

Assessing core outcomes in graduates: Psychometric evaluation of the Paediatric Intensive Care Unit - Nursing Knowledge and Skills Test

Author

Long, Debbie A, Mitchell, Marion L, Young, Jeanine, Rickard, Claire M

Published

2014

Journal Title

Journal of Advanced Nursing

DOI

[10.1111/jan.12241](https://doi.org/10.1111/jan.12241)

Rights statement

© 2014 Blackwell Publishing. This is the pre-peer reviewed version of the following article: Assessing core outcomes in graduates: Psychometric evaluation of the Paediatric Intensive Care Unit – Nursing Knowledge and Skills Tes, Journal of Advanced Nursing, Volume 70, Issue 3, pages 698–708, 2014, which has been published in final form at [dx.doi.org/10.1111/jan.12241](https://doi.org/10.1111/jan.12241).

Downloaded from

<http://hdl.handle.net/10072/55861>

Griffith Research Online

<https://research-repository.griffith.edu.au>



Assessing core outcomes in graduates: Psychometric evaluation of the Paediatric Intensive Care Unit – Nursing Knowledge and Skills Test.

Journal:	<i>Journal of Advanced Nursing</i>
Manuscript ID:	JAN-2013-0260.R1
Manuscript Type:	Original Article
Keywords:	Critical Care, Child Nursing, Psychometric Testing, Staff Development, Instrument development
Category:	Nursing

SCHOLARONE™
Manuscripts

Copy

ABSTRACT

Aim. To develop and psychometrically test the Paediatric Intensive Care Unit – Nursing Knowledge and Skills Test - a multiple choice test for measuring the key nursing knowledge and skills required for safe, competent practice.

Background. Intensive care graduate nurse residency or orientation programs are key strategies in the development of safe and competent practitioners. Essential to these programs are an evaluation of knowledge and skills. Multiple-choice examinations provide a valuable way of evaluating broad knowledge and skills, however there has been limited work in this area to date.

Design. Psychometric evaluation.

Methods. The instrument was administered to 79 nurses from four paediatric intensive care units in Australia and New Zealand over an 18 month period between 2008 and 2010. Internal consistency using Kuder-Richardson 20, item analysis, and construct validity using the “known groups” technique were explored.

Results. Kuder-Richardson 20 reliability estimate for the 109 item test was 0.85. Instrument scores were significantly higher amongst nurses with postgraduate education, and more years of paediatric intensive care experience. Item difficulty indices ranged from .08 to 1, with a mean item difficulty of .66. Item discrimination ranged from .2 to .8.

Conclusion. Testing of the instrument demonstrated encouraging psychometric properties. With additional refinement, this tool could provide educators and managers with an instrument to assist in the assessment of knowledge and skill acquisition. The instrument requires further testing in different samples of paediatric intensive care nurses to enable validation in other settings and cross-cultural comparisons.

SUMMARY STATEMENT

Why is this research needed?

- New nursing graduates in critical care environments require specialised orientation programs to facilitate the development of safe and competent practitioners.
- In paediatric intensive care graduate programs, recruitment and retention outcomes have been successfully reported, however professional outcomes such as knowledge and skill acquisition warrant further attention.
- While performance assessments are useful in the evaluation of graduate nurse competence, multiple choice examinations and their contribution to overall competency assessment have received little attention.

What are the key findings?

- This is the first published paediatric intensive care nursing specific tool to be rigorously developed and psychometrically evaluated.
- The initial development and testing of the Paediatric Intensive Care Unit-Nursing Knowledge and Skills Test suggests a reliable and valid measure of paediatric intensive care nursing knowledge and skills required for safe, competent practice following a graduate nurse program.
- Levels of paediatric intensive care experience and education are independently associated with higher test scores.

1
2
3 **How should the findings be used to influence policy/ practice/ research/**
4
5 **education?**
6

- 7
- 8 • The Paediatric Intensive Care Unit – Nursing Knowledge and Skills Test could
9 facilitate the achievement of both curriculum and professional development
10 learning objectives for nurses caring for children in the paediatric intensive care
11 environment.
12
 - 13 • Given the cost of delivering specialised orientation programs, knowledge and
14 skill acquisition are important outcomes to report, along with recruitment and
15 retention figures, when determining the success of graduate nurse programs.
16
 - 17 • Further exploration of the knowledge and skills levels of paediatric intensive care
18 nurses internationally would be useful in establishing a benchmark and
19 understanding variations in education and practice.
20
21
22
23
24
25
26
27
28
29
30
31

32 **Keywords:** paediatric intensive care, graduate nurse, examination, multiple-choice,
33 psychometrics, item analysis
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Historically, nurses entered intensive care after a period of time spent refining basic skills in ward areas. With the current shortage of nurses, particularly in critical care areas, recruitment of experienced nurses into critical care through these traditional channels has been problematic (Shelton 2003). The impact of this shortage has serious implications for critical care nursing because the sickest patients require the highest nurse-to-patient ratio and are likely to need the attention of the most highly trained nurses. To address this shortage, critical care areas are employing graduate nurses directly from their undergraduate nursing programs. This raises some challenges for nurse educators as both graduates and the specialty area of intensive care require unique approaches to their orientation programs.

Ultimately, the overall goal of the orientation program is to prepare and retain safe competent practitioners, encourage critical thinking, foster integration into the system and profession, and promote the skills of lifelong learning (Boyle 1998, Dunn 1992, Thomason 2006). Outcomes of intensive care unit (ICU) orientation programs have been predominantly measured in the literature through recruitment, retention and core nursing competencies. There is clear evidence that many ICUs have developed comprehensive orientation programs to successfully attract and retain nursing staff. To ensure the provision of safe, quality care in our ICUs, the profession must also ensure that the novice nurses recruited via orientation programs are provided with adequate specialty preparation for practice. ICU orientation programs are resource intensive and costly, therefore it is important that professional development and quality of care outcomes are also rigorously evaluated.

Background

Evaluation is vital to good practice in delivering a cycle of continuously improving education in nursing (Lambert 2012, O'Shea 2002). As with any educational program, systematic program evaluation is also required for orientation programs to measure educational effectiveness and assure public accountability (Pross 2010). The greatest effect of orientation program evaluations lies in the ability to demonstrate the transference of learning to the bedside.

Professional development outcome measures reported in the literature include: self-reporting of knowledge and confidence (Cavanaugh and Huse 2004), critical thinking (Kaddoura 2010, Square 2010), pre and post-testing of knowledge and clinical assessment tasks (Chestnutt and Everhart 2007, Duvall 2009, Hall and Marshall 2006, Morris et al. 2009, Peterson and Van Buren 2006, Square 2010, Wynd 2002) and competency based assessments (Rashotte and Thomas 2002). Many authors report improved professional development in the areas of critical thinking, confidence, knowledge and skills (Cavanaugh and Huse 2004, Square 2010, Stefanski and Rossler 2009), however most authors fail to describe the measurement instruments used in reaching these results.

As teaching, learning and evaluating are intimately related, any evaluation must be closely aligned with the curriculum and provide adequate, timely and meaningful feedback to nurses wherever possible. Just as nursing graduates are provided with many different types of learning experiences from classroom to clinical setting, multiple methods must also be utilised to evaluate their learning across all aspects of intensive care nursing (Mahara 1998, Watson et al. 2002, Redfern et al. 2002,

1
2
3 Murrells et al. 2009). Combining multiple methods of evaluation adds breadth and
4
5 depth to determining safe, competent practice (Mahara 1998).
6
7

8
9
10 Several generic tools have been developed which measure perceived competence
11 (Watson et al. 2002) and evaluate broad domains such as assessment, planning and
12 evaluation. Despite the reflective benefits of completing self-assessments (Cowin et
13 al. 2008, Meretoja et al. 2004), these generic tools can fail to capture the context
14 specific knowledge and skills that characterise clinical practice in speciality areas
15 (McGrath et al. 2006). While Objective Structured Clinical Examinations (OSCEs),
16 simulation, and Clinical Performance Assessment Tool (CPAT) assessments are all
17 reported as useful for specific clinical objectives, they may be inadequate or
18 inappropriate if a broad range of knowledge and skills need to be assessed. Written
19 tests, particularly multiple choice questions, can be extremely valuable and reliable
20 for assessing knowledge acquisition that is prerequisite for problem solving, decision
21 making and critical thinking required in competent, safe practice (Downing and
22 Yudkowsky 2009, Haladyna 1997). This method of assessment, however, has received
23 minimal attention in the paediatric intensive care literature, possibly as a result of the
24 time and energy required to develop and evaluate the tests.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 Some ICU orientation programs list the American Association of Critical Care's
46 Essentials of Critical Care Orientation (ECCO) (or its paediatric version) as
47 containing MCQ tests for evaluation purposes, however minimal details are provided
48 regarding the test construction or its reliability and validity (American Association of
49 Critical-Care Nurses 2010, American Association of Critical-Care Nurses 2011,
50 Chestnutt and Everhart 2007, Friedman et al. 2011, Morris et al. 2007, Peterson and
51
52
53
54
55
56
57
58
59
60

1
2
3 Van Buren 2006). Furthermore, the costs involved in implementing and maintaining
4 the ECCO program often make it uneconomical for small or independent
5 organisations (Duvall 2009). The Basic Knowledge Assessment Tool for Pediatric
6 Intensive Care Nurses (PEDS-BKAT) is the only existing independent assessment
7 instrument for the paediatric intensive care unit (PICU) population (Runton and Toth
8 1998). However, since the PEDS-BKAT4's development in 1998, there has been
9 limited reported reliability or validity testing. Long, Young, Rickard, and Mitchell
10 (2012) tested the PEDS-BKAT4 and found that the instrument was not a reliable and
11 valid measure of basic PICU nursing knowledge in the Australian and New Zealand
12 context. The authors suggested that some of the questionable reliability and validity
13 reflected variations in practice between countries and therefore further research into
14 the knowledge and skills of paediatric intensive care nurses in Australia and New
15 Zealand would be extremely useful (Long et al. 2012).
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

34 **THE STUDY**

36 **Aim**

37
38 The aim of this study was to test the psychometric properties of the Paediatric
39 Intensive Care Unit – Nursing Knowledge and Skills Test (PICU-NKST) in a national
40 sample of **Australian and New Zealand** PICU nurses.
41
42
43
44
45
46
47

48 **Methodology**

49
50 A prospective methodological study was conducted to evaluate the psychometric
51 properties of the 109 item PICU-NKST, and included internal consistency reliability,
52 item analysis and construct validity.
53
54
55
56
57
58
59
60

Participants

A convenience sample of staff from four PICUs in Australia and New Zealand were invited to participate in the psychometric testing of the PICU-NKST, via an email to each Nurse Unit Manager. These four hospitals represent over 50% of PICU admissions in the region each year (Australian and New Zealand Intensive Care Society 2010). Participants were eligible if they were permanently employed as a registered nurse in a clinical, education or management role within the PICU. PICU nurses who met the inclusion criteria were approached by a nurse researcher or nurse educator in each participating unit. After informed consent was obtained, participants completed the test under 'exam conditions' in their PICU using a pencil and the mark sense sheet provided, and an information sheet describing their demographic characteristics. The completed test and information sheet were placed in a sealed envelope by the participant and given to the test invigilator (nurse researcher or nurse educator). Data was collected over an 18 month period between 2008 and 2010.

Instrument

Downing and Haladyna's (1997) model for item validity evidence guided the methodology of the PICU-NKST development, whereby an initial practice analysis was conducted to establish content validity (Long et al. 2013). Using information provided from the practice analysis, a content by process test plan was developed which outlined the patient care problems and the activities undertaken by PICU nurses. Using a weighting analysis (Raymond and Neustel 2006) and further consensus from 15 nurse educators, the final test plan outlines the generation of 109 items assessing PICU nursing knowledge and skills (Figure 1), called the Paediatric Intensive Care Nursing Knowledge and Skills Test (PICU-NKST).

1
2
3
4
5 Development of the PICU-NKST items were guided by Haladyna, et al.'s (2002)
6 principles for effective item writing. Consensus following distribution of the draft of
7 109 items to the panel of four independent experts indicated minor item writing flaws
8 which required further revision. Due to word limitations, further information of the
9 PICU-NKST's item development is detailed in Long (2012).
10
11
12
13
14
15
16
17

18 **Data Analysis**

19
20 Data were analysed using the statistical program SPSS version 17 (SPSS
21 Incorporated, Chicago, IL, USA). Both descriptive and inferential analyses were used
22 in the psychometric evaluation of the PICU-NKST. Continuous data were calculated
23 as means and standard deviations, categorical data as counts and percentages.
24
25 Statistical significance was set at $p < .05$.
26
27
28
29
30
31
32
33

34 Internal consistency reliability was assessed using the Kuder Richardson 20 (KR-20)
35 estimate. A KR-20 of .7 to .79 was considered the acceptable range for initial low-
36 stakes test development (Axelson and Kreiter 2009). Item responses were recoded
37 dichotomously for analysis (0 = incorrect, 1 = correct). Quality of the test was further
38 examined using standard error of measurement and 90% score band (confidence
39 intervals) calculations: $s\sqrt{1-r}$ (where s = standard deviation, r = reliability); and $1.65 \times$
40 SEM, respectively.
41
42
43
44
45
46
47
48
49
50

51 To assess the performance of individual items in the PICU-NKST an item analysis
52 was undertaken, including item difficulty, item discrimination and distractor analysis.
53
54 The following formula was used to calculate the item difficulty: $D = c/n$ (where D is
55
56
57
58
59
60

1
2
3 item difficulty, c is the number of correct responses, and n is the number of
4 respondents) (Downing 2009). The higher the item difficulty, the easier the question
5 is. Item difficulty indices were also recoded in the following categories: easy ($>.8$);
6 moderate (.4 to .8); and difficult ($<.4$) (Tarrant and Ware 2010, Osterlind 1998). Item
7 discrimination was calculated with the following formula: $ID = (a-b)/n$ (where ID is
8 item discrimination, a is the response frequency of the upper quartile (75th percentile
9 and above), b is the response frequency of the lower quartile (25th percentile and
10 below), and n is the number of respondents in the upper quartile (75th percentile and
11 above)) (Downing 2009). A negative value means that participants receiving a low
12 score in the test tended to select the correct option more than a higher-scoring
13 participant. Conversely, a positive value for this index means that higher-scoring
14 respondents were more likely to select the response more often (Downing 2009). Item
15 discrimination indices were recoded into the following categories: poor ($<.10$); low
16 (.10 to .19); acceptable (.20 to .29); good (.30 to .39); and excellent (0.40) (Tarrant
17 and Ware 2010, Ebel and Frisbie 1986). **An ideal test item is one in which each
18 incorrect option (distractor) is chosen by some examinee's who do not know the
19 content being tested.** Distractor performance was evaluated using two criteria to
20 define non-functioning distractors: those chosen by fewer than 5% of participants and
21 those with a positive discrimination index (Tarrant and Ware 2010, Haladyna and
22 Downing 1988).

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49 To evaluate the construct validity of the PICU-NKST, the “known groups” technique
50 was used to test the relationship between test scores and two pertinent nurse
51 characteristics: level of education and years of PICU experience. For the level of
52 education known groups’ analysis, the sample was divided into five participant
53
54
55
56
57
58
59
60

1
2
3 groups: diploma in nursing, bachelor of nursing, postgraduate certificate, postgraduate
4
5 diploma, and master's degree. In the experience known groups' analysis, the sample
6
7 was divided into three groups: those with less than two years of PICU experience,
8
9 those with 2-5 years experience, and those with more than 5 years of experience.
10
11 Analysis of Variance (ANOVA) was used to test **if there was sufficient evidence to**
12
13 **suggest that** PICU nurses with higher levels of education would score significantly
14
15 higher on the PICU-NKST than nurses with undergraduate education only; and if
16
17 nurses with increased PICU experience would achieve significantly higher PICU-
18
19 NKST scores than those with little or no PICU experience.
20
21
22
23
24

25 **Ethical considerations**

26
27 Permission to conduct this study was sought and obtained from the Human Research
28
29 Ethics Committees of the university and participating hospitals. Three hospitals
30
31 provided a waiver for the need to provide written consent, where completion and
32
33 return of the test implied consent. One hospital required written consent. Both
34
35 information and consent forms reiterated the voluntary nature of participation and that
36
37 participants may withdraw at any time without comment or penalty.
38
39
40
41
42

43 **RESULTS**

44 **Participants**

45
46
47 A total of 79 PICU nurses participated in the psychometric testing of the PICU-NKST.
48
49 The mean age of the sample was **37.9** years (SD 9.1). Table 1 describes the
50
51 demographic attributes of the sample. Females comprised the largest percent of
52
53 respondents with over half working full time within the PICU. The majority of
54
55 respondents worked clinically as Registered Nurses.
56
57
58
59
60

1
2
3
4
5 The scores for the test approximated a normal distribution, with a mean score for the
6 PICU-NKST of 72.1 (SD 10.7). The minimum score was 48 and the maximum score
7 was 90 out of a possible score range of 0 to 109. The 90% score band for estimating
8 an individual's true score was ± 6.9 .
9
10
11
12

13 14 15 16 **Internal Consistency Reliability**

17
18 Analysis yielded a KR-20 reliability estimate of .85, with a standard error of
19 measurement of 4.2. Fifty-eight items demonstrated low item-total correlations with
20 values less than .3.
21
22
23
24
25
26

27 **Item analysis**

28
29 PICU-NKST items ranged in difficulty from .08 (very hard) to 1 (very easy), with a
30 mean difficulty of 0.66 (SD .24). More than half of the items were classified as
31 moderately difficult (0.4-0.8) (See Table 2). The discriminating power of test items
32 ranged from -.2 (negative discrimination) to .8 (excellent discrimination). Ten
33 discrimination power indices were negative values. Approximately half of the items
34 fell within the excellent and good categories of discrimination, with a further third
35 falling into the acceptable and low categories (See Table 2). There were a total of 327
36 distractors in the 109 item test (3 distractors per item), of which 42.5% of distractors
37 were functioning (See Table 3). Eighty-two distractors (25.1%) contained both low
38 frequency selection and positive discrimination. Eighty-three of the 109 items
39 contained at least one functioning distractor (Table 3).
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Construct validity

There was a significant effect of postgraduate levels of education on PICU-NKST scores, $F(4,72) = 6.25, p < .001$. Hochberg's GT2 post-hoc comparison of the five groups indicate that participants who had completed a master's degree ($M = 82.89$, 95% CI [77.49, 88.29]) or a postgraduate diploma ($M = 76.94$, 95% CI [72.13, 81.74]) scored significantly higher on the PICU-NKST than the bachelor of nursing group ($M = 66.55$, 95% CI [62.25, 70.86]), $p < .001$ and $.009$ respectively. All other comparisons were not significant. The mean difference in PICU-NKST scores between nurses who held a master's degree and a bachelor degree was 14.09 points.

Across the total PICU-NKST scores, group differences based on experience were statistically significant at $p < .001$. The mean scores for the three subgroups were 64.43, 73.11, and 76.20, respectively for groups with increasing PICU experience of < 2, 2-5, and > 5 years. Pearson's r between the total PICU-NKST scores and the continuous variable for years of PICU experience was $.54, p < .001$. PICU experience accounted for 29% of the variance in PICU-NKST scores. PICU-NKST scores were not found to be related to other critical care experience (for example, adult or neonatal ICU, emergency) ($r = .057, p = .631$). As shown in Table 4, group differences were significantly different, and in the predicted direction.

DISCUSSION

The aim of this study was to develop and evaluate the psychometric properties of an instrument to measure paediatric intensive care nursing knowledge and skill. The PICU-NKST is a 109 item multiple choice test of knowledge and skill that takes approximately 90 minutes to administer and is quickly and easily scored. Internal

1
2
3 consistency reliability was high and item analysis demonstrated adequate item
4
5 difficulty and discrimination. Moreover, the PICU-NKST has the ability to discern
6
7 group differences based on years of PICU experience and educational qualifications.
8
9 To our knowledge, this is the first study that has comprehensively developed and
10
11 psychometrically tested an instrument that measures the broad knowledge and skill
12
13 domains required of paediatric intensive care nursing practice.
14
15
16
17

18
19 In this study, the reliability estimate was high, yet a large proportion of item-total
20
21 correlations were low. Traditionally a test is considered more reliable, the more
22
23 homogenous the items. Boyle and others however argue against the issue of
24
25 homogeneity, especially in a test measuring a breadth of knowledge and skills, such as
26
27 the PICU-NKST (Boyle 1991, Kline 1986, McDonald 1981). Boyle specifically
28
29 argues that moderate to low item homogeneity is actually preferred to ensure a broad
30
31 coverage of the particular PICU nursing knowledge and skill constructs being
32
33 measured (Boyle 1991). Removal of the items with low item-total correlation would
34
35 therefore result in a reduction in the breadth of measurement and inadvertent
36
37 disregard for the heterogeneity of PICU nursing practice. In addition, the SEM of the
38
39 PICU-NKST scores was 4.2. Assuming a PICU nurses score on the PICU-NKST was
40
41 70 and the SEM was 4.2, there are nine out of ten chances (90%) that the nurse's true
42
43 score would fall between 63 and 77. Given that possible PICU-NKST scores range
44
45 from 0 to 109, this 90% confidence interval is small, and is further evidence of the
46
47 acceptable reliability of the PICU-NKST.
48
49
50
51

52
53
54 Item analyses results of the test also support its acceptable validity. In this study, ten
55
56 indices of discriminating power were negative, indicating that nurses who had overall
57
58
59
60

1
2
3 lower scores performed better on the item than nurses who had overall higher scores
4
5 on the test. Negatively discriminating test items have been shown to detract from
6
7 some of the important psychometric characteristics of the overall test and reduce the
8
9 validity (Downing 2009). As such, the items with negative discrimination indices **or**
10
11 **no functioning distractors** will need to be revised or removed to support test integrity
12
13 in future iterations of the PICU-NKST. One possibility for negative discrimination is
14
15 the concept of guessing, whereby test-takers might take the risk to guess on items
16
17 which are more difficult or ambiguous (Considine and Thomas 2005). Further
18
19 analysis of item performance across different nursing experience categories may be
20
21 required to explore guessing in low- and high-scoring groups. Several items that had
22
23 positive indices however, fell in the low to poor discrimination category. These items
24
25 may be ambiguous or misleading, or have distractors that are too nearly correct. In
26
27 reviewing these items for retention in the test it is also important to evaluate the
28
29 corresponding item difficulty. A large proportion of items met the minimum
30
31 acceptable value of .20 or greater (Ebel and Frisbie 1986, Tarrant and Ware 2010) and
32
33 will therefore be retained in the test.
34
35
36
37
38
39

40
41 In this study, half of the items were moderately difficult. Knowing the difficulty of the
42
43 items helps to avoid making a test so hard or so easy that it fails to provide much
44
45 information about individual test takers. Also, an item that proves to be much harder
46
47 or easier than anticipated may be flawed in some way. In making a judgement about
48
49 whether an item is too easy or hard however, the context of the test needs to be taken
50
51 into consideration. Within the PICU-NKST and ICU orientation program context, for
52
53 example, there may be some core knowledge and skills that everyone is expected to
54
55 know and inclusion of the item will be important, regardless of its difficulty. In
56
57
58
59
60

1
2
3 addition to evaluating an individual item for their index of difficulty, the mean
4
5 difficulty level can also be useful in determining the appropriateness of a test's
6
7 difficulty. For the PICU-NKST, the mean level of difficulty was .66, which is within
8
9 the .40 to .80 range considered acceptable (Tarrant and Ware 2010). Ghiselli,
10
11 Campbell, and Zedeck (1981) state that it may be unrealistic to expect all items to
12
13 achieve difficulty levels within this range. As such, while some test items were not
14
15 within the moderate range of difficulty, an acceptable mean difficulty indicates that
16
17 the instrument in total is "not too easy" and "not too hard".
18
19

20
21
22 An effective distractor is an incorrect response choice that low test scorers tend to
23
24 select as opposed to high test scorers. In this study, one quarter of distractors were
25
26 chosen by less than five percent of nursing respondents and had poor discrimination.
27
28 Whilst four-option MCQs remain popular in local low-stakes nursing tests, recent
29
30 research has demonstrated that three-option items offer many advantages over four-
31
32 option tests, including: less time required to construct items; less testing time
33
34 required; difficulty index increases making items easier; improvements in
35
36 discrimination indices; and increases in reliability (Rodriguez 2005, Tarrant and Ware
37
38 2010). In this study, removing the least functioning distractor of each item may
39
40 improve discrimination but not alter overall test reliability (Rodriguez 2005, Tarrant
41
42 and Ware 2010). Further to this, a reduction in the average testing time of 90 minutes,
43
44 by removal of the least functioning distractor, would prove beneficial in some
45
46 instances where time constraints exist (Tarrant and Ware 2010, Haladyna and
47
48 Downing 1993).
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Given that both nursing knowledge and skill acquisition are generated from
4 instruction and are grounded in practice (Benner 1984), it had been anticipated that
5 more experienced nurses and those with higher education would achieve a higher
6 score on the PICU-NKST. Responses were as predicted with master's prepared nurses
7 achieving the highest PICU-NKST scores. However, the bachelor of nursing group
8 scored the lowest on the test, lower than nurses with a diploma, which may reflect the
9 generalist preparation received by graduate nurses in Australia and New Zealand since
10 hospital based diploma programs have been discontinued. A further explanation,
11 unable to be explored statistically due to small numbers, may relate to the possibility
12 that the diploma nurses had more years of experience (depending on individual states,
13 diploma programs were discontinued as long as 15 years ago). Findings also
14 confirmed that nurses with more PICU experience scored significantly higher on the
15 test. These results are supported by Benner's (1984) theory of expertise development
16 whereby education and experience form the trajectory foundations for skill acquisition
17 and expertise. The fact that PICU nurses' years of experience had a stronger
18 relationship with overall scores than other critical care experience confirms that the
19 test measures a construct unique to PICU. Both McGaghie (1980) and Eraut (1994)
20 argue that in describing professional knowledge, it needs to be unique and have
21 discrete boundaries and the current study's results support this. Manley and Garbett
22 (2000) however, suggest that experience on its own does not result in expertise, but is
23 founded on learning and processing from those experiences to develop subsequent
24 practice. Similarly, in addition to undergraduate training, experience needs to be re-
25 consolidated by further learning, and the role of postgraduate education in this
26 learning is extremely important in the overall development of expertise.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Limitations

Whilst the PICU-NKST is not intended as a direct measure of PICU nursing competence, it does provide information about those aspects considered critical to the concept of competence, which are knowledge and skill. Accordingly, PICU-NKST content is specific to Australia and New Zealand and represents the current dynamic nature of PICU nursing. Mastery of knowledge and skills, however, does not guarantee successful performance in practice, but major gaps in this mastery would presumably be a serious limitation in the practice of the profession. Furthermore, whilst the PICU-NKST would prove to be most informative with graduate nurse groups, the sample for psychometric testing of the PICU-NKST incorporated a broad range of PICU nursing experience. The inclusion of a wide range of nursing experiences did allow, however, the contrasting of characteristics of nurses with limited PICU exposure with the characteristics of nurses with the opposite state, that is, experience and further education. Given the absence of an instrument to assess knowledge and skill in this population, both the product and the process of developing the instrument have therefore represented a unique and critical advancement in PICU nursing education and assessment.

PICU nursing knowledge and skill test scores were examined in a small convenience sample of nurses, representing approximately 20% of the PICU nursing population in the region (total population unknown). Financial and staffing concerns at the time of the study prevented further recruitment to the sample. Given the constraints of clinical populations, item analysis is frequently performed with small populations (<100) (Ferketich 1991) and it is plausible to analyse tests with as few as 30 subjects (Downing and Yudkowsky 2009). Therefore, the sample of 79 participants provided

1
2
3 sufficient information for informing item performance in a final operational test,
4
5 however caution needs to be taken when deciding to discard items based on a single
6
7 small sample test (Ferketich 1991). Establishing test-retest reliability would
8
9 strengthen the applicability of the PICU-NKST, however a strong learning effect
10
11 could pose a difficulty and careful consideration would need to be given to the re-test
12
13 timeframe. Finally, the test in this study was developed to assess competent level
14
15 practice. Whilst it could potentially be used across all expertise levels, nursing
16
17 practice at an expert level would inform a practice analysis differently, and therefore
18
19 new test blueprints would be required. The test in this study was developed as there
20
21 was no other instrument to inform knowledge and skill levels of PICU nursing
22
23 practice in Australia and New Zealand.
24
25
26
27
28
29

30 **CONCLUSION**

31
32 As with other health professional education, the effective measurement of knowledge
33
34 and skill is an important component of both nursing education and practice. Given the
35
36 introduction of graduate nurses into specialty areas such as paediatric intensive care, it
37
38 is crucial that educators and managers have a method of measuring the learning needs
39
40 and core outcomes required of educational programs. In the development of any
41
42 assessment methods, such as multiple choice examinations, it is important that these
43
44 methods are scrutinised for their effectiveness.
45
46
47
48
49

50 Psychometric testing of the 109 multiple-choice-question PICU-NKST demonstrated
51
52 evidence of acceptable reliability and validity. With some modification to a few items
53
54 to improve discrimination between low- and high-scorers, this instrument will provide
55
56 educators and managers alike with an instrument to assist in the overall assessment of
57
58
59
60

1
2
3 knowledge and skill acquisition, and curriculum and professional development
4
5 learning objectives.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Review Copy

References

- 1
2
3
4
5 American Association of Critical-Care Nurses (2010) Essentials of Pediatric Critical
6 Care Orientation.
7
8 American Association of Critical-Care Nurses (2011) Essentials of Critical Care
9 Orientation. American Association of Critical-Care Nurses.
10 Australian and New Zealand Intensive Care Society (2010) Report of the Australian
11 and New Zealand Paediatric Intensive Care Registry 2009.
12 Axelson, R. & Kreiter, C. (2009) Reliability. In *Assessment in Health Professions*
13 *Education*(Downing, S. and Yudkowsky, R. eds.) Routledge, New York.
14 Benner, P. (1984) *From novice to expert. Excellence and power in clinical nursing*
15 *practice.*, Addison-Wesley Publishing Company, California.
16 Boyle, G. (1991) Does item homogeneity indicate internal consistency or item
17 redundancy in psychometric scales? *Personality and Individual Differences*,
18 **12**(3), 291-294.
19 Boyle, M., Butcher, R. & Kenney, C. (1998) Study to validate the outcome goal,
20 competencies and educational objectives for use in intensive care orientation
21 programs. *Australian Critical Care*, **11**(1), 20-24.
22 Cavanaugh, D. & Huse, A. (2004) Surviving the nursing shortage: Developing a
23 nursing orientation program to prepare and retain intensive care unit nurses.
24 *The Journal of Continuing Education in Nursing*, **35**(6), 251-256.
25 Chestnutt, B. & Everhart, B. (2007) Meeting the needs of graduate nurses in critical
26 care orientation: Staged orientation program in surgical intensive care. *Critical*
27 *Care Nurse*, **27**(3), 36-51.
28 Considine, J. & Thomas, S. (2005) Design, format, validity and reliability of multiple
29 choice questions for use in nursing research and education. *Collegian*, **12**(1),
30 19-24.
31 Cowin, L., Hengstberger-Sims, C., Eagar, S., Gregory, L., Andrew, S. & Rolley, J.
32 (2008) Competency measurements: testing convergent validity for two
33 measures. *Journal of Advanced Nursing*, **64**(3), 272-277.
34 Downing, S. (2009) Statistics of Testing. In *Assessment in Health Professions*
35 *Education*(Downing, S. and Yudkowsky, R. eds.) Routledge, New York.
36 Downing, S. & Haladyna, T. (1997) Test item development: Validity evidence from
37 quality assurance procedures. *Applied Measurement in Education*, **10**(1), 61-
38 82.
39 Downing, S. & Yudkowsky, R. (2009) *Assessment in Health Professions Education*,
40 Routledge, Mahwah, New Jersey.
41 Dunn, S. (1992) Orientation: The transition from novice to competent critical care
42 nurse. *Critical Care Nurse Quarterly*, **15**(1), 69-77.
43 Duvall, J. (2009) From novice to advanced beginner: The critical care internship.
44 *Journal for Nurses in Staff Development*, **25**(1), 25-27.
45 Ebel, R. & Frisbie, D. (1986) *Essentials of educational measurement*, Prentice Hall,
46 Englewood Cliffs, NJ.
47 Eraut, M. (1994) *Developing professional knowledge and competence.*, Falmer Press,
48 London.
49 Ferketich, S. (1991) Aspects of item analysis. *Research in Nursing and Health*, **14**,
50 165-168.
51 Friedman, M., Cooper, A., Click, E. & Fitzpatrick, J. (2011) Specialized new graduate
52 RN critical care orientation: Retention and financial impact. *Nursing*
53 *Economic*, **29**(1), 7-14.
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Ghiselli, E., Campbell, J. & Zedeck, S. (1981) *Measurement Theory for the Behavioural Sciences*, W.H. Freeman and Company, San Francisco.
- Haladyna, T. (1997) *Writing Test Items to Evaluate Higher Order Thinking.*, Allyn and Bacon, Boston.
- Haladyna, T. & Downing, S. (1988) Functional distractors: Implications for test-item writing and test design. In *Annual Meeting of the American Educational Research Association* New Orleans, LA.
- Haladyna, T. & Downing, S. (1993) How many options is enough for a multiple-choice test item? *Educational and Psychological Measurement*, **53**(4), 999-1010.
- Haladyna, T., Downing, S. & Rodriguez, M. (2002) A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, **15**(3), 309-333.
- Hall, D. & Marshall, R. (2006) Evaluation of a 16-week critical care internship program using a staff development program effectiveness evaluation tool. *Journal for Nurses in Staff Development*, **22**(3), 134-143.
- Kaddoura, M. (2010) Effect of the Essentials of Critical Care Orientation (ECCO) program on the development of nurses' critical thinking skills. *The Journal of Continuing Education in Nursing*, **41**(9), 424-432.
- Kline, P. (1986) *A Handbook of Test Construction: Introduction to Psychometric Design.*, Methuen, New York.
- Lambert, N. (2012) Ban happy sheets! - Understanding and using evaluation. *Nurse Education Today*, **32**, 1-4.
- Long, D. (2012) Development of a valid and reliable measure of paediatric intensive care knowledge and skills. In *School of Nursing and Midwifery, Faculty of Health*, Vol. Doctor of Philosophy Griffith University, Brisbane, pp. 323.
- Long, D., Young, J., Rickard, C. & Mitchell, M. (2012) Measuring paediatric critical care nursing knowledge in Australia and New Zealand: How the Basic Knowledge Assessment Tool for Pediatric Nurses (PEDS-BKAT4) performs. *Australian Critical Care*. 10.1016/j.aucc.2012.02.004.
- Long, D., Young, J., Rickard, C. & Mitchell, M. (2013) Analysing the role of the PICU nurse to guide education of new graduate nurses. *Nurse Education Today*. 10.1016/j.nedt.2013.01.016.
- Mahara, M. (1998) A perspective on clinical evaluation in nursing education. *Journal of Advanced Nursing*, **28**(6), 1339-46.
- Manley, K. & Garbett, R. (2000) Paying Peter and Paul: Reconciling concepts of expertise with competency for a clinical career structure. *Journal of Clinical Nursing*, **9**(3), 347-359.
- McDonald, R. (1981) The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, **34**, 110-117.
- McGaghie, W. (1980) The evaluation of competence: Validity issues in the health professions. *Evaluation and the Health Profession*, **3**(3), 289-320.
- McGrath, P., Fox-Young, S., Moxham, I., Anastasi, J., Gorman, D. & Tollefson, J. (2006) Collaborative voices: ongoing reflections on nursing competencies. *Contemporary Nurse*, **22**(1), 46-58.
- Meretoja, R., Isoaho, H. & Leino-Kilpi, H. (2004) Nurse Competence Scale: Development and psychometric testing. *Journal of Advanced Nursing*, **47**(2), 124-133.

- 1
2
3 Morris, L., Pfeifer, P., Catalano, R., Fortney, R., Hilton, E., McLaughlin, J., Nelson,
4 G., Palamone, J., Rabito, R., Wetzel, R. & Goldstein, L. (2007) Designing a
5 comprehensive model for critical care orientation. *Critical Care Nurse*, **27**(6),
6 37-60.
- 7
8 Morris, L., Pfeifer, P., Catalano, R., Fortney, R., Nelson, G., Rabito, R. & Harap, R.
9 (2009) Outcome evaluation of a new model of critical care orientation.
10 *American Journal of Critical Care*, **18**(3), 252-262.
- 11 Murrells, T., Robinson, S. & Griffiths, P. (2009) Assessing competence in nursing.
12 *Nursing Management*, **16**(4), 18-19.
- 13 O'Shea, K. (2002) *Staff Development Nursing Secrets*, Hanley & Belfus, Philadelphia.
- 14 Osterlind, S. (1998) *Constructing test items: Multiple choice, constructed response,*
15 *performance, and other formats.*, Kluwer Academic Publishers, Boston.
- 16 Peterson, K. & Van Buren, K. (2006) Implementing Essentials of Critical Care
17 Orientation. *Critical Care Nursing Quarterly*, **29**(3), 218-230.
- 18 Pross, E. (2010) Promoting excellence in nursing education (PENE): Pross evaluation
19 model. *Nurse Education Today*, **30**(6), 557-561.
- 20
21 Rashotte, J. & Thomas, M. (2002) Incorporating educational theory into critical care
22 orientation. *Journal of Continuing Education In Nursing*, **33**(3), 131-137.
- 23 Raymond, M. & Neustel, S. (2006) Determining the Content of Credentialling
24 Examinations. In *Handbook of Test Development*(Downing, S. and Haladyna,
25 T. eds.) Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey.
- 26 Redfern, S., Norman, I., Calman, L., Watson, R. & Murrells, T. (2002) Assessing
27 competence to practise in nursing: a review of the literature. *Research Papers*
28 *in Education*, **17**(1), 51-77.
- 29 Rodriguez, M. (2005) Three options are optimal for multiple choice items: A meta-
30 analysis of 80 years of research. *Educational Measurement, Issues and*
31 *Practice*, **24**(2), 3-13.
- 32
33 Runton, N. & Toth, J. (1998) Introducing the Basic Knowledge Assessment Tool for
34 Pediatric Critical Care Nursing (PEDS-BKAT4). *Critical Care Nurse*, **18**(3),
35 67-72.
- 36 Shelton, D. (2003) The nursing shortage: Its impact on the care of children. *Newborn*
37 *and Infant Nursing Reviews*, **3**(1), 27-31.
- 38 Square, N. (2010) Modeling clinical applications in intensive care settings for nursing
39 orientation. *Advances in Neonatal Care*, **10**(6), 325-329.
- 40 Stefanski, R. & Rossler, K. (2009) Preparing the novice critical care nurse: A
41 community-wide collaboration using the benefits of simulation. *The Journal of*
42 *Continuing Education in Nursing*, **40**(10), 443-451.
- 43 Tarrant, M. & Ware, J. (2010) A comparison of the psychometric properties of three-
44 and four-option multiple choice questions in nursing assessments. *Nurse*
45 *Education Today*, **30**(6), 539-543.
- 46
47 Thomason, T. (2006) ICU nursing orientation and postorientation practices. *Critical*
48 *Care Nursing Quarterly*, **29**(3), 237-245.
- 49 Watson, R., Stimpson, A., Topping, A. & Porock, D. (2002) Clinical competence
50 assessment in nursing: A systematic review of the literature. *Journal of*
51 *Advanced Nursing*, **39**(5), 421-431.
- 52
53 Wynd, C. (2002) Evidence-based education and the evaluation of a critical care
54 course. *The Journal of Continuing Education in Nursing*, **33**(3), 119-125.
- 55
56
57
58
59
60

	Cardiac	Invasive lines and monitoring	Respiratory	Airway management	Drug administration and monitoring	Renal	Assessment	Additional	Patient care problem System n	Patient care problem System %
	6/7. Pacing (4) 13. CPR (4)	14. Arterial line (10) 15. ABG (1) 19. RA/CVP (1) 20. LA lines(2) 40. ICP (4)	24. Mech Vent (18) 26. HFOV (1) 29. Nitric oxide (1)	31. Airway M'ment(10) 32. Intubation (1) 34. Retape ETT (2) 36. Suctioning (2)	45. NMBA Admin (1) 46. NMBA Monitor (1) 47. Sedation Admin (1) 48. Sed Monitoring (1) 49. Inotropes (17) 11. Phospho. Inhib (4)	41. PD (1) 42. CVVHDF (1)	5. Assessment (3) 1. ECG (2) 4. HaemoD Monit (1)	Pathophysiology (5) Communication (4) Deteriorating Child (2) Developmental (3) Death (1) Feeding (1)		
Cardiovascular TGA	2 Pacing	2 LA lines			1 Inotropes 2 Phospho Inhib	1 PD	1 ECG			
VSD	2 Pacing				2 Inotropes		1 Assessment	1 Pathophysiology		
Cardiac Failure			1 Mech Vent		1 Inotropes 2 Phospho Inhib					
SVT			1 Mech Vent				1 ECG		21	19
Respiratory Asthma		1 Arterial line	2 Mech Vent	1 Airway M'ment				1 Pathophysiology		
Bronchiolitis		2 Arterial line	2 Mech Vent	1 Airway M'ment 1 Retape ETT 1 Suctioning						
Pneumonia		1 Arterial line 1 ABG	2 Mech Vent 1 Nitric oxide 1 HFOV	1 Airway M'ment 1 Suctioning	1 Inotropes 1 NMBA Monitor 1 Sedation			1 Communication		
Respiratory Failure	1 CPR	1 Arterial line	1 Mech Vent	1 Airway M'ment	2 Inotropes 1 Sed ation				30	28
Injury Burns		1 Arterial line	1 Mech Vent	1 Airway M'ment	1 Inotropes			2 Pathophysiology		
Immersion			1 Mech Vent		2 Inotropes			1 Communication 1 Pathophysiology		
Ingestion				1 Airway M'ment	1 Inotropes			1 Communication		
Trauma - head		4 ICP			1 Inotropes 1 NMBA Admin				20	18
Renal HUS			1 Mech Vent		2 Inotropes	1 CVVHDF			4	4
Neurological Meningitis		1 Arterial line	2 Mech Vent	1 Retape ETT	1 Inotropes					
Seizures			2 Mech Vent	2 Airway M'ment					9	8
Gastrointestinal Gastro		1 Arterial line					2 Assessment	2 Deteriorating Child 3 Developmental	8	7
Miscellaneous DKA		1 Arterial line		1 Airway M'ment			1 Haemod Monitor			
Resp Arrest	3 CPR		1 Mech Vent	1 Airway M'ment 1 Intubation	1 Inotropes			1 Communication 1 Death		
Shock - septic		1 Arterial line 1 RA/CVP	1 Mech Vent		1 Inotropes			1 Feeding	17	16
Activity Total	8	18	20	15	25	2	6	16	109	100

Legend		Abbreviations
Pink Shaded Area:	Lists PICU nursing activity followed by number of items for this activity in brackets	CPR Cardiopulmonary resuscitation
Yellow Shaded Area:	Lists Patient Care Problems - Systems followed by diagnoses	ABG Arterial blood gas
Green Shaded Area:	Number of items represented in patient care problem (system) areas	RA/CVP Right atrial/ Central venous pressure
Blue Shaded Area:	Total number of PICU nursing activities	LA Left atrial
		ICP Intracranial pressure
		HFOV High frequency oscillator ventilator
		NMBA Neuromuscular blocking agent
		PD Peritoneal dialysis
		CVVHDF Continuous veno-venous haemodiafiltration
		ECG Electrocardiogram

Figure 1: Paediatric Intensive Care Unit – Nursing Knowledge and Skills Test (PICU-NKST) Test Plan

Table 1. Demographic attributes of PICU Nurses (N=79).

PICU Nurse Demographic	n	(%)
Gender		
Female	62	79.5
Type of Employment		
Full time	54	69.2
Part time	24	30.8
PICU Work Activity		
Clinical	70	89.7
Education	5	6.4
Management	1	1.3
Other	2	2.6
Highest Educational Attainment		
Diploma in Nursing	5	6.5
Bachelor of Nursing	27	35.1
Postgraduate Certificate	20	26.0
Postgraduate Diploma	16	20.8
Master	9	11.7
Specialty of Postgraduate Qualification		
PICU	25	32.1
ICU/Critical Care	5	6.4
Paediatrics/Child Health	4	5.1
Education	1	1.3
Cardiothoracic	1	1.3
Business/Management	3	3.8
General – Not otherwise specified	6	7.7
Title of Nursing Position		
RN*	49	64.5
CN/CNE/CNF*	19	25.0
CNC/NUM/CNS/NE*	8	10.5
	Mean	SD
Age (years)	37.9	9.1
PICU experience (years)	7.5	6.5
Other critical care experience (years)	1.9	3.3
Total critical care experience (years)	9.3	7.5
Total nursing experience (years)	13	8.6
Average hours worked in PICU per week (hours)	35	7.1

*RN=Registered Nurse, CN=Clinical Nurse, CNE=Clinical Nurse Educator, CNF=Clinical Nurse Facilitator, CNC=Clinical Nurse Consultant, NUM=Nurse Unit Manager, CNS=Clinical Nurse Specialist, NE=Nurse Educator.

NB. Where n do not add to 79, there was missing data.

Table 2. Summary of item analysis for PICU-NKST items (Total N=109)

	n	Percent (%)
<u>Discrimination</u>		
Excellent (+0.4 to +1.0)	44	40.4
Good (+0.3 to +0.39)	14	12.8
Acceptable (+0.2 to +0.29)	19	17.4
Low (+0.1 to +0.19)	16	14.7
Poor (<+0.1)	6	5.5
Unsatisfactory (-0.01 to -1.00)	10	9.2
<u>Difficulty</u>		
Difficult (0.01 to 0.40)	17	15.6
Moderate (0.4 to 0.80)	56	51.4
Easy (0.80 to 1.0)	36	33.0
<u>Combination Difficulty and Discrimination</u>		
Type 1	16	14.7
Type 2	47	43.1
Type 3	9	8.3
Type 4	27	24.8
Type 5	10	9.2

Table 3. Summary of Distractor performance for PICU-NKST items.

	Number (Total N=327)	Percent (%)
Less than 5% selected distractor	65	19.9
Discrimination ≥ 0	41	12.5
Frequency < 5% & Discrimination ≥ 0	82	25.1
Functioning distractors	139	42.5
	Number (Total N=109)	Percent (%)
Functioning distractors per item		
None	26	23.9
One	39	35.8
Two	32	29.4
Three	12	11.0
Functioning distractors per item (mean)	1.28	

Review Copy

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 4. PICU-NKST scores by years of PICU experience and level of education.

	PICU-NKST Score Mean			
	Mean	SD	F	<i>p</i>
Years of PICU experience				
< 2 years (<i>n</i> = 23)	64.4	10.9		
2–5 years (<i>n</i> = 17)	73.1	7.8		
> 5 years (<i>n</i> = 39)	76.2	9.4	11.0	<.0001
Highest level of education				
Diploma (<i>n</i> = 5)	68.8	12.3		
Bachelor (<i>n</i> = 27)	66.5	10.8		
Postgraduate Certificate (<i>n</i> = 20)	71.9	8.4		
Postgraduate Diploma (<i>n</i> = 16)	76.9	9.0		
Masters (<i>n</i> = 9)	82.9	7.0	6.2	<.0001

1
2
3 Dear Editor and Staff

4
5 Many thanks for your valuable feedback which will assist in improving the quality of our
6 manuscript.

7
8 Please find our response to the Reviewer's comments below.
9

10 -----
11
12 **EDITORIAL STAFF COMMENTS:** Please remove p values from the abstract. As previously
13 notified, if your paper is accepted, changes will be required for the figure.

14
15 **P values removed, as requested.**

16
17 **EDITOR'S COMMENTS:**

18 The concerns about your paper lie in your statistical treatment of the data. You may want to
19 consider seeking statistical assistance in both the analysis and presentation of the results.

20
21 **We have consulted a biostatistician who has helped us clarify our reporting.**

22
23 **REVIEWER COMMENTS:**

24
25 **Reviewer: 1**

26 Comments to the Author

27 Thank you for submitting your paper. Interesting area of work but you have a large number
28 of errors in the paper which made it a difficult read at times.

29
30 1. Top of page 11 – you imply you tested an experimental hypothesis i.e. “.. was used to
31 test if ..” all statistical hypothesis testing only ever tests the null hypothesis – you could have
32 used text of null or you should have said “...(ANOVA) was used to assess if there was
33 sufficient evidence to suggest that PICU nurses with higher...”

34
35 **Changed, as suggested.**

36
37 2. Page 11 Result para you state mean 37 years with SD 9.06 and yet in your table mean
38 is 38 years and SD is 9.1.

39
40 **Discrepancy in rounded, corrected.**

41
42 3. I also feel you should inform the reader in the next para on page 11 that the possible
43 range for the score was 0 to 109

44
45 **Changed, as suggested.**

46
47 4. Top of page 12 – the 90% or indeed and % CI is not usually presented in that way and
48 I have no idea where you got the 6.9 from nor does the reader. I tried $1.6449 \cdot SD / \sqrt{n}$ with
49 $sd = 10.7$ and $n=79$, but that gave me about 1.98 and even using a t-distribution I cannot see
50 how it could get as high as 6.2 – this needs an explanation.

51
52 **Reworded to include 'score band'. Further explanation provided in methods section. See
53 further discussion in points 5, 10 and 11.**

54
55
56 5. Page 12 KR -20 para – assuming the SEM was correctly calculated at 4.2 it appears
57 later you used this value as the SE when calculating a 90% CI but the SE is simply
58 SD / \sqrt{n} – I think there may be some confusion here around what an SEM and an SE are.
59
60

The authors use the definition outlined in the Standards for Educational and Psychological Testing, which defines the Standard Error of Measurement (SEM) as *the standard deviation of a hypothetical distribution of measurement errors that arises when a given population is assessed via a particular test or procedure* [1]. SEM is an adequate measure if a general statistic is required for describing the likely accuracy of the score achieved by a randomly chosen examinee, and described by some as a better measure of quality than reliability [2]. The equation for the SEM can be expressed as:

$$\text{SEM} = S \sqrt{(1-r)} \quad [2-4]$$

Whereby S = standard deviation of the test, and

r = reliability of the test

Therefore, if the PICU-NKST has a standard deviation of 10.7 and a KR-20 reliability estimate of .85, the SEM would be calculated as follows:

$$\text{SEM} = S \sqrt{(1-r)} = 10.7 \sqrt{(1-.85)} = 10.7 \sqrt{.15} = 10.7 (39) = 4.17 \text{ Rounded to 4.2.}$$

6. Page 12 middle para line 32 – “Almost half” implies one nearly reaching half. Suggest “Approximately half” or “slightly more than half”

Changed, as suggested.

7. Page 12 last para – no idea why you have $r=0.51$ in ANOVA result along with the F and p-values. What is it? Do you mean the R^2 that comes out of ANOVA??

For clarity, this has been removed.

8. Foot of page 12 onto top of page 13 – I think you really need to go and completely check this result text and the Table 4 entries as they do not match. Your page 13 text states BN mean was 68.80 yet that in the table is the Diploma mean – which is correct? Makes a big difference to other text – you do however say in Discussion that Dip mean was higher than BN mean! Also is the reader left to assume that all of the other Tukey p-values were > 0.05 ? If the 68.8 was significantly different from 2 others surely the 66.5 was also significant!!!

Mean value written in error, corrected. Hochberg's GT2 was actually used for post hoc comparison as the group sizes were very different. The Diploma group only contained 5 nurses, compared to 27 in the BN group. Even though the Diploma and BN scores were similar (68.8 and 66.5, respectively), the results only indicate a difference between BN and postgraduate diploma and master's education.

9. Middle para page 13 where did the 30% come from – you talk about correlation and yet % variation explained is from a regression analyses – did you fit a regression model? I would have expected the % explained to be the square of the correlation which gave me 29.16 but this is possibly a rounding issue.

Percent of variation explained from the correlation, which was rounded to 30. Changed to 29%.

10. Page 14 main para – two issues – firstly as previously mentioned you appear to be using 4.2 and SE when it is the SEM and secondly the phrase “there are nine out of ten

chances (90%) that ..” The definition of a 90% CI is that there is a 90% probability that the actual true mean score will be in that interval not that a nurse’s score will be in it!

We confirm that we used SEM not SE. As outlined by the Standards for Educational and Psychological Testing [1], the authors also used the SEM to calculate confidence intervals (also known as score bands or confidence bands) for an obtained (raw) score.

One can interpret a 68% confidence interval as the score range around an obtained score that includes an examinee’s true score 68% of the time. Multiplying SEM by 1.65 will provide the needed value to construct a 90% confidence interval around an obtained score. The value 1.65 is appropriate for a 90% confidence interval because for distribution, the scores that fall within 1.65 deviations of the mean will include 90% of all the scores in the distribution (SEM x 1.96 for 95% CI, SEM x 2.58 for 99% CI) [3, 4]. Therefore, 90% confidence intervals can be calculated as follows:

$$90\%CI = \text{observed score} \pm (1.65 \times 4.2) = \text{observed score} \pm 6.93.$$

To avoid reader confusion by using the group mean, the hypothetical observed nurses raw score has been changed to 70. Therefore, the true score for this nurse would fall between 63 and 77. Using these calculations, all texts citing SEM calculations refer to an individuals obtained score and true score, not the actual true mean score [Eg. 3, 4, 5].

11. Same para just to remind you the 90% CI to me would not be 65.2 to 79 as this is using plus or minus 6.9 units and that appears to have come from 1.6449×4.2

Please see explanations and calculations provided above, however calculations were correct. Manuscript text has now been edited using an example of an examinee’s observed score of 70, to avoid confusion with the group mean score.

12. Foot of page 16 “However, the bachelor...” this contradicts previous text

Statement is correct, previous error corrected.

13. Table 2 – the Discrimination block only has 108 items and yet % appear to be from 109 – needs explained or is it a typo?

Error corrected.

14. Table 3 – Number heading states clearly that N = 327 and this does apply to the first 4 lines but the lower block is 109 for % calculations so you need to either put a new heading in or break the table.

Table 3 edited, as suggested.

15. Table 4 – very poor layout in my view – think it would look better with the two section one below the other and as said before entries do not match text

Table 4 edited, as suggested. Entries checked and correct.

Reviewer: 2

Comments to the Author

Thank you for writing this very interesting manuscript. I have few comments for you to consider:

1
2
3 * pg 7, line 44: would add sample of PICU nurses IN AUSTRALIA AND NEW ZEALAND
4 because it was said above that there was a lack of research in this population.
5

6 **Changed, as suggested.**
7

8 * pg 12, line 45 refers to Table 3. I am not clear about what the None Functioning distractors
9 per item means. For the N of 26, does this mean that all of the participants got the answers
10 correct for all questions? If there were no functioning distractors, then was the question
11 deemed "bad" and not used further?
12

13 **The use of and definition of distractors has been clarified in the Methods section. Each item
14 or question of the PICU-NKST has four plausible answers. One of these four is the correct
15 answer; the other three are defined as distractors. Therefore, the PICU-NKST has
16 $109 \times 3 = 327$ distractors. Non-functioning distractors in this study were defined by two criteria:
17 distractors chosen by less than 5% of nurses AND distractors with a positive discrimination
18 index. Theoretically, the more functioning distractors there are, the better the question.
19 Practically however, in most well designed achievement tests, more than two functional
20 distractors are rare and do not necessarily contribute to more effective measurement of
21 achievement (Haladyna & Downing, 1988). With regards to your question about N of 26, this
22 refers to 26 of the 109 items not having any functioning distractors. Changes to Table 3
23 headings, as suggested by the previous reviewer will clarify this for the reader. These items
24 should have the three distractors reviewed and improved. These items, of course, would
25 need to undergo further psychometric testing, as does the whole PICU-NKST over time. This
26 has been clarified in the Discussion.
27
28
29**

30 **Reviewer: 3**

31 Comments to the Author

32 Nicely done.
33
34

35 **Again, we would like to thank you for your valuable time and comments.**
36

37 **References**
38

- 39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
1. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, *Standards for educational and psychological testing* 1999, Washington, DC: American Educational Research Association.
 2. Tighe, J., et al., *The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessment than is reliability: an analysis of MRCP (UK) examinations*. BMC Medical Education, 2010. **10**: p. 40-48.
 3. Downing, S. and R. Yudkowsky, *Assessment in Health Professions Education* 2009, Mahwah, New Jersey: Routledge.
 4. Harvill, L., *Standard error of measurement*. Educational Measurement: Issues and Practice., 1991. **10**(2): p. 181-189.
 5. Bishop, D., *Standard Error of Measurement*, in *Technical Assistance Paper* 1996, Bureau of Exceptional Education and Student Services: Florida.