

## Review

# 'Failure to fail' in nursing – A catch phrase or a real issue? A systematic integrative literature review



Lynda J. Hughes<sup>a, \*</sup>, Marion Mitchell<sup>b, c, d</sup>, Amy N.B. Johnston<sup>b, c, e</sup>

<sup>a</sup> Griffith University, School of Education and Professional Studies, Mt Gravatt, Qld, 4111, Australia

<sup>b</sup> Griffith University, School of Nursing and Midwifery, Nathan, Qld, 4111, Australia

<sup>c</sup> Centre for Health Practice Innovation, Menzies Health Institute Queensland, Griffith University, Nathan, Qld, 4111, Australia

<sup>d</sup> Nurse Practice Development Unit, Princess Alexandra Hospital, Ipswich Rd, Woolloongabba, Qld, 4102, Australia

<sup>e</sup> Gold Coast University Hospital, Emergency Department, 1 Hospital Blvd, Southport, Qld, 4215, Australia

## ARTICLE INFO

## Article history:

Received 21 July 2015

Received in revised form

27 June 2016

Accepted 27 June 2016

## Keywords:

'Failure to fail'

Undergraduate nurse

Grade inflation

Fitness for practice

## ABSTRACT

'Failure to fail' is the allocation of pass grades to nursing students who do not display satisfactory clinical performance. This issue can have significant implications for individual students and assessors involved, as well as for nursing professionalism and patient safety. The aim of this systematic integrative literature review was to determine what is currently known about the issue of 'failure to fail' within undergraduate nursing programs. A literature search of five databases up to May 2015 was conducted to identify primary research papers. The search yielded 169 papers of which 24 met the inclusion criteria. The majority of papers had moderate or good methodological rigour, with most of the literature originating from the Northern Hemisphere. Five main themes emerged: failing a student is difficult; an emotional experience; confidence is required; unsafe student characteristics; and university support is required to fail students. The results suggest that 'failure to fail' is a real issue in tertiary facilities, with many complex facets. Given the costs of nurse education and the potential social and professional costs of poor quality nursing graduates, further rigorous research is required in this area.

© 2016 Elsevier Ltd. All rights reserved.

## Contents

1. Introduction .....	55
2. Background .....	55
3. Aim .....	55
4. Methods .....	55
4.1. Search strategy .....	55
5. Results .....	60
5.1. Geographical distribution .....	60
5.2. Research design .....	60
5.3. Quality of research .....	61
5.4. Significant themes .....	61
5.5. Failing a student is difficult .....	61
5.6. Emotional process .....	61
5.7. University support .....	61
5.8. Confidence .....	61
5.9. Unsafe characteristics .....	61
5.10. Other themes .....	61
6. Discussion .....	61
7. Implications for practice .....	62

\* Corresponding author.

E-mail address: [Lynda.Hughes@griffithuni.edu.au](mailto:Lynda.Hughes@griffithuni.edu.au) (L.J. Hughes).

8. Limitations .....	62
9. Conclusions .....	63
Acknowledgements .....	63
References .....	63

## 1. Introduction

In nursing, the concept of ‘failure to fail’ is used in the literature to describe allocation of pass grades to nursing students who do not display satisfactory clinical practice. This systematic literature review collated primary research on ‘failure to fail’ and grade inflation in nursing preregistration programs. It has been identified that practising nurses are concerned that students can pass assessment of competencies and not be competent in fundamental nursing skills (Butler et al., 2011). If this is indeed the case, it is of significant concern as when a student achieves an accredited nursing qualification, they are deemed ‘competent’ to practice and this should equate to being able to perform to an acceptable professional and university standard (Fotheringham, 2010; Hunt et al., 2012). This is the view and reasonable assumption of the public, academics, colleagues and most importantly, potentially vulnerable patients. Accordingly, the International Council of Nurses (ICN) outlines safety as the most important principle in the assessment of nursing students (ICN, 2006). Whilst literature abounds with assessment of competence issues, there is an apparent paucity of research on ‘failure to fail’ in the nursing literature. This systematic review was conducted to make a meaningful contribution to what is known about ‘failure to fail’ in nursing through a methodological analysis, evaluation and presentation of past primary research.

## 2. Background

Competence in nursing is vital for safe practice and maintaining a high nursing proficiency which is in the interest of the general public (Nurse and Midwifery Board of Australia, 2013). Competence assessment involves forms of measurement, judgement and interpretation of students by different and/or multiple assessors (Calman et al., 2002; Fotheringham, 2010; Norman et al., 2002; Oermann et al., 2009). Furthermore, assessors are required to use professional judgement, drawing inferences and using tacit knowledge (Nurse and Midwifery Board of Australia, 2013). Failure on the grounds of unsatisfactory clinical performance has been reported to be quite rare (Butler et al., 2011; Hunt et al., 2012), with apparent reluctance from assessors to fail student nurses even when their practice was questionable or unsatisfactory (Duffy, 2003). Despite the potential professional consequences of ‘failing to fail’ there is relatively little published research exploring it.

This systematic review identifies what is currently known about ‘failure to fail’ in assessment of clinical practice within undergraduate nursing programs. This enables identification of the breadth, depth, type and quality of literature available on ‘failure to fail’ in undergraduate nursing to determine the extent of the problem and whether patient safety is compromised. Evaluation of existing literature will contribute to nursing knowledge, and highlight areas for further research.

## 3. Aim

Assessment of competence is a complex phenomenon and continues to be problematic. There appears to be evidence

emerging from the literature suggesting there is a reluctance to fail students of nursing who demonstrate unsatisfactory clinical practice. The aim of this systematic review is to determine: *What is currently known about ‘failure to fail’ in clinical practice within undergraduate nursing programs?*

## 4. Methods

The Mixed Method Appraisal Tool (MMAT), a critical appraisal tool developed for the concomitant review of qualitative, quantitative and mixed methods studies, was employed (Pace et al., 2012; Pluye et al., 2011), to conduct the quality analysis. An advantage of the MMAT is its unique ability to provide a rich, detailed and practical assessment of quantitative and qualitative research quality through the use of one tool (Pace et al., 2012; Pluye et al., 2009). The MMAT has an accompanying document which provides guidelines for use allowing for standardised interpretation of items (Pluye et al., 2011). A limitation of the MMAT is that it does not provide an appraisal for systematic reviews. The AMSTAR appraisal tool was chosen to appraise the quality of the systematic reviews for this study, as it was developed specifically to critically appraise the methodological quality of these reviews (Shea et al., 2007a) and is the only tool that has been validated for this use (Smith et al., 2011). AMSTAR has been reported as having good agreement, reliability, construct validity and feasibility in assessing the quality of systematic reviews (Shea et al., 2007b, 2009). Application of critical appraisal tools like these provides information for consumers of research, determining the veracity of results and the validity of transference of results to their particular application (Katrak et al., 2004).

### 4.1. Search strategy

A comprehensive search was conducted across the relevant health electronic databases including: CINHAI Plus, MEDLINE, ProQuest Nursing and Allied Health Source, Scopus and Informit as outlined in Fig. 1. These data bases were chosen as they effectively cover the vast majority of health and clinical journals. Key search terms and MESH headings were used in combinations and are outlined in Fig. 1.

Reference management software was used to manage and sort the records. After removal of 98 duplicates from the 267 papers sourced from the multiple databases, the remaining 169 papers were assessed by the first author for inclusion/exclusion based on the pre-set criteria, the content relevance of the title and abstracts. Seventeen full-text articles that met all the criteria were reviewed. Google Scholar was used to examine the reference lists from those seventeen papers for further suitable papers. Selection was undertaken based on the pre-defined inclusion criteria. From this process, seven additional papers were identified. Of the final 24 papers that met the selection criteria, seven were drawn from three studies (DeBrew and Lewallen, 2014; Lewallen and DeBrew, 2012; Luhanga et al., 2008a, 2008b, 2008c; Seldomridge and Walsh, 2006; Walsh and Seldomridge, 2005). For the purpose of this review, the methodological quality was assessed once for each study,

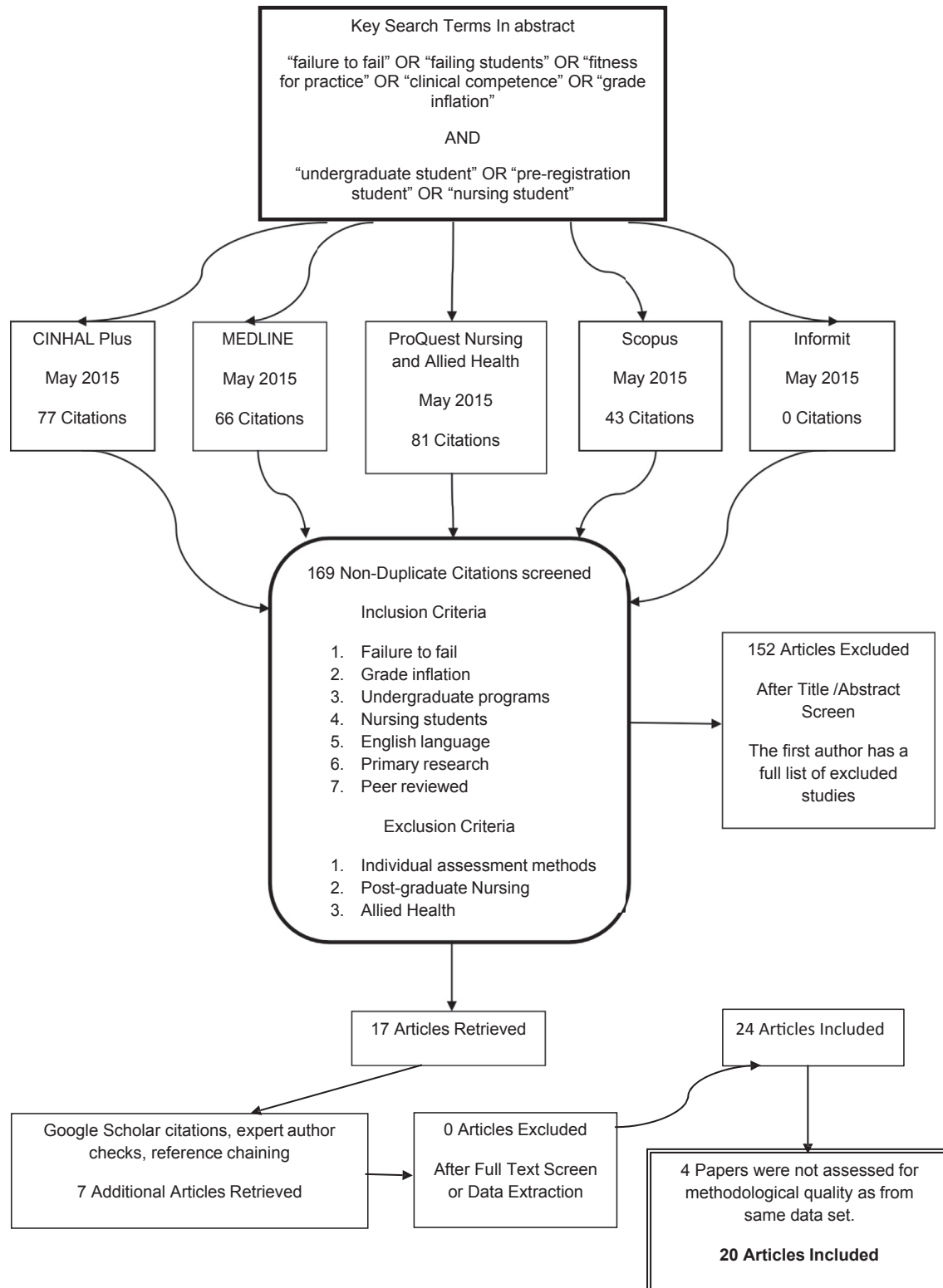


Fig. 1. Prisma Flow Chart of methodological processes used (Moher et al., 2009).

however the findings for all published papers were included as they reported different outcomes. This reduced the actual study count to twenty. The process used to refine and evaluate the records was recorded for each screening stage according to the PRISMA Statement (Moher et al., 2009) and is illustrated in Fig. 1. Each of the

twenty studies were independently assessed for quality by two authors (L.H. reviewed all articles and M.M. and A.J. each reviewed half of the articles) using either the MMAT or the AMSTAR tools. All three authors participated in quality appraisal and an inter-rater comparison of the two appraisers for each study was undertaken

**Table 1**

A summary of 24 studies which meet the review inclusion criteria on failure to fail and grade inflation in undergraduate nursing programs.

Author/country	Study design	Sample	Purpose	Main findings	Limitations	Quality score
Black et al., 2014 U.K.	Gadamer's Hermeneutics – from PhD	Assessors	Explore assessors' experiences of failing students.	Moral courage is required. Three themes identified: the personal price; a sense of professional responsibility; accountability.	Geographically limited. †	MMAT 100%
Brown et al., 2012 U.K.	Non-experimental survey	Assessors	Establish assessment practice at the University of West of Scotland.	18% of assessors passed failing students; 58% gave benefit of doubt; 8% believe the university would overturn; 4.5% lack confidence; most participants would contact the university however 5.5% did not respond.	Questions asked about University conducting the survey. Pilot tested but no example provided. No discussion of reliability. †	MMAT 50%
DeBrew and Lewallen, 2014 ¶ U.S.A	Qualitative Description	Assessors	To describe decision making regarding evaluation of a failing student.	Decisions difficult to quantify & ambiguous; process is difficult; assessors are gatekeepers & draw on experiences to make decision; communication most common reason by assessors for failing students.	Sample size, location and lack of diversity. The selection of individual participants from the selected schools was not explicit. No use of reflexivity. †	MMAT 50%
Lewallen and DeBrew, 2012			Characteristics of safe & unsafe students.	Characteristics that differentiate successful & unsuccessful are: communication; preparation; functioning in the clinical area		
Donaldson and Gray, 2012 U.K.	Systematic review	N/A	Exploring issues of grade inflation in clinical practice.	Benefits of grading practice are debated in literature. Reasons for grade inflation: Student related, assessor related, student assessor relationship and the tool. Use of rubrics is one way to address grade inflation.	Some papers were of low quality & opinion pieces. ‡	AMSTAR 6/11
Duffy, 2003 U.K.	Grounded Theory – from PhD	Assessors	Uncover assessors' experiences of failure to fail.	Assessors pass students despite student being weak. Patterns identified during students program. More students fail theory than practice. Decision to fail is difficult. Benefit of doubt given.	‡	MMAT 100%
Heaslip and Scammell, 2012 U.K.	Survey	Assessors & Students	To explore the issue of grading in practice.	Assessors were confident in grading practice. Assessors (70.5%) found descriptors useful but students	Small scale evaluation only. Questionnaire was not tested after stage one development with no example provided. ‡	MMAT 50%

(continued on next page)

Table 1 (continued)

Author/country	Study design	Sample	Purpose	Main findings	Limitations	Quality score
Hunt et al., 2012 U.K.	Survey – from PhD	Universities	Identify if pre-registration nursing students rarely fail practical assessments.	did not. Feedback was important with mentors (92.2%) feeling they gave it but students (12.5%) said they did not. Failing takes confidence, 17.9% lacked confidence. Discrepancies between theoretical (4%) & practical failure (0%) rates exist in U.K. Universities who failed no students in practice: Year 1 53%; Year 2 47%; Year 3 73%.	Consideration given to those that chose not to participate. Survey was deemed effective after pilot testing however no details were given. ‡	MMAT 50%
Jervis and Tilki, 2011 U.K.	Qualitative	Assessors	Why are assessors failing to fail?	Assessment is complex particularly in borderline students. Failing students is hard. Difficult to assess attitudes, interest and motivation. Lack of confidence to fail.	The authors outlined relationship with participants however no evaluation in relation to data analysis & findings was evident. ‡	MMAT 75%
Killam et al., 2011 Canada	Integrative review	N/A	Examine literature on unsafe nursing students.	Three themes identified: Ineffective interpersonal interactions; knowledge and skill incompetence; unprofessional image.	Potential for grey literature to not be identified. ‡	AMSTAR 6/11
Larocque and Luhanga, 2013 Canada	Qualitative Descriptive	Assessors	What are the perceptions of assessors on failure to fail?	Failing is difficult. Support is required for students & assessors. Failure has consequences for program, student & assessor. Personal, professional & structural reasons exist for not failing students.	Small sample size. Predetermined criteria such as previous experience present, but does not specify other criteria. No mention of self or reflexivity. ‡	MMAT 50%
Luhanga et al., 2014a,b Canada	Grounded Theory	Assessors	Strategies for managing unsafe students from assessor perspective.	Student success is primary concern. Major themes: recognise red flags; strategies for managing unsafe practice; strategies for success; decisions to fail are difficult; support and guidance required for all involved.	Sample size. The aim was to explore processes of managing unsafe students however experience with unsafe students was not a selection criteria. ‡	MMAT 75%
Luhanga et al., 2014a,b Canada	Qualitative Description	Assessors	Explore the issue of failure to fail in Canadian professional education programs.	Failing is difficult process. Academic & emotional support is required for students & assessors. Failure has consequences for program, student & assessor.	The paper did not outline rationale with predetermined criteria clearly. No consideration of self or reflexivity. ‡	MMAT 50%

Table 1 (continued)

Author/country	Study design	Sample	Purpose	Main findings	Limitations	Quality score
Luhanga et al., 2008a Canada	Grounded Theory – from PhD	Assessors	Processes assessors use to manage unsafe student practices.	Personal, professional & structural reasons exist for not failing students. Support for assessors is required, assessors felt guilt and self doubt, assessors require ongoing university support	The credibility of the researcher is briefly referred to however no discussion of self in relation to data analysis or reflexivity. ‡	MMAT 75%
2008b			Why is there a failure to assign failing grades?	Preceptors are gatekeepers. Preceptors acknowledged students pass without appropriate experience because: preceptor inexperience; students personal cost; extra workload; feelings of guilt; lack of tools & time to evaluate; nursing shortage.		
2008c			How assessors manage and deal with unsafe students.	Four categories of unsafe students: inability to demonstrate knowledge & skills, attitude problems, unprofessional behaviour & poor communication skills. Early identification & intervention is critical.		
Msiska et al., 2015 Malawi	Hermeneutical Phenomenology	Students	Explore the clinical learning experience of Malawian student nurses.	Bias present in clinical assessments, if the relationship is good, the evaluation is better and variations in clinical settings as learning environments.	No mention of self in relation to data analysis and reflexivity. ‡	MMAT 75%
Paskausky and Simonelli, 2014 U.S.A.	Descriptive Correlation Study	Students	Examine the discrepancy of final exam results with clinical grades.	Students perform better in practical grades. 90% had grade discrepancy of 5 or more (½ letter grade), 70% had 10 points (1 letter grade), 18% had 2 letter grade higher.	Single size and specific sample type. ‡	MMAT 100% §
Rittman and Osburn, 1995 U.S.A.	Qualitative Case Study	Preceptor	Understanding the experience of preceptoring an unsafe student.	Two major themes: know the student to enable planning of learning experience to ensure patient safety; Watchful listening to get a feel for the level of competence.	No mention of self in relation to data analysis & reflexivity. ‡	MMAT 75%
Scanlan and Care, 2004 Canada	Case Study	Faculty of nursing		Grade inflation exists. Students		MMAT 75%

(continued on next page)

Table 1 (continued)

Author/country	Study design	Sample	Purpose	Main findings	Limitations	Quality score
			To investigate if grade inflation occurs in the researchers faculty.	may overestimate skills. In clinical courses 90% of students achieved B+ or above with 60% A or A+. Final placement 80% A or A+ with only 3% of students getting B or lower.	No mention of self in relation to data analysis & reflexivity. †	
Seldomridge and Walsh, 2006 €	Quantitative Document Analysis	Students	Identify if leniency could account for grade inflation.	Grade inflation for clinical practice is significantly more than for theory. 95% of students grade A or B with 5% at level C.	Location, one university. ‡	MMAT 100% §
Walsh and Seldomridge, 2005 U.S.A.			Examine the relationship between grades in theory & practice.	Grade inflation for clinical practice is significantly more than for theory across all courses.		
Susmarini and Hayati, 2011 Indonesia	Qualitative Description	Clinical Facilitators	What is the experience of grade inflation?	Grade inflation is an issue. Three themes explored: causal factors; impact; solution.	The inclusion criteria are unclear. No data analysis section in the paper. No mention of self or reflexivity. †	MMAT 25%
Tanicala et al., 2011 U.S.A. & Canada	Qualitative	Nurse Educators	To examine views of what constitutes unsafe behaviours.	One major theme: context and patterns. Five Sub themes: safety; communication; thinking; ethics and standards.	Limited diversity & data from student & administrators. No mention of self in relation to data analysis & reflexivity. †	MMAT 75%

Table Notes: † = conflict of interest declared in paper, ‡ = no conflict of interest declared in paper, § = one criteria was not relevant thus the score was 100% of relevant criteria, ¶ = same data set as Lewallen and DeBrew 2012, € = same data set as Walsh and Seldomridge 2005, MMAT = Mixed Method Appraisal Tool.

until consensus in the final score was achieved.

## 5. Results

### 5.1. Geographical distribution

Of the twenty studies, the majority ( $n = 18$ ) were conducted in the Northern Hemisphere. Seven studies were conducted in the United Kingdom and four in the USA. It is important to note that results from one Canadian study were published across three

papers, and one author was involved in all but one of the studies. One study combined results from the USA and Canada. One study was reported from each Indonesia and Malawi.

### 5.2. Research design

Both qualitative and quantitative approaches were used to examine 'failure to fail'. There were also two systematic reviews conducted around this topic. One systematic review drew on information from 147 papers sourced from 1999 to 2009 around the

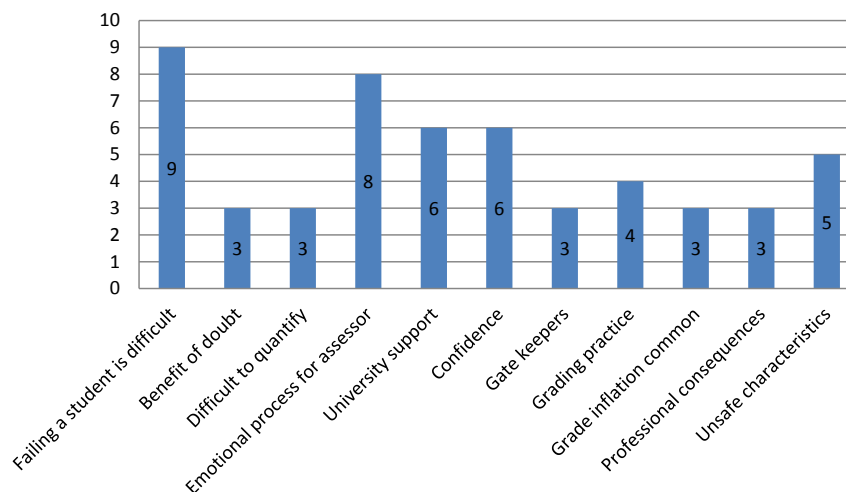


Fig. 2. Significant themes identified in the studies.

issues of grading in practice. However, this review included text and opinion pieces which may introduce bias as they are not based on primary research. The second systematic review was conducted in 2010 and synthesised data on unsafe nursing students in clinical practice. The authors identified twelve relevant manuscripts which included theses, theoretical articles and research articles.

Survey design and qualitative descriptive design were the two most favoured approaches to examining the issue. The use of grounded theory in numerous research reports supports the notion that very little is known about this phenomenon. The majority of the studies used a single site and a 'snapshot' of the issue being explored.

### 5.3. Quality of research

The methodological quality of both review studies was moderate with an AMSTAR score 7 out of 11 (see Table 1) (Canadian Agency for Drugs and Technologies in Health, 2011; Sequeira-Byron et al., 2011). The AMSTAR, however, was limited as many of the criteria were *not applicable* to these reviews. For example combining studies to test for heterogeneity is not applicable in the reviews included in this paper. Despite this limitation in the quality appraisal tool, the ability to categorise quality assessment ratings into ranges allows for easier interpretation for consumers of research and policy makers (Sequeira-Byron et al., 2011). There was a quality range in the remaining studies examined using the MMAT (see Table 1). In most of the qualitative papers reviewed, there was a distinct absence of consideration of the researchers' influence in the research process, particularly in relation to data analysis and the research findings.

### 5.4. Significant themes

Analysis of the themes related to 'failure to fail' in the reviewed literature highlighted five recurrent elements: failing a student is difficult; emotional process for assessor; university support; confidence; and unsafe characteristics of nursing students. Other themes presented in the literature are shown in Fig. 2. For clarity within this paper, individuals (preceptors, mentors, clinical supervisors, sign-off mentors, lecturers, and faculty) who assess students in practice are referred to collectively as 'assessors'.

### 5.5. Failing a student is difficult

That failing a student is a difficult process was identified in nine papers (Black et al., 2014; Brown et al., 2012; DeBrew and Lewallen, 2014; Duffy, 2003; Jarvis and Tilki, 2011; Larocque and Luhanga, 2013; Luhanga et al., 2014a,b; Luhanga et al., 2008b). All papers which identified this as a factor involved assessors as the sample.

### 5.6. Emotional process

Failing a student was found to be an emotional process for assessors in eight papers (Black et al., 2014; DeBrew and Lewallen, 2014; Duffy, 2003; Larocque and Luhanga, 2013; Luhanga et al., 2014a,b; Luhanga et al., 2008a, 2008b). Duffy's study (2003) in particular, identified that there were significant emotional consequences for both assessors and students when a student nurse was failed.

### 5.7. University support

University support was highlighted in six papers (Brown et al., 2012; Duffy, 2003; Larocque and Luhanga, 2013; Luhanga et al., 2014a,b; Luhanga et al., 2008b). It was identified to be essential,

and impacts on the experience both from the assessor's and student's point of view.

### 5.8. Confidence

Confidence on behalf of the assessors to fail a student nurse who is not performing was specifically identified in six papers (Black et al., 2014; Brown et al., 2012; Duffy, 2003; Heaslip and Scammell, 2012; Jarvis and Tilki, 2011; Luhanga et al., 2008b). Lack of confidence was an underlying theme present throughout the literature.

### 5.9. Unsafe characteristics

Five papers specifically sought to identify characteristics of unsafe practice in students (Killam et al., 2011; Lewallen and DeBrew, 2012; Luhanga et al., 2014a,b; Luhanga et al., 2008a, 2008c). The research identified that this information would help contribute to identification of borderline or unsafe students so that early intervention could be implemented.

### 5.10. Other themes

Other themes identified through this systematic review included: benefit of the doubt; assessors as gatekeepers to the profession; grade inflation in nursing and its impacts; biases in assessments whereby if the student nurse is nice they pass; and that communication and early intervention are good strategies to avoid failure to fail.

## 6. Discussion

This systematic review examines the research literature on what is known about failure to fail in undergraduate nursing programs. The literature confirms that 'failure to fail' is a real issue and is reported in the nursing literature. Like assessment generally, failure to fail emerges as a complex phenomena. A total of twenty studies released in 24 papers, were reviewed, with varying research quality. The majority of papers were either of moderate or good quality according to the critical appraisal tools utilised in this review. Generally the papers failed to acknowledge concepts of reflexivity and focused bias, with the qualitative studies lacking a clear philosophical perspective underpinning the research. It is important, particularly when undertaking qualitative descriptive research, that researchers explicitly identify their own assumptions regarding the research question and the assumptions the researcher makes about the research topic (Caelli et al., 2003). Declaring this at the outset helps to establish rigour and trustworthiness. Given the very limited information about 'failure to fail', it is essential that published data and conclusions from these data have veracity. Significant policy and process decisions around student assessment can have far reaching implications on the nursing profession and ultimately patient safety. These must be made on the basis of quality evidence.

The complexity of competence assessment is not a geographically isolated concern although each geographical area has its own methods and approaches to address competence assessment. Given the geographical coverage in this literature is mostly limited to North America and the United Kingdom, it remains to be seen if the issue of 'failure to fail' exists outside this geographical area. Further research should investigate whether failure to fail is common in other parts of the world including Asia and Europe. The pattern from this review suggests that this issue is likely to be a widespread problem.

Failing a student is perceived to be a difficult process (Black



et al., 2014; Brown et al., 2012; DeBrew and Lewallen, 2014; Duffy, 2003; Jervis and Tilki, 2011; Larocque and Luhanga, 2013; Luhanga et al., 2014a,b; Luhanga et al., 2008b). All papers which identified 'difficulties' as a factor in failure involved assessors as the sample. However, little objective data were presented to substantiate such difficulties or to identify specific 'difficult' components. Without these, development of future interventions to reduce such difficulties will remain challenging.

Studies also identified that failing a student often provoked strong emotional responses from assessors and was therefore, sometimes, avoided (Black et al., 2014; DeBrew and Lewallen, 2014; Duffy, 2003; Larocque and Luhanga, 2013; Luhanga et al., 2014a,b; Luhanga et al., 2008a, 2008b). Emotions reported by the assessors ranged from self doubt, to guilt, to failure being against the "caring" nature of the nursing profession. Several studies also discussed that the emotional response of the assessor was related to the personal responses from and consequences for the student (Duffy, 2003; Luhanga et al., 2014a,b). What was described as 'moral courage' was required for assessors to fail students (Black et al., 2014).

There seemed to be an assumption or perception from assessors that failure is also an emotional and harrowing experience for students. Whilst this may be a reasonable assumption, this needs to be examined by evaluating the experience of the students. Interestingly, there was only one study which used students as the sample, conducted in Malawi, which explored students' clinical learning experiences and the grading of students in practice (Msiska et al., 2015). Whilst this study used students as the sample group, the researchers did not explicitly examine perceptions of 'failure to fail' from the perspectives of students. The student's perspective appears to be missing in existing literature.

The confidence required to fail a student nurse appeared to be a theme present throughout the literature (Black et al., 2014; Brown et al., 2012; Duffy, 2003; Heaslip and Scammell, 2012; Jervis and Tilki, 2011; Luhanga et al., 2008b). This is particularly surprising in the United Kingdom as, following on from Duffy's 2003 study, a large educational initiative involving empowering assessors was implemented. It would be important to examine the factors around assessor confidence and the education and training that occurred in the early 2000's to see if it improved confidence.

The lack of staff perception of university support in failing students emerged in various facets, with some studies identifying that some assessors believed that the university would overturn their decision to fail, as well as a perceived pressure to pass students (Brown et al., 2012; Duffy, 2003; Luhanga et al., 2014a,b; Luhanga et al., 2008b). This aspect of research was limited, however, in that many studies conducted in this area were from single university sites and used the same data sets which may impact on the ability to generalise. No study reported actual instances of university support of staff or students in clinical assessment. Further exploration of university appeals systems and the patterns and distribution of fails across courses could be used to objectively confirm or refute such assumptions.

Other papers sought to identify the characteristics of unsafe students (Killam et al., 2011; Lewallen and DeBrew, 2012; Luhanga et al., 2014a,b; Luhanga et al., 2008a). This information could help contribute to early identification of borderline or unsafe students so that early intervention could be provided. That the identification of students who just 'got through' or were given the 'benefit of the doubt' did not occur early enough in their clinical learning to enable implementation of 'rescue strategies' was identified in two studies (Duffy, 2003; Luhanga et al., 2008a). Further research examining what aspects of the student nurse's clinical practice was deemed unsatisfactory to identify if there are links between student factors and assessment practices would be beneficial. Explicit identification of these factors may contribute to the development of

interventions to reduce failure to fail and support students in achieving success.

The notion of assessors as 'gatekeepers to the profession' was formalized in the United Kingdom by requiring sign-off mentors making the final decision about a students' fitness for practice (Hunt et al., 2012; Larocque and Luhanga, 2013; Tanicala et al., 2011). This embodies and formalizes the concept that assessors have an obligation to the profession and to the public, to care for the well-being of patients by ensuring nursing students are safe and competent to practise. What is yet to be clearly determined though, is what is known about patient safety amongst nursing students given the context of students passing who reportedly should not pass clinical courses. Furthermore, the literature demonstrates that grade inflation in clinical courses does occur and is a widespread issue (Donaldson and Gray, 2012; Hunt et al., 2012; Paskausky and Simonelli, 2014; Scanlan and Care, 2004; Seldomridge and Walsh, 2006; Susmarini and Hayati, 2011; Walsh and Seldomridge, 2005), and thus, may also impact on graduate nurse employment and ultimately patient safety.

Another negative ramification of grade inflation could be an overestimation of student perception of their own skill level, limiting or constraining their interest and/or ability to effectively reflect on their own practice, to self assess and to continue to learn. Exploration of a model whereby (over) confident students are corrected and even warned by subsequent assessors, react to this with indignation and even outrage, making it harder for assessors to maintain the 'moral courage' to fail them, allowing them to slip through with passes and thus the cycle repeats would be an interesting focus for future studies.

## 7. Implications for practice

In health care there is an increasingly widespread demand for graduate nurses to be 'fit for practice'. Further examination is required to fully understand this complex concept of 'fitness' and to ensure patients are safe. Further research is required to identify if 'failure to fail' is a universal problem in undergraduate nursing programs or restricted to those areas where research has thus far been conducted, perhaps stimulated by anecdotal or incidental observations of failure to fail. Further exploration of the aspects nursing students are actually failing in, when they do fail, clinical courses would support the development of effective future interventions. The majority of current literature focuses on how failing students is a difficult process for the assessors. Further research should be conducted in conjunction with students who have failed or nearly failed. This should include those students that have been placed on learning contracts or referred and subsequently passed, as these students may give clearer insight into what areas are identified as problematic and then the process/es by which they have overcome that particular issue.

## 8. Limitations

There are concerns about the common use of single populations used to obtain datasets in many of the papers, particularly given the relative scarcity of research conducted on this important phenomenon. A few expert authors in the field publish multiple papers, some from the same dataset, limiting the scope of focus used to address this significant phenomenon. This may bias review findings.

Only English language papers were reviewed so there may be research that has been conducted in non-English countries that would further illuminate this complex issue. Doctoral theses were also excluded from this review which may have excluded some emerging experts in this area. Only nursing literature was included

in this review, and it is acknowledged that there are many similarities across professional practice disciplines including nursing, medicine, allied health and teaching which may have contributed to the overall power of this review. The results should be interpreted with caution due to the variety in quality of the included papers.

## 9. Conclusions

This paper systematically reviewed the literature and synthesised the knowledge on 'failure to fail' in undergraduate nursing students. It provides policy makers, administrators, academics and those who assess student nurses in clinical practice, vital information regarding the real issue of 'failure to fail' and provides clear directions for future research. There is sufficient evidence in the literature, albeit of mixed quality, to establish that 'failure to fail' is indeed a real and significant issue. There are not, however, sufficient papers for quality triangulation of contributing factors and to understand the influences of the key stakeholders, students, assessors and the broader clinical consumer. There are many facets of 'failure to fail' in undergraduate nursing programs and more quality research exploring this complex phenomenon is required.

## Acknowledgements

The authors would like to acknowledge David Geelan for his advice and comments during the final phase of the manuscript development.

## References

- Black, S., Curzio, J., Terry, L., 2014. Failing a student nurse: a new horizon of moral courage. *Nurs. Ethics* 21 (2), 224–238. <http://dx.doi.org/10.1177/0969733013495224>.
- Brown, L., Douglas, V., Garrity, J., Kim Shepherd, C., 2012. What influences mentors to pass or fail students. *Nurs. Manag. - U. K.* 19 (5), 16–21.
- Butler, M.P., Cassidy, I., Quillinan, B., Fahy, A., Bradshaw, C., Tuohy, D., Tierney, C., 2011. Competency assessment methods - tool and processes: a survey of nurse preceptors in Ireland. *Nurse Educ. Pract.* 11 (5), 298–303. <http://dx.doi.org/10.1016/j.nepr.2011.01.006>.
- Calman, L., Watson, R., Norman, I., Redfern, S., Murrells, T., 2002. Assessing practice of student nurses: methods, preparation of assessors and student views. *J. Adv. Nurs.* 38, 516–523. <http://dx.doi.org/10.1046/j.1365-2648.2002.02213.x>.
- Caelli, K., Ray, L., Mill, J., 2003. 'Clear as mud': toward greater clarity in generic qualitative research. *Int. J. Qual. Methods* 2 (2), 1–13.
- DeBrew, J.K., Lewallen, L.P., 2014. To pass or to fail? Understanding the factors considered by faculty in the clinical evaluation of nursing students. *Nurse Educ. Today* 34 (4), 631–636. <http://dx.doi.org/10.1016/j.nedt.2013.05.014>.
- Donaldson, J.H., Gray, M., 2012. Systematic review of grading practice: is there evidence of grade inflation? *Nurse Educ. Pract.* 12 (2), 101–114.
- Duffy, K., 2003. Failing Students: a Qualitative Study of Factors that Influence the Decisions Regarding Assessment of Students' Competence in Practice. Caledonian Nursing and Midwifery Research Centre Glasgow.
- Fotheringham, D., 2010. Triangulation for the assessment of clinical nursing skills: a review of theory, use and methodology. *Int. J. Nurs. Stud.* 47 (3), 386–391. <http://dx.doi.org/10.1016/j.ijnurstu.2009.09.004>.
- Heaslip, V., Scammell, J.M.E., 2012. Failing underperforming students: the role of grading in practice assessment. *Nurse Educ. Pract.* 12 (2), 95–100.
- Hunt, L.A., McGee, P., Gutteridge, R., Hughes, M., 2012. Assessment of student nurses in practice: a comparison of theoretical and practical assessment results in England. *Nurse Educ. Today* 32 (4), 351–355.
- International Council of Nurses, 2006. Continuing Competence as a Professional Responsibility and a Public Right. International Council of Nurses, Geneva.
- Jervis, A., Tilki, M., 2011. Why are nurse mentors failing to fail student nurses who do not meet clinical performance standards? *Br. J. Nurs.* 20 (9), 582–587.
- Katrak, P., Bialocerkowski, A., Massy-Westropp, N., Kumar, V.S., Grimmer, K., 2004. A systematic review of the content of critical appraisal tools. *BMC Med. Res. Methodol.* 4 (1), 22.
- Killam, L.A., Luhanga, F., Bakker, D., 2011. Characteristics of unsafe undergraduate nursing students in clinical practice: an integrative literature review. *J. Nurs. Educ.* 50 (8), 437–446. <http://dx.doi.org/10.3928/01484834-20110517-05>.
- Larocque, S., Luhanga, F.L., 2013. Exploring the issue of failure to fail in a nursing program. *Int. J. Nurse Educ. Scholarsh.* 10 <http://dx.doi.org/10.1515/ijnes-2012-0037>.
- Lewallen, L., DeBrew, J., 2012. Successful and unsuccessful clinical nursing students. *J. Nurs. Educ.* 51 (7), 389–395. <http://dx.doi.org/10.3928/01484834-20120427-01>.
- Luhanga, F., Koren, I., Yonge, O., Myrick, F., 2014a. Strategies for managing unsafe precepted nursing students: a nursing faculty perspective. *J. Nurs. Educ. Pract.* 4 (5), p116.
- Luhanga, F., Yonge, O., Myrick, F., 2008a. Precepting an unsafe student: the role of the faculty. *Nurse Educ. Today* 28 (2), 227–231. <http://dx.doi.org/10.1016/j.nedt.2007.04.001>.
- Luhanga, F., Yonge, O.J., Myrick, F., 2008b. Failure to assign failing grades": issues with grading the unsafe student. *Int. J. Nurs. Educ. Scholarsh.* 5 (1), 1–14.
- Luhanga, F., Yonge, O., Myrick, F., 2008c. Hallmarks of unsafe practice: what preceptors know. *J. Nurs. Staff Dev.* 24 (6), 257–264. <http://dx.doi.org/10.1097/01.NND.0000342233.74753.ad>.
- Luhanga, F.L., Larocque, S., MacEwan, L., Gwewerere, Y.N., Danyluk, P., 2014b. Exploring the issue of failure to fail in professional education programs: a multidisciplinary study. *J. Univ. Teach. Learn. Pract.* 11 (2). <http://ro.uow.edu.au/jutlp/vol11/iss2/3>.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The, P.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6 (7), e1000097. <http://dx.doi.org/10.1371/journal.pmed.1000097>.
- Msiska, G., Smith, P., Fawcett, T., Munkhondya, T.M., 2015. Where is the grade coming from? Problems and challenges in evaluating the clinical performance of nursing students. *Open J. Nurs.* Vol. 05No.05 12. <http://dx.doi.org/10.4236/ojn.2015.55050>.
- Norman, I.J., Watson, R., Murrells, T., Calman, L., Redfern, S., 2002. The validity and reliability of methods to assess the competence to practice of pre-registration nursing and midwifery students. *Int. J. Nurs. Stud.* 39, 133–145. doi:S0020-7489(01)00028-1.
- Nurse and Midwifery Board of Australia, 2013. Framework for assessing national competency standards for registered nurses, enrolled nurses and midwives. Sourced from: <http://www.nursingmidwiferyboard.gov.au/Codes-Guidelines-Statements/Codes-Guidelines/Framework-for-assessing-national-competency-standards.aspx>.
- Oermann, M., Yarbrough, S.S., Saewert, K.J., Ard, N., Charasika, M., 2009. Clinical evaluation and grading practices in schools of nursing: national survey findings part II. *Nurs. Educ. Perspect.* 30, 352–357. <http://dx.doi.org/10.1043/1536-5026-30.6.352>.
- Pace, R., Pluye, P., Bartlett, G., Macaulay, A.C., Salsberg, J., Jagosh, J., Seller, R., 2012. Testing the reliability and efficiency of the pilot mixed methods appraisal tool (MMAT) for systematic mixed studies review. *Int. J. Nurs. Stud.* 49 (1), 47–53. <http://dx.doi.org/10.1016/j.ijnurstu.2011.07.002>.
- Paskausky, A.L., Simonelli, M.C., 2014. Measuring grade inflation: a clinical grade discrepancy score. *Nurse Educ. Pract.* 14 (4), 374–379. <http://dx.doi.org/10.1016/j.nepr.2014.01.011>.
- Pluye, P., Gagnon, M.-P., Griffiths, F., Johnson-Lafleur, J., 2009. A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *Int. J. Nurs. Stud.* 46 (4), 529–546. <http://dx.doi.org/10.1016/j.ijnurstu.2009.01.009>.
- Pluye, P., Robert, E., Cargo, M., Bartlett, G., O' Cathain, A., Griffiths, F., Boardman, F., Gagnon, M.P., Rousseau, M.C., 2011. Proposal: a mixed methods appraisal tool for systematic mixed studies reviews. Retrieved on 15th May 2015 from: <http://mixedmethodsappraisaltoolpublic.pbworks.com>.
- Rittman, M.R., Osburn, J., 1995. An interpretive analysis of precepting an unsafe student. *J. Nurs. Educ.* 34 (5), 217–221.
- Scanlan, J., Care, W., 2004. Grade inflation: should we be concerned? *J. Nurs. Educ.* 43 (10), 475–478.
- Seldomridge, L.A., Walsh, C.M., 2006. Evaluating student performance in undergraduate preceptorships. *J. Nurs. Educ.* 45 (5), 169–176.
- Sequeira-Byron, P., Fedorowicz, Z., Jagannath, V.A., Sharif, M.O., 2011. An AMSTAR assessment of the methodological quality of systematic reviews of oral healthcare interventions. *J. Appl. Oral Sci.* 19, 440–447.
- Shea, B., Grimshaw, J., Wells, G., Boers, M., Andersson, N., Hamel, C., Bouter, L., 2007a. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med. Res. Methodol.* 7 (1), 10.
- Shea, B.J., Bouter, L.M., Peterson, J., Boers, M., Andersson, N., Ortiz, Z., Grimshaw, J.M., 2007b. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One* 2 (12). <http://dx.doi.org/10.1371/journal.pone.0001350>.
- Shea, B.J., Hamel, C., Wells, G.A., Bouter, L.M., Kristjansson, E., Grimshaw, J., Boers, M., 2009. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J. Clin. Epidemiol.* 62 (10), 1013–1020.
- Smith, V., Devane, D., Begley, C., Clarke, M., 2011. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med. Res. Methodol.* 11 (1), 15.
- Susmarini, D., Hayati, Y., 2011. Grade inflation in clinical stage. *Am. J. Health Sci. (AJHS)* 2 (1), 21–28.
- Tanicala, M.L., Scheffer, B.K., Roberts, M.S., 2011. Pass/fail nursing student clinical behaviors phase I: moving toward a culture of safety. *Nurs. Educ. Perspect.* 32 (3), 155–161.
- Walsh, C., Seldomridge, L., 2005. Clinical grades: upward bound. *J. Nurs. Educ.* 44 (4), 162–168.