al. 2013; Fazio et al. 2013; Pratt et al. 2013). This resulted in giving "the benefit of the doubt to students who were less than competent" (Luhanga et al. 2010). Assessors doubted their own judgment or ability, often questioning if they had the skills to evaluate appropriately (Cleland et al. 2008), or to manage angry or upset trainees (Dudek et al. 2005; Earle-Foley et al. 2012; Bush et al. 2013):

> The dread that skills would not be good enough to manage a situation where a student was angry, upset or disagreed with their assessment were evident. The temptation not to fail under these circumstances is acknowledged. (Carr et al. 2010)

Assessors reported having little formal training in evaluating trainees (Rutkowski 2007; Bush et al. 2013; Fazio et al. 2013). Staff who had not received adequate training "might not appreciate the consequences of passing underperforming students who could be helped by remediation or by being failed" (Bush et al. 2013). Over half the assessor participants in Heaslip's study wished for more education on managing failing students (Heaslip & Scammell 2012). Assessors expressed difficulty with the inherent inconsistencies in subjective clinical evaluations, (Vezeau & McAllister 2009; Jervis & Tilki 2011; Brown et al. 2012; Deegan et al. 2012; Earle-Foley et al. 2012; Bush et al. 2013; Guerrasio et al. 2014;) and discomfort with the skill of identifying the specific behaviors that warranted a failing grade, such as non-cognitive skills, attitudes, (Luhanga et al. 2008; Jervis & Tilki 2011) or professionalism (Bush et al. 2013; Guerrasio et al. 2014):

> Professional behavior occurs along a spectrum and the point at which a student is no longer competent is not clearly defined. (Guerrasio et al. 2014)

Assessors struggled with uncertainty about the expected standards for trainees at different levels of the educational continuum (Bush et al. 2013), and expressed concern over

the lack of appropriate, objective or explicit evaluation tools (Williams et al. 2005; Duffy 2006; Rutkowski 2007; Luhanga et al. 2008; Luhanga et al. 2010; Jervis & Tilki 2011; Brown et al. 2012; Bush et al. 2013). There was a sense that the grading system could be unfair, and assessors did not want to draw undeserved negative attention to struggling trainees (Bush et al. 2013; Fazio et al. 2013; Larocque & Luhanga 2013). Furthermore, being unaware of the processes for reporting and failing a trainee played a role in failing to fail:

> We weren't aware of the procedures... that if we had a weak student that we had to contact the university early enough... when we did contact them it was too late to fail the student. (Duffy & Hardicre 2007)

There was also a recognition that failure to fail could stem from a lack of supporting documentation (Dudek et al. 2005; Williams et al. 2005; Cleland et al. 2008; Bush et al. 2013; Guerrasio et al. 2014):

> Often we don't do a good enough job of recording performance. Then if the student challenges you, you do not have a leg to stand on because you cannot recall specific incidents. (Dudek et al. 2005)

Barrier 5: institutional culture

In addition to considerations at the individual level, assessors described barriers at the institutional culture level. In making decisions about reporting underperforming trainees, assessors considered the current shortages of health professionals (Rutkowski 2007; Luhanga et al. 2008; Earle-Foley et al. 2012), and the institution's potential loss of financial support (Cleland et al. 2008; Larocque & Luhanga 2013). Pressure from the institution concerned with its reputation was also mentioned (Cleland et al. 2008; Fazio et al. 2013):

> There was also perceived pressure from the university to pass students at one of the schools; this was associated with issues of finance and reputation, accompanied by the belief that faculty with high failure rates would deter potential students. (Cleland et al. 2008)

Lack of support from the institution was reflected in its endorsement, be it implicit or explicit, of allowing failing trainees to progress or of grade inflation (Gopee 2008; Brown et al. 2012; Earle-Foley et al. 2012; Bush et al. 2013; Guerrasio et al. 2014):

> They [participants] were particularly concerned about the current lack of discussion that, according to one senior academic, meant there was no coherent approach so that some failing students continued to progress. (Duffy 2006; Bush et al. 2013)

Assessors commonly expressed feelings of anger and disappointment directed towards previous mentors for not failing the trainee or for failing to act upon concerns:

> I feel that it's kind of all been dumped on me and I'm picking up the pieces!...because previous mentors had failed to do so [to fail the trainee]. (Black et al. 2014)

This failure to act may be present if assessors are not aware of where the responsibility for failure lies (Vezeau & McAllister 2009), or if they feel the decision to fail is not theirs to make (Rutkowski 2007; Gopee 2008; Vezeau & McAllister 2009; Fazio et al. 2013)

Assessors suggested that part of the justification for passing a less-than-competent trainee was pressure from the academic institution to pass or the institution's history of overturning decisions to fail (Duffy 2006; Rutkowski 2007; Brown et al. 2012; Jervis & Tilki 2011; Larocque & Luhanga 2013; Pratt et al. 2013; Black et al. 2014; DeBrew & Lewallen 2014; Guerrasio et al. 2014). The latter generated a sense of betrayal and loss of trust in the university:

> It's frustrating. You think "I'm not taking another student... You know my opinion is not valued. I'm telling you this person is unsafe to be out there." And it's just not listened to. (Larocque & Luhanga 2013)

Barrier 6: consideration of available remediation for the trainee

Assessors' reluctance to fail a trainee reflected concern about the availability of remediation options available to the trainee (Dudek et al. 2005; Duffy & Hardicre 2007; Bush et al. 2013; Fazio et al. 2013; Guerrasio et al. 2014):

> Many participants felt that they could not fail a trainee if remediation was not available to them. They felt it was their responsibility to provide remediation and, if unable to do so, they might not fail that person. (Dudek et al. 2005)

Assessors reported dissatisfaction with the available remediation options, lamenting that "the remedial interview with a senior academic and being told to "pull their socks up" is not a sufficient response" (Bush et al. 2013), and that timing of remediation was important since assessors wanted to ensure that there would be sufficient time for remediation or that "a student could exit medicine with dignity" (Cleland et al. 2008).

Enabler 1: duty to patients, to society, and to the profession

Responsibility towards patient safety, duty to the public, and moral integrity were leading themes in supporting assessors in failing an underperforming trainee (Dudek et al. 2005; Scholes & Albarran 2005; Cleland et al. 2008; Bush et al. 2013; Carr et al. 2010; Luhanga et al. 2010; Watling et al. 2010; Jervis & Tilki 2011; Earle-Foley et al. 2012; Pratt et al. 2013; Black et al. 2014).

> Participants identified a sense of responsibility as the main motive to fail a trainee: to the public to ensure safety, to the profession to protect its reputation, and to the trainee to allow them the opportunity for remediation. (Dudek et al. 2005)

Assessors who failed underperforming trainees felt duty bound to the profession to uphold standards (Rutkowski 2007; Luhanga et al. 2008; Carr et al. 2010; Fitzgerald et al. 2010; Pratt et al. 2013). Assessors are gatekeepers to the profession (Fitzgerald et al. 2010; Black et al. 2014), and so must demonstrate professional accountability (Rutkowski 2007). As such, not acting ethically on encountering or detecting underachieving students is not condonable (Duffy & Hardicre 2007):

> Apart from protecting the student and patients, there was an overpowering sense of obligation to protect the nursing profession from incompetent and unsafe practitioners. Failing to fail an incompetent student was seen as putting the profession into disrepute, a belief that united the experience: You've got to uphold the profession ... I'm proud to be a registered nurse ... to belong to a profession that I'm very proud of, and I want that profession to have standards. (Black et al. 2014)

Moral integrity emerged as a powerful force driving assessors to fail a failing trainee (Scholes & Albarran 2005; Gopee 2008; Pratt et al. 2013; Black et al. 2014):

> The strength to fail was characterized by having the emotional and psychological integrity to manage their moral stress and demonstrate moral integrity: "I think sometimes it takes somebody quite tough to actually fail somebody and I appreciate that. (Black et al. 2014)

Enabler 2: support from the institution, from colleagues, and assessor development

Institutional endorsement for failing a trainee was an important contributor for enabling assessors to fail an underperforming trainee (Scholes & Albarran 2005; Brown et al. 2012; DeBrew & Lewallen 2014). Assessors felt supported if they could consult with other colleagues who may have made similar judgments and could confirm or support their observations, particularly if the trainee exhibited a pattern of behavior, rather than an isolated incident (Carr et al. 2010; Luhanga et al. 2010):

> If you see a trail of destruction, you have confidence because someone else had a problem. It is not just me. I can handle this. (Jervis & Tilki 2011)

The institution was described as a valuable source of support for assessors when the institution provided assessor development to recognize poor performance and to be able to describe what was observed (Scholes & Albarran 2005):

> Mentors need adequate preparation during mentorship training for dealing with the potentially failing student... we need to acknowledge our responsibility. We are not clear enough about their [evaluator] role, the processes involved, basically what to do if you have a problem student, how to deal with it... I think we need to address it in our mentorship programmes. (Duffy 2006)

A strong assessment system with established criteria further clarified and supported assessors' decision to fail, and helped to build confidence in judgment decisions (Duffy 2006; Carr et al. 2010; Brown et al. 2012; Bush et al. 2013). Clear institutional policies and procedures, administrative support and recommendations on how to manage a failure (Carr et al. 2010) supported assessors and may further minimize the potential distress from failing a trainee (Pratt et al. 2013):

> The student made legal threats...this lecturer was aware of the tremendous pressure placed on the mentor and asserted that if it had not been for the support... this mentor may well have "buckled under the pressure" and given a satisfactory assessment. (Duffy 2006)

Enabler 3: opportunities for trainees after failing

Assessors indicated that offering support and guidance to the failing trainee, offering them time to reflect on their performance and to grow from the failure experience, enabled the assessors to come to terms with the difficult task of failing a student (Dudek et al. 2005; Gopee 2008; Carr et al. 2010; Larocque & Luhanga 2013; Pratt et al. 2013):

> Despite the negative association with failing it was agreed that maximizing the positive elements that facilitated the student succeeding in future is essential. (Carr et al. 2010)

Discussion

This systematic review consolidates our understanding of the barriers that contribute to the failure to fail phenomenon, and the enablers that support overcoming this challenge. It highlights the personal and professional turmoil faculty members face when failing an underperforming trainee, and the important role the institution plays in setting a climate that supports faculty members who anticipate, are in the process of, or in the aftermath of failing a trainee. Assessors feel unprepared or unsure of what to document, how to document it, how to articulate subjective impressions of a failing performance. Assessors are also unsure of how their assessment fits into their institution's process for addressing a failing trainee. The literature clearly suggests that an assessor's sense of duty and moral integrity along with institutional support and opportunities for trainee remediation can help overcome the barriers reported. Though our observations are based on relatively few reports, the decisive factor of duty to patients and the profession merits further exploration. We found that these barriers and enablers are universally reported across health professions and trainee levels, and have been consistently reported over the past decade.

With this understanding of the "failure to fail" phenomena, we can now move forward to find interventions and solutions to these barriers. Some of this work has begun. For example, Dudek et al. designed faculty development interventions to teach assessors how to articulate their concerns on end of rotation assessment forms (Dudek et al. 2013; Dudek & Dojeiji 2014;). Others have called for more explicit assessor training in delivering difficult messages to underperforming trainees, (Jervis & Tilki 2011) and for investing support systems to address the mental fatigue experienced when failing a trainee (Carr et al. 2010; Pratt et al. 2013). Others have called for more explicit training of educators in ethical competence to uphold the standards and ethics of their profession, as a significant enabler of failing a trainee is a deeply rooted sense of duty (Earle-Foley et al. 2012; Gopee 2008; Luhanga et al. 2010; Black et al. 2014). In the UK, the nursing profession has implemented formalized training and education of mentors (Duffy 2006; Nursing and Midwifery Council (UK) 2008)

It may be beneficial for institutions to acknowledge that not all students will or should graduate from their health professions training. Since no admissions system is perfect, it stands to reason that there will be trainees who will fail to meet expectations or standards of competency. Institutions can support assessors by offering faculty development aimed at preparing assessors for the inevitability of failing a student.

On a larger scale, governing bodies in both undergraduate and graduate medical education have embraced programs that create more explicit statements of trainee expectations such as the milestones (Swing et al. 2013) and entrustable professional activities (Association of American Medical Colleges 2014). Initiatives such as these have the potential to create a shared mental model across assessors

of what a trainee should be able to do. With clearer expectations of performance, it may be easier for supervisors to report clinical performance that is not meeting standards.

Conclusion

To our knowledge, this is the first systematic review of the medical, nursing, and dentistry education literature that consolidates the available knowledge on the failure to fail phenomenon. The challenge our community now faces is determining how to best tackle the identified barriers and to harness the power of enablers to resolve the failure to fail problem.

Disclosure statement

The views expressed herein are those of the authors and do not necessarily reflect those of the United States of America's Department of Defense or other federal agencies.

Funding

This work was supported by Society of Directors of Research in Medical Education.

Glossary

Assessor: broad term to include individuals who supervise and teach trainees in clinical practice and are responsible for their assessment; the term includes faculty, preceptor, attending, clinical mentor, supervisor, evaluator.

Notes on contributors

Monica Yepes-Rios, MD, FACP, is an Assistant Professor at Department of Medicine, Uniformed Services University of the Health Sciences. She has research interests in communication skills and assessment and remediation in undergraduate and graduate medical education.

Nancy Dudek, MD, MEd, is an Associate Professor in the Faculty of Medicine at the University of Ottawa.

Rita DuBoyce, Col, USAF, MC, is the Senior Director of Clinical Education at the Val G. Hemming Simulation Center. Col DuBoyce has been affiliated with the Center since 2009 and is deeply involved in simulation education programs at the Uniformed Services University, particularly programs within the Department of Medicine.

Jerri Curtis, ***MD,*** is currently the Associate Dean for GME at the Uniformed Services University of the Health Sciences. She also serves as the Executive Director and Designated Institutional Official for the National Capital Consortium. She recently retired from the US Navy after a 30-year career. Dr. Curtis is board certified in Neonatal-Perinatal Medicine.

Rhonda J. Allard, MLIS has more than 20years of medical library experience. Her area of expertise includes consumer health, clinical librarianship, and evidence-based medicine instruction. She holds a Masters in Library and Information Science from the University of Pittsburgh and a Health Information Technology Certificate from Weill Cornell Medical School.

Lara Varpio, PhD, is an Associate Professor at Uniformed Services University of the Health Sciences. Her program of research investigates the many kinds of teams involved in health professions education (e.g., inter-professional clinical care teams, health professions education scholarship units, teams of research collaborators, etc.). She uses qualitative methodologies, informed with theories from the social sciences and humanities, to build practically relevant theory-based knowledge.

ORCID

Monica Yepes-Rios http://orcid.org/0000-0001-8270-8608
Rita Duboyce http://orcid.org/0000-0002-7710-0064
Rhonda J. Allard http://orcid.org/0000-0001-9021-3672

References

Association of American Medical Colleges. 2014. Core entrustable professional activities for entering residency (CEPAER); [cited 2015 Jan 14]. Available from: www.aamc.org/cepaer.

Black S, Curzio J, Terry L. 2014. Failing a student nurse: a new horizon of moral courage. Nursing Ethics. 21:224–238.

Brown L, Douglas V, Garrity J, Shepherd CK. 2012. What influences mentors to pass or fail students. Nurs Manag (Harrow). 19:16–21.

Bush HM, Schreiber RS, Oliver SJ. 2013. Failing to fail: clinicians' experience of assessing underperforming dental students. Eur J Dent Educ. 17:198–207.

Carr J, Heggarty H, Carr M, Fulwood D, Goodwin C, Walker WW, Whittingham K. 2010. Reflect for success: recommendations for mentors managing failing students. Brit J Commun Nurs. 15:594–596.

CASP UK. 2013. CASP Checklists; [cited 2015 Jan 14]. Available from: http://www.casp-uk.net/#!casp-tools-checklists/c18f8.

Cleland J, Leggett H, Sandars J, Costa MJ, Patel R, Moffat M. 2013. The remediation challenge: theoretical and methodological insights from a systematic review. Med Educ. 47:242–251.

Cleland JA, Knight LV, Rees CE, Tracey S, Bond CM. 2008. Is it me or is it them? Factors that influence the passing of underperforming students. Med Educ. 42:800–809.

Cohen GS, Blumberg P, Ryan NC, Sullivan PL. 1993. Do final grades reflect written qualitative evaluations of student performance? Teach Learn Med. 5:10–15.

Cohen GS, Henry NL, Dodd PE. 1990. A self-study of clinical evaluation in the McMaster clerkship. Med Teach. 12:265–272.

DeBrew JK, Lewallen LP. 2014. To pass or to fail? Understanding the factors considered by faculty in the clinical evaluation of nursing students. Nurse Education Today. 34:631–636.

Deegan J, Burton T, Rebeiro G. 2012. Clinical assessment and the benefit of the doubt: What is the doubt? Austr J Adv Nurs. 30:42–48.

Dudek N, Dojeiji S. 2014. Twelve tips for completing quality in-training evaluation reports. Med Teach. 36:1038–1042.

Dudek NL, Marks MB, Bandiera G, White J, Wood TJ. 2013. Quality in-training evaluation reports-does feedback drive faculty performance? Acad Med. 88:1129–1134.

Dudek NL, Marks MB, Regehr G. 2005. Failure to fail: the perspectives of clinical supervisors. Acad Med. 80:S84–S87.

Duffy K. (2006) Weighing the balance: a grounded theory study of the factors that influence the decisions regarding the assessment of students' competence in practice. Thesis. Scotland, UK: Glasgow Caledonian University.

Duffy K, Hardicre J. 2007. Supporting failing students in practice. 1: Assessment. Nurs Times. 103:28–29.

Earle-Foley V, Myrick F, Luhanga F, Yonge O. 2012. Preceptorship: using an ethical lens to reflect on the unsafe student. J Prof Nurs. 28:27–33.

Fazio SB, Papp KK, Torre DM, Defer TM. 2013. Grade inflation in the internal medicine clerkship: a national survey. Teach Learn Med. 25:71–76.

Fitzgerald M, Gibson F, Gunn K. 2010. Contemporary issues relating to assessment of pre-registration nursing students in practice. Nurse Educ Pract. 10:158–163.

Gopee N. 2008. Assessing student nurses' clinical skills: the ethical competence of mentors. Int J Therapy Rehabil. 15:401–407.

Guerrasio J, Garrity MJ, Aagaard EM. 2014. Learner deficits and academic outcomes of medical students, residents, fellows, and attending physicians referred to a remediation program, 2006–2012. Acad Med. 89:352–358.

Hatala R, Norman GR. 1999. In-training evaluation during an internal medicine clerkship. Acad Med. 74:S118–S120.

Heaslip V, Scammell JM. 2012. Failing underperforming students: the role of grading in practice assessment. Nurse Educ Pract. 12:95–100.

Horvath K. 2010. To fail or not to fail: that is the question. Arch Surg. 145:857

Jervis A, Tilki M. 2011. Why are nurse mentors failing to fail student nurses who do not meet clinical performance standards? Br J Nurs. 20:582–587.

Larocque S, Luhanga FL. 2013. Exploring the issue of failure to fail in a nursing program. Int J Nurs Educ Scholarsh. 10:1–8.

Luhanga F, Myrick F, Yonge O. 2010. The preceptorship experience: an examination of ethical and accountability issues. J Prof Nurs. 26:264–271.

Luhanga F, Yonge OJ, Myrick F. 2008. "Failure to assign failing grades": issues with grading the unsafe student. Int J Nurs Educ Scholarsh. 5: Article8

Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 151:264–269, w264.

New South Wales Nurses Association. 2010. Nurse mentors still unwilling to fail students. Lamp (Sydney, Australia). 67:13.

Nursing and Midwifery Council (UK). 2008. The code: professional standards of practice and behaviour for nurses and midwives; [cited 2015 Jan 14]. Available from: https://www.nmc.org.uk/standards/code/.

Pratt M, Martin L, Mohide A, Black M. 2013. A descriptive analysis of the impact of moral distress on the evaluation of unsatisfactory nursing students. Nurs Forum. 48:231–239.

Rutkowski K. 2007. Failure to fail: assessing nursing students' competence during practice placements. Nurs Stand. 22:35–40.

Scholes J, Albarran J. 2005. Failure to fail: facing the consequences of inaction. Nurs Crit Care. 10:113–115.

Speer AJ, Solomon DJ, Ainsworth MA. 1996. An innovative evaluation method in an internal medicine clerkship. Acad Med. 71: S76–S78.

Swing SR, Beeson MS, Carraccio C, Coburn M, Iobst W, Selden NR, Stern PJ, Vydareny K. 2013. Educational milestone development in the first 7 specialties to enter the next accreditation system. J Grad Med Educ. 5:98–106.

Tonesk X, Buchanan RG. 1987. An AAMC pilot study by 10 medical schools of clinical evaluation of students. J Med Educ. 62:707–718.

Vezeau TM, McAllister LE. 2009. The experience of nursing faculty with students who fail. Commun Nurs Res. 42:235–235.

Watling CJ, Kenyon CF, Schulz V, Goldszmidt MA, Zibrowski E, Lingard L. 2010. An exploration of faculty perspectives on the in-training evaluation of residents. Acad Med. 85:1157–1162.

Williams RG, Dunnington GL, Klamen DL. 2005. Forecasting residents' performance – partly cloudy. Acad Med. 80:415–422.