**RESEARCH ARTICLE**

# Failing to Fail nursing students among mentors: A confirmatory factor analysis of the Failing to Fail scale

Liv Bachmann [iD] | Cecilie Katrine Utheim Groenvik | Kari Westad Hauge | Signe Julnes [iD]

Faculty of Health Sciences and Social Care at Molde University College, Specialized University in Logistics, Molde, Norway

**Correspondence**
Signe Julnes, Faculty of Health Sciences and Social Care at Molde University College, Specialized University in Logistics, Molde, Norway.
Email: signe.julnes@himolde.no

## Abstract

**Aim:** The aim was to explore the psychometric properties with respect to the internal consistency reliability of the subject-specific questionnaire "Failing to Fail."

**Design:** Cross-sectional study.

**Methods:** Exploratory factor analysis with varimax rotation. A confirmatory factor analysis was used to examine the factor structure of the "Failing to Fail" scale. The sample included 336 Norwegian nurse mentors.

**Results:** The confirmatory factor analysis confirmed a five-factor structure of the "Failing to Fail" scale with adequate model fit. The factors were named as: (a) Insufficient mentoring competence; (b) Insufficient support in the working environment; (c) Emotional process dominates the assessment; (d) Insufficient support from the university; and (e) Decision-making detached from learning outcomes. The scale proved to be feasible to test whether mentors are Failing to Fail nursing students. The confirmatory factor analysis model supported the predictive validity of the "Failing to Fail" scale.

**KEYWORDS**
confirmatory factor analysis, exploratory factor analyses, Failing to Fail, nurse mentors

## 1 | INTRODUCTION

The primary role of the nurse mentor is not only to support the professional development of students, but also to prevent failing students from becoming registered nurses to safeguard the public from incompetent practitioners (Luhanga, Yonge, & Myrick, 2008a). "Failing to Fail" is to pass a nursing student who does not display satisfactory clinical performances (Hughes, Mitchell, & Johnston, 2016). Research has examined whether mentors who assess practical competence are Failing to Fail students in their assessment (Black, 2011; Brown, Douglas, Garrity, & Shepard, 2012; Duffy, 2003,2006; Gainsbury, 2010; Hauge, 2015; Hughes et al., 2016; Luhanga et al.,

2008a). The purpose of this article was to explore the psychometric properties of a subject-specific "Failing to Fail" questionnaire.

### 1.1 | Background

Failing to Fail student nurses in practical assessments is an international issue (Hunt, McGee, Gutteridge, & Hughes, 2016). Two literature reviews (Hauge, 2015; Hughes et al., 2016) indicate that most of the research on Failing to Fail is conducted with qualitative studies. A survey of Brown et al. (2012) found that 18% (N = 1790) of the mentors confirmed that they had passed a failing student. In other words, 18% of the mentors had the experience of Failing to Fail. In

a frequently referenced online survey conducted by the Nursing Times magazine (Gainsbury, 2010), 37% (N = 1945) of the mentors confirmed having passed underachieving students. Gainsbury included both if the mentor actually meant that the student should have failed and if they had been concerned about the student's competence and attitudes. Gainsbury (2010) indicated a significantly higher incidence than reported by Brown et al. (2012).

Both Hauge (2015) and Hughes et al. (2016) conducted a literature review to investigate the phenomenon Failing to Fail. Hughes et al. (2016) suggested that failure to fail is an issue with complex facets and identified five factors related to the failure to fail: the difficulty of failing a student, the emotional process for nurse mentors, the need for confidence, unsafe student characteristics and the importance of university support. Hauge (2015), in turn, identified six factors: a lack of knowledge and confidence, challenges related to documentation, personal attitudes and beliefs, lack of support, the scarcity of time and fear of losing face. When we see what is laid in the two authors' different factors, a four-factor model illuminates mentors' reasons for Failing to Fail: lack of competence, lack of support, emotional consequences for nurse mentors and personal attitudes and beliefs.

Hauge (2015) identified no studies from Scandinavia when investigating whether nursing students pass their clinical studies without achieving the learning outcomes. Hughes et al. (2016) recommended further research to identify whether "failure to fail" is a universal problem in undergraduate nursing programmes. The literature search also examined a lack of instruments measuring "Failing to Fail" within a nurse mentor context. The research group in pedagogy (Molde University College) developed a subject-specific questionnaire to investigate whether Norwegian mentors are Failing to Fail nursing students (Hauge et al., 2019). The questionnaire also investigated the reasons influencing their decisions. Brown et al. (2012) inspired the questionnaire. The research group discussed and assessed the relevance of the items. We retained some of the items from Brown et al. (2012). In addition, construction of the questionnaire was based on a theory-based methodology (Araï & Martinussen, 2010; Haraldsen, 1999) and qualitative studies (Duffy, 2003; Hauge, 2015; Rutkowski, 2007). The aim of the item construction was to ensure that the items covered different reasons for Failing to Fail. To strengthen content validity and conceptual validity, face validity was ensured by a reference group of six nurses with experience in mentoring. The reference group assessed the clarity and relevance of the questionnaire through providing written feedback. Afterwards, the reference group and representatives from the research group discussed the feedback. The reference group confirmed that the items were clear and relevant for the nurse context and they suggested adding two items to further investigate the Failing to Fail phenomenon.

We tested the questionnaire in a pilot study in 2016. An online questionnaire was distributed to nurse mentors (N = 336) from hospitals, nursing homes and home-based nursing care. The number of responders was 122 (36.3%), and out of these responders, 25 nurse mentors confirmed that they had failed a failing student. The reliability analysis indicated that the questionnaire measured if mentors are Failing to Fail and reasons for their decision. Because of the responses from the survey participants, we made minor changes in the questionnaire, such as clarification if mentors had further education in mentoring at the time they failed to fail a student. We also supplemented with two more items focusing on the education institution (Hauge et al., 2019).

The aim of the current study was to explore the psychometric properties with internal consistency reliability of the subject-specific questionnaire "Failing to Fail" (FTF). The emerging research questions were to study the factor structure of the questionnaire, to estimate the factor model and to study the model fit to our questionnaire data.

## 2 | METHOD

### 2.1 | Design

This cross-sectional study used the questionnaire FTF. Data on gender, age, number of years of practical nursing and serving as mentors and further education were also collected. If a respondent answered "yes" to the question "Have you ever passed a student you actually thought should fail?", they had to answer another 30 items connected to the reasons for failing the student, based on the last time they passed a student they actually thought should fail. Each item was answered on a five-point Likert scale with the following answer options: 1 = disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = somewhat agree and 5 = agree. The items related to the reasons for their decision were structured under four different topics. These topics were reasons related to personal focus, reasons related to student focus, reasons related to colleague and management focus and focus on the educational institution. The internal reliability coefficient for the items in the pilot study was very good (Cronbach's alpha = 0.89). The reference group and the respondents confirmed that the items were clear and relevant for their context, which indicated that the construct validity was acceptable (Hauge et al., 2019).

### 2.2 | Sample

We recruited a convenience sample of nurse mentors from one university hospital and from nursing homes and home-based nursing care in another municipality in Norway. We collected the data from October 2016–January 2017. A cover letter sent to the managers by email gave detailed information about the aim of the study. The managers distributed this information and the questionnaire to eligible participants by email. They distributed the online questionnaire to 2,380 nurse mentors. We sent reminders by email four times.

### 2.3 | Ethical consideration

We received permission from the University College to conduct the study. The management in the hospitals, nursing homes and home-based care granted approval for the study. We informed the respondents that their responses were anonymous and that their participation was voluntary. Returning the completed online questionnaire implied

their informed consent. According to the Norwegian Social Science Data Services (NSD), Research Ethics Committee approval was not required. It was neither required to report the study to NSD because no personal data were traceable to an email or IP addresses.

## 2.4 | Analysis

To explore and test a model of FTF, we performed both exploratory and confirmatory factor analyses. We proposed a model using exploratory factor analysis (EFA) to support dimensionality and interpretation of the factors. Then, we tested a four-factor model, based on the reviews of previous studies (Hauge, 2015; Hughes et al., 2016), in a confirmatory factor analysis (CFA) framework. The Stata software, version 15, was used to analyse the data. Descriptive analyses described the items and the demographic variables. An exploratory factor analysis (EFA) explored the feasibility of performing a factor analysis of the subject-specific questionnaire FTF. The Kaiser–Melkin–Olkin (KMO) measure of sampling adequacy was 0.71 ($p < 0.001$), well above the recommended value of 0.50 (Williams, Brown, & Onsman, 2012). A significant result of Bartlett's test of sphericity ($p < 0.001$) confirmed the appropriate correlation among variables and was adequate for factor analysis (Williams et al., 2012). We performed an exploratory factor analysis with orthogonal varimax rotation. An eigenvalue greater than one was used as an extraction criterion to extract factors. Since the significance of a factor loading depends on the sample size, we set the cut-off for factor loadings at 0.40 (Mehmetoglu & Jakobsen, 2016). A one-sample Kolmogorov–Smirnov test (Field, 2009) examined the distributions of the items in the scale. The scores were significantly different from a normal distribution, but according to Jolliffe (2002), an EFA does not require normality.

CFA tested the model fit of the extracted factor model. Three absolute fit indices tested the model fit of the factorial model (Table 1): chi-square statistic, standardized root mean squared residual (SRMR) and root mean square error of approximation (RMSEA). Absolute fit indices measure how well a factor analysis model fits the sampled data. We evaluate the internal consistency reliability of the subscales and the total scale by using Cronbach's alpha.

## 3 | RESULTS

### 3.1 | Demographics

A total of 561 mentors responded the questionnaire, resulting in a response rate of 23.6%. Furthermore, 93 of the mentors (16.6%)

had ever passed a student they actually thought should fail. Table 2 describes the frequency distribution of the items. Only 20 observations (scores) out of 14 601 scores were missing. To avoid a reduced number of observations in the analysis, we chose to substitute missing data. We substituted missing data by generating data randomly, according to the marginal distributions of the variables. The results with and without compensation for missing data were very similar. We based further analyses on a sample of 93. Most respondents were female (87.2%), and the average age was 41.74 years. The mean length of time practising nursing was 14.17 years; 14.9% of the mentors have education credits in mentorship, and they have mentored 11.16 students.

### 3.2 | Exploratory factor analysis

In our study, the eigenvalue criterion suggested a solution with six factors. The sixth factor model was rejected in favour for the five-factor model because the sixth factor consisted of two items and explained approximately 6% of the variance in the original items. The scree plot suggested that the point of diminishing returns was a five-factor solution. A five-factor structure was considered more meaningful and explained 80.05% of the variance in the original items (Table 3). This finding indicated a five-factor structure as the best solution. In Table 4, we can see the varimax-rotated factor loadings matrix. For more clear presentation of the results, numbers in bold indicate items with loadings above 0.40.

### 3.3 | Confirmatory factor analysis

Based on the initial EFA model, among 30 items, we deleted three items because of no loading above 0.40 in the five-factor solution. Four items loaded on more than one factor. "I did not feel confident to handle the situation," "I was concerned about running into conflict with the student" and "I didn`t have enough strength to handle the situation" loaded on factor one and factor three. These items were retained in factor one with the highest loading and best interpretation. Additionally, the item "I gave the student the benefit of the doubt" loaded on factor one and factor five. We retained this item in factor five based on the highest loading and best interpretation. After running a "Cronbach's alpha if item deleted" test on each factor, we deleted one of the items in factor two despite of a loading above 0.40. During this process, the questionnaire was reduced from 30–26 items. A confirmatory factor model was estimated, and several goodness-of-fit tests were performed, with the aim of investigating whether the modified "Failing to Fail" factor model was acceptable.

Each factor had a set of 4–8 items. We examined the items in each factor to identify a theoretical connection to develop a common substantial heading. The factors were named: (a) Insufficient mentoring competence; (b) Insufficient support in working environment; (c) Emotional process dominates the assessment; (d) Insufficient support from the university; and (e) Decision-making detached from learning outcomes. The internal consistency reliability (Cronbach's

**TABLE 1** Absolute goodness-of-fit indices (N = 93)

| Indicators | Parameter estimates | Standard range[a] |
| --- | --- | --- |
| Chi-square/*df* | 1.86 | <2–5 |
| SRMR | 0.11 | ≤0.10 |
| RMSEA | 0.096 | ≤0.10 |

[a]Mehmetoglu and Jakobsen (2016).

**TABLE 2** The frequency distribution of the items

| | Disagree % | Somewhat disagree % | Neither agree or disagree % | Somewhat agree % | Agree % |
|---|---|---|---|---|---|
| q17_1 Failing a student was in conflict with my role as a a caring person | 75.3 | 8.6 | 3.2 | 12.9 | 0.0 |
| q17_2 I couldn`t prove my concerns were valid | 22.6 | 28.0 | 10.8 | 29.0 | 9.7 |
| q17_3 I wasn`t sufficiently prepared | 25.8 | 16.1 | 15.1 | 32.3 | 10.8 |
| q17_4 I didn`t feel confident to handle the situation | 29.0 | 15.1 | 16.1 | 26.9 | 12.9 |
| q17_5 I thought failing a student was a personal defeat as a mentor | 81.7 | 10.8 | 2.2 | 4.3 | 1.1 |
| q17_6 I discovered the problem too late | 25.8 | 14.0 | 14.0 | 33.3 | 12.9 |
| q17_7 I was concerned I would be labelled a "bad" mentor | 86.0 | 8.6 | 1.1 | 2.2 | 2.2 |
| q17_8 I was concerned about running into conflict with the student | 55.9 | 12.9 | 9.7 | 14.0 | 7.5 |
| q17_9 I wasn`t familiar with the formal procedures for failing students | 32.3 | 18.3 | 16.1 | 22.6 | 10.8 |
| q17_10 I was uncertain what to document | 23.7 | 18.3 | 14.0 | 32.3 | 11.8 |
| q17_11 I was uncertain how to document | 29.0 | 16.1 | 12.9 | 31.2 | 10.8 |
| q17_12 I felt it was difficult to prioritize this in addition clinical tasks | 24.7 | 10.8 | 9.7 | 29.0 | 25.8 |
| q17_13 I didn`t have enough strength to handle the situation | 36.6 | 17.2 | 20.4 | 17.2 | 8.6 |
| q18_1 I liked the student | 45.2 | 19.4 | 28.0 | 3.2 | 4.3 |
| q18_2 I didn`t want to hurt the student`s feelings | 48.4 | 19.4 | 18.3 | 10.8 | 3.2 |
| q18_3 I didn`t want to expose the student to the consequences of failing | 46.2 | 18.3 | 11.8 | 19.4 | 4.3 |
| q18_4 I gave the student the benefit of the doubt | 15.1 | 10.8 | 9.7 | 46.2 | 18.3 |
| q18_5 The student didn`t put the patient`s life at risk | 7.5 | 8.6 | 16.1 | 23.7 | 44.1 |
| q18_6 The student did his/her best | 17.2 | 22.6 | 25.8 | 25.8 | 8.6 |
| q18_7 The student was aware of her/his own problems and wanted to improve | 29.0 | 21.5 | 11.8 | 28.0 | 9.7 |
| q18_8 There had been a lack of continuity in the mentoring because of holidays, sick leave etc. | 51.6 | 11.8 | 10.8 | 10.8 | 15.1 |
| q19_1 I was concerned about what my colleagues would think about me | 83.9 | 9.7 | 5.4 | 1.1 | 0.0 |
| q19_2 I was concerned that my nursing manager would not support my decision | 82.8 | 6.5 | 8.6 | 2.2 | 0.0 |
| q19_3 I experienced a lack of support from my colleagues | 74.2 | 14.0 | 3.2 | 6.5 | 2.2 |
| q19_4 I experienced a lack of support from my nursing manager | 76.3 | 10.8 | 9.7 | 2.2 | 1.1 |
| q19_5 The organizational culture made me feel it was my my personal responsibility to make the decision | 66.7 | 12.9 | 10.8 | 3.2 | 6.5 |
| q20_1 I believed the liaison lecturer would persuade me to to pass the student | 31.2 | 11.8 | 24.7 | 24.7 | 7.5 |
| q20_2 I experienced a lack of support from the liaison lecturer during the clinical placement | 19.4 | 15.1 | 19.4 | 31.2 | 15.1 |
| q20_3 I experienced a lack of support from the liaison lecturer in the decision | 17.2 | 8.6 | 20.4 | 20.4 | 33.3 |
| q20_4 I have previously experienced a lack of support in similar situations | 46.2 | 10.8 | 17.2 | 11.8 | 14.0 |

alpha coefficients) for the factors ranged from 0.65–0.85: factor 1:0.85; factor 2:0.81; factor 3:0.75; factor 4:0.75; and factor 5:0.65. The overall internal consistency reliability (Cronbach's alpha) was 0.84.

The chi-square statistic for the model produced a statistically significant finding ($\chi^2$ = 537.8, $df$ = 289, $p$ < 0.001). However, since the chi-square statistic is sensitive to large samples, it is preferable to use the standardized version, $\chi^2/df$. The size of this statistic was 1.86, with acceptable ratios being between 2.0–5.0. The standardized root mean squared residual (SRMR) was 0.11, with the preferable size being below 0.10. The root mean square error of approximation was

RMSEA = 0.096, also with a preferable size below 0.10. Taking the small number of observations into consideration, these fit indices suggested that the five-factor model of FTF was acceptable. The factor loadings presented in Figure 1 were all significant at $p$ < 0.05. The factor loadings varied between 0.40–0.85 (Figure 1).

## 4 | DISCUSSION

This study presented the mapping of the topic of Failing to Fail in a Norwegian nurse mentor context. The analyses identified and

**TABLE 3** Eigenvalues, percentage of explained variance and cumulative percentage of explained variance

| Factor | Eigenvalue | Percentage of variance | Cumulative percentage of variance |
|--------|-----------|------------------------|-----------------------------------|
| 1 | 6.302 | 35.97 | 35.97 |
| 2 | 2.784 | 15.89 | 51.86 |
| 3 | 1.935 | 11.04 | 62.91 |
| 4 | 1.607 | 9.17 | 72.08 |
| 5 | 1.396 | 7.97 | 80.05 |

tested the factor structure of the FTF by using EFA and CFA tests. Beginning with the EFA, we extracted five factors from the FTF. We modified our model by statistical analyses and theoretical insights, an approach that increases the chances of discovering a true model (Kline, 2011). Gorsuch (1983) claims that five respondents per variable would be enough for a reliable factor analysis, thus indicating a sample size of approximately 150 respondents in our study. Despite a sample of 93, we did an EFA to explore the research field. According to Wetzel (2012), factor analysis methods could explore validity in studies with fewer than 100 respondents.

A CFA examined the Cronbach's alpha in the subscales, which ranged from minimally acceptable (0.65) to very good (0.85) (Hair, Black, Babin, & Anderson, 2010). Even though subscale five had a minimal acceptable Cronbach's alpha, we chose to retain this subscale because it measured a different part of the phenomenon not measured in other subscales. The total Cronbach's alpha without subscale 5 had a negative impact on the overall Cronbach's alpha value. The overall Cronbach's alpha for the five-subscale model with 26 items was very good (0.84). According to Nunnally and Bernstein (1994), this value indicates that the internal consistency of the scores in the FTF was promising. They recommend 0.70 as an acceptable value for the early stages of the testing internal consistency of new research tools. This preliminary finding suggests that the questionnaire is a reliable tool for the measurement of FTF. A high internal consistency value for an instrument is essential because it indicates that rates appear to assess the items in a consistent way.

The chi-square test examined acceptable fit for the model. However, we know that the chi-square statistic is sensitive to sample size and any trivial mismatch can achieve statistical significance (Kline, 2011). This test fits small samples (Ringdal & Wiborg, 2017). On the other hand, small samples may obscure poor fit and yield less precise estimates of the parameters in CFA (West, Taylor, & Wu, 2012). Other fit indices indicated that SRMR (<0.1) provides support for an acceptable fit (Wang & Wang, 2012).

We labelled each factor according to the traits of the items. Factor one had eight items associated with challenges related to documentation, the mentor not being sufficiently prepared, the mentor not feeling confident or having enough strength to handle the situation and the mentor not being familiar with the formal procedure and feeling it was difficult to prioritize mentoring in addition to clinical tasks. According to Schau (1998), professionals are competent when

they are fit or qualified for what they do. We labelled the factor "Insufficient competence in mentoring." This result is consistent with results in previous research (Duffy, 2006; Black, 2011; Brown et al., 2012; Hauge, 2015; Jervis & Tilki, 2011; Luhanga, Yonge, & Myrick, 2008b; Rutkowski, 2007). The second factor contained five items and was labelled "Insufficient work environment support," which reflected concerns about what colleagues would think, concerns that the nursing manger would not support the mentor's decision, a lack of support from colleagues or nurse manager and the organizational culture making the mentor feel it was her/his personal responsibility to make the decision. Luhanga et al. (2008b) reported that some mentors had doubts about their decision and therefore had a great need for support.

The third factor was labelled "Emotional process dominates the assessment." It contained five items related to being labelled as a bad mentor or running into conflict with the student, liking the student, not wanting to hurt the student's feelings and not wanting to expose the student to the consequences of failing. Our findings are consistent with those of previous reports by Brown et al. (2012), Cleland, Knight, Rees, Tracey, and Bond (2008), Dudek, Marks, and Regehr (2005) and Luhanga et al. (2008b). Factor four contained items such as a lack of support from the liaison lecturer during the clinical placement in this decision or in similar previous situations, as well as believing that the liaison lecturer would persuade the mentor to pass the student. We labelled this factor "Insufficient support from university." A review by Hughes et al. (2016) also highlighted university support. The final factor reflected "Positive assessment detached from learning outcomes" and contained four items: giving the student the benefit of the doubt, the student not putting patients' lives at risk, the student doing his/her best and the student being aware of his/her own problem and wanting to improve. These findings are consistent with those of Cleland et al. (2008), Duffy (2006) and Jervis and Tilki (2011). The overall aim of nursing is to improve quality of care, thereby ensuring safety (Tella et al., 2013). Amsrud, Lyberg, and Severinsson (2015) found that clinical supervision influences the development of skills that are important for patients' safety and care. The fact that mentors state that students did not put patients' lives at risk and that they gave students the benefit of the doubt as reasons to allow them to pass is alarming with regard to patient safety.

The covariance between the factors in our model (Figure 1) is low. There is a significant and negative covariance between factor 4–5, meaning that insufficient support from the university in factor 4 will increase the decision-making detached from learning outcomes in factor 5. This finding indicates that support is important if we want to reduce the occurrence of mentors Failing to Fail nursing students. Hunt et al. (2016) investigated what enable mentors to fail underperforming students. They revealed that mentors needed to be secure to fail a student nurse and that mentors used a three-stage decision-making process to be ascertain in their decision. Furthermore, they identified the link lecturer as the key person in this process and that emotional support was more important than appraisal, instrumental and informational support. According to

**TABLE 4** Matrix of factor analysis with varimax rotation (30 items)

| Item | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 |
|------|----------|----------|----------|----------|----------|
| q17_1 | 0.05 | 0.31 | 0.36 | −0.03 | −0.24 |
| q17_2 | **0.65** | 0.05 | −0.10 | 0.07 | 0.06 |
| q17_3 | **0.56** | 0.13 | 0.24 | −0.16 | −0.12 |
| q17_4 | **0.70** | 0.23 | 0.36 | −0.06 | −0.11 |
| q17_5 | 0.04 | **0.43** | 0.31 | 0.08 | 0.03 |
| q17_6 | 0.25 | −0.00 | 0.14 | −0.29 | 0.07 |
| q17_7 | −0.03 | 0.24 | **0.52** | 0.22 | −0.01 |
| q17_8 | 0.35 | 0.17 | **0.60** | 0.19 | −0.22 |
| q17_9 | **0.40** | 0.24 | 0.20 | −0.19 | −0.26 |
| q17_10 | **0.85** | 0.06 | 0.03 | −0.01 | 0.18 |
| q17_11 | **0.77** | 0.07 | −0.02 | −0.09 | 0.05 |
| q17_12 | **0.45** | 0.06 | 0.29 | −0.02 | 0.23 |
| q17_13 | **0.56** | 0.33 | 0.44 | 0.05 | 0.05 |
| q18_1 | −0.27 | 0.15 | **0.49** | 0.09 | 0.16 |
| q18_2 | 0.12 | 0.16 | **0.76** | −0.03 | 0.17 |
| q18_3 | 0.19 | 0.16 | **0.67** | −0.26 | 0.15 |
| q18_4 | 0.36 | 0.04 | 0.23 | −0.18 | **0.44** |
| q18_5 | 0.11 | −0.02 | 0.19 | −0.13 | **0.53** |
| q18_6 | 0.20 | −0.09 | 0.05 | −0.06 | **0.55** |
| q18_7 | 0.03 | 0.13 | 0.07 | −0.09 | **0.60** |
| q18_8 | 0.12 | 0.22 | 0.05 | −0.12 | 0.28 |
| q19_1 | 0.04 | **0.75** | 0.33 | 0.02 | −0.12 |
| q19_2 | 0.11 | **0.73** | 0.15 | −0.10 | −0.07 |
| q19_3 | 0.13 | **0.55** | −0.11 | −0.05 | 0.18 |
| q19_4 | 0.31 | **0.64** | 0.20 | 0.21 | 0.09 |
| q19_5 | 0.23 | **0.63** | 0.02 | 0.01 | 0.25 |
| q20_1 | 0.08 | 0.21 | 0.06 | **0.63** | 0.11 |
| q20_2 | −0.02 | −0.06 | −0.01 | **0.72** | −0.12 |
| q20_3 | −0.09 | −0.09 | −0.06 | **0.75** | −0.19 |
| q20_4 | −0.06 | 0.19 | 0.09 | **0.55** | 0.20 |

*Note.* The numbers in bold indicate items with loadings above 0.40.

Hunt et al. (2016), four key attributes characterized a link lecturer who fulfilled the role: accessibility, approachability, authoritative knowledge about assessment and willingness to act as an emotional anchor. Our study revealed that insufficient support from the university increased mentors' decision-making detaches from learning outcomes, which is in line with Hunt et al. (2016). In other words, it appears that the support from the link lecturer is very central in preventing mentors from Failing to Fail underperforming students.

However, it seems to be a difference in Hunt et al. (2016) and our findings.

Hunt et al. (2016) emphasize the importance of emotional support because mentors feel emotionally stressful to stand in the assessment process of an underperforming student. In our study, we find no significant covariance between the factor "Emotional process dominate the assessment" and the factor "Decision-making detaches from learning outcomes." Based on our finding, it is difficult to conclude that emotional support is most important in preventing mentors from Failing to Fail underperforming students.

## 4.1 | Further development of the FTF scale

It will be of special interest to investigate whether large-scale studies can confirm the preliminary findings of the FTF as a multifactorial phenomenon (Five-factor model). As Nunnally and Bernstein (1994) note, "Most measures should be kept under constant surveillance to see if they are behaving as they should" (p. 87).

## 4.2 | Strengths and limitations

The possibility of making inaccurate predictions or assumptions is normal in small-scale studies (Van Teijlingen & Hundley, 2002). Therefore, we must interpret the results within the limitations set by
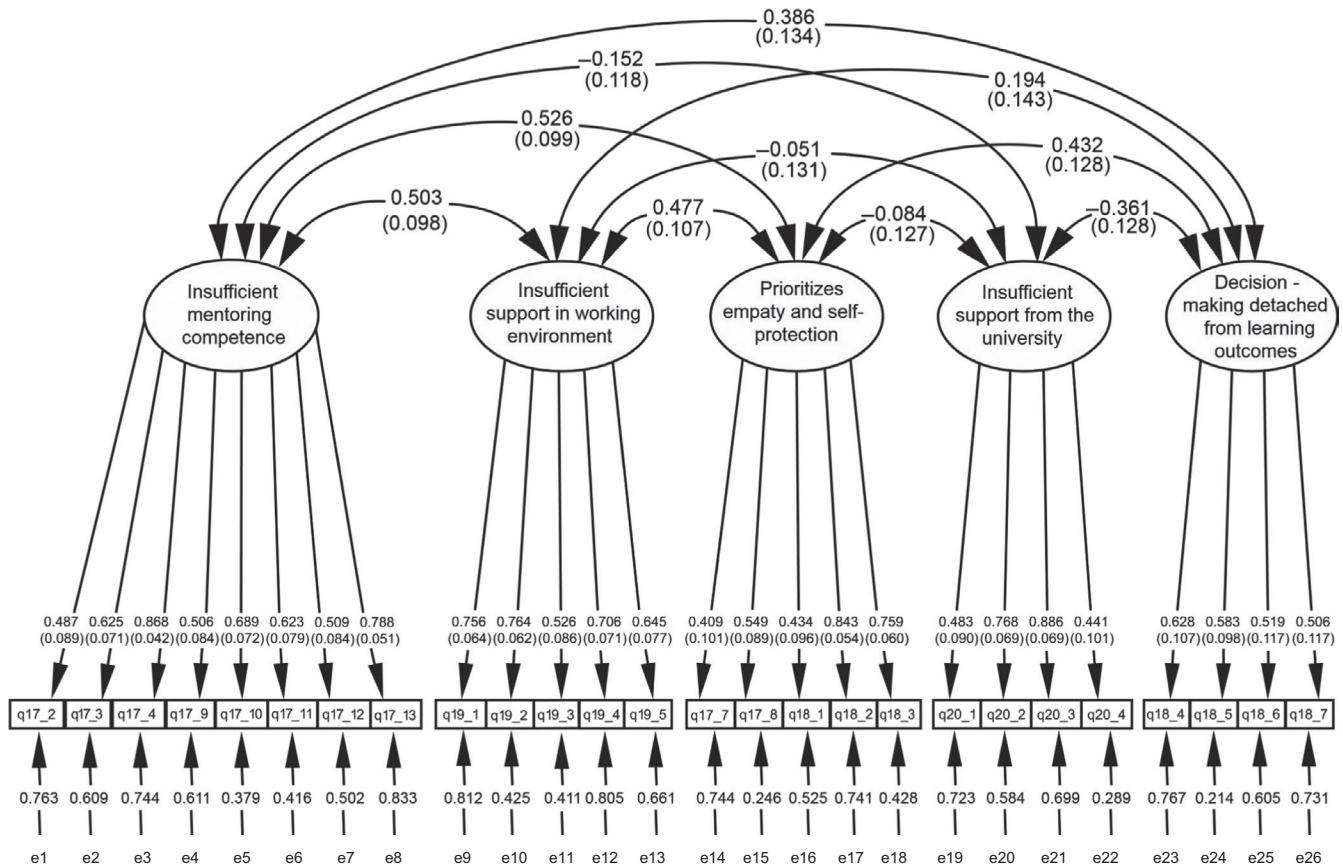
**FIGURE 1** The factor structure of Failing to Fail

the small-scale design of our research. Even though we had a small sample, 16.6% of nurse mentors confirmed that they had passed a student they thought should have failed, a result nearly consistent with the 18% found by Brown et al. (2012).

An important question in the development of a questionnaire is the feasibility of the questionnaire. The FTF scale was easy to administer and able to be completed within 5 minutes. Low rates of missing data can indicate that the questionnaire was easy to understand. This indicates a good feasibility. The fact that the respondents came from different parts of both municipal and specialist health services increases the likelihood that the participants were representative of nurse mentors. The demographic details of the participants' gender in the sample that responded to the questionnaire ($N$ = 561) and in the subsample of mentors who had passed a student they thought should fail ($N$ = 93) roughly reflect the Norwegian nurse population (Statistics Norway, 2017). Our model indicated that support is important. This finding increased external validity because it pointed out an important correlation between educational support and mentors' decision-making detached from learning outcomes.

## 5 | CONCLUSION AND IMPLICATIONS

In conclusion, the FTF scale is feasible for testing if mentors are Failing to Fail nursing students and their reasons for their decision.

The result of the CFA confirmed the factor structure of the FTF scale and demonstrated adequate model fit. This finding provides additional evidence for the internal consistency reliability of the scale. The CFA model supported the predictive validity of the FTF scale. We presume that the validated FTF scale will help professionals in educational institutions and nurse mentors identify the areas to avoid Failing to Fail students. Future research could further test the psychometric properties of the FTF scale and cross-validate the scale on mentors in other countries.

## CONFLICT OF INTEREST

The authors declare there is no conflict of interest.

## AUTHOR CONTRIBUTION

BL, GUKC, HWK and JS: Study design, data collection, data analysis and manuscript writing.

## ORCID

_Liv Bachmann_ (ID) https://orcid.org/0000-0002-4984-412X

_Signe Julnes_ (ID) https://orcid.org/0000-0003-0637-9333

## REFERENCES

Amsrud, K. E., Lyberg, A., & Severinsson, E. (2015). The influence of clinical supervision and its potential for enhancing patient safety – Undergraduate nursing students' views. _Journal of Nursing Education and Practice_, _5_(6), 87–95. https://doi.org/10.5430/jnep.v5n6p87

Araï, D., & Martinussen, M. (2010). _Kvantitativ forskningsmetodologi i samfunns- og helsefag [Quantitative research methodology in social- and health science]_. Bergen, Norway: Fagbokforlaget.

Black, S. (2011). Being a mentor who fails a pre-registration nursing student in their final placement: Understanding failure. PhD Thesis London South Bank University. Retrived from https://my.rcn.org.uk/_data/assets/pdf_file/0008/458234/Black_S_PhD_Thesis_August_2011.pdf

Brown, L., Douglas, V., Garrity, J., & Shepard, C. K. (2012). What influences mentors to pass or fail students. _Nursing Management_, _19_(5), 16–21. https://doi.org/10.7748/nm2012.09.19.5.16.c9260

Cleland, J., Knight, L. V., Rees, C. E., Tracey, S., & Bond, C. M. (2008). Is it me or is it them? Factors that influence the passing of underperforming students. _Medical Education_, _42_(8), 800–809. https://doi.org/10.1111/j.1365-2923.2008.03113.x

Dudek, N. L., Marks, M. B., & Regehr, G. (2005). Failure to fail: the perspectives of clinical Supervisors. _Academic Medicine_, _80_(Supplement), S84–S87. https://doi.org/10.1097/00001888-200510001-00023

Duffy, K. (2003). Failing students: a qualitative study of factors that influence the Decisions Regarding Assessment of students` Competence in Practice. Glasgow Caledonian University. Caledonian Nursing and Midwifery Research Centre School of Nursing, Midwifery and Community health.

Duffy, K. (2006). Weighing the Balance: A grounded theory study of the factors that influence the decisions regarding the assessment of student's competence in practice. Unpublished PhD Thesis, Glasgow: Glasgow Caledonian University.

Field, A. P. (2009). _Discovering statistics using SPSS_ (3rd ed.). London, UK: Sage.

Gainsbury, S. (2010). Mentors passing students despite doubts over their ability. _Nursing times_, _106_(16), 2–3.

Gorsuch, R. (1983). _Factor analysis_. Hillsdale, NJ: L. Erlbaum Associates.

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). _Multivariate data analysis: Global edition_ (7th ed.). Upper Saddle River, NJ: Pearson Higher Education.

Haraldsen, G. (1999). _Spørreskjemametodikk: Etter Kokebokmetoden [Survey methodology]_. Oslo, Norway: Ad Notam Gyldendal.

Hauge, K. W. (2015). Why are mentors failing to fail students who do not meet clinical performance standard? _Nordic Journal of Nursing Research_, _35_(1), 20–28. https://doi.org/10.1177/0107408314560325

Hauge, K. W., Brask, O. D., Bachmann, L., Bakken, H., Bergum, I. E., Grønvik, C. K. U., ... Ulvund, I. (2019). Are mentors failing to fail nurse students in practice? _Klinisk Sygepleje_, _33_(1), 23–29. http://doi.otg/10.18261/issn.1903-2285-2019-01-03

Hughes, L. J., Mitchell, M., & Johnston, A. N. B. (2016). "Failure to fail" in nursing – A catch phrase or a real issue? A systematic integrative literature review. _Nurse Education in Practice_, _20_, 54–63. https://doi.org/10.1016/j.nepr.2016.06.009

Hunt, L. A., McGee, P., Gutteridge, R., & Hughes, M. (2016). Failing securely: The processes and support which underpin English nurse mentors' assessment decisions regarding under-performing students. _Nurse Education Today_, _39_, 79–86. https://doi.org/10.1016/j.nedt.2016.01.011

Jervis, A., & Tilki, M. (2011). Why nurse mentors failing to fail student nurses who do not meet clinical performance standards? _British Journal of Nursing_, _20_(9), 582–587. https://doi.org/10.12968/bjon.2011.20.9.582

Jolliffe, I. T. (2002). _Principal component analysis_ (2nd ed.). New York, NY: Springer.

Kline, R. (2011). _Principles and practice of structural equation modelling_ (3rd ed.). New York, NY: Guilford.

Luhanga, F., Yonge, O. J., & Myrick, F. (2008a). "Failure to assign failing grades": Issues with grading the unsafe student. _International Journal of Nursing Education Scholarship_, _5_(1), 1–14. https://doi.org/10.2202/1548-923X.1366

Luhanga, F., Yonge, O. J., & Myrick, F. (2008b). Precepting an unsafe student: The role of the faculty. _Nurse Education Today_, _28_(2), 227–231. https://doi.org/10.1016/j.nedt.2007.04.001

Mehmetoglu, M., & Jakobsen, T. G. (2016). _Applied statistics using Stata: A guide for the social sciences_. London, UK: Sage.

Nunnally, J. C., & Bernstein, I. H. (1994). _Psychometric theory_ (3rd ed.). New York, NY: McGraw-Hill.

Ringdal, K., & Wiborg, Ø. (2017). _Lær deg Stata. Innføring i statistisk dataanalyse [Learn Stata. Introduction to statistical analyses]_. Bergen, Norway: Fagbokforlaget.

Rutkowski, K. (2007). Failure to fail: Assessing nursing students' competence during practice placements. _Nursing Standard_, _22_(13), 35–40. https://doi.org/10.7748/ns2007.12.22.13.35.c6299

Schau, G. M. (1998). _Gode fagfolk vokser. Personlig kompetanse som utfordring [Flourishing professional. Personal competence as a challenge]_. Oslo, Norway: Cappelen Akademisk Forlag.

Statistics Norway. (2017). Tabell: 07938: Personer med helse- og sosialfaglig utdanning, etter kjønn og alder [Persons with health- and social education, age and gender]. Retrieved from https://www.ssb.no/sok?sok=07938

Tella, S., Liukka, M., Jamookeeah, D., Smith, N.-J., Partanen, P., & Turunen, H. (2013). What do nursing students learn about patient safety? An integrative literature review. _Journal of Nursing Education_, _53_(1), 7–13. https://doi.org/10.3928/01484834-20131209-04

Van Teijlingen, E., & Hundley, V. (2002). The importance of pilot studies. _Nursing Standard_, _16_(40), 33–36. https://doi.org/10.7748/ns.16.40.33.s1

Wang, J., & Wang, X. (2012). _Wiley Series in Probability and Statistics_ (Vol. 12). Chichester, UK: John Wiley & Sons.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), _Handbook of structural equation modeling_ (pp. 209–231). New York, NY: Guilford Press.

Wetzel, A. P. (2012). Factor analysis methods and validity evidence: A review of instrument development across the medical education continuum. _Academic Medicine_, _87_(8), 1060–1069. https://doi.org/10.1097/ACM.0b013e31825d305d

Williams, B., Brown, T., & Onsman, A. (2012). Exploratory factor analysis: A five -step guide for novices. _Australasian Journal of Paramedicine_, _8_(3), 1–12.