BMC Medical Education

RESEARCH

Open Access

Why do instructors pass underperforming students? A Q-methodology study

Chunlin Liu[1†], Jananey Rajagopalan[2†], Bruce Wainman[2], Sarah Wojkowski[3], Joanna Pierazzo[4] and Noori Akhtar-Danesh[4*]

Abstract

Background Formal evaluations are an integral part of a student's learning and encourage students to learn and help instructors identify students' weaknesses. Over the past few decades there have been growing concerns that instructors and evaluators are passing students who do not meet expectations. This phenomenon, in which instructors pass students who do not meet expectations, has been referred to as "failure-to-fail". In this study, we used Q-methodology to identify instructors' justifications for failure-to-fail.

Methods A Q-methodology study was conducted to identify the major viewpoints of instructors at a Canadian university. A by-person factor analysis with principal component factor extraction and Varimax rotation was used. The analysis was conducted using the QFACTOR program in Stata. A Cohen's effect size of 0.80 was used to identify distinguishing statements.

Results Fifty seven instructors participated in this study. Through a by-person factor analysis, three factors representing three viewpoints emerged: Intrinsically Motivated, Extrinsically Motivated, and Administratively & Emotionally Deterred. The Intrinsically Motivated group perceived mental barriers that prevented them from failing students. They strongly disagreed that they experienced pressure from either students or their schools to pass students. The Extrinsically Motivated believed that their higher-ups and the university encouraged them to pass all students. They perceived discomfort associated with defending their reasons for failing students and were concerned that failing students would damage their own career advancements. The Administratively & Emotionally Deterred group believed that the process of failing a student was stressful and exhausting. They disagreed that a failed student is a result of the instructor's own inadequate guidance or mentorship.

Conclusions This study identified three distinctive viewpoints that outline areas of consideration for addressing the failure-to-fail mechanism. More transparent discussions within schools, as well as identifying solutions, are required to create systems that ensure educational and professional standards are maintained. Further replication of this study in various disciplines may be used to determine whether these findings are consistent in different fields.

Keywords Q-methodology, Failure-to-fail, By person factor analysis, Evaluation

[†]Chunlin Liu and Jananey Rajagopalan contributed equally to this work as the first authors.

*Correspondence:
Noori Akhtar-Danesh
daneshn@mcmaster.ca
Full list of author information is available at the end of the article

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

Formal evaluations are an integral part of a student's learning. They encourage students to learn and help instructors identify students' weaknesses so that learning can be tailored to address these concerns [1]. However, there have been growing concerns over the past decades that instructors and evaluators are passing students who do not meet expectations—both in the academic and clinical course context [2–6]. The phenomenon in which instructors pass students who do not meet all expectations of clinical performance has been referred to in the literature as "failure-to-fail" [3, 6].

The term "failure-to-fail" (FTF) was first used by Lankshear [7] to describe the phenomenon that English nursing tutors were passing nursing students who did not demonstrate all required competencies. Since then, FTF has been observed and studied in other health disciplines such as medicine, physiotherapy, midwifery, and social work [5, 8–11]. Within these studied disciplines, FTF appears to be widespread among instructors. In one study, Hauge et al. [12] reported that up to 58.2% of nursing instructors had passed an underperforming second-year nursing student. FTF is also a global phenomenon and has been observed in healthcare educational programs based in Canada, Australia, Indonesia, the United Kingdom, and the United States [11, 13–16].

Additional evidence for FTF comes from the growing concern of grade inflation observed across educational institutions. In 2004, Scanlan & Care [17] identified in a Canadian school of nursing that up to 80% of the class received a grade of A or A + in a theoretical course, while only 3% of the class received a grade of B or lower. In 2008, Docherty & Dieckmann [18] surveyed full-time and part-time nursing faculty based in the United States and showed that 43% of respondents had given nursing students a higher grade than they deserved at least once. In addition, the recent global COVID-19 pandemic introduced unprecedented changes to all fields of education, which led to reports of further grade inflation in undergraduate courses [19, 20].

Both personal and institutional factors contribute to FTF. On a personal level, delivering bad news to others can be distressing [21, 22]. This tendency to avoid delivering bad news is known as the "MUM Effect," which, in an educational context, manifests as a reluctance to fail students [11, 21]. The MUM Effect describes the hesitation to communicate negative information, driven by the discomfort of delivering bad news and a desire to maintain a positive relationship with others [22, 23]. At an institutional level, instructors may pass an underperforming student to avoid negative evaluations or further escalation. Negative evaluations from students can cause distress for instructors [24]. Additionally, university administrators often use student feedback to assess an instructor's performance. To prevent receiving a negative evaluation, instructors might inflate grades and pass underperforming students [25].

A 2016 systematic review on FTF in medicine, nursing, and dentistry identified common barriers and enablers that instructors experience when assessing a student [6]. The authors identified six broad barriers: (1) assessor's professional considerations (i.e., increased workload, lower instructor evaluations), (2) assessor's 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 current literature outlines the various barriers instructors face when deciding whether to fail a student. However, instructors may be more influenced by specific barriers depending on their unique contexts, backgrounds, and experiences. For instance, one instructor might be more affected by institutional culture, while another might find the administrative process of failing a student more challenging. The literature lacks detailed profiles of different instructors and their unique reasons for failing to fail. To address this gap, research should better describe and categorize instructors' mindsets regarding FTF, while considering their diverse experiences and motivations. By categorizing instructors into different FTF profiles based on which barriers are most relevant to them, the literature can more accurately identify the unique barriers most pertinent to each instructor. A more descriptive profile for instructor mindsets would also allow institutions to implement more precise changes to better address FTF.

To date, most studies investigating FTF have used either a Likert scale or qualitative approaches such as interviews [2, 15, 26, 27]. While these approaches provide evidence to support the prevalence of FTF and identify a thematic consensus among instructors, they have limitations. First, a qualitative approach provides an overview of all factors that contribute to FTF but does not provide insight into which factor is most pertinent to instructors. Second, qualitative assessments such as interviews can be prone to biases. For example, during interviews, the interviewee may alter some of his or her response or omit some perspectives due to social desirability tendencies, particularly when answering sensitive questions [28]. The interviewees provide open-ended responses that may not fully address the questions, and their responses can vary between individuals, making it difficult to identify any general themes or trends [29]. The use of Likert scale assessments also has limitations. The mean and median are typically used to interpret results from Likert data, and although the mean and median are intuitive, they do not reflect the entire breadth and diversity of responses

[30, 31]. To overcome the aforementioned limitations, Q-methodology may be an approach that could be used to study the phenomenon of FTF [32].

Q-methodology combines qualitative and quantitative approaches to study subjectivity [33]. Previously, Q-methodology has been successfully applied to investigate the barriers to technology uptake and course evaluation [32, 34, 35]. However, there is no current published work that uses Q-methodology to study FTF. When studying sensitive topics like FTF, where participants may fear judgment or embarrassment, this research approach can be particularly valuable. Using Q-methodology, Instructors' responses can be categorized using a framework, and Q-study results offer a systematic method for understanding instructors' thought patterns and concerns related to FTF. The current study employed Q-methodology to identify patterns, themes, and potential rationales among nursing and rehabilitation sciences instructors regarding FTF.

Materials and methods

In this section, we first provide a brief review of Q-methodology and then describe the different steps of our study based on a Q-methodology framework.

An overview of Q-methodology

As a combination of qualitative and quantitative methods, Q-methodology was introduced in 1935 by William Stephenson [36]. Usually, the main objective in a Q-methodology study is to identify patterns of thought among study participants, not to estimate their numerical distributions. The methodology is used to explore human perceptions and interpersonal relationships by identifying similarities and differences in perceptions between groups [37]. The different stages of a Q-methodology study include identifying all possible statements related to the topic of the study (concourse), selecting a representative list of statements from the concourse (known as a Q-sample), designing a grid (Q-sort table) for data collection, and analysis and interpretation [33].

The concourse can be assembled from the literature, previous Q-studies, or by collecting statements from potential study participants. A representative sample of the concourse, known as the Q-sample, is then selected, and a grid (Q-sort table) with quasi-normal distributions is developed for data collection (see Fig. 1 for an example). The number of cells in the Q-sort table equals to the number of statements in the Q-sample. The Q-sort table includes a rating scale across the top that depending on the number of the statements in the Q-sample can vary, for example, from −3 to + 3 to −6 to + 6. The range and distribution of the Q-sort table are quite arbitrary, have a negligible effect on the results of the study, and can be altered for the convenience of the participants [38]. The Q-sort table is used for data collection and each completed Q-sort table is known as a Q-sort. The statements and the Q-sort table are provided to participants, and they are asked to rank order (sort) each statement relative to the other statements. Using the Q-sort table, participants rank statements from those they most agree with to those they most disagree with.

As mentioned, in a Q-methodology study, the main objective is to identify the range of opinions, not their numerical distributions. Therefore, sample size is not a determining factor and low response rates do not bias the study results [39]. Usually, a sample size of 40–60 participants is sufficient for meeting the statistical requirements of Q-methodology [38]. The quantitative component of a Q-methodology study includes a by-person factor analysis of the Q-sorts to classify (factorize) participants into different groups, so that each factor includes participants with similar views or perceptions regarding the topic of

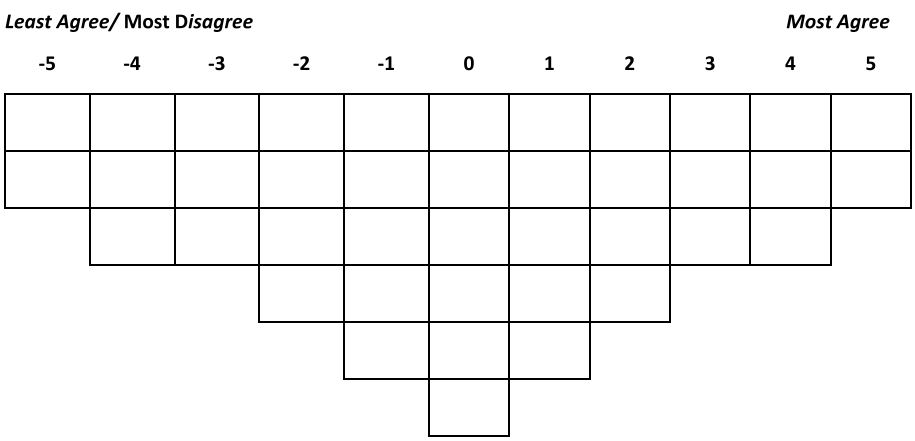

Fig. 1 Q-sort table with 40 cells used for data collection

the study. Furthermore, each factor is typically described based on its distinguishing statements. A statement for each factor is identified as a distinguishing statement if its score for the factor is significantly different from its scores from the other factors [33].

Face validity of a test or scale indicates how appropriate it is to assess the concept that it desires to measure [40, 41]. The face validity of statements used in a Q-methodology study is assessed by using the exact wording of the statements from participants and the literature. However, the statements are slightly edited for grammar and readability [42]. The content validity of statements is assessed by domain experts [42]. The Q-sorting process provides an opportunity for participants to express their inner subjective views, and there is no external criterion to evaluate or judge an individual's response or feeling to a statement [39]. Therefore, a participant's completed Q-sort is regarded as a valid expression of his or her view. In addition, several studies reported high test–retest reliability (≥ 0.80) for the Q-sorting process [38, 43].

Stages of Q-methodology in the current study

Concourse and Q-sample development

The first step of Q-methodology is to develop a series of statements that reflect viewpoints on a topic [33]. To survey the understanding of FTF, a comprehensive literature review was conducted, including previously published case studies, surveys, and reviews [6, 9, 13, 15, 26, 27, 44–46]. After reviewing the existing literature, a list of 82 statements, known as the concourse, was collected. These statements were reviewed by the research team to identify similarities and differences, including repeating themes or unclear statements [33]. Through continual discussion between the authors, the initial list was decreased to a final sample of 40 statements. The final set of statements is referred to as the Q-sample and can be found in Additional file 1. These statements represented various broad viewpoints on the topic of FTF. A five-question demographic questionnaire investigating age, gender, school, years of experience and class size was also developed to better understand the potential influence of these variables on attitudes toward FTF.

Participants

Participants were recruited by email from the School of Nursing (SON) and the School of Rehabilitation Science (SRS) at a Canadian university in 2022. Potential participants, including full-time and part-time faculty members, and clinical instructors from both schools were recruited through email. The respondents were required to have performed classroom and/or clinical instruction activities at any point in the curriculum, with no minimum number of hours or previous teaching experience. The email contained a secure link with the consent form, Q-sample, Q-sort table, instructions, and a short demographic questionnaire. Participation in the study was voluntary and confidential. Participants were offered an opportunity to enter a draw for one of two $50 gift cards as compensation for their participation.

The Q-sort table and data collection

In this study, we developed a Q-sort table with 40 cells for data collection (Fig. 1). The Q-sample, Q-sort table and instruction page were presented to the participants. Participants were first asked to separate the statements into three sections: statements they agreed with, disagreed with, and a section for the rest (neutral). Any number of statements could be placed in these three sections. Participants were then instructed to sort the statements from the three sections into the Q-sort table (Fig. 1), ranking Q-sample statements relative to each other (based on degree of agreement/disagreement). A statement placed in a negative number cell suggested that the participant disagreed (or agreed less) with the statement; and a statement placed in a positive cell suggested agreement. Each of the 40 statements had to be assigned to a single cell. Participants were also asked to provide optional qualitative feedback to justify their most extreme rankings (e.g., -5 and + 5) (Additional file 2).

Data analysis/interpretation

After the data collection, a by-person factor analysis was performed on the Q-sorts using the "qfactor" command in Stata [47]. Unlike ordinary factor analysis, which is based on the correlation between variables, traits, or statements, a by-person factor analysis is based on the correlations between participants, represented by Q-sorts in Q-methodology [42]. Q-sorts that correlate significantly with each other form a group, known as a "factor", resulting in several factors of similar Q-sorts. Therefore, each factor represents a group of individuals with similar views, feelings, or experiences in relation to the theme of the study. We used a principal component factor extraction with Varimax rotation to identify the factors. The factor scores, as z-scores, for each factor were calculated using a regression method to determine the score for each statement. These scores were then converted back to the original Q-sort format from -5 to + 5. For example, the two statements with the highest scores were assigned + 5, the next three highest were assigned + 4, and so forth. Factors are interpreted based upon their distinguishing statements and consensus statements (statements with similar scores between factors) as well as qualitative and demographic data. A Cohen's effect size of 0.80 was used for identifying the distinguishing statements [48].

Table 1 Summary of Participant Demographics ($n=57$)

Variable	*n* (%)
Age	
< 30 years	6 (11.8)
30 -39 years	15 (29.4)
40 -49 years	17 (33.3)
≥ 50 years	13 (22.8)
Unreported	6 (10.5)
Years of experience	
< 10 years	28 (49.1)
≥ 10 years	23 (40.4)
Unreported	6 (10.5)
Class size	
≤ 10	18 (31.6)
11–50	13 (22.8)
51–100	12 (21.0)
> 100	6 (10.5)
Unreported	8 (14.0)

57 participants responded to the demographic questions

Ethical considerations

Ethical approval was obtained from the Hamilton Integrated Research Ethics Board (HiREB# 15,784). We confirm that all methods were carried out in accordance with relevant guidelines and regulations. The participants provided informed consent to participate in this study. Participation in the study was optional, confidential, and anonymous. At the end of the survey, the participants were offered an opportunity to enter a draw for one of two $50 gift cards as compensation for their participation.

Results

In total, 57 instructors from the School of Nursing and the School of Rehabilitation Sciences responded to the survey. Participant's age ranged from less than 30 years old ($n=6$) to older than 50 years old ($n=13$). Twenty-eight participants had more than 10 years of teaching experience, and 23 participants had 10 or less years of experience. The class sizes of the participants ranged from small (less. than or equal to ten) to large (more than 100 students). Eighteen participants reported that they taught small classes with fewer than or equal to 10 students, while 6 participants reported that they taught large classes with more than 100 students. For a more thorough breakdown of demographic information, see Table 1.

Factors

Using a by-person factor analysis, three factors, or 'salient viewpoints', emerged. This analysis loaded participants into individual factors by associating their responses with the themes most strongly represented by a single factor. The three factors encompassed 51 participants, and 6 participants did not load significantly on any of these three factors and were excluded from further analysis. The three factors were named by the research team based on their distinguishing statements: (1) Intrinsically Motivated, (2) Extrinsically Motivated, (3) Administratively & Emotionally Deterred. There was no statistically significant association between the factors and the demographic variables (e.g., age, years of experience, class size). Although we collected the name of the school to which each participant belonged to, to ensure respondent anonymity, we chose not to report this information for each factor, as in many cases, the frequency of the participants from each school was less than six.

Factor 1: intrinsically motivated

Seventeen participants loaded on Factor 1, the Intrinsically Motivated group of which nine participants were from the SON and eight were from the SRS. The distinguishing statements for this factor are presented in Table 2. Instructors who loaded onto this factor reported that they perceived mental barriers that prevented them from failing a student. In general, these instructors believe that they should not fail a student because an inept student is a result of their own inadequacy (Statement #8: + 5). They felt guilty for failing a student because they understood that no student was perfect (Statement #7: + 4). They were concerned that failing a student would damage the student's future (Statement #14: + 3) and felt that failing a student would contribute to the instructor's sense of self-doubt (Statement #23: + 2).

Instructors in the Intrinsically Motivated group strongly disagreed that they experienced pressure from either students or the school to either pass or fail a student (Statements #1 and #17: -5, -4, respectively). They also strongly disagreed with the idea that their own university would not support them in their decision to fail a student (Statement #38: -5). They were not concerned with students appealing to their decision or feedback (Statements #13: -3).

Factor 2: extrinsically motivated

Twenty participants loaded on Factor 2, the Extrinsically Motivated group; 12 from the SON and 6 from the SRS (for two participants the school is not known). Their distinguishing statements are presented in Table 3. The Extrinsically Motivated group felt external influences from "higher-ups" and the university encouraging them to pass all students (Statement#1: + 5), and that the university does not normally support their

Table 2 Distinguishing statements for the Intrinsically Motivated group (Factor 1)

Statement		Factor 1	Factor 2	Factor 3
8	If I fail a student, I failed myself because I did not provide enough guidance/mentorship	**5**	0	-5
7	I feel guilty for failing students, because I know no one is perfect	**4**	-1	0
14	I worry that my feedback for the student will jeopardize their future	**3**	-5	0
23	Failing students gives me a sense of self doubt	**2**	-1	-3
19	I believe passing an incompetent student is a disservice to the broader society	**2**	5	5
24	Failing students goes against the "caring" nature of my profession	**1**	-3	-2
20	I feel that there is inadequate academic/emotional support for students when they fail a course	**1**	-2	4
27	I feel uncomfortable defending my reasons for failing a student	**0**	2	-4
11	I do not have adequate time/energy to properly evaluate student performance	**-1**	-4	2
13	I worry that students will appeal my decision/feedback	**-3**	3	1
17	My students pressure me to pass them	**-4**	0	0
1	My higher-ups/the university encourage me to find reasons to pass all students	**-5**	5	-2
38	I do not feel that the University of my faculty supports my decision to fail a student	**-5**	4	-1

Table 3 Distinguishing statements for the Extrinsically Motivated group (Factor 2)

Statement		Factor 1	Factor 2	Factor 3
1	My higher-ups/the university encourage me to find reasons to pass all students	-5	**5**	-2
38	I do not feel that the University of my faculty supports my decision to fail a student	-5	**4**	-1
26	The appeal process regrading student work puts my credibility on the line	-3	**3**	-1
27	I feel uncomfortable defending my reasons for failing a student	0	**2**	-4
30	I worry that failing students would negatively impact my reappointment or tenure aspirations	-4	**2**	-3
8	If I fail a student, I failed myself because I did not provide enough guidance/mentorship	5	**0**	-5
20	I feel that there is inadequate academic/emotional support for students when they fail a course	1	**-2**	4
31	I worry that failing students will create an uncomfortable learning environment for other students	1	**-3**	1
11	I do not have adequate time/energy to properly evaluate student performance	-1	**-4**	2
34	Due to the current shortages of health professionals, I find it difficult to fail students	0	**-5**	1
14	I worry that my feedback for the student will jeopardize their future	3	**-5**	0

decision to fail a student (Statement#38: + 4). They felt that when students appeal against their decision, their credibility will be damaged (Statements #26: + 3). They felt uncomfortable defending their reason for failing a student (Statements #27: + 2) and were concerned that failing a student would damage their own career advancements (Statements #30: + 2). The Extrinsically Motivated group did not believe that failing a student would jeopardize the student's future (Statements #14: -5), nor did they believe that failing a healthcare student is wrong amidst a healthcare professional shortage (Statements #34: -5). They did not feel that limited time or energy affected their evaluation of the students (Statements #11: -4), and they did not believe that failing a student would create an uncomfortable learning environment (Statements #31: -3). Finally, instructors in this group did not believe that inadequate academic and emotional support for students affects their evaluation process (Statements #20: -2).

Factor 3: administratively & emotionally deterred

Fourteen participants loaded on Factor 3, the "Administratively & Emotionally Deterred" group; 11 participants from the SON and two from the SRS. For one participant the school is not known. They strongly agreed that they wanted to avoid dealing with the additional workload and increased time commitment of failing a student (Statement #29: + 5). They felt that the process of failing a student is.

stressful and exhausting (Statement #9: + 4). They strongly disagreed that a failed student is a result of the instructor's own inadequate guidance or mentorship (Statement #8: -5) and that they avoid failing a student because they do not know what type of information to document to support their decision to fail a student (Statement #25: -5). The distinguishing statements for this group are presented in Table 4.

Table 4 Distinguishing statements for the Administratively & Emotionally Deterred group (Factor 3)

Statement		Factor 1	Factor 2	Factor 3
29	Failing a student involves extra amounts of work and time which leads to a reluctance to fail students	-1	1	**5**
9	Having to evaluate student performance is stressful and exhausting for me	0	-2	**4**
20	I feel that there is inadequate academic/emotional support for students when they fail a course	1	-2	**4**
28	I fear being threatened with legal action when failing students	-4	-1	**3**
21	I feel that there is inadequate emotional support in place for me when I give a student a failing grade	-2	0	**3**
11	I do not have adequate time/energy to properly evaluate student performance	-1	-4	**2**
35	I worry that I may appear being biased if I my assessments if I disliked an underperforming student	-2	-4	**2**
1	My higher-ups/the university encourage me to find reasons to pass all students	-5	5	**-2**
33	I am reluctant to fail an upper-year student, as they've already come so far in their education."	3	1	**-2**
5	I worry that my colleagues will view me negatively if I fail students	0	1	**-3**
39	I do not fail students as I often lack enough supporting evidence for this judgment	1	1	**-4**
27	I feel uncomfortable defending my reasons for failing a student	0	2	**-4**
25	I do not know what type of information should be documented to support my impression that a student is performing poorly	2	1	**-5**
8	If I fail a student, I failed myself because I did not provide enough guidance/mentorship	5	0	**-5**

Consensus statements

There were seven statements that all participants agreed or disagreed with each of these statements to a similar extent (Table 5). The statement that all participants agreed strongly with was: "I need to feel 100% confident when failing a student in case I need to justify my decision. When I am not 100% sure, I rather be cautious and pass them". The statements all participants felt neutral about are: (a) "I worry that students will view me negatively if I give critical feedback"; (b) "I worry that failing students gives my program a bad reputation". The statement that participants disagreed with the most was: "I feel like I do not have the adequate communication skills to properly communicate my constructive feedback for students".

Discussion

This study highlighted broad reasons why instructors in nursing and rehabilitation sciences may fail to fail underperforming students. Using Q-methodology, we categorized instructors' experiences with FTF into three factors: Intrinsically Motivated, Extrinsically Motivated, and Administratively & Emotionally Deterred groups. These factors represent three main perspectives on FTF from an instructor perspective, and instructors in the same factor have similar considerations and barriers in terms of student evaluation.

Our results demonstrated that all the instructors felt confident in identifying underperforming students. The three groups of instructors reported little challenges with objectively evaluating a student's performance (Statement #36). This finding is inconsistent with some previous observations that instructors fail to fail because it is difficult to objectively evaluate students [6, 14, 49]. This may be because of the types of instructors who participated in this study. For example, when instructors evaluate nursing students, they evaluate the student against a list of competencies outlined by the nursing governing body [50]. However, the language used to describe the competencies may not be well defined or written with

Table 5 List of consensus statements

Statement		Factor 1	Factor 2	Factor 3
40	I need to feel 100% confident when failing a student in case I need to justify my decision. When I am not 100% sure, I rather be cautious and pass them	3	4	4
4	I worry that students will view me negatively if I give critical feedback	0	0	1
37	Due to fear of breaching student confidentiality, I find it difficult to seek the help of resources, when I fail students	-3	-1	0
18	I feel like the student should have been failed earlier, and I do not want to be the bad person	1	2	0
22	I worry that failing students gives my program a bad reputation	-2	-2	-1
36	I find it difficult to objectively evaluate a student's performance	-1	-1	-2
16	I feel like I do not have the adequate communication skills to properly communicate my constructive feedback for students	-2	-2	-4

academic jargon, hence allowing room for subjective interpretation by the instructor [50]. The inconsistency between our results and those of previous studies may be attributed to differences in the evaluation guidelines, rubrics, and training levels of the instructors across different universities.

The findings of our study support the hypothesis that there are broad internal, external, or systematic factors that may contribute to FTF. First, instructors across all factors agreed that they needed to be certain a student they are failing deserves to fail. When they were uncertain, the tendency was to pass the student who was potentially an underperforming student (Statement #40: +3, +4, +4), giving students "the benefit of the doubt" [6, 51]. Our results show that "the benefit of the doubt" affects all factors, suggesting that it is a pertinent contributor to FTF for all instructors.

Student's perception of the instructor often contributes to FTF. Previous research has shown that instructors avoid failing students because they do not want students to view them negatively [6, 44]. Instructors value students' perceptions of them for two proposed reasons: professional relationships and ranking. Students and instructors often form a friendly bond that can prevent instructors from failing a student [51, 52]. Students may also leave critical reviews for their instructors – an action that may impact their instructor's future ability to secure promotions and a tenure in their institutions [6, 15, 53]. As a result, instructors avoid awarding failing grades to students to avoid poor reviews [25, 53]. In the current research, we show that most instructors felt indifferent toward the role of the student-instructor dynamic in FTF (Statement #4: 0, 0, 1). The difference between our findings and past literature may also be attributed to different university policies related to how teaching evaluations are used in consideration of instructor career progress. Historically, students' evaluations of faculty have been collected by institutions to assess the appropriateness of faculty to receive promotions [54]. However, there is growing recognition that students' evaluations of faculty do not fully indicate their ability to teach [54]. Consequently, more universities are placing less emphasis on students' evaluations of faculty and finding other ways to evaluate the competency of the faculty, such as peer reviews [55].

Black et al. [13] reported that some instructors experience "moral stress" when they fail a student. Instructors who experience moral stress attribute the student's failure to their own inadequacy and may question whether they have done enough to support the student, or if they have been too harsh [13]. In our research, the Intrinsically Motivated group reported experiencing similar feelings of moral stress. They avoided failing a student because they perceived that failing a student was the result of their own lack of support and guidance (+5). This sentiment was supported by qualitative comments left by the instructors:

> *"I must teach the content and if a student is not getting the content, I can feel like I have failed them. Especially in a clinical placement setting"*
> *"I pride myself in providing clinical instruction that is understandable and catered to the student's learning needs, and not performing to the level that they are expected to may reflect instruction that does not resonate with their learning needs."*

As the Intrinsically Motivated group feel they are to blame for an underperforming student, they also experience a sense of guilt (+4) and self doubt (+2) when evaluating an underperforming student. The combination of responsibility, guilt, and self-doubt creates a sense of moral distress among instructors [13]. The experience of moral distress evokes stress, anxiety, and discomfort among instructors [2–4, 9, 13]. Consequently, instructors avoid failing students to minimize their experience of moral stress. Our results showed that among the three factors, instructors in the Intrinsically Motivated group were most influenced by this sense of moral stress, responsibility, and guilt. They also avoid failing a student because they believe that no student is perfect (+4). Past research has shown that many instructors understand that students are at the beginning of their education journey, and that they will have more time in the future to improve their skills [13]. Consistent with the notion that instructors avoid failing early-stage trainees is the fact that first-year students fail less often than do more senior students [12].

Instructors in the Intrinsically Motivated group also avoided failing underperforming students since they worried that a failing grade would jeopardize the student's academic and professional future. The concern for students' futures aligns with previous findings [9, 56] and may stem from the rapport that develops between instructors and students over time. The concern for students may also be attributed to the lack of information an instructor has regarding the process to failing and supporting student remediation:

> *"I am not sure what would really happen if I were to fail a student. Would it mean that they are never able to practice? Would they have another chance in a future placement?"*

When instructors are uncertain about the consequences of failing, they are more likely to pass underperforming students by giving them the benefit of the doubt. Our findings suggest that to help instructors evaluate

students more accurately, universities and programs should offer additional education to instructors about the learning resources available to students who may be falling behind.

In many aspects, instructors in the Extrinsically Motivated group held opposing beliefs to the instructors in the Intrinsically Motivated group. Specifically, compared to the Intrinsically Motivated group, the Extrinsically Motivated group were much less concerned that failing a student may jeopardize their future (-5) and they strongly agreed that pressures from the university or students influenced their ability to fail a student (+5). Our results indicated that the Extrinsically Motivated group generally avoided failing students because they felt pressured by the university or the students. For example, they strongly felt that the university encouraged them to find reasons to pass students (+5); they worried the university would not support them in their decision to fail a student (+4); and they were concerned with students appealing their decision (+3). This concern was captured in the written comment provided by an instructor:

> *"[There is] zero support to have students out of sequence and any appeals seem to automatically be granted".*
> *"I was in a situation where I didn't believe the student had the skills to pass the final placement. The University discouraged me from failing the student."*

The Intrinsically Motivated and Extrinsically Motivated groups also differed in their perspectives on the university's role in evaluation. The Extrinsically Motivated group were concerned that the university would not support their decision to fail a student (+4). In contrast, the Intrinsically Motivated group did not report feeling pressured by the university (-5). The contrasting sentiments between these two groups further illustrated that the rationale of FTF for one instructor may be completely different from the rationale of another instructor. Our findings on how external pressure fosters FTF are consistent with previous research suggesting that instructors worry about the school overturning their decision to fail a student [15, 57].

The barriers faced by the Extrinsically Motivated group highlight the importance of instructors receiving support from their program and the broader university. When an underperforming student is identified, they perceive that the university should support the decision made by the instructor [6, 14, 44, 58]. However, it may not be clear to the university how well the instructor managed the failure and supported the student with a plan for success.

Factor 3 were identified as the Administratively & Emotionally Deterred group. Many studies across different institutions have shown that failing a student translates to more administrative work for the instructors [15, 45, 57, 59–62]. As instructors become overwhelmed with work, they may be discouraged by the perceived additional work associated with failing a student. This sentiment was echoed by participants in the current study:

> *"Failing a student means supporting them in remediation which is a lot of extra time/work/stress above and beyond an already over-full schedule. The pressure when managing a student's remediation is intense because you don't want them to perform poorly on placement (and impact the relationship with community clinicians) or have to leave the program."*

Our results indicate that the Administratively & Emotionally Deterred group avoided failing students because they perceived that failing a student would result in additional work for them (+5). The process of evaluating, and the possibility of re-evaluating failed students are also stressful for instructors in this group (+4). The perceived stress associated with evaluation may stem from issues with the current evaluation process. This concern aligns with the concerns of instructors at other institutions who also reported evaluation tools and rubrics can be vague and inadequate for determining whether a student has demonstrated the appropriate knowledge or skills [50, 51], while some instructors feel that they have not received adequate training to properly assess students [61, 63, 64]. When instructors have unclear assessment rubrics and inadequate training, they may find student evaluation to be stressful or anxiety-inducing, which may contribute to the stressful experience of failing a student.

Instructors in the Administratively & Emotionally Deterred group expressed significant concerns about students appealing their failing grades, a fear well-documented in the literature [6, 15, 60, 62, 64]. Several reasons may explain why instructors may be reluctant to engage in the appeal process. They might feel uncomfortable defending their decision to fail a student before the university and fearing that doing so could damage their credibility [45, 64]. There is also a prevalent fear that during the appeal process, the university might side with the student rather than the instructor. This lack of trust in the appeal system leads many instructors to pass underperforming students to avoid the potential complications and stress of a grade appeal [6, 9].

Finally, it is important to recognize that the responsibilities of nursing and rehabilitation science faculty are constantly evolving. For example, the recent COVID-19 pandemic imposed mandatory changes on nursing and rehabilitation practitioners, and these changes may have contributed to FTF. Due to social distancing guidelines during the COVID-19 pandemic, many instructors were

forced to teach virtually. The shift to online learning translated to nursing and rehabilitation sciences instructors relying on online and simulation-based educational experiences [65, 66]. For many instructors, there were increases in faculty workload, decreases in support mechanisms, and changes to grading metrics [66]. In addition to these teaching challenges, the COVID-19 pandemic also introduced more clinical work and responsibilities for nursing faculty [67]. It is possible that the increase in clinical work and virtual teaching models both contributed to the growing prevalence of fatigue or burnout among faculty members. As a result, it may be that faculty members are more likely to pass an underperforming student to avoid the extra administrative tasks.

Findings of the current study have significant implications for healthcare educational practices and policies. Identifying intrinsic, extrinsic, and administrative deterrents highlights the different mindsets of instructors and emphasizes the need for a diverse approach to alleviating FTF. To address FTF, all stakeholders including students, instructors, and the educational institution need to work together. For instance, instructors and the institution can work together to identify unnecessary administrative work when failing a student and work together to streamline the process. Educational institutions can further support instructors by developing clear and well-defined evaluation rubrics, as well as providing comprehensive training for instructors to enhance their confidence in student assessments. Institutions are also encouraged to build an academic support system for both students and instructors. For students, there should be a system to help them overcome learning challenges and provide the possibility for remediation. For instructors, the system should ensure they are aware of additional academic support resources and remedial opportunities for students.

Although the current literature mostly describes the instructor's role in FTF, it is important to remember that universities and other regulatory bodies influence student evaluation. In Canada, universities oversee the evaluation of learners and determine whether students have met competency expectations. The university program is accountable to various agencies, such as the College of Nurses and Accreditation Agencies, to ensure that all students are fulfilling the core competencies. Additionally, instructors are also regulated and held accountable by their professional college (e.g., College of Nurses of Ontario, College of Physiotherapists of Ontario) to ensure that all licensed healthcare professionals are fulfilling their professional obligations. Ultimately, it is a complex system of setting professional competencies and determining whether a student has met the degree requirements from the university's perspective. All shareholders within the system, including the instructors, the university, and the government bodies, have a role in reducing FTF and developing a fair evaluation system.

Limitations

The current study did not distinguish between academic-based and clinical teaching faculty. In this study, academic faculty refers to individuals who are responsible for teaching specific courses in the university setting and may have larger class sizes. On the other hand, clinical faculty are involved in supervising and instructing students in clinical settings and are focused on clinical skills and decision making—often in small groups or in a one-to-one ratio at a clinical site. To better protect the privacy of our respondents, we did not collect additional information on the classes that the instructors taught.

Conclusions

This study explored the perceptions of instructors at a Canadian university as to why they are hesitant to fail students. The results of this work identified three distinctive viewpoints: Intrinsically Motivated, Extrinsically Motivated, and Administratively & Emotionally Deterred, which were named based on their distinguishing statements. This research outlines areas of consideration for addressing the failure-to-fail mechanism. More transparent discussions within schools and identifying solutions are recommended to create systems that maintain educational and professional standards. Further replications of this study in various disciplines may be used to determine whether these findings are consistent across different fields.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-024-06126-2.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

The authors gratefully acknowledge Joshua Mitchell for his assistance in formatting and distributing the online data collection survey.

Authors' contributions

NA, BW, and SW conceived the study idea and the research design. CL and JR drafted the first draft. NA analysed the data. All authors reviewed and revised the manuscript and approved the final version.

Funding

The authors received no funding for this project.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the Hamilton Integrated Research Ethics Board (HiREB# 15784). We confirm that all methods were carried out in accordance with relevant guidelines and regulations. The participants provided informed consent to participate in this study. Participation in the study was optional, confidential, and anonymous. At the end of the survey, the participants were offered an opportunity to enter a draw for one of two $50 gift cards as compensation for their participation.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Temerty Faculty of Medicine, University of Toronto, Toronto, Canada. [2]Faculty of Health Sciences, McMaster University, Hamilton, Canada. [3]School of Rehabilitation Sciences, McMaster University, Hamilton, Canada. [4]School of Nursing, McMaster University, 1280 Main St. West, Room 2J21D, Hamilton, ON L8S 4K1, Canada.

Received: 23 February 2024 Accepted: 3 October 2024
Published online: 14 October 2024

References

1. Schildkamp K, van der Kleij FM, Heitink MC, Kippers WB, Veldkamp BP. Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. Int J Educ Res. 2020;103:101602.
2. Bachmann L, Groenvik CKU, Hauge KW, Julnes S. Failing to Fail nursing students among mentors: A confirmatory factor analysis of the Failing to Fail scale. Nurs Open. 2019;6(3):966–73.
3. Hughes LJ, Mitchell M, Johnston AN. "Failure to fail" in nursing - A catch phrase or a real issue? A systematic integrative literature review. Nurse Educ Pract. 2016;20:54–63.
4. Luhanga F, Yonge OJ, Myrick F. "Failure to assign failing grades": issues with grading the unsafe student. Int J Nurs Educ Scholarsh. 2008;5:Article8.
5. Tonesk X, Buchanan RG. An AAMC pilot study by 10 medical schools of clinical evaluation of students. J Med Educ. 1987;62(9):707–18.
6. 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):1092–9.
7. Lankshear A. Failure to fail: the teacher's dilemma. Nurs Stand. 1990;4(20):35–7.
8. Cleland JA, Roberts FE. "I Don't Want Them Working Alongside Me", or Assessing Performance of Physiotherapy Students on Clinical Placement. Health and Social Care Education. 2013;2(1):3–5.
9. Duffy K. Deciding to fail. The Journal of Practice Teaching and Learning. 2013;11(3):36–58.
10. Finch J, Tedam P. Failure to fail or fast tracking to failure: a critical exploration of social work placements. Soc Work Educ. 2023:1–16. https://doi.org/10.1080/02615479.2023.2236153.
11. Scarff CE, Bearman M, Chiavaroli N, Trumble S. Keeping mum in clinical supervision: private thoughts and public judgements. Med Educ. 2019;53(2):133–42.
12. Hughes LJ, Mitchell ML, Johnston ANB. Just how bad does it have to be? Industry and academic assessors' experiences of failing to fail – A descriptive study. Nurse Educ Today. 2019;76:206–15.
13. Black S, Curzio J, Terry L. Failing a student nurse. Nurs Ethics. 2014;21(2):224–38.
14. DeBrew JK, Lewallen LP. To pass or to fail? Understanding the factors considered by faculty in the clinical evaluation of nursing students. Nurse Educ Today. 2014;34(4):631–6.
15. Larocque S, Luhanga FL. Exploring the issue of failure to fail in a nursing program. Int J Nurs Educ Scholarsh. 2013;10(1):115–22.
16. Susmarini D, Sri Hayati Y. Grade Inflation In Clinical Stage. Am J Health Sci (AJHS). 2011;2(1):21–8.
17. Scanlan JM, Care WD. Grade Inflation: Should We Be Concerned? J Nurs Educ. 2004;43(10):475–8.
18. Docherty A, Dieckmann N. Is There Evidence of Failing to Fail in Our Schools of Nursing? Nurs Educ Perspect. 2015;36(4):226–31.
19. Karadag E. Effect of COVID-19 pandemic on grade inflation in higher education in Turkey. PLoS ONE. 2021;16(8):e0256688–e0256688.
20. Tillinghast JA, Mjelde DJW, Yeritsyan A. COVID-19 and Grade Inflation: Analysis of Undergraduate GPAs During the Pandemic. SAGE Open. 2023;13(4):21582440231209110.
21. Adkins DA, Aucoin JW. Failure to fail - Factors affecting faculty decisions to pass underperforming nursing students in the clinical setting: A quantitative study. Nurse Educ Pract. 2022;58:103259.
22. Dibble JL. Breaking Good and Bad News: Face-Implicating Concerns as Mediating the Relationship Between News Valence and Hesitation to Share the News. Commun Stud. 2014;65(3):223–43.
23. Rosen S, Tesser A. On Reluctance to Communicate Undesirable Information: The MUM Effect. Sociometry. 1970;33(3):253–63.
24. Moore S, Kuol N. Students evaluating teachers: exploring the importance of faculty reaction to feedback on teaching. Teach High Educ. 2005;10(1):57–73.
25. Stroebe W. Student Evaluations of Teaching Encourages Poor Teaching and Contributes to Grade Inflation: A Theoretical and Empirical Analysis. Basic Appl Soc Psychol. 2020;42(4):276–94.
26. Gingerich A, Sebok-Syer SS, Larstone R, Watling CJ, Lingard L. Seeing but not believing: Insights into the intractability of failure to fail. Med Educ. 2020;54(12):1148–58.
27. Jervis A, Tilki M. Why are nurse mentors failing to fail student nurses who do not meet clinical performance standards? Br J Nurs. 2011;20(9):582–7.
28. Bergen N, Labonté R. "Everything Is Perfect, and We Have No Problems": Detecting and Limiting Social Desirability Bias in Qualitative Research. Qual Health Res. 2020;30(5):783–92.
29. Jackson KM, Trochim WMK. Concept mapping as an alternative approach for the analysis of open-ended survey responses. Organ Res Methods. 2002;5(4):307–36.
30. Smithson J, Birks M, Harrison G, Nair CS, Hitchins M. Benchmarking for the effective use of student evaluation data. Qual Assur Educ. 2015;23(1):20–9.
31. Steyn C, Davies C, Sambo A. Eliciting student feedback for course development: the application of a qualitative course evaluation tool among business research students. Assess Eval High Educ. 2019;44(1):11–24.
32. Brewer-Deluce D, Sharma B, Akhtar-Danesh N, Jackson T, Wainman BC. Beyond Average Information: How Q-Methodology Enhances Course Evaluations in Anatomy. Anat Sci Educ. 2020;13(2):137–48.
33. Akhtar-Danesh N, Baumann A, Cordingley L. Q-methodology in nursing research: A promising method for the study of subjectivity. West J Nurs Res. 2008;30(6):759–73.
34. Mackinnon C, Akhtar-Danesh N, Palombella A, Wainman BC. Using Q-methodology to determine students perceptions of interprofessional anatomy education. Anat Sci Educ. 2022;15:877–85.
35. Valaitis R, Akhtar-Danesh N, Eva K, Levinson A, Wainman B. Pragmatists, positive communicators, and shy enthusiasts: three viewpoints on Web conferencing in health sciences education. J Med Internet Res. 2007;9(5):e39.
36. Stephenson W. Correlating persons instead of tests. Character Pers. 1935;4:17–24.
37. Chinnis AS, Summers DE, Doerr C, Paulson DJ, Davis SM. Q methodology:a new way of assessing employee satisfaction. J Nurs Adm. 2001;31(5):252–9.
38. Brown SR. Political Subjectivity: Application of Q Methodology in Political Science. New Haven: Yale University Press; 1980.
39. Brown SR. A primer on Q methodology. Operant Subjectivity. 1993;16:91–138.
40. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A practical guide to their development and use, 6th ed. Oxford: Oxford University Press; 2024.
41. Holden RB. Face validity. In: Weiner IB, Craighead WE, editors. The Corsini Encyclopedia of Psychology. 4th ed. Hoboken, New Jersey: Wiley; 2010. p. 637–8.
42. Valenta AL, Wigger U. Q-methodology: definition and application in health care informatics. J Am Med Inform Assoc. 1997;4(6):501–10.

43. Dennis KE. Commentary: Looking at Reliability and Validity Through Q-Colored Glasses. Operant Subjectivity. 1992;16(1/2):37–44.
44. Docherty A. Failing to Fail in Undergraduate Nursing: Understanding the Phenomenon. Nurs Educ Perspect. 2018;39(6):335–42.
45. Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. Acad Med. 2005;80(10 Suppl):S84–87.
46. Guraya SY, van Mook W, Khoshhal KI. Failure of faculty to fail failing medical students: Fiction or an actual erosion of professional standards? J Taibah Univ Med Sci. 2019;14(2):103–9.
47. Akhtar-Danesh N. qfactor: A command for Q-methodology analysis. Stand Genomic Sci. 2018;18(2):432–46.
48. Akhtar-Danesh N. Using Cohen's Effect Size to Identify Distinguishing Statements in Q-methodology. Open J Appl Sci. 2018;8(2):73–9.
49. Ziring D, Frankel RM, Danoff D, Isaacson JH, Lochnan H. Silent witnesses: Faculty reluctance to report medical students' professionalism lapses. Acad Med. 2018;93(11):1700–6.
50. Almalkawi I, Jester R, Terry L. Exploring mentors' interpretation of terminology and levels of competence when assessing nursing students: An integrative review. Nurse Educ Today. 2018;69:95–103.
51. Hughes LJ, Johnston ANB, Mitchell ML. Human influences impacting assessors' experiences of marginal student performances in clinical courses. Collegian. 2018;25(5):541–7.
52. Dobbs S. Why some nurse educators are reluctant to fail students. Nurs N Z (Wellington, NZ : 1995). 2017;23(1):12–5.
53. Fontana JS. Nursing Faculty Experiences of Students' Academic Dishonesty. J Nurs Educ. 2009;48(4):181–5.
54. Hornstein HA. Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. Cogent Educ. 2017;4(1):1304016.
55. Quenette A, Rybas N. Evaluating instructor quality: Moving away from student evaluations to peer review. New Dir Teach Learn. 2023;2023(173):67–74.
56. Luhanga FL, Larocque S, MacEwan L, Gwekwerere YN, Danyluk P. Exploring the Issue of Failure to Fail in Professional Education Programs: A Multidisciplinary Study. J Univ Teach Learn Pract. 2014;11(2):17–41.
57. Luhanga F, Yonge O, Myrick F. Precepting an unsafe student: The role of the faculty. Nurse Educ Today. 2008;28(2):227–31.
58. Finch J, Taylor I. Failure to Fail? Practice Educators' Emotional Experiences of Assessing Failing Social Work Students. Soc Work Educ. 2013;32(2):244–58.
59. Carr J, Heggarty H, Carr M, Fulwood D, Goodwin C, Walker W, Whittingham K. Reflect for success: recommendations for mentors managing failing students. Br J Community Nurs. 2010;15(12):594–6.
60. 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):800–9.
61. Rutkowski K. Failure to fail: assessing nursing students' competence during practice placements. Nurs Stand. 2007;22(13):35–40.
62. 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):1157–62.
63. Bing-You R, Varaklis K, Hayes V, Trowbridge R, Kemp H, McKelvy D. The feedback tango: An integrative review and analysis of the content of the teacher⇓learner feedback exchange. Acad Med. 2018;93(4):657–62.
64. Bush HM, Schreiber RS, Oliver SJ. Failing to fail: clinicians' experience of assessing underperforming dental students. Eur J Dent Educ. 2013;17(4):198–207.
65. Hoseini-Beidokhti M, Abaeian G, Abedini-Baghbadorani S, Latifi SM. Quality and Effectiveness of Virtual Education From the Viewpoint of Rehabilitation Professors and Students of Ahvaz Jundishapur University of Medical Sciences During the COVID-19 Pandemic. Iran Rehabil J. 2022;20(3):441–8.
66. Leaver CA, Stanley JM, Goodwin Veenema T. Impact of the COVID-19 Pandemic on the Future of Nursing Education. Acad Med. 2022;97:S82–9.
67. Sacco TL, Kelly MM. Nursing Faculty Experiences During the COVID-19 Pandemic Response. Nurs Educ Perspect. 2021;42(5):285–285.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.