Measuring paediatric intensive care nursing knowledge in Australia and New Zealand: How the Basic Knowledge Assessment Tool for pediatric critical care nurses (PEDS-BKAT4) performs

Author
Long, Debbie, Young, Jeanine, Rickard, Claire, Mitchell, Marion

Published
2013

Journal Title
Australian Critical Care

DOI
https://doi.org/10.1016/j.aucc.2012.02.004

Copyright Statement
Copyright 2012 ACCCN. Published by Elsevier. This is the author-manuscript version of this paper. Reproduced in accordance with the copyright policy of the publisher. Please refer to the journal website for access to the definitive, published version.

Downloaded from
http://hdl.handle.net/10072/50791
MEASURING PAEDIATRIC CRITICAL CARE NURSING KNOWLEDGE IN AUSTRALIA AND NEW ZEALAND: HOW THE BASIC KNOWLEDGE ASSESSMENT TOOL FOR PEDIATRIC NURSES (PEDS-BKAT) PERFORMS.

Debbie Long a,b,*
Jeanine Young c,d
Claire M Rickard e
Marion L Mitchell f

aClinical Nurse Consultant, Paediatric Intensive Care Unit, Royal Children’s Hospital, Herston Road, Herston, Brisbane, Queensland 4069, Australia
bPhD Candidate, School of Nursing and Midwifery, Griffith University, Nathan, Brisbane, Australia
cDirector of Nursing Research, Royal Children’s Hospital, Brisbane, Australia
dAdjunct Professor, Research Centre for Clinical and Community Practice Innovation, Griffith University, Nathan, Australia
eProfessor, Research Centre for Clinical and Community Practice Innovation, Griffith University, Nathan, Australia
fAssociate Professor and Principal Research Fellow, Research Centre for Clinical and Community Practice Innovation, Griffith University, Nathan, and Princess Alexandra Hospital, Australia

*Corresponding author. Tel.: +61 7 3636 9204; fax.: +61 7 3636 1990.
E-mail address: debbie.long@griffithuni.edu.au

Keywords:
Knowledge
Multiple-choice questions
PEDS-BKAT
Paediatrics
Critical care

Acknowledgement:
DL would like to thank Dr Mary Murray and Dr Jenniefer Barr for their contributions to the original study concept.
Abstract:
Validated professional knowledge measures are limited in paediatric intensive care unit (PICU) nursing. The Basic Knowledge Assessment Tool for Pediatric Intensive Care Nurses (PEDS-BKAT) measures knowledge, however content and practice differences exist between various PICUs. The study aim was to evaluate the PEDS-BKAT in the Australian and New Zealand setting. A panel of 10 experts examined item and scale content validity. Items were evaluated for 31 evidence-based item writing flaws and for cognitive level, by a 4-person expert panel. Thirty-six PICU nurses completed the PEDS-BKAT, with reliability and item analysis conducted. Mean item content validity was 0.70, and 43% of items had content validity less than 0.8. Overall (Scale) content validity was 0.71. Thirty-five percent of items were classified as flawed. Thirty-five percent of items were written at the ‘knowledge’ level, and 58% at ‘understanding’. The mean PEDS-BKAT score was 60.8 (SD=9.6), KR-20 reliability 0.81. The mean item difficulty was 0.62, and the mean discrimination index was 0.23. The PEDS-BKAT was not a reliable and valid measure of basic PICU nursing knowledge in Australian and New Zealand. Further research into the types of knowledge and skills required of PICU nurses in this setting are needed to inform the development of a future tool.
INTRODUCTION

Learning and development for nurses is crucial for ongoing improvements in the quality of patient care (1). Identifying the learning needs of employees and evaluating the results of staff development are therefore essential elements of any nursing practice environment (2-4). Evaluation of nurses’ knowledge and skills must employ methods that are appropriate. Important considerations include the level of learner, the learning context and the specific learning objectives. There appears to be an abundance of information about how to design and develop tools for assessing knowledge and skills (5), but existing tools may already be available and appropriate to meet the needs of a unit. When choosing to use an existing tool educators need to consider the quality of the tool. Is there a tool that claims to measure what you would like to measure? Does the tool claim to measure the concept of interest? Would the tool produce consistently reliable and valid results in the planned clinical setting?

Nurse managers and educators are increasingly required to evaluate the quality and efficacy of nursing care provided to their patients. The goal of any staff development education, particularly orientation programs, is to develop a nurse who is capable of safe, effective and competent practice. In an applied discipline such as nursing, clinical expertise is acquired through formal knowledge by way of the educational process in the first instance, which is further advanced with clinical experience (6). Several authors (1-4, 7) suggest that staff assessment is a vital part of the education process, even those undertaken in-house, and as such, deserves attention during staff development planning.

Many areas of nursing practice can be evaluated with tools that assess overall competence, to specific areas such as knowledge of a particular body system or procedural skills. Regardless of the area of practice requiring assessment, finding an appropriate tool for this evaluation
can often appear a daunting task (8). In addition, assessments may use various methods including rating scales, checklists, objective structured clinical examinations, short answer and multiple choice question (MCQ) examinations, to name a few. Whilst each method of assessment has its own advantages and disadvantages, no one test format provides exhaustive information for judging staff development or performance (1).

MCQs are used to measure knowledge as an endpoint in education and are credited with a number of advantages (9). They are objective, so variation in marking due to subjective factors is eliminated and they are easy to mark, although the questions themselves still have to be scrutinised for bias. Ease of marking is likely to be a significant advantage for overburdened staff (5, 10). In addition, measurable comparisons within and between assessments can be made over time to assess changes in an individual or group’s learning. MCQs can be efficient if responses take less time to complete, and therefore it is possible to test a greater range of content. Nevertheless, there are a number of potential disadvantages associated with MCQs. Several criticisms have been levelled at the MCQ, including: the ability to only assess the recall of fact; encouragement of rote memorisation and guessing if answer not known; inability to demonstrate the thinking process; and the time required to construct well-written MCQ items (often up to four hours each) (5, 9, 11, 12). These advantages and disadvantages need to be considered before selecting MCQs as a method of choice for assessment of learning and competence.

As the knowledge base of nursing rapidly expands and changes, particularly in intensive care care, so do the standards of practice set out by professional bodies. Professional standards of nursing practice offer quality statements relative to education in the service and academic settings. Competency development is viewed as one means by which the profession can self-
evaluate and enhance its accountability to the public. In 1996 the Australian College of Critical Care Nurses (ACCCN) developed competency standards for Australian specialist level critical care nurses (13). These standards do not reflect beginner or competent levels of practice but address competencies of the wider critical care nursing community, and are not specific to paediatric intensive care unit (PICU) nursing practice. As such, the ACCCN standards are intended as a guide for the best possible total care of the critically ill patient and provide goals for nurses to work toward, rather than representing basic standards for safe practice. Whilst the competency standards were never intended to be used to directly assess clinical practice, many would consider them useful in defining optimum performance. Fisher et al. (14) recently examined how well the elements and competency statements that comprise the competency standards adequately measure the construct of competence in the critical care environment. They found that statistically there was no support for the current structure of the ACCCN competencies because the elements did not fit uniquely to a single competency, but were multidimensional and loaded across several competencies. Consequently, this reinforces the argument that the ACCCN competency standards (11) were not intended for evaluation, therefore other methods for evaluating critical care nurses’ performance need to be explored.

Within our unit, the situation arose where the nurse educator wished to evaluate the knowledge of new graduate nurses following an orientation period in the PICU. This need was also reflected in several on-line forums, Children’s Hospitals Australiasia Nurse Unit Manager Forum, PICU Nurse International Forum and PICUPhD Forum. Database searches were conducted for evaluation tools that assessed content in the paediatric intensive care domain and with a focus on multiple choice format. The Basic Knowledge Assessment Tool for Pediatric Intensive Care Nurses (PEDS-BKAT) was identified as the only existing tool for the PICU nursing population (15).
The PEDS-BKAT is a 100 item test which measures basic knowledge in paediatric intensive care nursing (15). These items measure content related to the following areas of practice: cardiovascular, monitoring lines, pulmonary, neurology, endocrine, renal, gastrointestinal/parenteral, and other. The PEDS-BKAT contains 96 multiple choice and four short answer questions that measure both the recall of basic information and the application of basic knowledge in practice situations. The copyrighted tool was originally tested on 105 PICU nurses across six United States of America (USA) states and found to be reliable (internal consistency) with a Cronbach’s coefficient alpha of 0.86 (15). Since the PEDS-BKAT was developed in 1998, there have been no further reported reliability or validity testing. Whether this assessment was appropriate to use in an Australian and New Zealand setting needed to be established. Considerable differences exist however, in the educational content and practices between North American and Australian and New Zealand paediatric intensive care settings. The appropriateness of the PEDS-BKAT in providing valid and reliable data when measuring knowledge for local nurses was considered, particularly in light of previous studies which demonstrated that setting is a key factor in determining tool validity (16).

**STUDY AIM**

This study evaluated whether the PEDS-BKAT was a valid and reliable tool for assessing paediatric intensive care nurses’ knowledge in the Australian and New Zealand setting.

**METHODS**

The methods used for this study included an evaluation of content validity, cognitive level and adherence to item writing principles, reliability and item analysis. Figure 1 provides a
Flow diagram to summarise the methods used in this study. Downing and Haladyna (10) outline the importance of validity studies in tool development and recommend each of these areas be addressed in supporting a specific interpretation or use of a tool and its score. The need for validity evidence is also well established and justified by the American Educational Research Association, the American Psychological Association and the National Council for Measurement in Education (17). The original license to use the PEDS-BKAT was purchased in 1998.

**Content Validity**

Content validity is the degree to which the items on a tool adequately represent the universe of content (18). Polit, Beck and colleagues (19, 20) outline several methods of quantifying experts’ degree of agreement regarding the content relevance of an instrument. These include averaging experts’ ratings of item relevance and using a pre-established criterion of acceptability, using coefficient alpha to quantify agreement on item relevance by three or more experts, and computing a multi-rater kappa coefficient. The content validity index (CVI) was used to determine item (I-CVI) and scale (S-CVI(Ave)) validity in this study. Polit, Beck and colleagues propose that with greater than three experts and clear instructions regarding the rating task, the CVI has advantages over the aforementioned methods with regards to focus on agreement of relevance and consensus, and provision of both item and scale information (19, 20).

**Sample:** A panel of 20 experts was identified through a professional network comprising nationally and locally recognised professionals based on the following criteria:

1. PICU Nurse Educator
2. At least one representative from each PICU in Australia and New Zealand.
These nurses were recognised for their expertise in PICU nursing staff orientation and education and were thus able to assess the basic knowledge domain necessary for the delivery of safe PICU nursing care.

**Data Collection:** The experts were asked to rate each PEDS-BKAT item for its relevance to basic knowledge of PICU nurses in their unit (1-satisfactory, 2-satisfactory with minor changes, 3-not satisfactory in current format, 4-not satisfactory as not within domain of basic knowledge). Internet-distributed surveys were utilised to facilitate timely responses, given the wide geographical distribution across Australia and New Zealand. The internet survey for this study was designed following the principles for constructing web surveys as outlined by Dillman (21). Dillman’s principles represent an attempt to ease the tension that exists between applying traditional questionnaires and computer logic and include: constructing web questionnaires so they scroll from question to question, providing a PIN number for limiting access only to people in the sample, and using graphical symbols or words that convey a sense of where the respondent is in the completion process, but avoid those that require significant increases in computer resources (21).

**Analysis:** An I-CVI was computed as the number of experts giving a rating of either 1 (satisfactory) or 2 (satisfactory with minor changes), divided by the total number of experts. S-CVI (Ave) was computed as the average of the I-CVIs for all items on the scale. Polit and Beck (19) suggest that this calculation places the focus on average item quality rather than on average expert performance. Furthermore, they propose that any I-CVI greater than 0.78 would be considered excellent and the goal for the S-CVI should be 0.90 or higher (20).

**Cognitive Level and Adherence to Standard Item Writing Principles**
MCQs are often criticised for not reflecting the cognitive levels and degrees of sophistication that nurses require in practice. Many authors agree that MCQs can assess higher cognitive functions such as analysis and problem solving (5, 22-25). In addition, one of the most common problems affecting MCQs is the presence of item writing flaws, despite item writing guidelines being well documented (26-29). Haladyna et al. (27) summarised current recommendations concerning item writing and generated a list of 31 principles to guide item writing. Examples of principles in this list include avoidance of: “all of the above” or “none of the above”; unfocused stems; and negative words in the stem. Using these principles, Downing (26) demonstrated that ineffective item formats can make MCQs up to 15% more difficult, thereby disadvantaging test takers and reducing the validity of examinations. It was therefore considered important to include item formatting in this assessment.

**Sample:** Four raters, local experts in the content area of PICU nursing but not the previously described nurse educators, independently rated the PEDS-BKAT items. Collectively these four experts had significant clinical PICU experience and represented interests in PICU management, research and education.

**Data Collection:** The expert raters were asked to classify 96 of the PEDS-BKAT items (the four short answer items were not assessed) against the 31 principles of effective item writing (27). A *standard* item was operationally defined as any item that did not violate one or more of Haladyna et al.’s principles. A *flawed* item was operationally defined as an item that violated one or more of these principles (26). Items were classified as either standard or flawed. If the item was classified as flawed, the type of rule violation was recorded. The experts were also asked to rate which cognitive level each item addressed: recall, understanding, problem solving or critical thinking. To ensure consistency in reporting, the
experts were provided with definitions of each cognitive level and examples of a MCQ assessing the same content area but at differing cognitive levels.

**Analysis:** Descriptive statistics using counts and percentages were computed. For the 96 PEDS-BKAT items, three separate scales were computed: a total scale which reflected the characteristics of the test as it was administered, a standard scale which reflected the characteristics of a hypothetical test that included only the unflawed items, and a flawed scale which reflected the characteristics of a hypothetical test that included only the flawed items (26).

**Reliability and Item Analysis**

Improvements in curriculum can be guided by how learners perform on tests, providing the test is valid and reliable for the context (5). Therefore reliability and item analysis are important considerations for evaluating the PEDS-BKAT in the Australian and New Zealand setting. Reliability is defined as the degree to which a measure gives consistent or reproducible results when applied in different situations (30). There are a number of statistical formulae for estimating the reliability of an exam. The Kuder-Richardson Formula 20 (KR-20), for example, calculates the internal consistency reliability coefficient for measures with dichotomous choices (eg. correct/incorrect), and is based on the number of test items, the proportion of the responses to an item that are correct, the proportion of responses that are incorrect and the spread of scores. A high KR-20 coefficient (>0.90) indicates a test that measures largely the same characteristics or domains of knowledge, with a level of 0.70 or greater being considered as generally acceptable (5, 10, 30). Item analysis is a process which examines student responses to individual test items in order to assess the quality of those items and of the test as a whole. Item analysis is especially valuable in improving items
which will be used again in later tests, and can be used to eliminate ambiguous or misleading items in a single test administration. In addition, item analysis is valuable for identifying specific areas of course content which need greater emphasis or clarity (10).

Sample: The study consisted of a convenience sample of 36 PICU nurses who were in the process of completing the Graduate Certificate Course in Paediatric Intensive Care Nursing. A graduate certificate course reflects what is considered fundamental, formalised education in specialty areas in Australia and was therefore appropriate to evaluate nurses enrolled at this level of their professional development. The PICU nurses were all permanently employed in one of the three campuses caring for critically ill children in one Australian city.

Data Collection: The PEDS-BKAT test (15) was administered prior to the commencement of the graduate certificate course under exam-like conditions. Nurses recorded their responses on a computerised mark sense sheet. There were four options with one correct response per question. No marks were deducted for incorrect answers.

Analysis: For the three scales described in the previous section on adherence to item writing principles (total, standard and flawed), the following data were computed: raw test scores; Kuder-Richardson 20 reliability (KR-20); mean item discrimination; and mean item difficulty. Mark sense sheets were analysed by a university’s computer system running the Exam System II software package, created by NCS Pearson.

Ethics
Ethics approval was granted by each of the eight participating hospitals and the university. Consent was either written or implied by return of the particular survey/test. Each of the three
phases contained an information cover sheet which outlined that: 1) participation in the study was voluntary, 2) participants were free to withdraw before or during the study without comment or penalty, and 3) participants would not be prejudiced as a result of their answers. In addition, participants were made aware of the purpose, nature and risks/benefits of the research. The participants who completed the PEDS-BKAT were advised that their scores would only be known to the researchers and would not influence their employment within their respective institutions. All information provided by the participants was anonymous and confidential and all records were kept separate and a coding system, known only to the researchers, was used.

**RESULTS**

**Content Validity**

Ten of the 20 experts selected participated in the content validity study by completing the internet survey (50% response rate). Demographics for the participating experts are outlined in Table 1. These experts were predominantly Nurse Educators with critical care qualifications and an average of 20 years of nursing experience. I-CVIs ranged from 0.00 to 1.00, with a mean I-CVI of 0.70. Twenty one items had an I-CVI of 1.00. Forty-three percent (43%) of I-CVIs were less than 0.8. S-CVI (Ave) was 0.71.

**Cognitive Level and Adherence to Standard Item Writing Principles**

The four participating expert independent raters held positions in paediatric nursing research (n=2), PICU nursing management (n=1) and PICU nursing education positions (n=1), and had an average of 23 years of nursing experience. Thirty-four of the 96 items (35%) were classified as flawed by the expert raters. The use of a negatively worded stem was the most frequent flaw (49%). Items where choices were not independent, more than one choice could
have been the correct answer, and items which were overly specific or overly general contributed to a further 43% of the flawed items. Of the 100 items, 35% were written at the recall/knowledge level; 58% were at the understanding level; and 7% were at the problem solving level.

Reliability and Item Analysis
Of the participating 36 PICU nurses the mean years of nursing practice was 12.4 (SD 6.3). The median length of PICU experience was 6.1 years (min 0.9, max 30 years). The mean PEDS-BKAT score for all nurses was 60.8 (SD=9.6) out of a possible score of 96. The scores ranged from 43 to 82. Kuder-Richardson coefficient (KR-20) was used to measure how well all the items on the PEDS-BKAT worked together and was 0.81. The mean item difficulty was 0.62, and the mean discrimination index was 0.23. Values ranged from 0.08 to 1.00 for item difficulty and from -0.19 to -0.72 for item discrimination. Two of the items did not discriminate (indices of 0) and 12 items had negative indices, indicating that nurses who received a low total score chose the correct answer to these items more often than the nurses who scored higher overall. In addition, 15% of items on the PEDS-BKAT had discrimination indices greater than 0.4.

Table 2 shows the summary psychometric characteristics of the overall PEDS-BKAT test and its two subsets: standard and flawed. Comparing the standard (62 items) and the flawed (34 items) scales, the observed KR-20 reliability was 0.79 versus 0.56. The standard scale mean item difficulty was 0.65 and the flawed scale mean item difficulty was 0.55. The mean item discrimination was equal for both the standard and flawed scales (0.25).

DISCUSSION
This study identified that in its current form, the PEDS-BKAT is not a reliable and valid measure of basic nursing knowledge for Australian and New Zealand paediatric intensive care nurses. Forty-three percent of items scored an I-CVI of less than 0.8. Polit et al. (20) suggest that I-CVI should be greater than 0.78 to be considered an excellent item. Reasons for low I-CVIs and subsequent low S-CVI (Ave) could include a) item not considered a measure of basic knowledge for PICU nurses, and b) content covered by item is not practiced by Australia and New Zealand PICU nurses. Given that reliability is a necessary, but not sufficient, condition for validity, validity results must be weighted with importance when determining a tool’s feasibility of use (31). In this study, whilst a reliability coefficient of 0.81 could be considered adequate, the low content validity results from this study suggest that the PEDS-BKAT should be used with caution in our context. Others have had similar experiences when validating tools outside of the environment for which they were developed. Schaller and James (16) examined the nutritional knowledge of nurses in Australia and questioned whether the theoretical concepts seen as relevant by the overseas developers of their nutrition tool were aligned with those perceived as relevant in Australia. Similar investigations of the adult version of the BKAT in Australia have also questioned the suitability of its items in the Australian context (32).

Content for the PEDS-BKAT is based on an adult version of the tool, the Basic Knowledge Assessment Tool (BKAT). The BKAT was developed in the USA in 1984 through a review of literature, interviews with nursing staff, and an expert panel of nursing clinicians and educators. Content for the PEDS-BKAT was further developed from clinical practice and a panel of experts in paediatric intensive care nursing. Of the original 100 items of the adult BKAT (version 4), 20 were replaced with new items specific to paediatric intensive care, and minor changes in the wording of 10 additional items were made so that they applied correctly
to paediatrics. No changes were made to the remaining 70 items (15). Downing and Haladyna (10) recommend that test content should be aligned with curriculum standards or rigorous analysis of current practice, neither of which were described in the original development of the PEDS-BKAT (15). Since this study was undertaken, a subsequent version of the PEDS-BKAT has since been developed (33), containing 90 items, however limited reliability and validity data is available and this is not reported in the peer reviewed literature.

The frequency of item writing flaws found in the PEDS-BKAT (35%) is not unexpected, given that well constructed MCQ items can be time-consuming and difficult to write (11). Other studies have demonstrated that item writing flaws are not unique to nursing and have occurred in up to 75% of questions from examinations and instructor guides accompanying textbooks (12, 26, 28, 29). In this study, flawed items contributed to questions being 10% points more difficult than standard items (non-flawed). Downing (26) concluded in his study of medical student examinations that up to 15% of students could be incorrectly classified as failed when they should have been classified as passes, due solely to flawed item formats. In the case when examinations are high-stakes, this false negative rate seems unreasonably high (26).

Very few questions in the PEDS-BKAT were found to assess cognitive ability at the problem-solving level. In a professional discipline such as nursing, clinicians are required to process large amounts of complex information and to act responsibly upon that information. Ideally, test questions should be written to reflect the level of sophistication at which clinicians are expected to practice (28); nurses who take exams requiring rote recall (knowledge level) do not necessarily perform at levels requiring higher order thinking that is associated with functioning well in the clinical setting (28, 34). Many authors advocate that
test items should match the intended cognitive levels outlined in learning objectives (23-25). Masters et al. (34) examined the quality of 2913 MCQs in nursing and found that 72% of test bank questions were written at knowledge and comprehension levels. Tarrent et al. (28) also examined the quality of 2770 MCQs in pre-registration and post-registration nursing courses and found 91% of questions were written at the recall and comprehension level. Whilst it is not expected that these types of assessment should have all questions testing at the higher cognitive domains, unfortunately the literature does not provide any guidance as to the appropriate proportion of higher cognitive questions. Tarrent et al. (28) stated that tests containing less than 20 percent of higher order thinking items were difficult to justify, even in undergraduate programs. If the desired outcome of intensive care orientation and transition programs is for nurses to utilise critical thinking skills to inform sound clinical judgements, then test questions written only at knowledge and comprehension levels may not be a valid assessment of PICU nurses’ learning. We propose that in postgraduate PICU nursing, a measure of knowledge and skills needs to include closer to 50 percent of higher order thinking questions.

The mean item difficulty of the PEDS-BKAT was 0.62. In norm-referenced testing, an item difficulty index of 0.5 is considered standard because the ability of an item to differentiate between higher and lower achieving students depends on variability. However, Nunnally and Bernstein (30) suggest that when guessing becomes a factor, the index should be greater than 0.60. Further to this, Nunnally and Bernstein (28) recommend that multiple choice items with four or more alternatives have an index of 0.35-0.80. Using Nunnally’s criteria, the mean item difficulty for the overall scale has a good item index.
In this study, the mean item discrimination for the PEDS-BKAT has positive discrimination, indicating that students who received a high total score chose the correct answer for a specific item more often than the students who had a low overall score. However, the size of the mean discrimination is small (30). In examining the individual items on the test, discrimination indices ranged from -0.19 to 0.72. Two items did not discriminate with indices of 0. Twelve items had negative indices indicating that students who received a low total score chose the correct answer to these items more often than the students who scored higher overall. Only 54% of items on the PEDS-BKAT had discrimination indices greater than 0.3. Based on these indices, almost half of the test items would require revision, modification or replacement if they were to be used in the Australian and New Zealand context.

Nurses in this study achieved a mean PEDS-BKAT score of 60.8, with only two nurses scoring higher than Runton and Toth’s (15) reported mean of 79. Two interpretations of these results could be that participating nurses demonstrated poorer knowledge in the domain of paediatric intensive care, or the PEDS-BKAT is not suitable for use in the Australian and New Zealand context. Questions could also be raised about the differences between Toth’s sample and the sample in this study, and the effect of practice change between the development of the PEDS-BKAT and the testing of this tool in the Australian and New Zealand setting. Given the combined results of this study, it would appear that two options remain: modify the PEDS-BKAT for use in the Australian and New Zealand setting, with permission; or develop a new tool. Considering the framework for which the PEDS-BKAT was developed and the extensive modifications which would be required, we suggest it would be preferable to start with a mapping of what Australian and New Zealand PICU nursing practice is currently and then utilise this information to inform the development of a new tool.
Strengths and Limitations

To our knowledge, this is the first study to evaluate the reliability and validity of the PEDS-BKAT since its development. Findings from this study are consistent with others which demonstrate the difficulties with interpreting test scores outside of the country with which they are developed. Content validity was examined using a rigorous process and provided both overall scale and item level data. Furthermore, extensions to traditional instrument psychometric testing provided a comprehensive approach to assessing reliability and validity. These included assessment of cognitive levels and adherence to item writing principles, both of which have been demonstrated to influence scores if items are constructed poorly. Generalisability of the findings from this study, however, may be limited by several factors. First, this study examined content validity of the PEDS-BKAT in Australia and New Zealand with ten nurse educators. Although we assessed content validity with a representative sample, there is an assumption that the role of the nurse educator had a good grounding in the educational principles and needs of PICU nurses. Additionally, as we examined PICU nursing knowledge scores in a small convenience sample in one state of Australia and New Zealand, it is also possible that our analysis suffers from some selection bias and that our findings do not accurately reflect the PICU nursing population in the region. Finally, only one form of reliability analysis was undertaken.

CONCLUSIONS AND RECOMMENDATIONS

Nursing education is an expensive undertaking which aims to educate nurses who will continually strive to provide patient care of a high quality. Continual evaluation of effectiveness of nursing education therefore is essential to assess knowledge deficits, plan educational sessions, and determine ways to promote quality nursing care. It is of paramount importance that multiple choice assessments used in nursing are valid and reliable if nursing,
as a profession, is to produce credible results that may be used to change nursing practice or methods of nursing education.

Paediatric intensive care nursing is a growing specialty which requires evaluation of its educational programs and outcomes in each different practice environment. Measurement of this needs to reflect the complex environment and processes of the PICU, which includes assessment of higher order thinking. This study found that almost half of the PEDS-BKAT items did not reach content validity levels of 0.8. Thirty five percent of the items met criteria for multiple choice item writing violations and over 90% of the items were written at the knowledge and understanding level. Further, nurses in this study scored lower than previously reported means and items discriminated poorly. It is suggested that some of these results reflect variations in practice in Australia and New Zealand. Therefore research into the knowledge and skills of paediatric intensive care nurses in Australia and New Zealand and the profession’s own expectations of knowledge and skill levels would be extremely useful. This would help determine a PICU nurse knowledge and skill standard for graduating and practicing nurses and could assist in the design of a new assessment tool. Nurse Educators could then design appropriate educational courses to maximise knowledge levels, improve skills and ultimately enhance patient care.
References

Figure 1. Flow diagram of methods

Reliability and Validity Assessment of the PDE-BKAT

Content Validity
- Assessed by PICU nurse educators
- Each item assessed for suitability of use in Australia & New Zealand using 4-point scale (satisfactory, satisfactory with minor changes, not satisfactory in current format, not satisfactory as not within domain of basic knowledge)
- Content validity index generated for each item (I-CVI) and total tool (S-CVI (Ave))

Cognitive Level and Adherence to Item Writing
- Assessed by PICU experts
- Each item assessed for level of cognition using 4-point scale (recall, understanding, problem solving, critical thinking)
- 96 items assessed (4 items not MCQ format) for adherence to item writing principles using 31 evidence based standards (scored as standard or flawed)

Reliability and Item Analysis
- PICU nurses tested
- Administration under test conditions using marked sense sheets
- Raw and mean scores reported
- Three scales reported on: Total, Standard, and Flawed test
- Reliability (Kuder-Richardson-20)
- Item analysis reported as item difficulty and item discrimination

Table 1. Demographics of Nurse Educators evaluating content validity

<table>
<thead>
<tr>
<th></th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of educator</td>
<td></td>
</tr>
<tr>
<td>Nurse Educator</td>
<td>70 (7)</td>
</tr>
<tr>
<td>Clinical Nurse Educator/Facilitator</td>
<td>30 (3)</td>
</tr>
<tr>
<td>Current education position</td>
<td></td>
</tr>
<tr>
<td>Job shared</td>
<td>60 (6)</td>
</tr>
<tr>
<td>Full time</td>
<td>40 (4)</td>
</tr>
<tr>
<td>Postgraduate Critical Care Qualification</td>
<td>90 (9)</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Postgraduate PICU Qualification</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>60 (6)</td>
</tr>
<tr>
<td>Length of time in any nursing practice (years): – mean (SD)</td>
<td>20.6 (6.1)</td>
</tr>
<tr>
<td>Length of time in previous education positions (years): – mean (SD)</td>
<td>1.7 (1.8)</td>
</tr>
<tr>
<td>Length of time in current education position (years): – mean (SD)</td>
<td>4.1 (3.2)</td>
</tr>
<tr>
<td>Hours/fortnight worked in current education position: – mean (SD)</td>
<td>47.0 (18.7)</td>
</tr>
</tbody>
</table>
Table 2. Psychometric characteristics of the PEDS-BKAT

<table>
<thead>
<tr>
<th></th>
<th>Original Test (n=96)</th>
<th>Standard Items (n=62)</th>
<th>Flawed Items (n=34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean test score (SD)</td>
<td>60.8 (9.6)</td>
<td>79</td>
<td>56</td>
</tr>
<tr>
<td>Range of test scores</td>
<td>43.0-82.0</td>
<td>0.65</td>
<td>0.55</td>
</tr>
<tr>
<td>KR20 Reliability</td>
<td>0.81</td>
<td>0.79</td>
<td>0.56</td>
</tr>
<tr>
<td>Mean Difficulty</td>
<td>0.62</td>
<td>0.65</td>
<td>0.55</td>
</tr>
<tr>
<td>Mean Discrimination</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>