



Validation of Assessment Tool for Cranial Ultrasound Performance

Pattama Tanaanantarak

**A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor
of Philosophy in Health Sciences**

Prince of Songkla University

2024

Copyright of Prince of Songkla University



Validation of Assessment Tool for Cranial Ultrasound Performance

Pattama Tanaanantarak

**A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor
of Philosophy in Health Sciences**

Prince of Songkla University

2024

Copyright of Prince of Songkla University

Thesis Title Validation of Assessment Tool for Cranial Ultrasound Performance

Author Miss Pattama Tanaanantarak

Major Program Health Sciences

Major Advisor*Pasuree S.*

Digitally signed : 2024.05.23 9:49:31 +07:00

(Assoc. Prof. Dr. Pasuree Sangsupawanich)

Examining Committee :

Digitally signed : 2024.06.05 23:07:25 +07:00

.....airperson
(Assoc. Prof. Dr. Cherdsak Iramaneerat)*Pasuree S.*

Digitally signed : 2024.05.23 9:49:33 +07:00

.....ommittee
(Assoc. Prof. Dr. Pasuree Sangsupawanich)

Digitally signed : 2024.06.08 13:37:48 +07:00

.....ommittee
(Assoc. Prof. Dr. Osaree Akaraborworn)*P. Thongsuksai*

Digitally signed : 2024.06.10 9:58:26 +07:00

.....ommittee
(Assoc. Prof. Paramee Thongsuksai)

The Graduate School, Prince of Songkla University, has approved this Thesis
as Fulfillment of the requirements for the Doctor of Philosophy in Health Sciences



Digitally signed : 2024.06.10 11:56:28 +07:00

(Asst. Prof. Dr. Kawinbhat Sirikantisophon)
Acting Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate's own investigations. Due acknowledgement has been made of any assistance received

Pasuree S......Signature
Digitally signed : 2024.05.23 9:49:35 +07:00

(Assoc. Prof. Pasuree Sangsupawanich)

Major Advisor

Miss Pattama Tanaanantarak

Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

Miss Pattama Tanaanantarak
Candidate

ชื่อวิทยานิพนธ์	Validation of Assessment Tool for Cranial Ultrasound Performance
ผู้เขียน	น.ส. ปัทมา ธนอนันตรักษ์
สาขาวิชา	วิทยาศาสตร์สุขภาพ
ปีการศึกษา	2566

บทคัดย่อ

บทนำ:

เนื่องจากความก้าวหน้าอย่างรวดเร็วในด้านสรีรศาสตร์และการดูแลทารกแรกเกิด การทำอัลตราซาวด์สมองเป็นวิธีการมาตรฐานในการประเมินความผิดปกติของสมองในทารกแรกเกิดที่คลอดก่อนกำหนด การเพิ่มจำนวนผู้เชี่ยวชาญในการทำอัลตราซาวด์สมองให้เพียงพอจึงมีความจำเป็น โดยต้องมีแบบประเมินที่เที่ยงตรงและเหมาะสม การศึกษานี้มีวัตถุประสงค์เพื่อสร้างและตรวจสอบความเที่ยงตรงของแบบประเมินประสิทธิภาพของการทำอัลตราซาวด์สมอง

วิธีการศึกษา:

การศึกษานี้ใช้วิธี modified Delphi เพื่อสร้างแบบประเมินสำหรับการประเมินประสิทธิภาพของการทำอัลตราซาวด์สมอง เริ่มจากผู้วิจัยรวบรวมคุณลักษณะที่เหมาะสมสำหรับการประเมินการทำอัลตราซาวด์สมองใส่ในแบบประเมินเบื้องต้นหลังการทบทวนวรรณกรรม จากนั้นคณะผู้เชี่ยวชาญในการทำอัลตราซาวด์สมองซึ่งเป็นรังสีแพทย์เฉพาะทางเด็กแก้ไขโดยใช้กระบวนการ Delphi จำนวนสามรอบจนได้คุณลักษณะที่จะประเมินครบแล้ว ลำดับต่อมาได้มีการเพิ่มสเกลรูปกริในแบบประเมินสำหรับการให้คะแนน แบบประเมินที่มีสเกลรูปกริจะส่งให้คณะผู้เชี่ยวชาญพิจารณาอีกครั้งเพื่อยืนยันรับรองแบบประเมินขั้นสุดท้ายทำให้ได้แบบประเมินประสิทธิภาพการทำอัลตราซาวด์สมองใหม่ หลังจากนั้นแบบประเมินประสิทธิภาพการทำอัลตราซาวด์สมองใหม่จะถูกประเมินความเที่ยง โดยผู้ประเมินสองคนนำแบบประเมินมาใช้ประเมินการทำอัลตราซาวด์สมองของทารกแรกเกิดโดยแพทย์ประจำบ้านรังสีวินิจฉัยจำนวน 27 คนที่บันทึกการทำอัลตราซาวด์สมองในวิดีโอคลิปแบบประเมินจะถูกประเมินความเที่ยงและร้อยละของความเห็นพ้องในการใช้แบบประเมินให้คะแนนการทำอัลตราซาวด์สมองในวิดีโอคลิปอันเดียวกัน

ผลการศึกษา:

รังสีแพทย์เฉพาะทางเด็กจำนวน 17 คนที่ทำงานในสภาพแวดล้อมที่แตกต่างกันเข้าร่วมในคณะผู้เชี่ยวชาญของการศึกษานี้ ความถูกต้องของเนื้อหาในแบบประเมินสร้างขึ้นจากการที่ผู้เชี่ยวชาญทุกคนซึ่งเป็นรังสีแพทย์เฉพาะทางเด็กมีความเชี่ยวชาญอย่างสูงในความรู้ความสามารถของการทำอัลตราซาวด์สมอง หลังจากแบบประเมินผ่านกระบวนการ Delphi ทั้งสามรอบ แบบประเมินมีคุณลักษณะที่ใช้ประเมินการทำอัลตราซาวด์สมองทั้งหมด 14 รายการ สุดท้ายแล้วกลายเป็นแบบประเมินที่มีคุณลักษณะที่ใช้ประเมินการทำอัลตราซาวด์สมองทั้งหมด 15 รายการ หลังจากนั้นแบบประเมินเพิ่มใช้สเกลรูปรีก และแบ่งออกเป็น 3 ส่วนของการประเมิน ได้แก่ ขั้นตอนการเตรียมการ การใช้เครื่องอัลตราซาวด์ และประสิทธิภาพของการทำอัลตราซาวด์สมอง เพื่อให้ผู้ประเมินประเมินได้ง่ายขึ้น ความเที่ยงของแบบประเมินถูกทดสอบด้วย Kappa ของ Cohen ค่า Kappa และร้อยละของความเห็นพ้องระหว่างผู้ใช้แบบประเมินสองคนส่วนใหญ่อยู่ในระดับปานกลางถึงเกือบสมบูรณ์แบบ ค่า Kappa เท่ากับ 0.42–0.93 และ ร้อยละของความเห็นพ้องเท่ากับ 77.8–100%

สรุปผลการศึกษา:

การศึกษานี้สร้างแบบการประเมินการทำอัลตราซาวด์สมองที่ผ่านการตรวจสอบเป็นครั้งแรก โดยใช้วิธี modified Delphi แบบประเมินขั้นสุดท้ายซึ่งแบ่งออกเป็นสามส่วนและมีสเกลรูปรีก ใช้งานง่ายและสามารถเชื่อถือได้

Thesis Title	Validation of Assessment Tool for Cranial Ultrasound Performance
Author	Ms. Pattama Tanaanantarak
Major Program	Health Sciences
Academic year	2023

ABSTRACT

Introduction

Due to rapid advances in obstetrics and neonatal care, cranial ultrasound (US) has become the standard method for evaluating intracranial abnormalities in preterm neonates. To develop sufficient numbers of skilled cranial ultrasonologists, evaluating them using a consistent and suitable assessment tool is necessary. The objective of this study is to create an evaluation tool for cranial US performance.

Materials and Methods

We used a modified Delphi method to produce a quality assessment tool for evaluating cranial US performance. The researcher developed the initial evaluation tool items following a literature review and then the expert panel modified them using a three-round Delphi process. Rubric scales for grading assignments were added when consensus regarding generated items was reached. Experts were consulted to confirm the final assessment tool using a rubric scale. Two raters evaluated cranial US performance of 27 residents in video clips using the assessment tool. Reliability and percent agreement were assessed.

Result

Seventeen paediatric radiologists working in varying environments participated in our expert panel. The proposed evaluation tool's content validation was facilitated by all experts who were paediatric radiologists. Following three rounds of the Delphi

process, the initial 14-item assessment form became a final 15-item form. A rubric scale was used in the final form, which was divided into 3 sections: the preparation phase, US machine operation, and cranial US performance, to make it easier for raters to assess in the field. Interrater reliability was evaluated with Cohen's Kappa. The Kappa value and percent agreement between two raters for most items in the tool was moderate to almost perfect (0.42–0.93) and 77.8–100%, respectively.

Conclusion

This study produced the first validated cranial US assessment tool using a modified Delphi method. The final assessment form, which was divided into three sections and included a rubric scale, was simple to use and potentially reliable.

ACKNOWLEDGEMENT

I would like to express my gratitude to the following people:

To the thesis examining committees, Assoc. Prof. Dr. Cherdsak Iramaneerat, Assoc. Prof. Dr. Osaree Akaraborworn, Assoc. Prof. Paramee Thongsuksai, and Assoc. Prof. Dr. Pasuree Sangsupawanich for their valuable comments and suggestions.

To my advisor, Assoc. Prof. Dr. Pasuree Sangsupawanich for their valuable guidance, comments, and continuous support throughout my research.

To all pediatric radiologists in Thailand for their contributions to develop this assessment tool.

To secretary and my classmates at Health science PhD program, Faculty of Medicine, Prince of Songkla University for all your support.

To Asst. Prof. Srila Samphao for your comments and suggestions.

Finally, to my family who their support and encourage.

Pattama Tanaanantarak

Contents

Contents	Page
List of Tables	8
List of Figures	9
List of Abbreviations and Symbols	11
Chapter	
Chapter 1: Introduction	15
1.1 Background and Rationale	15
1.2 Objectives of the study	16
1.3 Literature review	16
1.3.1 Cranial ultrasound	16
1.3.2 Cranial ultrasound pathologic findings in preterm babies	32
1.3.3 Assessment tool in medical education	35
1.3.4 Assessment tool development	44
1.3.5 Principle of assessment tool evaluation	57
Chapter 2: Methodology	64
2.1 Study design	64
2.2 Study procedure	64
2.2.1 Phase I: Assessment tool development	64
2.2.2 Phase II: Assessment tool performance evaluation	67
2.3 Statistic analysis	70
Chapter 3: Results	71
3.1 Phase I: Assessment tool development	71
3.1.1 Participants	71
3.1.2 Delphi process	72
3.2 Phase II: Assessment tool performance evaluation	80
3.2.1 Subjects characteristics	80
3.2.2 Content validity	82

Contents	Page
3.2.3 Interrater reliability	82
3.2.4 Construct validity	84
3.2.5 Internal consistency reliability	87
Chapter 4: Discussion	88
Chapter 5: Conclusion	92
References	93
Appendices	103
Appendix A: Study flow	103
Appendix B: Time schedule of the study	104
Appendix C: Record forms	105
Appendix D: Initial cranial US assessment form	107
Appendix E: Initial cranial US skill assessment form with rubric score, example	108
Appendix F: Dummy table 1: Predictive factors information	109
Appendix G: Dummy table 2: Summary DOPS scores and interrater reliability	110
Appendix H: Item generation	111
Appendix I: The ethics committee approval	147
Vitae	156

Contents	Page
List of tables	
1.1 Cranial US schedule for routine surveillance of preterm infants with uncomplicated clinical course and should be intensified if clinically indicated or in the presence of anomalies detected on cranial US	17
1.2 Grading of intraventricular hemorrhage	33
1.3 Grading of periventricular leukomalacia	34
1.4 The main differences between three types of assessment	36
1.5 Assessment methods with their domain, purpose of assessment, strengths, and limitations	42
1.6 Literatures of ultrasound skill assessment tool development	45
1.7 Advantages and disadvantages of three, four, and five levels of performance	51
1.8 The use of neutral and non-neutral choice on the Likert scale	52
1.9 Example of holistic rubrics	54
1.10 Example of analytic rubrics	55
1.11 The advantages and disadvantages of rubrics	56
1.12 A 2x2 table of sample results in example comparing two different raters rate of presence or absence of a finding on imaging	61
1.13 Cohen's Kappa interpretation	62
3.1 Characteristic of experts in expert panel	71
3.2 Response of experts to Delphi process	72
3.3 Items in initial assessment tool	73
3.4 Initial items with mean scores and distributions in Delphi round 1	74
3.5 Items with mean scores and distributions in Delphi round 2	76
3.6 Final cranial US performance assessment form with rubrics	78

Contents	Page
3.7 Characteristic of neonates	80
3.8 Characteristic of Radiologic residents	81
3.9 Interrater reliability of the assessment tool	83
List of figures	
1.1 The ultrasound machine	20
1.2 Details of each button function of main panel of ultrasound machine	21
1.3 The sector/convex/curvilinear 5-8 MHz and linear 3-11 MHz probes	22
1.4 The optimal depth of cranial US includes from top of brain to foramen magnum	23
1.5 a: Six standard coronal planes on cranial US through anterior fontanelle in mid sagittal view; b: five standard sagittal planes on cranial US through anterior fontanelle in coronal view	25
1.6 First coronal images	25
1.7 Second coronal images	26
1.8 Third coronal images	27
1.9 Fourth coronal images	28
1.10 Fifth coronal images	29
1.11 Sixth coronal images	29
1.12 First sagittal images	30
1.13 Second and Fourth sagittal images	31
1.14 Third and Fifth sagittal images	31
1.15 Summary of standard coronal images for cranial US	32
1.16 Summary of standard sagittal images for cranial US	32
1.17 Miller's pyramid	37
1.18 The typical three rounds Delphi process	47
1.19 Example of rubric structure	50

Contents	Page
1.20 Selection of suitable inter- and intrarater reliability coefficients depending on the type of data	59
1.21 Five-point Likert scale example with varying degrees of disagreement	61
2.1 Guidelines of the minimum sample size requirements for Cohen's Kappa	68
2.2 Merge video footage	69
3.1 Example of Google form of rating scale for initial items in cranial US assessment form	74
3.2 Box and whisker plot of total score and PGY 1, 2, and 3 of radiologic residents	84
3.3 Box and whisker plot of total score and experience of cranial US cases of radiologic residents	85
3.4 Box and whisker plot of overall performance and PGY 1, 2, and 3 of radiologic residents	86
3.5 Box and whisker plot of overall performance and experience of cranial US cases of radiologic residents	87

LIST OF ABBREVIATIONS AND SYMBOLS

IVH	=	Intraventricular hemorrhage
PVL	=	Periventricular leukomalacia
US	=	Ultrasound
NICU	=	Neonatal intensive care unit
AIUM	=	American Institute of Ultrasound in Medicine
GA	=	Gestational age
HUS	=	Head ultrasound
NEC	=	Necrotizing enterocolitis
PHVD	=	Post-hemorrhagic ventricular dilation
WMI	=	White matter injury
CDI	=	Color Doppler images
MHz	=	Megahertz
TGC	=	Time-gain compensation curve
MI	=	Mechanical index
TI	=	Thermal index
GM	=	Germinal matrix
WPBA	=	Workplace-based assessment
MCQ	=	Multiple choice question
SAQ	=	Self-assessment questionnaire
OSCE	=	Objective structured clinical examination
Mini-CEX	=	Mini-clinical evaluation exercise
DOPS	=	Direct observation of procedural skills
MEQ	=	Modified essay question
RAND Corporation	=	Research and Development Corporation
IQ	=	Intelligence Quotient
Kw	=	Kappa coefficient
ICC	=	Intraclass correlation coefficient
K	=	Kappa
NaN	=	Not a Number
SE _k	=	Standard error of kappa

REC	=	Research Ethics Committee
PGY	=	Postgraduate year
HD	=	High definition

LIST OF PAPER

This thesis is based on the following paper:

Paper 1: Pattama Tanaanantarak, MD, Supika Kritsaneepaiboon, MD, Vorawan Charoonratana, MD, Chomanad Chittchang, MD, Kanokporn Chutiwongthanaphat, MD, Preeyacha Pacharn, MD, Kriengkrai Iemsawatdikul, MD, Nichanan Ruangwattanapaisarn, MD, Suwadee Eng-Chuan, MD, Thitiporn Junhasavasdikul, MD, Atchara Mahayosnond, MD, Nipaporn Tewattanarat, MD, Piyaporn Kitpraphasiri, MD, Teerasak Phewplung, MD, Thipsumon Tangsiwong, MD, Pasuree Sangsupawanich, MD, PhD. A cranial ultrasound performance assessment tool: development and validation, Medical Ultrasonography, Dec 2023,1-8. DOI: 10.11152/mu-4307. (Indexed by Medline/Pubmed, Scopus, WOS (SCIE). Impact factor = 1.7)

Subject Răspuns: Problem about artical fee paying
From Dana Fodor <dfodor@umfcluj.ro>
To: pattama tanaanantarak
<tanaanantarak@yahoo.com>
Cc: medultrasonography@gmail.com
<medultrasonography@gmail.com>
Date 11 Mar at 12:13

Dear Dr Pattama Tanaanantarak,

you ahve the permission for reprinting the paper.

Best regards,

Prof Dr Daniela Fodor
Editor sef "Medical Ultrasonography"
Clinica Medicala II
UMF "Iuliu Hatieganu" Cluj-Napoca

Show trimmed content ▾

Chapter 1

Introduction

1.1 Background and rational

Nowadays, obstetrics and neonatal care is evolving dramatically, resulting in increasing the survival rate of premature babies. These premature babies are prone to have intracranial complications, particularly intraventricular hemorrhage (IVH) and periventricular leukomalacia (PVL). However, the symptoms and signs are often ambiguous.¹ Since 1970s, ultrasound (US) has been a widespread tool to screen intracranial abnormalities in premature babies. US can demonstrate major abnormalities within the brain and these findings are related to long-term prognosis. Long-term abnormalities following IVH and PVL, frequently detected in premature babies, should be ranged from minor cognitive disorder to severe mental retardation.^{1,2}

The advantages of US include cheap cost, safe procedure, no radiation exposure, no sedation required and available at bedside, especially in critically-ill patients.¹ At present, cranial US has become a part of patient care, particularly in neonatal intensive care unit (NICU). However, the quality and accuracy of US are operator dependent, requiring skills and experiences. Moreover, the operators affectively conduct the cranial US are still scarce.²

Teaching or training of cranial US is crucial in order to adequately increase the numbers of expertise. In fact, learning of cranial US as well as US of other parts of the body has no affective schemes. In Songklanagarind Hospital, training of cranial US has been designed as workplace-based learning, starting from observing the pediatric radiologist experts. Next, the learners perform the cranial US under supervision for 10 or more cases as recommended by the institute. Then they will be allowed to independently perform the cranial US. In Croatia, 25 cases under supervision is the number recommended in order to independently practice.³ A logbook is one of an assessment method used to record the number of observing cases or procedural performed. The advantage of the logbook is a user-friendly assessment format. However, the disadvantages include low reliability and accuracy due to self-assessment process with no guarantee of its accuracy.⁴ Moreover, only the number of

cases does not reflect the learner's performance and competency.^{3,5} Currently, there is no standard recommendation for assessing the learner's performance in cranial US. Therefore, the appropriate assessment tool needs to be developed and validated.

1.2 Objectives of the study

Phase I: To develop new assessment tool in cranial US skill performance

Phase II: To determine the validity and reliability of the developed new assessment tool in cranial US skill performance

1.3 Literature reviews

1.3.1 Cranial US

Cranial US is a useful method for evaluating the infant brain. It is the initial imaging to check for brain abnormalities in newborns. It offers a number of benefits including safety; no

ionizing radiation and no requiring sedation, a quick and painless approach, inexpensive, availability at the bedside with repetitions, and the lack of absolute contraindication to cranial US. The indications for cranial US in accordance with the consensus reached in 2020 by the American Institute of Ultrasound in Medicine (AIUM) are shown as follows.⁶

- Abnormal increase in head circumference
- Hemorrhage or parenchymal abnormalities in preterm and term infants
- Ventriculomegaly (hydrocephalus)
- Vascular abnormalities
- Suspected hypoxic ischemic injury (hypoxic ischemic encephalopathy)
- Patients on hypothermia, extracorporeal membrane oxygenation, and other support machines
- Congenital malformations
- Signs or symptoms of a central nervous system disorder (e.g., seizures, facial malformations, macrocephaly, microcephaly, and intrauterine growth restriction)
- Congenital or acquired brain infection

- Suspected or known head trauma
- Craniosynostosis
- Follow-up or surveillance of previously documented abnormalities, including prenatal abnormalities
- Screening before surgery

Premature newborns are now more likely to survive due to the rapid advancements in obstetrics and neonatal care. These premature infants are more likely to experience intracranial problems. Because the intracranial abnormality in these babies may frequently be asymptomatic yet have a significant impact on future brain development, the cranial US screening routine is created to rule out intracranial pathology. Table 1.1 shows screening cranial US recommendation in our institute.⁷

Table 1.1 Cranial US schedule for routine surveillance of preterm infants with uncomplicated clinical course⁷

GA	First imaging	Repeat imaging*	Time-corrected imaging
≤31+6 weeks	HUS 4-7 days postbirth	4–6 weeks postbirth	- Routinely for neonates born before 26 weeks - Not routinely for neonates born between 26+0 and ≤31+6 weeks†
≥32+0 to 36+6 weeks with additional risk factors‡	HUS 4-7 days postbirth	4–6 weeks postbirth, and only if first image is abnormal	Not routinely

GA, Gestational age; HUS, Head ultrasound; IVH, Intraventricular hemorrhage; NEC, Necrotizing enterocolitis; PHVD, Post-hemorrhagic ventricular dilation; PVL, Periventricular leukomalacia; WMI, White matter injury.

*If an abnormality (Grade 2 or higher IVH or WMI) is detected on first imaging, a repeat HUS should be performed 7–10 days later. If ventricular dilation or worsening IVH/WMI is detected, the frequency of HUS should be intensified (at least weekly initially and as clinically indicated thereafter). A repeat HUS should also be conducted in the weeks following acute illness (e.g., NEC or sepsis).

† HUS between 37 and 42 weeks corrected GA if previous moderate-to-severe anomalies (Grade 3 or higher IVH, PHVD, or Grade 3–4 PVL) on HUS, or presence of additional risk factors (e.g., critical illness requiring mechanical ventilation or vasopressors, NEC, major surgery).

‡ Additional risk factors for the late and moderately preterm infant include: need for critical care out of keeping with the usual neonatal course, complicated monochorionic twin pregnancy, microcephaly, complicated postnatal course: sepsis, NEC, major surgery, or abnormal neurological symptoms.

1.3.1.1 US skills development

The development of US skills needs to be understood prior to develop the cranial US skill assessment tool. The creation of US skills consists of a combination of complex motor and cognitive skills. Motor or hand skills need to be incorporated with visual ability. The US performers move their hand with the US probe while their eyes stay on the images on US screen. This process is called hand-eye coordination that has to work continuously with cognitive skills in order to interpret the findings on US screen, so-called visual-cognitive skills.⁸

The motor skill development model proposed by Fitts and Posner in 1967^{9,10} consists of 3 stages.

Stage 1: The cognitive phase, in this stage, the learners must be conscious at all times in order to plan their hand movement during US performing. It causes US process to be slow, not smooth, and with errors, the characteristics of novices to do new skills.

Stage 2: The associative phase, this is continuous from the cognitive phase. After the learners continue to practice, US will be done faster and smoother. Mistakes are also reduced, indicating an increased expertise for the learner.

Stage 3: The autonomous stage, the stage in which the learners become experts in their practice. This autonomous stage will happen when the learners practice continuously and extensively for a period of time until they can perform US expertly, quickly, smoothly, and correctly, by doing US almost automatically.

The visual-cognitive skills are an important skill in the process of performing cranial US. The performers have to correctly detect and interpret the results or diagnosis of the US images. In addition, it is necessary to convert two-dimension images into three-dimension images. The skills consist of two-step process started with a global impression followed by focal search in US. The performers should start

performing US in all areas of relevant following the standard protocol. During US, they have to concern about the characteristics on real time US images such as the color, shape, and symmetry. Then the performers have to decide whether the images are abnormal based on their knowledge and experience and finally giving the diagnosis. This skill in those who are the novices are different from the experts. The novices need a longer time, always pay their attention in unrelated location and usually require more information prior to make a diagnosis. However, after training, the learners will have more skilled, resulting in less time required, less attention in unrelated area, and give accurate diagnoses with only specific information.

1.3.1.2 Technical consideration¹¹⁻¹³

- 1. Pictures of US machine** using in this study and detail of button function of its main panel (figures 1.1 and 1.2)



Figure 1.1 The ultrasound machine: Canon Aplio A550 (Canon Medical System, Tokyo, Japan)

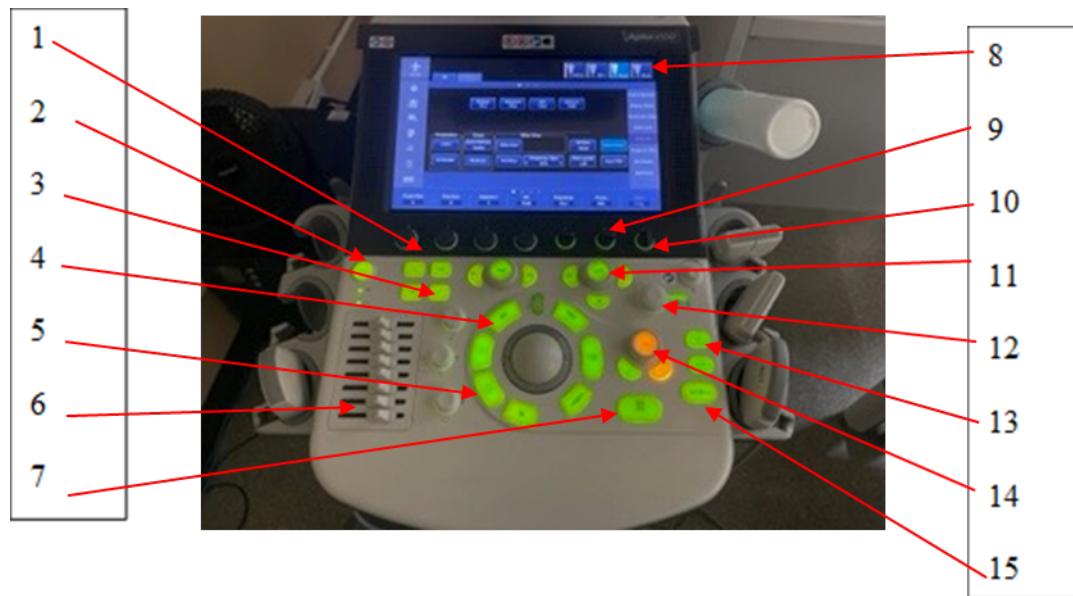


Figure 1.2 Details of each button function on the main panel of ultrasound machine

- 1 = Enable text annotations to be added to images
- 2 = Turns the power ON/OFF
- 3 = Starts basic measurement mode
- 4 = Switches trackball junctions or specify the cursor position
- 5 = Inserts a body mark
- 6 = Specifies the echo reception sensitivity for various depths calculated from the body surface
- 7 = Freezes the screen
- 8 = Chooses the probe on touch panel
- 9 = Sets the frequency
- 10 = Sets the focus
- 11 = Switches to CDI mode; the dial around the knob is used to adjust the gain in color mode.
- 12 = Specifies the depth and the zoom level; press to switch between depth and zoom modes
- 13 = Changes the layout format
- 14 = Switches to 2D mode; the dial around the knob is used to adjust the gain in 2D mode.
- 15 = Saves still images

2. Probe/Transducer

For best information collection, the right probe selection is crucial. To view deeper structures, a convex and low-frequency probe is used. Structures closer to the probe are characterized using the linear, high-frequency probe. Sector/convex/curvilinear probes at 5-8 MHz and linear probes at 3-11 MHz are the best options for the cranial US (figure 1.3). The type and frequency of the probe must be adjusted based on the anterior fontanelle's width and the head's size. The majority of newborn heads are suitable for the small curvilinear probe. Usage of both, a curvilinear and linear probe, is recommended, depending on the required structures to be evaluated.



Figure 1.3 The sector/convex/curvilinear 5-8 MHz and linear 3-11 MHz probes

3. Parameters

3.1 Frequency

Frequency needs to be modified based on the analyzed structure and the child's age or size. For the initial scan, the sector/convex/curvilinear 7.5-10 MHz is sufficient for the entire brain parenchyma. The US images of the cerebrum, brainstem, cerebellum, and foramen magnum should be included. For additional scans, the linear probe 9–18 MHz is utilized for a more thorough assessment of superficial anatomical features.

3.2 Depth

The field of view of the scan should be set up at the beginning to include the maximal depth of the cranium in order to properly visualize all of the cranial contents (figure 1.4).



Figure 1.4 The optimal depth of cranial US from the top of the brain (white dots) to foramen magnum (white line)

3.3 Gain

A proper gain adjustment prevents missing significant discoveries or misdiagnosing normal findings as the disease.

3.4 Time-gain compensation curve (TGC)

Every scan should be started with a TGC adjustment to prevent abnormal parenchymal brightness in some brain areas that could be misconstrued for brain edema. The majority of US machines contain a button to modify it automatically to guarantee a uniform and balanced image from proper TGC curve adjustment.

3.5 Focus

Throughout the scan, the focus should be adjusted to the area of interest, starting from the posterior fossa then moved to other areas based on the scan's results.

4. Cranial US technical consideration and safety includes the items as follows:

- Only perform a head US when necessary
- Reduce scan time to a suitable minimum
- If a Doppler US is done that results in elevated mechanical index (MI) and thermal index (TI), utilize Doppler only when necessary, and avoid leaving the probe stationary for an extended period of time because soundwave energy may heat up the tissue.
- Anterior fontanelle is the main approach because of excellence of physiologic widen.
- The standard planes for cranial US are coronal and sagittal views.
- For the best evaluation, start the scan by attempting to obtain a symmetric, centered, and perpendicular image
- Analyze the images by comparing right to left
- When scanning, try not to exert too much pressure with the probe on the fontanelle
- To avoid missing significant findings, perform the scanning systematically and completely with paying close attention to all essential structures such as the corpus callosum, the fourth ventricle, and the cisterna magna
- When pushing the freeze button with one hand, keep the other hand firmly planted on the baby's head to prevent any unintentional reflexive motion artifacts that could result in fuzzy, unclear images

5. Standard planes and normal appearance of the brain appeared in US images

The main approach for a brain US scan is through the anterior fontanelle. Additional brain US scans are performed through the posterior, sphenoid, and mastoid fontanelle, especially when the pathologies in the cerebellum or posterior fossa are suspected. Standard places include coronal and sagittal views through the anterior fontanelle (figure 1.5 a and 1.5 b).

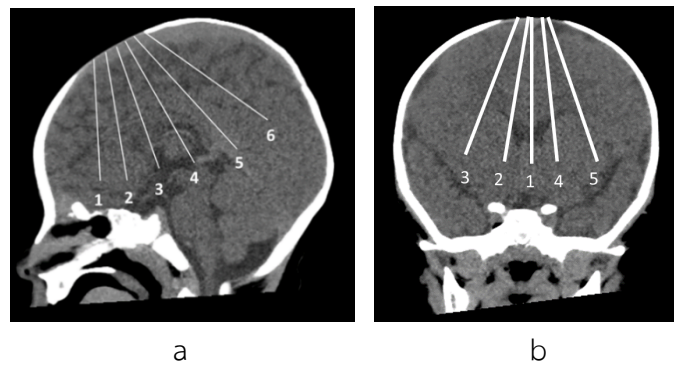


Figure 1.5 a: Six standard coronal planes on cranial US through anterior fontanelle in mid sagittal view; b: five standard sagittal planes on cranial US through anterior fontanelle in coronal view

5.1 Coronal views

When scanning through the anterior fontanelle and tilting the probe sequentially from anterior to posterior, six standard sections of the coronal images are obtained and explored (figure 1.5 a). The interhemispheric fissure in the middle should be the center of all symmetrical coronal images.

5.1.1 The first coronal section

The image is the frontmost part. It shows the frontal lobes and the caudal boundary is made up of the superior orbital wall (figure 1.6). The gray and white matter of frontal lobe are symmetrical. There is a clear interhemispheric fissure in the midline. The gray matter is typically hypoechoic and the white matter is typically hyperechoic.

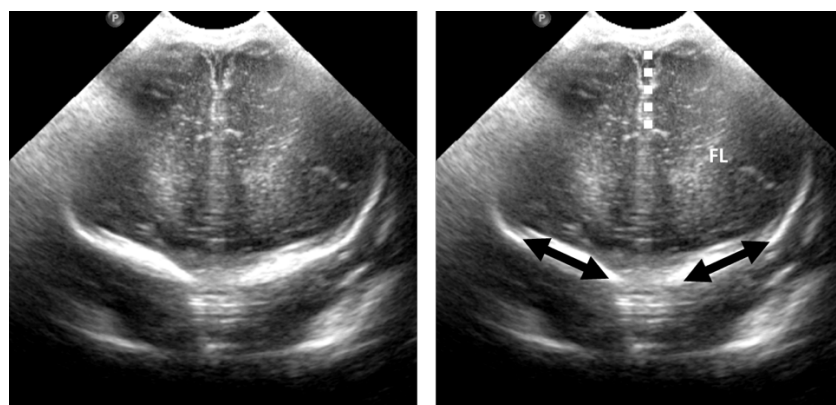


Figure 1.6 First coronal image: interhemispheric falx (white dot line), bilateral frontal lobes (FL), and bony orbital roof (double arrow).

5.1.2 The second coronal section

The image is made up of the middle cerebral artery, frontal lobes, cavum septi pellucidi, and frontal horns of the lateral ventricles. A useful marker is the arterial double contour and pulsation (figure 1.7). In this section, depending on gestational age (GA), normal findings include symmetry or slight asymmetry in ventricular size. The frontal horns of lateral ventricles can be collapsed or restricted ventricular cavities in the first 24 hours after birth, and slightly increased size of ventricular cavities beyond the first day of life.

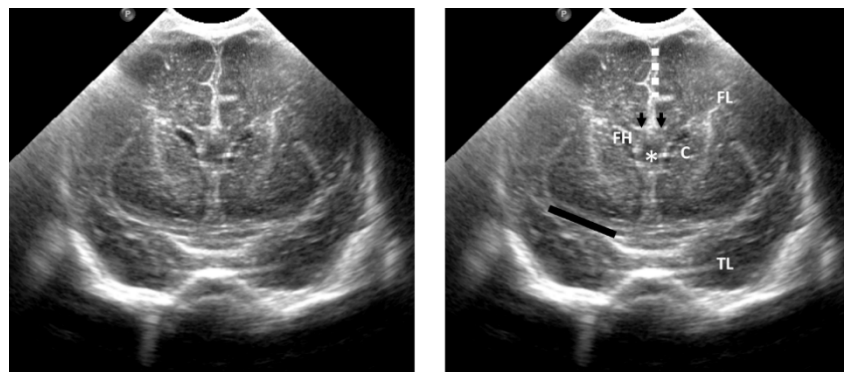


Figure 1.7 Second coronal image: a frontal horn of lateral ventricle (FH), cavum septum pellucidum (white star), caudate nucleus (C), middle cerebral artery (black line), corpus callosum (small black arrows), interhemispheric falx (white dot line), and temporal lobe (TL)

5.1.3 The third coronal section

This region, the level of the third ventricle, shows the sylvian fissure, deep gray matter (basal ganglia and thalami), bodies of the lateral ventricles, and third ventricles (figure 1.8). Preterm babies may have wider Sylvian fissures. Although almost symmetrical, the disease is not excluded by symmetry. The basal ganglia, particularly the thalami, should always be carefully examined since they can show edema, ischemia or hemorrhage when they become bilaterally hyperechoic. In order to rule

out germinal matrix (GM) hemorrhage in preterm infants, special attention should be focused on the caudo-thalamic groove in this view.

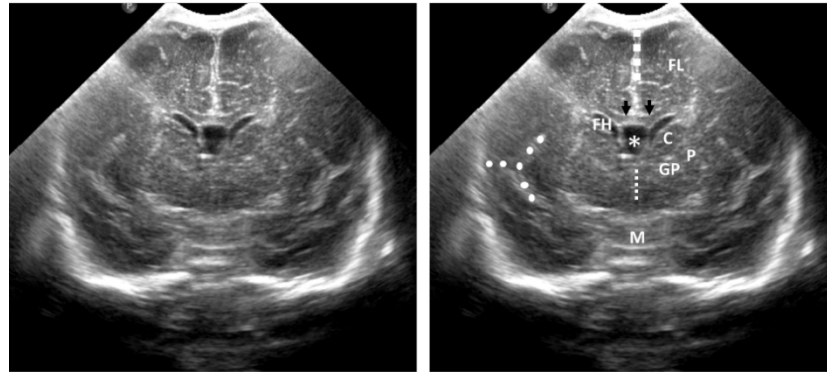


Figure 1.8 Third coronal image: a third ventricle (small white dot line), frontal horn of lateral ventricle (FH), cavum septum pellucidum (C), corpus callosum (small black arrows), caudate nucleus (C), globus pallidus (GP), putamen (P), midbrain (M), Sylvian fissure (white dot), frontal lobe (FL), and temporal lobe (TL)

5.1.4 The fourth coronal section

This section is the level of cerebellum. It contains a portion of the brain stem, vermis, cerebellar hemispheres, tentorium cerebelli, body of the lateral ventricles, brain parenchyma, and cisterna magna. It has a typical appearance of a Christmas tree or a lady dancing (figure 1.9). Symmetry or mild asymmetry of ventricular cavities size, choroid plexuses, and brain parenchymal echogenicity are all considered normal findings in this section. Scan through the posterior or mastoid fontanelles to rule out a hemorrhage in this level is necessary since the cerebellum displays a hyperechoic appearance due to tentorial cerebelli.

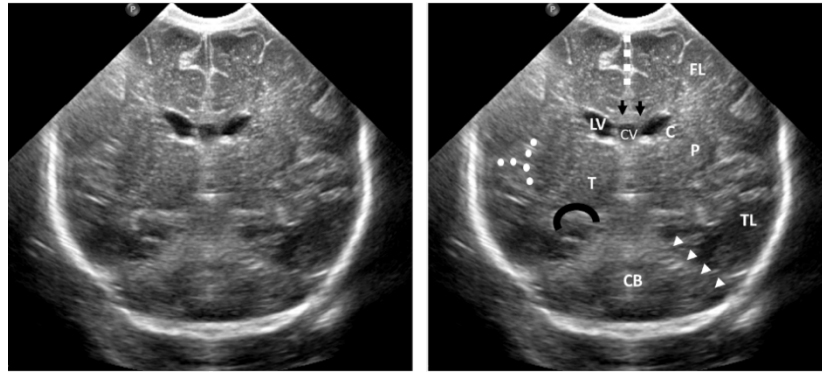


Figure 1.9 Fourth coronal image: a cerebellum (CB), tentorial cerebelli (white arrow heads), hippocampal fissure (black curve line), lateral ventricle (LV), cavum verge (CV), corpus callosum (white arrows), caudate nucleus (C), putamen (P), thalamus (T), Sylvian fissure (white dot), frontal lobe (FL), temporal lobe (TL), and interhemispheric falx (white dot line)

5.1.5 The fifth coronal section

The choroid plexuses, predominating in this picture, run through the ventricular trigone in this area. Laterally, the ventricular cavities split apart. The cortex, subcortical, and periventricular white matter of the parietal and temporal lobes are included. The cavum vergae between the lateral ventricles can be seen (figure 1.10). The symmetry of the periventricular white matter (peritrigonal blush) and the echogenicity of the choroid plexuses are considered normal findings in this region. Despite looking symmetrical, the periventricular white matter ought to be less echogenic than the choroid plexuses. If not, periventricular leukomalacia or another type of white matter pathology must be considered. If the lateral ventricles appear parallel, corpus callosum dysgenesis should be suspected.

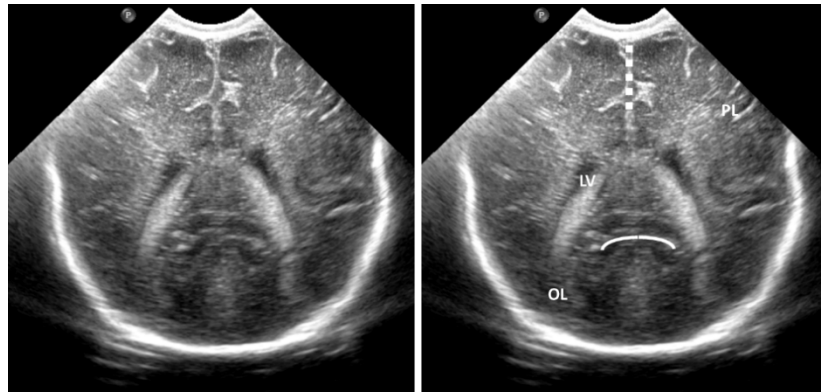


Figure 1.10 Fifth coronal image: a lateral ventricle (LV), interhemispheric falx (white dot line), parietal lobe (PL), occipital lobe (OL), and parieto-occipital sulcus (white curve line)

5.1.6 The sixth coronal section

The periventricular white matter and cortical gray matter are seen in this segment through the occipital lobes, together with the interhemispheric fissure in the middle (figure 1.11). Symmetry of the sulcation and periventricular white matter are typical findings in this area.

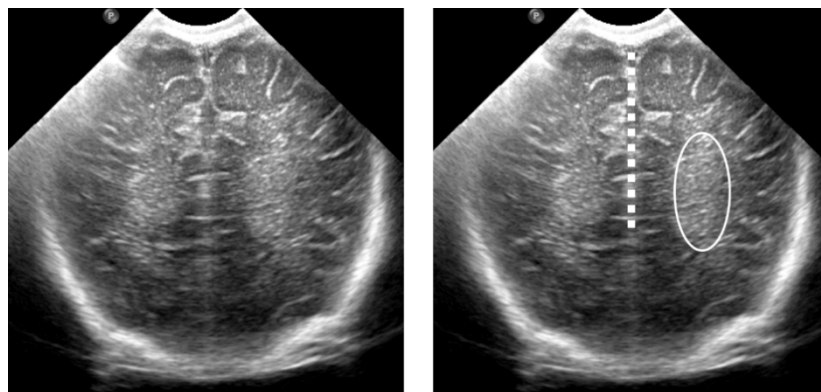


Figure 1.11 Sixth coronal image: a parieto-occipital white matter (oval shape) and interhemispheric falx (white dot line)

5.2 Sagittal views

A minimum of five standard sagittal images (figure 1.5 b) should be obtained by tilting the transducer from the midline to lateral, firstly to the right and secondly to the left.

5.2.1 The midline sagittal section

This image conveys a great deal of information and varieties of anatomical features. All areas from the top of the brain to foramen magnum must be included. The entire corpus callosum, third ventricle, brainstem, fourth ventricle, and cerebellar vermis are examples of normal anatomy (figure 1.12).

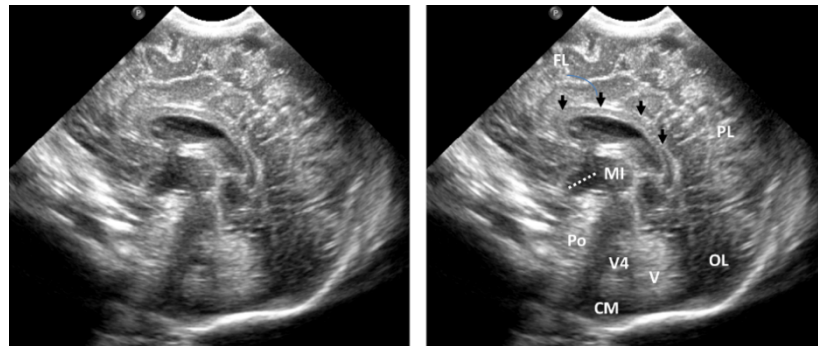


Figure 1.12 First sagittal image: a corpus callosum (small black arrows), third ventricle (small white dot line), fourth ventricle (V4), massa intermedia (MI), pons (Po), vermis (V), cisterna magna (CM), frontal lobe (FL), parietal lobe (PL), and occipital lobe (OL)

5.2.2 The parasagittal right and left sections

The lateral ventricle may be seen in these images together with the thalami, basal ganglia, and periventricular white matter (figure 1.13). The caudate nucleus is located anteriorly in this picture, while the thalamus is located posteriorly, separated by the caudothalamic groove. The lateral ventricles' choroid plexus frequently protrudes anteriorly and situates close into the caudothalamic groove. It should be assumed that the hyperechoic focus bulges in the caudothalamic groove represent GM bleeding.

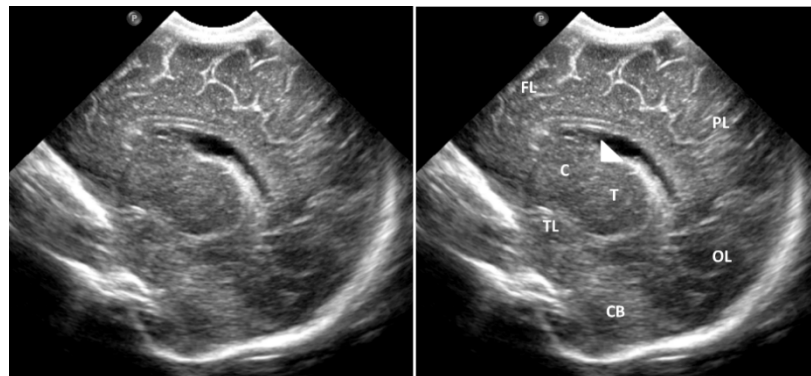


Figure 1.13 Second and fourth sagittal images: a caudate head (C), thalamus (T), caudothalamic groove (white triangle), frontal lobe (FL), parietal lobe, (PL), occipital lobe (OL), temporal lobe (TL), and cerebellum (CB)

5.2.3 The tangential parasagittal right and left sections

These sections show the periventricular white matter, sulci, and Sylvian fissure, laterally and superficially to the lateral ventricles (figure 1.14). The homogeneity of white matter, normal hypoechoic gray matter, appropriate sulcation in accordance with GA, and Sylvian fissures are the hallmarks of normal findings.

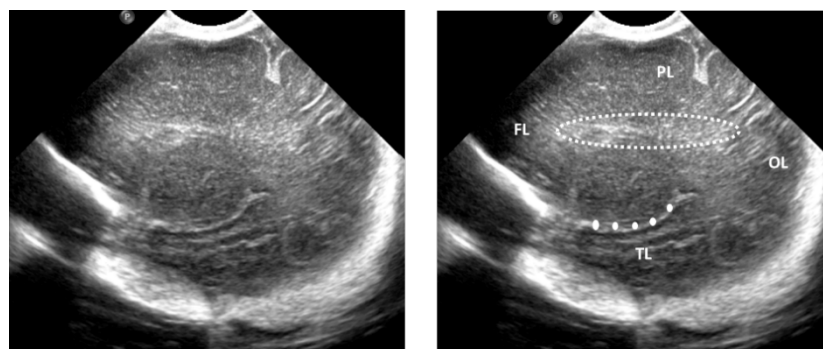


Figure 1.14 Third and fifth sagittal images: a Sylvian fissure (white dots), periventricular white matter (dot oval shape), frontal lobe (FL), parietal lobe (PL), occipital lobe (OL), and temporal lobe (TL)

Figure 1.15 and figure 1.16 summarize the standard coronal and sagittal views for cranial US, respectively

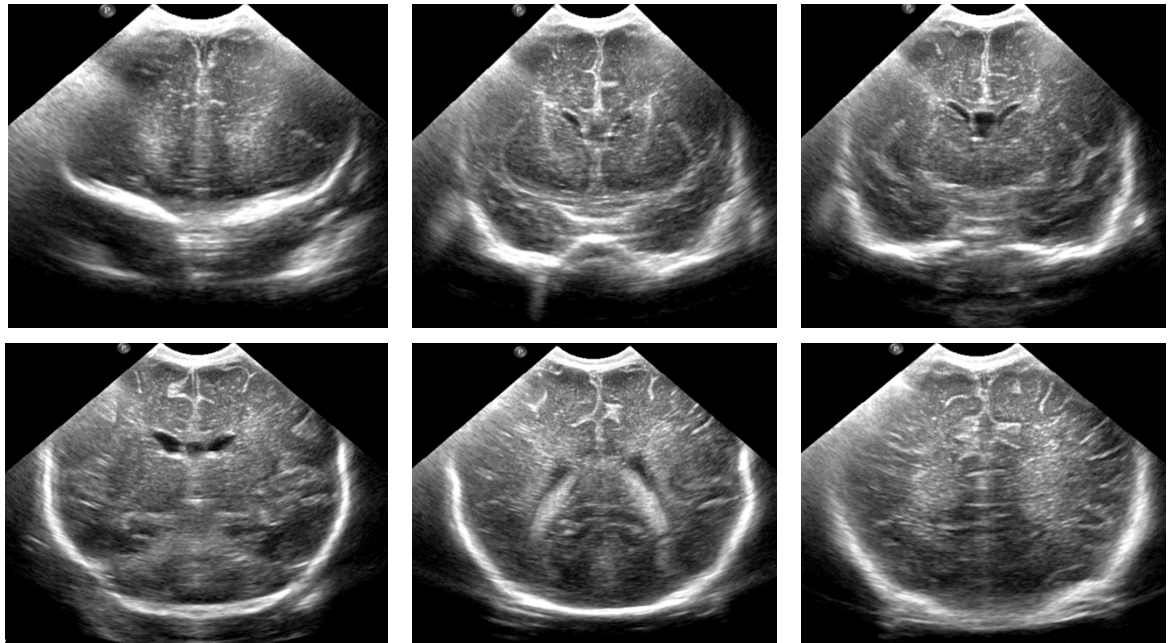


Figure 1.15 Summary of standard coronal views for cranial US

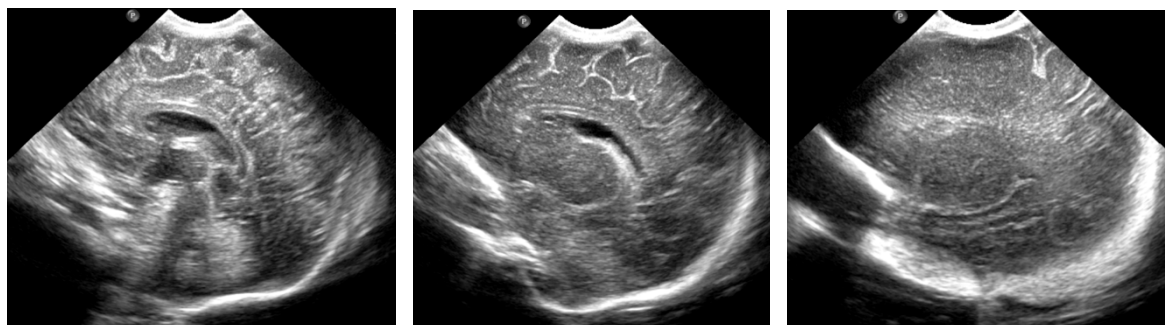


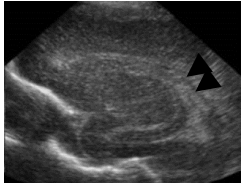
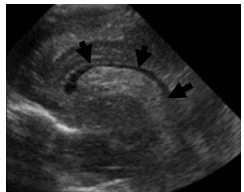
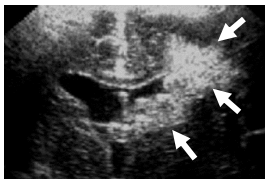
Figure 1.16 Summary of standard sagittal views for cranial US

1.3.2 Cranial US pathologic findings in preterm babies^{1,11}

US is an important tool in the diagnosis of IVH or subependymal hemorrhage, PVL or hypoxic ischemic encephalopathy of preterm, and hydrocephalus. These abnormalities are commonly found in preterm babies during second to third trimester

because their brain is vulnerable to injury from ischemia or hemorrhage. In preterm babies, IVH occurs after sudden hemodynamic change. The preterm baby's brain still has highly vascular germinal matrix located along the ventricular surface between the caudate nucleus and thalamus, especially at the caudothalamic groove. These vessels are fragile and prone to develop bleeding. Normally, these vessels involute when the babies are near term age, approximately 34-36 weeks GA. IVH is often found during the first 7 days after birth. The hemorrhage occurs at the caudothalamic groove and may extend into the ventricle or adjacent parenchyma. In US images, hemorrhage appears hyperechoic lesion as white color. IVH is divided into 4 grades according to Papile LA, et al in 1978⁸ (table 1.2). Grade 1 and grade 2 IVH show good prognosis, but grade 3 and grade 4 IVH may reveal long-term neurologic abnormalities.¹

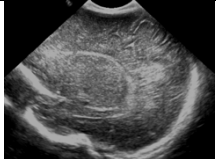
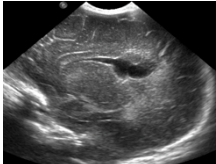
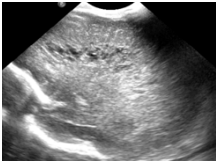
Table 1.2 Grading of intraventricular hemorrhage¹⁴

Papile grading	Description	US image
Grade 1	Subependymal (germinal matrix) hemorrhage	
Grade 2	IVH (hemorrhage fills less than 50% of lateral ventricle) without ventricular dilatation	
Grade 3	IVH with lateral ventricular dilatation	
Grade 4	IVH with intraparenchymal hemorrhage	

IVH, Intraventricular hemorrhage

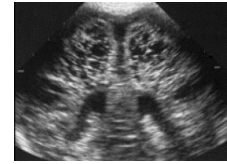
PVL causes by lacking of blood or oxygen at the periventricular white matter region within the brain tissue of preterm baby, especially before 33 weeks GA., unlike the ischemic brain in term infant that happens at the cortical- subcortical part. Heterogeneity of affected areas from ischemia between preterm and term brains are due to different locations of watershed areas in their brains. Watershed area of the brain is the area at risk to develop brain ischemia. PVL occurs at the periventricular region because it is a watershed area or junction of intracranial vessels including anterior, middle, and posterior cerebral arteries in preterm babies.⁵ In normal brain, the finding of periventricular white matter on US images has less echogenicity than the choroid plexus.¹⁵ When the PVL occurs, the affected brain tissue has an echogenicity equal to or greater than the choroid plexus. Grading of PVL, according to de Vries LS, et al in 1992¹⁶, is separated into 4 grades (table 1.3). Proper diagnosis of PVL is essential because cerebral palsy, intellectual impairment, and visual disturbance are more common in these preterm babies. Moreover, over 50% of children with PVL or grade 3 IVH will develop cerebral palsy.

Table 1.3 Grading of periventricular leukomalacia¹⁶

Vries grade	Description: PVL (increased periventricular echogenicity)	US image
Grade 1	Persisting more than 7 days	
Grade 2	Developing into small periventricular cysts (Swiss-cheese appearance)	
Grade 3	Developing into extensive periventricular cysts, occipital and fronto-parietal white matter	

Grade 4

In deep white matter developing into
extensive subcortical cysts



PVL, Periventricular leukomalacia

Although, the US has high accuracy in the diagnosis of cystic PVL, there are some limitations in the diagnosis of non-cystic or grade 1 PVL. In this case, it is necessary to follow-up cranial US in the next 7 days. Although the findings on US images of grade 4 IVH and PVL are shown an increased echogenicity, only grade 4 IVH has mass effect to adjacent structures.

1.3.3 Assessment tool in medical education

Several components of medical education work together in order to have the desired effects on students' learning. These comprise of instructional strategies, learning goals, and evaluations. Initially, the measurement is the main focus of evaluation in medical education. Nowadays, competencies need to be incorporated into the various parts of the current medical educational system. A well-designed assessment encourages student motivation, the growth of self-evaluation, learning retention, and learning transfer. Therefore, the selection of assessment methods is crucial.^{17,18}

1.3.3.1 Assessment types

There are three main types of assessment.¹⁹

1. Diagnostic assessment

The diagnostic examination is administered at the start of the unit or the course, as appropriate. Data on what the students already know about the subject is gathered through this exam. A learner's current understanding of or viewpoint on a subject or issue that will be covered in the course is determined by a set of written questions (multiple choices or short answers). The purpose of a diagnostic assessment is to provide an overview of where each student stands academically,

emotionally, or ideologically so that the teacher may make informed decisions about how to teach the new course material and the best teaching strategy.

2. Formative assessment

Prior to summative evaluation, formative assessment allows for the identification of gaps, misunderstandings, and growing understanding. Informal questions, practice tests, one-minute papers, and clearest/muddiest point activities are a few examples of formative assessment tools. Without the pressure of grades, formative assessment enables students to test their knowledge or put their talents to the test.

3. Summative assessment

Summative evaluation is used to evaluate learning at its conclusion. It is used to compare student learning to a baseline or standard. Summative evaluation's primary goal is to assign grades to students.

The summary of three types of assessments is shown in table 1.4.

Table 1.4 The main differences between three types of assessment¹⁹

	Diagnostic assessment	Formative assessment	Summative assessment
Feedback	Informal	Dialogue-based/comments, ungraded	Grades
Timing	Before the lesson	Throughout the learning process (small check-ins)	Closing of a unit or course
Rationale	Analyze prior experience, abilities, assumptions, etc.	Observe progress and give remarks	Compare the student's performance to a benchmark standard

Examples	Background knowledge survey and ungraded test	A variety of group projects in classroom, in-class writing assignments, practice exams, clearest/muddiest point exercises, and minute papers	Midterm and final multiple-choice tests, a final project, and a term paper
-----------------	---	--	--

1.3.3.2 Model of developing the assessment tool

In general, the characters that are usually evaluated in the learner's competency include knowledge, skills, and attitudes, in other words, cognitive, psychomotor, and affective domains in Bloom's Taxonomy in 1950-1970s.²⁰ For the development of cranial US skills, the assessment tool should be developed to assess the two aspects; knowledge or cognitive and skills or psychomotor.

Miller's pyramid is an example showing how learners have been evaluated (figure 1.17).²¹ Actions are defined in each of the pyramid's four tiers. These are "knows" for fundamental information, "knows how" for knowledge that has been applied, analyzed, or interpreted, "shows how" for actual application and practical demonstration in a simulated setting, and "does" for performance in a real-time setting.

The purpose of developing the cranial US assessment tool is to assess the top-level learners in the pyramid, "does". The assessment of actual performance in real situation, so-called workplace-based assessment (WPBA), is essential to quality management.²²

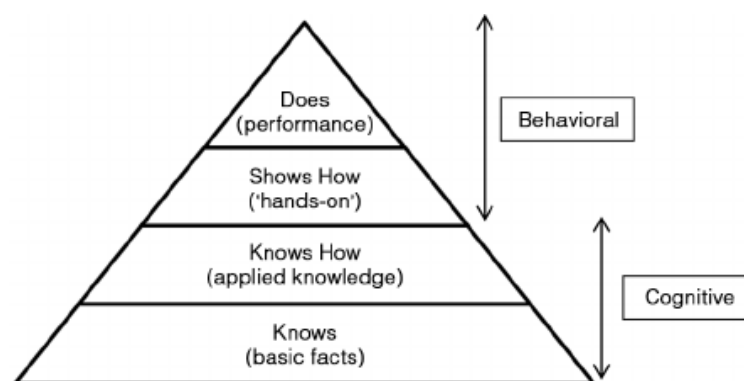


Figure 1.17 The Miller's pyramid²¹

The proper evaluation methods of each level according to Miller's pyramid is different. For example, assessment of knowledge or how to apply knowledge should be performed by using multiple choice questions (MCQs), extended matching item or short answer questions (SAQs). To evaluate the next level, "show how", multi-station objective structured clinical examination (OSCE) is appropriate. Finally, to evaluate the learner's ability at the top of the pyramid, "performance level or does", the optimal methods are mini-clinical evaluation exercise (mini-CEX), direct observation of procedural skills (DOPS), etc.^{23,24}

In general, a learner's abilities can be assessed using a variety of assessment methods. Each assessment method has advantages as well as limitations. The assessment methods are divided into three main sections including written assessments, assessments by supervising clinicians, and WPBA.^{17,18,24,25}

1. Written assessment^{17,18,24,25}

1.1 Multiple choice questions (MCQs) are flexible assessment formats that can be used to measure knowledge, abilities, values, and thinking skills. MCQs come in a variety of types, including the most popular single or one best option type question, true or false type question, multiple true or false questions, matching type question, the extended matching type question, and key feature question. A true or false question with multiple options is referred to as a multiple true or false question. Matching items type consists of stems and answers arranged in columns under matching instructions. The list of a theme, alternative lists, lead-in question, and case summary make up the extended matching questions. The key feature question focuses on critical decision on scenario-based test with limited number of questions. MCQs can offer several examination items that cover a wide range of content areas with relatively short time period causing great reliability.

1.2 Short answer questions (SAQs) resemble multiple choice questions without alternatives. The learner is supposed to give the response to an open-ended semi-structured question, not to choose the response. Due to their lengthy nature and sparser sampling, short answer questions have low dependability.

1.3 Essay questions have low reliability due to time-consuming and small sample.

2. Assessment by supervising clinicians^{17,18,24,25}

The format of a long case examination varies. In accordance with tradition, a student is given a block of uninterrupted and unobserved time, typically 30–45 minutes, to interview with and examine a non-standardized real patient. The student then discusses his or her findings and action plan with examiners who then conduct an unstructured oral examination with the student about the patient's illness and related issues. It has poor content validity, low reliability, and less consistency. Although, the reliability is poor, it can be raised by giving each student more cases. Moreover, observing the candidate while he or she is taking the history and performing the physical examination would be a more accurate way to judge the candidate's proficiency.

2.1 Objective structured clinical examination (OSCE) is the tool to examine psychomotor domain such as skills and clinical management. It uses standardized or trained patients and a sufficient sample size. It consists of multiple stations which focuses on each different task. Students go around a set of timed stations in rotation. The same schedule is followed by all candidates as they move sequentially between stations. OSCEs have used standardized patients as the primary assessment tool. Technical abilities have, however, also been evaluated using alternative assessment items, such as data interpretation exercises utilizing clinical cases and clinical scenarios with mannequins. A standardized checklist of specific behaviors or a global rating form are used by the observing faculty member, or patient to assess the student's performance. It enhances interrater reliability. The wide-ranging content and skill sample contributes to the high reliability. To ensure consistency, judges must undergo extensive training.

2.2 Multisource or 360-degree assessments are made up of questionnaires that numerous persons inside a person's sphere of influence fill out. Peers, members of the clinical team, and patients who may offer perceptions into the trainee's work habits, aptitude for teamwork, and interpersonal sensitivity serve as evaluators and fill out evaluation forms. The evaluations often include routine behaviors that are evaluated by reliable sources, and assessments

frequently correspond with future clinical and academic success. The results are rather reliable and consistency.

2.3 A portfolio is a structure that contains proof of the accomplishment of learning objectives across time. It is used to support learning and for formative and summative assessments. It usually consists of textual papers, but it can also have multimedia, video or audio recordings, photographs, and other types of images. It can also be kept in an electronic format. It might contain a record of the case or procedure logs, literature reviews, peer assessments, etc. Self-reflection or assessment is also included in portfolios. The best use of portfolio assessment is to measure the mastery of competencies that are challenging to test in other ways, such as practice-based improvement, the use of scientific evidence in patient care, professional behavior, and patient advocacy. Portfolio evaluation is closely tied to self-directed learning. The use of portfolios in formative evaluation and professional development is highly recommended. In this context, the portfolio for formative evaluation called developmental portfolio and the portfolio for summative evaluation called showcase portfolio. They are, however, fundamentally non-standardized.

2.4 Oral examination or case-based discussion is the process by which the examinee's patients are evaluated via a standardized oral examination. A skilled doctor interrogates the subject regarding the care given, probing for the rationale behind the diagnosis, the interpretation of the clinical findings, and the treatment recommendations. An established scoring formula and procedures are used by the examiner to rate the examinee. It results in poor content validity, low interrater reliability, and inconsistency in marking or bias.

2.5 Checklists include necessary or desired particular actions, behaviors, or steps that make up a more complicated competency. The presence of a checkmark or the word "yes" denotes the occurrence of the behavior or the accuracy or completeness of an option. The checklist's predefined order of steps, which is rarely the case in practice, is a drawback. The creation of a checklist necessitates agreement among a number of specialists regarding key behavior,

timing, and performance evaluation standards in order to assure the validity of the content and scoring rules.

2.6 Procedure, operative, or case logs or logbooks provide detailed information about each patient the student encountered, including the illnesses they were treated for and the procedures or surgeries they underwent. It keeps track of the number of continuous cases throughout time. It is helpful in figuring out the range of medical experience.

3. Workplace-based assessment (WPBA) ^{17,18,24,25}

3.1 Mini-clinical evaluation exercise (mini-CEX) is used when a supervising physician observes a learner-patient interaction in any healthcare setting. It requires short period of 10 to 20 minutes. During this time, the learner is expected to focus on getting a thorough history and physical examination. The learner then presents a diagnostic and treatment plan to the assessor. Following the use of a standardized form to grade the performance, educational feedback is given. It is an effective technique for formative evaluation. It also has good reliability comparable to structured examinations using standardized patients.

3.2 Direct observation of procedural skills (DOPS) is a technique used to assess trainees as they do a procedure. Informed consent, preoperative planning, analgesia, technical competence, aseptic technique, postoperative care, counseling, and communication competences are all assessed using this tool.

It is crucial to pick the appropriate tool for evaluating trainees. It is critical to understand that not all tools are appropriate in all situations. No single approach can evaluate every sort of knowledge and skills; all assessment techniques have limitations. Validity and reliability must be shown in an effective assessment. All stakeholders, including students, faculty, and licensing organizations, should be able to implement the effective assessment and must find it acceptable. Examiners must be properly trained in order to use some of the more recent assessment techniques. Ineffective assessments can encourage unhealthful learning methods.

The optimal assessment method of US skills performance could be done by using DOPS. DOPS are divided into two groups; DOPS with or without criterion. DOPS without criterion is easy to use, requires minimal training, and may be comprehensive, however, it has low reliability. DOPS with criterion has an excellent reliability, easy to use and requires minimal training. Moreover, DOPS with criterion can be used as formative assessment, summative assessment, and progress test. However, its coverage depends on the items being assessed.

Several assessment methods to evaluate learner's ability include their proposes, strengths, and limitations are described in table 1.5.^{17,24-27}

Table 1.5 Assessment methods with their domain, purpose of assessment, strengths, and limitations.^{17,24-27}

Methods	Domain	Purpose	Strengths	Limitations
Multiple choice questions (MCQs)	Knowledge, applied knowledge	Summative	Good reliability due to many content areas, short time	Difficult to write, cueing
Modified essay questions (MEQs)	Knowledge, applied knowledge	Summative	-	Limiting reliability due to smaller samples, time-consuming
Global rating	Shows how, performance	Summative and formative	Useful for assessing personal attitudes	Subjective, vary reliability
Portfolio	Knowledge	Summative and formative	Linked to self-directed learning	Non-standardized, time-consuming
Oral examination or VIVA	Knowledge, applied knowledge	Summative and formative	Feedback provided by experts	Subjective, bias, time-consuming

Methods	Domain	Purpose	Strengths	Limitations
Long essay questions	Knowledge	Summative	-	Low reliability
Short-answer questions	Applied knowledge	Summative and formative	Assess interpretation and problem-solving ability	Low reliability due to limited samplings, time-consuming, costly
Extended matching item	Knowledge, applied knowledge	Summative	-	-
Key feature test	Applied knowledge	Summative	Good reliability and validity when constructed guideline	Time-consuming
Long cases	Performance	Summative and formative	Face validity	Low reliability
OSCE	Shows how	Summative and formative	High reliability, consistency	Time-consuming, expensive
Mini-CEX	Shows how, performance	Summative and formative	Feedback provided by experts	Time-consuming
DOPS	Performance	Summative and formative	Excellent reliability, construct validity	Require trained evaluators
Checklists	Performance	Summative, formative	Good validity, reliability	Need trained evaluators
360-degree evaluation	Knowledge, applied	Formative	Assess range of competencies	Time-consuming, administratively demanding

Methods	Domain	Purpose	Strengths	Limitations
	knowledge, performance			
OSCE, Objective structured clinical examination; Mini-CEX, Mini-clinical evaluation exercise; DOPS, Direct observation of procedural skills				

1.3.4 Assessment tool development

Two main approaches to develop a new knowledge, guideline or tool include evidence-based methods and consensus development methods. Evidence-based methods are appropriate for developing new researches with sufficient evidences while consensus development methods are proper if there are incomplete or limited evidences.²⁸

For medical education, consensus development methods are often used in varieties of purposes, for examples, to define new or revised course curriculum, to develop lists in the assessment tool, to set competencies of students or learners and to develop educational resources, etc.²⁹

Consensus development methods consist of two popular methods; Delphi process and nominal group technique.^{29,30} Delphi technique includes a large number of participants, using the email questionnaire method, low cost, and avoided abusive dominance. However, the limitation includes no extensive discussion or debate, and time consuming. On the other hand, nominal group technique includes a small number of participants with spacious interactions and extensive discussion, quite expensive cost, potential for outstanding participants and concluded immediately in a short period of time.²⁹⁻³¹

The previous literatures which develop US skill assessment tools are described in table 1.6. The Delphi method and expert consensus method are commonly used when the purpose is to develop an assessment tool for US skills.

Table 1.6 Literatures of ultrasound skill assessment tool development³²⁻³⁹

Literatures	Methods
Endobronchial ultrasound skills and tasks assessment tool ³²	Expert consensus
Development of an assessment tool to measure students' perceptions of respiratory care education programs: Item generation, item reduction, and preliminary validation ³³	Expert consensus
International multispecialty consensus on how to evaluate ultrasound competence: A Delphi consensus survey ³⁴	Delphi method
Development and validation of an assessment tool for competency in critical care ultrasound ³⁵	Delphi method
A modified Delphi Process to define competencies for assessment leads supporting a doctor of pharmacy program ³⁶	Delphi method
The assessment of competency in thoracic sonography (ACTS) scale: validation of a tool for point-of-care ultrasound ³⁷	Expert consensus
Tools for measuring competency in clinical ultrasound: recommendations from the ultrasound competency word group ³⁸	Expert consensus
Development of a focused cardiac ultrasound image acquisition assessment tool ³⁹	Delphi method

1.3.4.1 Delphi method

Delphi method or Delphi technique is the method to solve the particular problem according to group of expert's consensus. The word "Delphi" is the name of the Greek town. The temple of Apollo locates in this town where once has been the home of the oracle. Originally, it is developed by the RAND Corporation in the United States in 1950s.⁴⁰ The Delphi project, likes the national defense study, has been used to predict the risk of atomic attack on United States. Later, the Delphi method was used in many areas such as business, education, or technology. Nowadays, the Delphi method is the procedure to discover the answer from expert's consensus in order to solve the specific problems.

The definition of Delphi according to Turoff M and Linstone HA in 1975⁴¹ is quoted that “Delphi may be characterized as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem”

Four characteristics of the Delphi method described by Rowe G, et al in 1991⁴² are shown as follows:

1. **Anonymity of participants:** This process is crucial because it prevents group guidance by someone due to its authority or reputation. Therefore, participants can comment and make decisions, independently.
2. **Iteration or repetition:** The Delphi process provides feedback to participants in the form of group comments, repeatedly.
3. **Controlled feedback:** The participants are allowed to explain or modify their comments until they reach a conclusion.
4. **Statistical group response:** Statistics are used to summarize group of comments.

The typical three rounds of Delphi process are shown in figure 1.18.⁴³ The Delphi process is the process with standard sequences as follows.

- *Create the research question:* The research question may originate from experience, literature review, or previous pilot studies.
- *Research design:* Design the research should be considered the feasibility of research.
- *Research sample:* This process includes selection of experts and size. In Delphi method, there is no definite role or consensus about expert's selection and its number. Four factors to be determined in expert's selection are knowledge and experience, voluntary participation, adequate time, and effective communication.

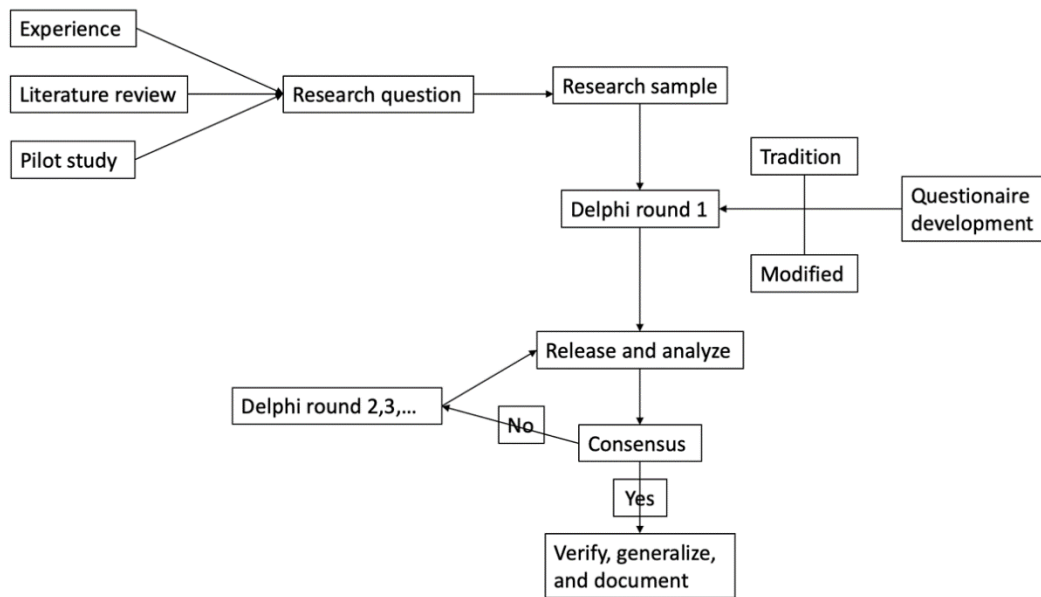


Figure 1.18 The typical three rounds Delphi process⁴³

Delphi round 1 questionnaires: After examining the literatures, the first round of the Delphi process can either involve brainstorming as part of the construction of a questionnaires (traditional Delphi process) or involve the revision of the initial, broad questionnaires based on the opinions of the experts (modified Delphi process).

- *Delphi pilot study:* The process is to test and adjust the preliminary questionnaires to be comprehensive and up-to-date.
- *Release and analyze the round 1 questionnaires:* The questionnaires will be sent to the participants. The participants are asked to complete or rate the questionnaires and return it to the researcher. Then, the researcher will analyze the obtained answer and show the distribution of the answer including any suggestions or comments.

Delphi round 2 questionnaires: In this round, the number of items in the questionnaires could be reduced or increased as well as improved according to the participant's suggestions.

- *Release and analyze the round 2 questionnaires:* The questionnaires will be sent back to the participants. It contains information on the distribution of

answers and suggestions or comments from round 1. The participants are again allowed to edit or change their answers.

Delphi round 3 questionnaires: This is a duplicated process which usually enhance specificity of questionnaires to the research in each round.

- *Release and analyze the round 3 questionnaires:* This repeated process will be stopped when the research questions have answers or conclusions.
- Verify, generalize, and document the research results.

1.3.4.2 Clarifying the assessment criteria^{19,44}

After items or characters of learners performing cranial US are completed using Delphi process. The next step is to build the level of mastery; the description of characteristics at each level of mastery. A clear description of characteristics or criteria is essential. Depending on how students interpret the expectations, the clarity of the assignment criteria influences both how easily and readily they accept the task and its quality. All as described above are a form of rubric.

The term "rubric" refers to a helpful assessment tool that explain all of the particular criteria that students' work should meet and specifies the degree of quality required to meet each criterion. These descriptions focus on the quality of performance, not the quantity of the procedure. Therefore, the rubric is a proper tool to evaluate skill performance as in this study, focusing on assessment of cranial US performance.

The steps to develop the rubric as a tool of assessment in this study are shown as follows.^{45,46}

1. **Learning objective:** The skills in each assignment must be assessed. Then, the task description or learning objective is the cranial US performance assessment.
2. **Performance criteria:** The observable criteria must represent those skills. In this study the criteria or items to be evaluated in cranial US assessment form should be developed using Delphi method.
3. **Performance levels:** Use the descriptors for each attribute separately to indicate the best and lowest performance levels for analytical rubrics.

4. **Performance level descriptions:** When using analytical rubrics, finish the rubric by providing additional points along the continuum from outstanding to poor for each attribute. For each intermediate level of performance, provide a description.

A rubric is an evaluation tool that delineates making criteria in great detail. It can be used to assess a variety of student work including written assignments, presentations, class participation, total grades, and skill performance evaluation.

A rubric offers a list of standards that highlights the crucial elements of the activity being planned for or assessed. The requirements and objectives for the assignment are made clearer via rubrics. When properly applied, rubrics can give students useful and timely feedback, promote critical thinking and self-evaluation, explain objectives to students, highlight the abilities required for any activity, and promote fair and uniform grading. A rubric is organized like a matrix and has four parts (figure 1.19):¹⁹

1. A task description: The performance being evaluated
2. Rows: The characteristics that will be evaluated
3. Columns: Levels of mastery or scales
4. Cells: A description of each characteristic at each level of competence or scale

The most crucial assessment criteria that will be used to assess students' work should be selected prior to start establishing the rubrics. Different assessment criteria for each assignment, discipline, and course level should be included in the rubrics. Descriptions of how each assessment criteria should be met by students will be designed when the most crucial assessment criteria are selected. As in this study, these descriptions are known as performance levels, level of mastery, or assessment standards. They can be portrayed using letter grades, a point total, or descriptions of various quality levels. The examples include "Outstanding-Competent-Require re-thinking", "1-10", "Beginning-Developing-Accomplished-Exemplary", or "Distinguished-Proficient-Emergent-Insufficient".¹⁹

The crucial question is how many mastery or performance levels ought to be on the assessment form. The vast majority of performance evaluation forms offer three, four, or five performance levels. It is uncommon for an evaluation system to merely have two levels of ratings and function on a pass/fail basis. Additionally, it is highly rare for an appraisal tool to require distinctions between levels that are any finer than five. The advantages and disadvantages of three, four, and five levels of performance are shown in table 1.7.⁴⁷

To overcome central tendency bias in performance evaluation, the optimal range of performance level or level of mastery is “four”. This will not allow them to be neutral. Moreover, in performance assessment tool, we need to force reviewers or raters to take a side such as pass/fail.

Task description

Cranial US performance assessment form with rubric

	4 (Distinguished)	3 (Proficient)	2 (Emergent)	1 (Insufficient)
1. Preparation phase (2 items)				
Identified patient	Done and confirm	Done	Wrong patient	Do not done
Proper positioning of the US machine	Right hand side, close to the patient	Right hand side	Other sides	Left hand side

Level of mastery/scale

	4 (Distinguished)	3 (Proficient)	2 (Emergent)	1 (Insufficient)
2. US machine operation (7 items)				
Selecting appropriated probe	Small curvilinear 8.5 MHz probe	Linear high frequency probe	Large curvilinear low frequency probe	Hockey-stick probe
Corrected probe handling	Complete	Inadequate	Improper	Incorrect
Applying of warm gel onto the probe	Put the optimal amount of gel onto the probe	Put gel onto the probe	Put gel on patient's head	Do not use gel

Description of each characteristic at each level of mastery

Characteristic

Figure 1.19 Example of rubric structure¹⁹

Although “five levels of performance” is the most useful popularity in appraisal form, however, central tendency bias is the main disadvantage. The central tendency bias is the propensity for respondents to rate the majority of items in the middle of the scale. Similar to how central tendency bias affects performance evaluation, there

is a propensity for reviewers or raters to emphasize average results rather than unique findings. Students who perform at their best might not feel valued if evaluations are made only on averages. This will seriously affect their motivation. On the other hand, even if they don't do a terrific performance, underperforming students might consistently get superior reviews. These individuals will pass up the opportunity for development that an honest review may have offered.

The neutral choice means a choice that has a midpoint, for example, three points scale, five points scale, and seven points scale. The non-neutral choice refers to the scale without a midpoint, such as two choices, four points scale, and six points scale. The use of neutral choice is popular in Likert scales which are commonly used in survey questionnaires. Five points scale is also the most commonly used. For the non-neutral choice is preferred in the practitioner question. In this case, four points scale is favored. The different use of neutral and non-neutral choice is present in the table 1.8. The evidence validity and reliability of scale with neutral choice have been inconsistent. Another issue to consider should be include the number of scales.⁴⁸

Table 1.7 Advantages and disadvantages of three, four, and five levels of performance.⁴⁷

Advantages	Disadvantages
Five levels	
- Allows for the finest levels of performance distinction - Increased conformity to the bell curve distribution - The majority of managers think they can distinguish between the five performance levels - Consistent with the standard "A-B-C-D-F" school grading system - Simplest rating system, requiring less training	- It could be more challenging for managers to explain how to achieve better performance levels - There are often only four levels used - Middle ratings are frequently seen negatively, such as average, mediocre, or "C-student" - Could promote a central tendency

Advantages	Disadvantages
Four levels	
- Excludes the center rating, which might be interpreted as "average" - The "central tendency" rating mistake is eliminated - May cause raters to be biased either positively or negatively - More precise distinctions than a three-level scale can be made	- It might not be possible to tell which employees need to be terminated and which ones need to improve. - May cause raters to be biased either positively or negatively - There are often only three levels used
Three levels	
- It is simple for managers to divide performance into three levels - If supervisors are given fewer options, they are more likely to be reliable and consistent - Some jobs might be evaluated more effectively on a "pass/fail" basis - Only three performance levels can be empirically demonstrated - A middle rating suggests expected performance rather than average performance - More consistent with TQM principles	- Potentially insufficiently fine-grained performance distinctions - Managers often make system changes by including plusses and minuses - Does not differentiate between those who need to be fired and those who can be improved - There are often only two levels used - Does not allow for identifying the truly exceptional 2-5%

TQM, Total quality management

Table 1.8 The use of neutral and non-neutral choice on the Likert scale.⁴⁸

	Include a midpoint	Omit a midpoint
Benefit	* Respondents are not compelled to agree or disagree, therefore, they are free to express their real neutral or indifferent opinion	* Prevents respondents from possibly misusing the midpoint

	Include a midpoint	Omit a midpoint
Problem	* Respondents may use the midpoint as a dumping ground when: - Lack sufficient knowledge of the survey's topic - Have mixed feelings about the subject - Have little interest in the subject - Believe their response depends on other variables - They strive to give responses that are more acceptable in society	* Respondents are not given the chance to express their neutral opinion * Making respondents choose a side could lead to biased results
When	* Due to their familiarity with the subject, respondents ought to be given the opportunity to provide a neutral opinion * For the purpose of statistical analysis, it is crucial to use the scale as an interval scale	* Respondents are: - Unfamiliar or uncomfortable with the survey topic - Unlikely to have developed an opinion on the topic - Facing intense social desirability pressures - Propensity to have satisficing behavior
Strategies to use	* Clarify survey questions since people often choose a midpoint when they're unsure of what the questions mean * Include extra alternatives like "Not applicable", "I'm not sure", or "It varies"	* Instead than imposing a choice on respondents, include "I don't know" or "Not applicable" alternative

Type of Rubrics¹⁹

There are two types of rubrics and of methods for evaluating students' efforts:

1. Holistic rubrics

One- dimensional or holistic rubrics that measure participants' overall performance on a task or item in accordance with set standards of performance. Performance descriptions are typically written in full sentences and paragraphs (table 1.9).¹⁹

Table 1.9 Example of holistic rubrics¹⁹

Score	Criteria
4 (80-100%)	Research paper displays thorough comprehension and accomplishment of the given goals. The thesis statement/argument is clearly expressed, intricate, and unique, and the writing does not devote an excessive amount of time to one point of development at the expense of advancing other topics in the paper's main body. Additionally, writing must be free of errors, clear of ambiguity, and read smoothly, creatively, and purposefully.
3 (70-79%)	Research paper displays thorough comprehension and accomplishment of the designated goals, The thesis statement or argument is presented, veers toward being complicated and original, and the writing exhibits accuracy and balance in building body points, yet it may occasionally display faults and errors in correctness. Although there are some mistakes and ambiguities in writing, it still reads clearly and coherently.
2 (60-69%)	Research paper displays partially understood and carried out of the specified goals. The thesis statement or argument is faintly presented and/or expected and unconfident, and the language is uneven in terms of balance in building body points, showing faults and errors in correctness. Additionally, writing frequently contains mistakes and ambiguities, which can make it read strangely and erratically.
1 (50-59%)	The research work shows a lack of comprehension and application of the given goals. The thesis statement or argument is oversimplified, unoriginal, or absent altogether and the writing is imbalanced in building body points, weak, and unfinished. Additionally, the writing is riddled with mistakes and ambiguities and reads incoherently.

2. Analytic rubrics

Two-dimensional or analytical rubrics, where the rows are the assessment criteria and the columns are the levels of achievement. allows raters to use a single rubric to evaluate participants' accomplishments across numerous categories. By adding up the criteria, raters can give each criterion a distinct weight or value and include an overall success (table 1.10).¹⁹

Table 1.10 Example of analytic rubrics¹⁹

Criteria	Level of mastery			
	Excellent (80-100%)	Good (70-79%)	Competent (60-69%)	Adequate (50-59%)
Knowledge of forms, conventions, terminology, and strategies relative to the importance of sources to subject	Displays thorough and incisive understanding of forms, conventions, terminology, and tactics in relation to the significance of source to the topic	Significant understanding of forms, conventions, terminology, and tactics in relation to the significance of sources to the topic	Demonstrates some familiarity of forms, conventions, terminology, and strategies relative to importance of sources to the topic	Lacked understanding of forms, conventions, terminology, and strategies relative to importance of sources to the topic
Critical and creative thinking skills	Makes excellent use of critical and creative thinking skills	Effectively employs critical and creative thinking skills	Modest effectiveness when applying critical and creative thinking skills	Restricted use of critical and creative thinking skills
Communication of information and idea	Communicates information and ideas with confidence and	Communicates information and ideas with great clarity	Communicates information and ideas with partial clarity	Communicates information and idea with lack clarity

Criteria	Level of mastery			
	Excellent (80-100%)	Good (70-79%)	Competent (60-69%)	Adequate (50-59%)
	a high level of clarity			
Quality of argument and writing	The argument is sophisticated and original and the writing is powerful, fluent, and creatively coherent	The argument is complicated and original and the writing is concise and cogent	The argument adopts a reasonable and expected stance and the wording is somewhat cohesive and understandable	The argument is straightforward and unoriginal and the wording is weak and inconsistent
Spelling and grammar	No errors	Some errors	A few errors	Several errors

The summation of advantages and disadvantages of rubrics are described in the table 1.11.⁴⁹

Table 1.11 The advantages and disadvantages of rubrics⁴⁹

Advantages	Disadvantages
Rubrics motivate students to exercise critical thought about their own performance and written work.	Because the rubric scale offers so many variables, using it could be a little challenging.
Additionally, by fostering metacognition, rubrics can help students become more perceptive learners.	Rubrics' language is not as straightforward as it should be.
Due to its emphasis on "criterion-referenced" rather than norm-referenced" grading, it aids educators in making more better decisions.	The lower scale might describe student achievement in a negative manner, discouraging students.
Learning objectives are made more clearer to students through rubrics, resulting in more fair treatment of students.	Each level needs to be easier to measure and monitor.
It lessens the chances that teachers will favor their students.	A more thorough specification is required.

1.3.5 Principle of assessment tool evaluation

1.3.5.1 Validity and reliability⁵⁰⁻⁵²

Validity and reliability are qualities that should be included in every well-designed evaluation tool. Therefore, each assessment tool used to measure the outcomes of the study must have its validity and reliability evaluated and reported.

Validity describes the correctness of a test or assessment tool. It refers to how effectively the evaluation captures the key outcome of interest. Validity is not a characteristic of the instrument per se, but rather of how the tool is used in a certain context with a given group of learners.

The validity is divided into 3 types:⁵⁰

1. **Construct validity** refers to the tool that can measure the intended construct.
2. **Content validity** refers to the tool that can accurately measure all aspects of the content. It offers a summary of the procedures followed in developing the tool, including who is responsible for doing so. In other words, national experts would give the tool more validity than local experts, and local experts would give the tool more validity than non-experts.
3. **Criterion validity** refers to the tool that can measure the same variables in comparison with the existing tool. Concurrent validity refers to the validity of the criterion when it is measured concurrently with the construct; predictive validity, on the other hand, refers to the validity of the criterion when it is measured at some point in the future after the construct has been measured.

Reliability of assessment tool means that it consistently produces the same outcomes when using with the same set of individuals and environments. In its simplest form, reliability refers to reliable or consistent outcomes. In determining validity, reliability is a component.

There are 3 types of reliability:⁵¹

1. **Test-retest reliability** means the stability of measurements at different times (overtime), one rater, and at least two administrations. Assuming the thing needs to be measured does not change over time. For example, the student who has high test score today, will have high test score next week. They are frequently evaluated using Pearson's r correlation. In general, a test-retest correlation of $+ .80$ or greater is considered to indicate a good reliability.
2. **Inter-rater reliability** means the stability of measurements when measured by different raters or observers using the same tool (across raters). Generally, it is estimated by percent agreement and kappa.
3. **Internal consistency reliability** means the stability of measuring instrument when divided its section (across items). It measures the degree of correlation between the answers. The scores of those items are correlated with each other. It is common practice to determine the correlation coefficients between the answers on your evaluation instrument using the Cronbach's alpha test of internal consistency. The set of items are evaluated the mean of all possible split-half correlation.

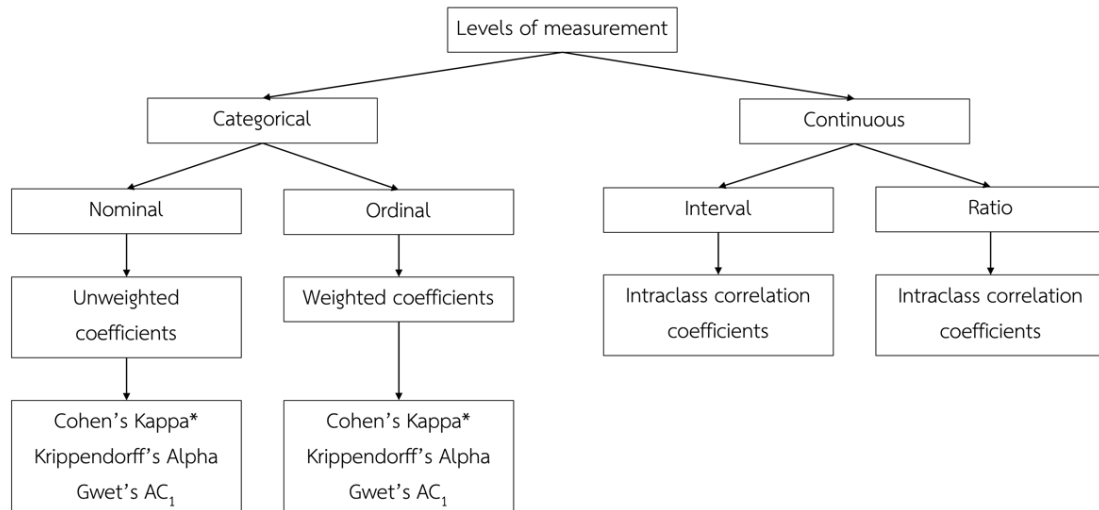
The assessment tool to be created in this study is targeted at the evaluation of cranial US performance as part of DOPS. It is important to evaluate how well the resident performed on the cranial US. The inter-rater agreement must be important in determining observer judgments of resident performance.

1.3.5.2 Choosing the right reliability statistic⁵³

The type of data should be taken into account when selecting the dependability statistic. Statistics' level of measurement separates the data into two primary categories; categorical data and continuous data. The nominal and ordinal data types are both present in the categorical data. Interval and ratio are the other two forms of continuous data. Depending on the kind of data we have gathered, the dependability coefficient needed for our analysis will change (figure 1.20).⁵³

Four levels of measurement are available:

1. **Nominal:** The data can only be categorized. There is no order between the categories. Examples are male/female, seen/not seen, marital status, and yes/no.
2. **Ordinal:** The data can be ranked and categorized. However, there is no how different of interval between ranking. For instance, a Likert scale rating, cancer stages, etc.
3. **Interval:** The data can be categorized, ranked, and distributed uniformly. The intervals between ranking are equal. There is no true zero point. For examples, ratings on a visual analog scale, IQ level, and temperature in Fahrenheit or Celsius.
4. **Ratio:** The information has a natural zero, can be categorized, ranked, and is equally spaced. Examples include test scores, body weight, height, or temperature in Kelvin.



* Cohen's Kappa can only be used where there are only two raters, all other measures are suitable for use with two or more raters.

Figure 1.20 Selection of suitable inter- and intra-rater reliability coefficients depending on the type of data to be analyzed⁵³

Reliability of nominal data

With binary data and two raters for inter-rater reliability or twice for intra-rater reliability, the simplest reliability test would be conducted. Cohen's Kappa, a measure of agreement that has been chance-adjusted, is the most often used coefficient. Similar options to Cohen's unweight Kappa include Krippendorff's alpha, Fleiss' Kappa, and Gwet's AC1, all of which can be applied in scenarios with three or more raters.⁵³

Reliability of ordinal data

Whenever categories can be arranged in a hierarchy, such as from weak to strong or low to high, this is referred to be ordinal data. Because it does not consider the size of the difference between data points and treats any discrepancy as a full disagreement, the basic Kappa coefficient previously mentioned is not appropriate for analyzing such data. However, when the data is ordinal, raters may provide identical scores, which still represents some degree of agreement, as in the case of Raters 1 and 2 in figure 1.21.⁵³

Cohen J developed a weighted version of his Kappa coefficient (K_w) to get over the problem that Kappa coefficients have with ordinal data analysis. With fewer restrictions than Cohen's Kappa and the potential to be extended to evaluate ordinal data, Krippendorff's alpha is a particularly adaptable reliability coefficient. The use of these metrics for analyzing ordinal data overlaps significantly with the usage of conventional correlation coefficients. When utilized in two-rater ordinal data studies, Krippendorff's alpha is mentioned as being equivalent to Spearman's r while the weighted Kappa is sometimes directly equivalent to a Pearson's r . The range of Krippendorff's alpha is between 0, which denotes utter lack of reliability, and 1, which denotes perfect reliability. Although it is even less frequently used than Krippendorff's alpha, Gwet's AC1 has a weighted variant for usage with ordinal data.⁵³⁻⁵⁵

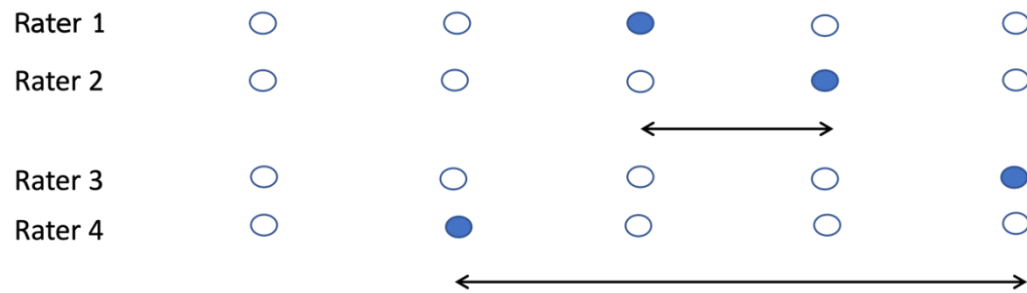


Figure 1.21 Five-point Likert scale example with varying degrees of disagreement. Raters 1 and 2's given scores differ by one point, however Raters 3 and 4's differ by three points, presumably indicating a far greater degree of disagreement⁵³

Reliability of continuous data

It is advised to use an intraclass correlation coefficient (ICC) to evaluate the reliability of inter-rater and intra-rater agreement for continuous data since it can account for systematic bias in measurements that decrease dependability.

A straightforward contingency table (table 1.12) can be created in this situation by simply organizing the data. Calculating an unweighted coefficient that corrects for chance agreement is the proper procedure for assessing agreement with such data.⁵³⁻⁵⁷

57

Table 1.12 A 2x2 table of sample results in example comparing 2 different raters rate of presence or absence of a finding on imaging⁵³⁻⁵⁷

Rater 2	Rater 1		Total
	Present	Absent	
Present	30 (a)	11 (b)	41 (g1)
Absent	8 (c)	23 (d)	31 (g2)
Total	38 (f1)	34 (f2)	72 (N)

The kappa (k) statistic formula⁵²: $k = \frac{P_o - P_e}{1 - P_e}$

Where P_o = observed (percent) agreement, $(a+d)/N$

P_e = agreement expected by chance, $((g1*f1)+(g2*f2))/N^2$

In our example

$$P_o = (30+23)/72 = 0.736$$

$$P_e = ((41*38)+(31*34))/72^2 = (1558+1054)/5184 = 0.504$$

$$K = (0.736-0.504)/(1-0.504) = 0.232/0.496 = 0.468$$

The kappa value ranges from -1 to +1. The value below 0, although unlikely occurred, means disagreement. The value of +1 indicates perfect agreement. The interpretation of kappa value is described in table 1.13. The accepted kappa value is of .80 or more.⁵⁶⁻⁵⁹

Table 1.13 Cohen's kappa interpretation⁵⁶⁻⁵⁹

Value of kappa	Level of agreement
0.00-0.20	None to slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Strong

After the kappa value is achieved, confidence interval is needed to confirm the value of obtained kappa. The smaller or narrower confidence interval represents more precise of agreement.

The confidence interval formula⁵⁸ is

$$k - 1.96 \times SE_k \text{ to } k + 1.96 \times SE_k$$

where: 1.96 is the constant value to multiply with the standard error of kappa (SE_k)

The standard error of kappa (SE_k) formula is

$$SE_k = \sqrt{\frac{p(1-p)}{n(1-p_e)^2}}$$

where: P represents the actual observed agreement, and P_e represents expected chance agreement

Pearson's correlation coefficient is evaluated in both construct validity and test-retest reliability. The construct validity is determined between correlation of total score or overall performance and number of previous cranial US. The correlation of test-retest reliability is analyzed, after the retest is performed at 1 month.

Chapter 2

Methodology

2.1 Study design

This study was an experimental study taken place at the neonatal intensive care unit (NICU) of Pediatrics department and ultrasound room of Radiology department, Faculty of Medicine, Prince of Songkla University. The study was approved by the Ethics Committee of Faculty of Medicine, Prince of Songkla University. The approval number was REC 62-342-7-1, dated 12-25-2019. This study was performed based on the ethical principles of the Declaration of Helsinki and The International Conference on Harmonization in Good Clinical Practice.

2.2 Study procedure

This study used mixed-methods of Delphi technique to develop a new cranial US assessment tool. The developed new assessment tool in cranial US skill performance was evaluated using the Delphi process for content validation and Cohen's Kappa with percent agreement for inter-rater reliability. The methods were divided into two phases; assessment tool development and assessment tool performance evaluation.

2.2.1 Phase I: Development of the new cranial US assessment tool⁴⁰⁻⁴³

The modified Delphi method with three steps procedures including development of initial cranial US assessment form, selecting group of experts into expert panel, and shaping and verifying the final assessment form was used to develop the new cranial US assessment tool.

First step: Generating items for the initial cranial US assessment form

The research question of this study was “How can we accurately assess the cranial US performance of radiologic residents, Department of Radiology, Faculty of medicine, Prince of Songkla University?” To answer the question, the cranial US skill assessment tool needed to be developed and validated.

Drafting the assessment items in the initial cranial US assessment form was initiated from the literature review.^{34,35,37,60-64} The literature review was collected by searching the Pubmed from 2011- 2020 with key words as ‘ultrasound’, ‘skill’, ‘assessment’, ‘tool’. Overall, 300 related articles were extracted. Of those, 265 were excluded due to unrelated or repetitive articles, leaving 35 articles to be reviewed.^{32,63-96} All aspects of general US skills that suitable for and significantly correlated with cranial US were inserted into the initial assessment form as initiated items. The detail of items generation was in the appendix H.

Second step: Selecting group of experts into expert panel

The selection criteria of the experts included pediatric radiologists currently practice in the pediatric radiologic field with an experience in neonatal cranial US for at least 3 years. All pediatric radiologists in Thailand who met the criteria were invited to participate in this study via the Google form. Those who were invited and would like to be part of the expert panel filled the Google form with maintaining anonymity.

Third step: Delphi process

The Delphi method is the method to seek the answer for specific problem from expert’s consensus, therefore, it is suitable for this study. The important aspects to be concerned in the Delphi method include anonymity of experts, iteration, feedback control, and statistical analysis of the results. The Delphi process is the iteration process and will be performed until reaching the consensus.

The Delphi process in this study included the iteration process as follows:

1. First round of the Delphi study: After the primary researcher developed the initial cranial US assessment form following the literature review, it was sent to the participated experts as Google form via emails. The participated experts were requested to score each item on 5-point rating score. The scores were ranged from 1 to 5; 1 meant the item was not relevant, 2 defines as the item as relevant but not important, 3 to 5 mean these items were relevant and important. 5 was equivalent to most important (3= little important, 4= moderately important, 5= absolutely important). The experts were also asked to give comments or opinions to eliminate some items or add other items including modified the initial items. This process might be a part of qualitative

assessment of content validity because repetitive nature of Delphi process and all selected experts have an extensive knowledge in cranial US.

2. Analysis of the Delphi round 1 result from experts: After first round of the Delphi process was completed, the mean scores and their distribution were evaluated. All items that had scores equal or more than 4 were kept and eliminated the lower mean scores. In each round, we allotted a total response time of 1 month, with email reminders sent to non-responders after 2 weeks.
3. Second round of the Delphi process: The cranial US assessment tool that eliminated the lower score items and added the new items from expert opinions was sent again to all experts to evaluate the mean scores and distribution of all experts. The experts were asked to re-rank their rating scores and provide opinions, with opportunity to change their scores after viewing group's replies of the previous round.
4. Analysis of the Delphi round 2 results, third round of the Delphi study, and analysis of the Delphi round 3 results were repeated processes which was stopped when the opinions were concluded without further suggestion.
5. Setting the performance levels came after completing the elements to determine performance criteria for cranial US. We settled on the following four performance levels: Distinguished, Proficient, Emergent, and Insufficient. We used these four levels with superior scoring performance dependability than fewer levels when making formative and summative decisions. For each level of each performance criterion, performance descriptions were created using rubrics, which offer information to help raters make decisions.
6. All finalized items in the cranial US assessment tool from third round of Delphi process were divided into three divisions: the preparation phase, US machine operation, and cranial US performance, and their sum was the overall performance score.
7. The final step involves research documentation, verification, and generalization of the final cranial US evaluation instrument with a rubric scale that was once more submitted to all experts for approval. This process resulted in a fully designed cranial US skill assessment tool.

2.2.2 Phase II: Determination of reliability of the new cranial US assessment tool⁹⁷⁻¹⁰¹

In this phase II, it was determined whether the final cranial US skill assessment tool was reliable. Two pediatric radiologists who were raters used the tool to assess the cranial US skill performance of radiologic residents as seen on video footage. Interrater reliability was assessed using the Cohen's Kappa.

Inclusion criteria:

1. The cranial US performers included PGY- 1, PGY- 2, and PGY- 3 radiologic residents (who have different levels of cranial US experience), Department of Radiology, Prince of Songkla University.
2. All babies who were scheduled to obtain the cranial US either at the US unit, Department of Radiology or admitted at the NICU, Department of Pediatrics, Faculty of Medicine, Prince of Songkla University.

Sample size:

The sample size was calculated with the formula by Flack et.al.¹⁰²

$$n \geq \left[\frac{z_{1-\alpha} \max \tau(\hat{\kappa} | \kappa = \kappa_0) + z_{1-\beta} \max \tau(\hat{\kappa} | \kappa = \kappa_1)}{\kappa_0 - \kappa_1} \right]^2$$

N= sample size

K₀= value of Kappa of null hypothesis = 0.3 or 0.4

K₁= value of Kappa of alternative hypothesis (from previous literature) = 0.7 or 0.8

Beta= power

Alpha= 0.05

Alternative hypothesis = two-sided

In this study, there were 4 categories of rating items with power (n^b) at 90.0% and alpha of 0.05. According to guidelines of the minimal sample size requirements for Cohen's Kappa⁵⁵ the minimal sample size was 26 (figure 2.1).

TABLE 1. Sample size calculation for kappa at category 2 until 7 when proportion each category is assumed at proportionate

Category	K ₁	K ₂	n ^a	n ^b	Category	K ₁	K ₂	n ^a	n ^b	Category	K ₁	K ₂	n ^a	n ^b
2×2	0.0	0.2	194	259	4×4	0.0	0.2	71	97	6×6	0.0	0.2	49	68
		0.3	85	113			0.3	32	44			0.3	23	32
		0.4	47	62			0.4	18	25			0.4	13	18
		0.5	29	38			0.5	12	16			0.5	8	12
		0.6	20	25			0.6	8	11			0.6	6	8
		0.7	14	17			0.7	6	7			0.7	4	6
		0.8	10	12			0.8	4	5			0.8	3	4
		0.9	7	8			0.9	3	4			0.9	2	3
0.3	0.3	0.4	698	927	0.3	0.3	0.4	348	465	0.3	0.3	0.4	283	380
		0.5	169	222			0.5	86	114			0.5	70	94
		0.6	72	94			0.6	37	49			0.6	31	41
		0.7	39	49			0.7	20	26			0.7	17	22
		0.8	23	28			0.8	12	15			0.8	10	13
		0.9	14	17			0.9	8	9			0.9	6	8
0.5	0.5	0.6	563	742	0.5	0.5	0.6	317	420	0.5	0.5	0.6	271	359
		0.7	133	171			0.7	76	98			0.7	65	85
		0.8	54	68			0.8	31	40			0.8	27	34
		0.9	27	32			0.9	16	19			0.9	14	16
0.7	0.7	0.8	363	471	0.7	0.7	0.8	223	290	0.7	0.7	0.8	196	255
		0.9	79	96			0.9	49	60			0.9	43	53

Figure 2.1 Guidelines of the minimal sample size requirements for Cohen's Kappa⁵⁵

Informed consent:

The final cranial US assessment tool from phase I was used to evaluate cranial US skills performance of radiologic residents after the information and objectives of the study were given to either radiologic residents or neonate's parents who participated in this study with written informed consent. Two raters were assigned to use the assessment tool and completed the form after standardization with instruction sheets.

Video recording:

All items of final cranial US assessment tool were used to evaluate the cranial US skills performance among radiologic residents recorded in the video footage. Video footage was recorded using two mobile video cameras by the researcher. The researcher was trained to record the video including angle of the camera and filed of

view in order to provide the same quality of video images. The first device recorded hand movement of the residents. The second device recorded the US screen and US pad to evaluate performer's skills of using US machine and ability to create the US images. The US performers were asked to wear gloves on both hands during the procedure. Prior to recording the video, the pediatric radiologic staff, other radiologic residents, and NICU nurses were requested not to comment, intervene, or give any suggestion during cranial US procedure. If the recorded video images revealed the name or face of the participants, it would be concealed. Two video-recorded from both cameras of the same baby were merged into one footage, synchronously, with high definition (HD) resolution (figure 2.2).



Figure 2.2 Merge video footage showed the view of resident's skill on US machine and view of resident's hand movement during US.

Cranial US procedure:

A small 8-5 MHz curvilinear transducer, the most suitable for assessing the anterior fontanel, was used to perform cranial US. The probe was placed at the anterior fontanel without pressure, using a gel applied to the probe to ensure a good contact between the probe surface and adjacent scalp skin. The baby was placed in supine position during examination. The cranial US was started with coronal view. Then, the

probe was rotated 90 degrees to perform US on sagittal view. The US images should be recorded as the standard cranial US protocol (as shown in the chapter 1).

Cranial US skill evaluation:

The cranial US skills performance of radiologic residents recorded on video footage was independently evaluated by two raters using the developed new assessment tool. Two raters were pediatric radiologists from different institutions either pediatric or university hospital.

2.3 Statistical analysis

Expert characteristics were listed as percentage values. Each item's mean score was assessed during the Delphi process. Cohen's Kappa (unweighted and weighted Kappa) was used to determine how reliable each assessment tool item was. The degree of agreement between two raters was also calculated. The construct validity was evaluated by comparing total score and level of residents by PGY, total score and experience of cranial US cases, overall performance and level of residents by PGY, and overall performance and experience of cranial US cases residents using t-test or ANOVA. A p value < 0.05 was indicated statistically significant. The internal consistency reliability was also determined using Cronbach's alpha coefficient. Utilizing R software 3.6.1, all data were analyzed.

Chapter 3

Results

3.1 Phase I: Assessment tool development

3.1.1 Participants

All 28 pediatric radiologists in Thailand have been invited to be the expert panel in this study. An online community of pediatric radiologists was used to distribute the invitation forms. Seventeen pediatric radiologists (60.7%) consent to participate in the Delphi procedure. Table 3.1 displays the characteristics of the experts in both Delphi round 1 and round 2. Majority of expert participants were female (82.3% and 84.6%), working as instructors (52.9% and 61.5%) in the University Hospital (70.6% and 61.5%). Over 50% of the experts experienced on the cranial US for more than 5 years.

Table 3.1 Characteristics of expert participants in round 1 and 2 (n=17 and n=13)

Characteristics	Number (%)	Number (%)
Gender		
Male	3 (17.7)	2 (15.4)
Female	14 (82.3)	11 (84.6)
Position		
Instructor	9 (52.9)	8 (61.5)
Assistance Professor	3 (17.7)	3 (23.1)
Associated Professor	3 (17.7)	2 (15.4)
Professor	2 (11.7)	0 (0)
Workplace		
University Hospital	12 (70.6)	8 (61.5)
Pediatric Hospital	2 (11.8)	2 (15.4)
General Hospital	2 (11.8)	2 (15.4)
Others	1 (5.8)	1 (7.7)
Experiences on cranial US (years)		
3-5	8 (47.1)	6 (46.2)
>5-10	4 (23.5)	3 (23.1)
>10	5 (29.4)	4 (30.7)

3.1.2 Delphi process

Three rounds of the Delphi process were described as below. The response rate of experts in round 1 was 60.7% and round 2-3 was 76.5% (table 3.2).

Table 3.2 Response of experts to Delphi process

Delphi round	No. sent out	No. responders (%)
1	28	17 (60.7)
2	17	13 (76.5)

Delphi round 1

From the literature review, 14 pertinent items were established and included in the initial evaluation form (table 3.3). These items included the position of US machine, selected appropriate probe, corrected handing of probe, applying of gel on probe, using proper functions of US machine, adjusting appropriate gain and frequency, adjusting depth and focus to include whole brain and appropriate area of focus, cleaning the probe after using, corrected placement of probe at anterior fontanel to starting US, corrected probe orientation, displaying cranial US approach, presenting of relevant brain structures according to standard protocol, identified the positive findings, and overall performance. All experts taken part in the Delphi round 1 received the Google form containing these 14 items via emails (figure 3.1). The mean scores and distributions of all items are shown in table 3.4. 'Additional US by employing linear probe' and 'identified patient' were two new items added from the expert recommendation in Delphi round 1.

Table 3.3 Items in the initial assessment tool

Items
1. Position of US machine, appropriately
2. Selection appropriate probe
3. Corrected handing of probe
4. Applying of gel on probe
5. Using proper functions of US machine
6. Adjusting appropriate gain and frequency
7. Adjusting depth and focus to include whole brain and appropriate area of focus
8. Cleaning probe after using
9. Corrected placement of probe at anterior fontanel to starting US
10. Corrected probe orientation
11. Displayed cranial US approach, systematically
12. Presenting of relevant brain structures according to standard protocol
13. Identified the positive findings (e.g. IVH, PVL, hydrocephalus, and anomaly)
14. Overall performance

แบบแสดงความคิดเห็นเกี่ยวกับแบบประเมินประสิทธิภาพการทำอัลตราซาวด์สมอง (cranial ultrasound assessment form)

การให้คะแนน 5= ช้อยนี้เกี่ยวข้องและสำคัญมาก, 4= ช้อยนี้เกี่ยวข้องและสำคัญปานกลาง, 3= ช้อยนี้เกี่ยวข้องและสำคัญน้อย, 2= ช้อยนี้เกี่ยวข้องและไม่สำคัญ, 1= ช้อยนี้ไม่เกี่ยวข้อง

*จำเป็น

โปรดให้คะแนนความสำคัญของแต่ละข้อที่ใช้ประเมินการทำอัลตราซาวด์สมอง *

	5 = absolutely important	4 = moderate important	3 = little important	2 = relevant but not important	1 = not relevant
Positioning of US machine, appropriately	☐	☐	☐	☐	☐
Selecting appropriate probe	☐	☐	☐	☐	☐
Correct handling of probe	☐	☐	☐	☐	☐

Presenting of relevant brain structures according to standard protocol

Identified the positive findings (e.g. IVH, PVL, hydrocephalus, and anomaly)

Overall performance

Comment and opinion (including add or modified items in assessment form) *

คำตอบของคุณ

ถัดไป

สร้างแบบฟอร์ม

นำเสนอด้วย Google Forms

เนื้อหาไม่ได้ถูกสร้างหรือแก้ไขโดย Google

Google Forms

Figure 3.1 The Google form of rating scale for initial items in cranial US assessment tool

Table 3.4 Initial items with mean scores and distributions in Delphi round 1.

Items	Distribution (n=17)					Mean scores
	5	4	3	2	1	
1. Positioned US machine, appropriately	7	9	1	0	0	4.35
2. Selected appropriate probe	16	1	0	0	0	4.94
3. Corrected handling of probe	14	3	0	0	0	4.82
4. Applying of gel on probe	10	7	0	0	0	4.59
5. Using proper functions of US machine	13	4	0	0	0	4.76
6. Adjusting appropriate gain and frequency	12	5	0	0	0	4.71
7. Adjusting depth and focus to include whole brain and appropriate area of focus	17	0	0	0	0	5.00

8. Cleaning probe after using	8	8	1	0	0	4.41
9. Corrected placement of probe at anterior fontanel to starting US	15	2	0	0	0	4.88
10. Corrected probe orientation	16	1	0	0	0	4.94
11. Displayed cranial US approach, systematically	11	6	0	0	0	4.65
12. Presenting of relevant brain structures according to standard protocol	13	4	0	0	0	4.76
13. Identified the positive findings (e.g., IVH, PVL, hydrocephalus, and anomaly)	17	0	0	0	0	5.00
14. Overall performance	13	4	0	0	0	4.76

Delphi round 2

The second round of Delphi process involved sending the experts the updated assessment form initiated from the Delphi round 1 including all of the original items with mean scores of 4 or more and the two newly added proposed items. The results of Delphi round 2 are shown in table 3.5. All items had average mean scores of 4 or more. The item of ‘additional US by using linear probe’ was suggested to change to ‘selected appropriate probe’. No more additional comment, recommendation or changed items from the experts in round 2. In this stage, there were final 15 items consisted of identified patient, position of US machine, selected appropriate probe, corrected handing of probe, applying of gel on probe, using proper functions of US machine, adjusting appropriate gain and frequency, adjusting depth and focus to include whole brain and appropriate area of focus, cleaning probe after using, corrected placement of probe at anterior fontanel to starting US, corrected probe orientation, displaying cranial US approach, presenting of relevant brain structures according to standard protocol, identified the positive findings, and overall performance.

Delphi round 3

All 15 finalized items were accepted by the experts without additional comments. The assessment form's rubric scale was created based on the characteristics of each item by divided into 3 sections: the preparation phase, US machine operation, and cranial US performance. Four degrees of competence including distinguished, proficient, emergent, and insufficient were established (table 3.6). The experts received the final assessment form with rubric scale for approval. In round 3, there were an agreement and verification by all experts.

Table 3.5 Items with mean scores and distributions in Delphi round 2

Items	Distribution (n=13)					Mean scores
	5	4	3	2	1	
1. Positioned US machine, appropriately	5	7	1	0	0	4.31
2. Selected the appropriate probe	12	1	0	0	0	4.92
3. Corrected probe handing	11	2	0	0	0	4.85
4. Appling of gel on the probe	10	3	0	0	0	4.77
5. Using proper functions of the US machine	10	3	0	0	0	4.77
6. Adjusting the appropriate gain and frequency	11	2	0	0	0	4.85
7. Adjusting the depth and focus to include the whole brain and appropriate area of focus	11	2	0	0	0	4.85
8. Cleaning the probe after using	7	5	1	0	0	4.46
9. Correctly placed the probe at the anterior fontanel to starting US	10	3	0	0	0	4.77
10. Corrected probe orientation	13	0	0	0	0	5.00
11. Displayed cranial US approach, systematically	8	4	1	0	0	4.53

12. Presenting the relevant brain structures according to standard protocol	9	4	0	0	0	4.69
13. Identified positive findings (e.g., IVH, PVL, hydrocephalus, and anomaly)	13	0	0	0	0	5.00
14. Overall performance	8	5	0	0	0	4.62
15. Identified the patient	10	3	0	0	0	4.77
16. Additional US using a linear probe	7	5	1	0	0	4.46

Table 3.6 Final cranial US performance assessment form with rubric scale

Cranial US performance assessment form with rubric scale				
	4 (Distinguished)	3 (Proficient)	2 (Emergent)	1 (Insufficient)
*Preparation phase (2 items)				
Identified patient	Done and confirmed	Done	Not done	Wrong patient
Properly positioned the US machine	Located to the right, close to the patient	Located to the right of the patient	Any side but tries to do with the dominant hand	Any side and operates with the non-dominant hand
*US machine operation (7 items)				
Selecting the appropriate probe	Small curvilinear 8-5 MHz probe	Linear high-frequency probe	Large curvilinear low-frequency probe	Hockey stick probe
Corrected probe handling	Complete	Inadequate	Improper	Incorrect
Applying warm gel onto the probe	Put the optimal amount of gel onto the probe	Put gel onto the probe	Put the gel on the patient's head	Did not use gel
Using proper buttons on the US machine	Familiar with operating	Operated with experience but was inconsistent	Operated with inexperience	Unable to operate
Adjusted gain and frequency, correctly	Appropriate	Inappropriate	Incorrect	Unable to operate

Cranial US performance assessment form with rubric scale				
	4 (Distinguished)	3 (Proficient)	2 (Emergent)	1 (Insufficient)
Adjusting the depth and focus to include the whole brain and appropriate area of focus	Optimum field of view, depth, and focus	Included the whole brain but inappropriate depth and focus	Included partial brain	Unable to operate
Cleaning the probe after using	Done			Not done
*Cranial US performance (5 items)				
Corrected placement of the probe at the anterior fontanel	Placed on the anterior fontanel with good contact	Placed on the anterior fontanel with poor contact	Place on another fontanel	Placed on other areas
Confirmed the side of brain orientation	Correct side with confidence	Confirmed several times	Not done	Wrong side
Displayed cranial US approach, systematically	Consistently systematic display	Some systematic display	Unsystematic display	Could not display
Presenting the relevant brain structures according to standard protocol	Followed standard protocol	Inconsistent	Presence of some images	Could not do

Cranial US performance assessment form with rubric scale				
	4	3	2	1
	(Distinguished)	(Proficient)	(Emergent)	(Insufficient)
Identified positive findings (e.g., IVH, PVL, hydrocephalus, or anomaly)	Completely identified positive findings	Identified some findings	Missed positive findings	Could not do

3.2 Phase 2: Assessment tool performance evaluation

3.2.1 Subject characteristics

Neonates

Twenty-seven neonates participated in this study. Mean age was 35 days (range, 8-80). Only 3 neonates were being supported by the ventilator. The US procedure was performed mostly at the NMCU (55.6%). The information is provided in table 3.7.

Table 3.7 Characteristics of neonates and places of performing US (n = 27)

	Number (%)
Neonates	
Ventilator support	3 (11.1)
Places to perform US	
NICU	10 (37.0)
NMCU	15 (55.6)
US room	2 (7.4)
Cranial US findings	
Normal	7 (26.0)
IVH	18 (66.7)
PVL	3 (11.1)
Other	4 (14.8)

Radiologic residents

In this study, 27 radiologic residents were participated. Two thirds of them were female. Almost half were enrolled in PGY3 training. Most of them (70%) had experienced at least five cranial US cases performed. Only 4 residents lacked of any cranial US training experience, while the remaining 23 had some kinds of training, mostly lecture and observation (table 3.8).

Table 3.8 Characteristic of radiologic residents (n = 27)

	Number (%)
Gender	
Male	
Female	9 (33.3)
	18 (66.7)
PGY	
Year 1	4 (14.8)
Year 2	11 (40.7)
Year 3	12 (44.5)
Number of cranial US performed (cases)	
0	6 (22.2)
<5	2 (7.4)
5-10	5 (18.5)
11-15	11 (40.8)
16-20	3 (11.1)
Experiences in cranial US training	
Never	4 (14.8)
Lecture	21 (77.8)
Observation	19 (70.4)
Workshop	5 (18.5)
Self-learning	9 (33.3)
Others	2 (7.4)

3.2.2 Content validity

The term "content validity" describes a method that can precisely assess all aspects of the content. Pediatric radiologists who were subject matter experts in cranial US made up the entire Delphi panel of experts. The Delphi process confirmed good content validity due to its iteration nature.

3.2.3 Interrater reliability

Cohen's Kappa (unweighted and weighted Kappa) and the percent agreement between two raters were used to calculate interrater reliability. Percent agreement represented the simplest way to measure agreement, but it does not account for variations in chance. While, Kappa measure of agreement above and beyond what would be predicted randomly. Most items in the assessment form in this research, except the item 'cleaned the probe after using', were ordinal data. Weighted Kappa was more appropriate to evaluate Kappa value. Each evaluation tool item's Kappa score ranged from 0.39 to 0.94 and 'NaN'. 'NaN' was abbreviation for Not a Number, which meant that the data's lack of variance prevented the calculation of the Kappa value. All items had a percent agreement that ranged from 37% to 100%. The detail of Cohen's Kappa (unweighted and weighted Kappa) and the percent agreement of each item are present in table 3.9. In all 15 items, 3 items: identification of patient, selected appropriate probe, and cleaning the probe after using had Kappa value of NaN with 100 percent agreement. Three items: properly positioned the US machine, presented the relevant brain structures according to standard protocol, and overall performance had almost perfect agreement (weighted Kappa value = 0.89-0.94 with 88.9-96.3 percent agreement). Corrected probe handling, adjusting the depth and focus to include the whole brain and appropriate area of focus, confirmed the side of brain orientation, and identified positive findings had substantial agreement (weighted Kappa value = 0.70-0.80 with 77.8-85.2 percent agreement). Three items: applying warm gel onto the probe, using proper buttons on the US machine, and corrected placement of the probe at the anterior fontanel, had moderate agreement (Kappa value = 0.42-0.60 with 70.4-81.5 percent agreement). Two items 'adjusted gain and frequency,

correctly’ and ‘displayed cranial US approach systematically’ had fair agreement, (Kappa value = 0.39 with 37.0 and 63.0 percent agreement, respectively).

Table 3.9 Interrater reliability of the assessment tool

Items	Unweighted Kappa	Weighted Kappa	% Agreement
1. Identification of the patient	NaN	NaN	100
2. Properly positioned the US machine	0.93	0.94	96.3
3. Selected appropriate probe	NaN	NaN	100
4. Corrected probe handing	0.69	0.80	81.5
5. Applying warm gel onto the probe	0.42	0.42	77.8
6. Using proper buttons on the US machine	0.44	0.60	70.4
7. Adjusted gain and frequency, correctly	0.15	0.39	37.0
8. Adjusting the depth and focus to include the whole brain and appropriate area of focus	0.55	0.76	77.8
9. Cleansing the probe after using	NaN	NaN	100
10. Corrected placement of the probe at the anterior fontanel	0.55	0.55	81.5
11. Confirmed the side of brain orientation	0.57	0.70	77.8
12. Displayed cranial US approach, systematically	0.27	0.39	63.0
13. Presenting the relevant brain structures according to standard protocol	0.82	0.89	88.9
14. Identified positive findings (e.g., IVH, PVL, hydrocephalus, and anomaly)	0.65	0.80	85.2
15. Overall performance	0.84	0.92	88.9

3.2.4 Construct validity

The radiologic residents were categorized according to their PGY and experience on cranial US. The total score and level of PGY of residents were analyzed. There were statistically differences in total score of PGY1 and PGY2 residents and PGY1 and PGY3 residents ($p = 0.023$ and 0.007 , respectively, figure 3.2). Further analysis according to total score and experience of cranial US was also determines. The results revealed significant statistically difference in total score and experience of cranial US <5 cases and ≥ 5 cases ($p = 0.034$, figure 3.3).

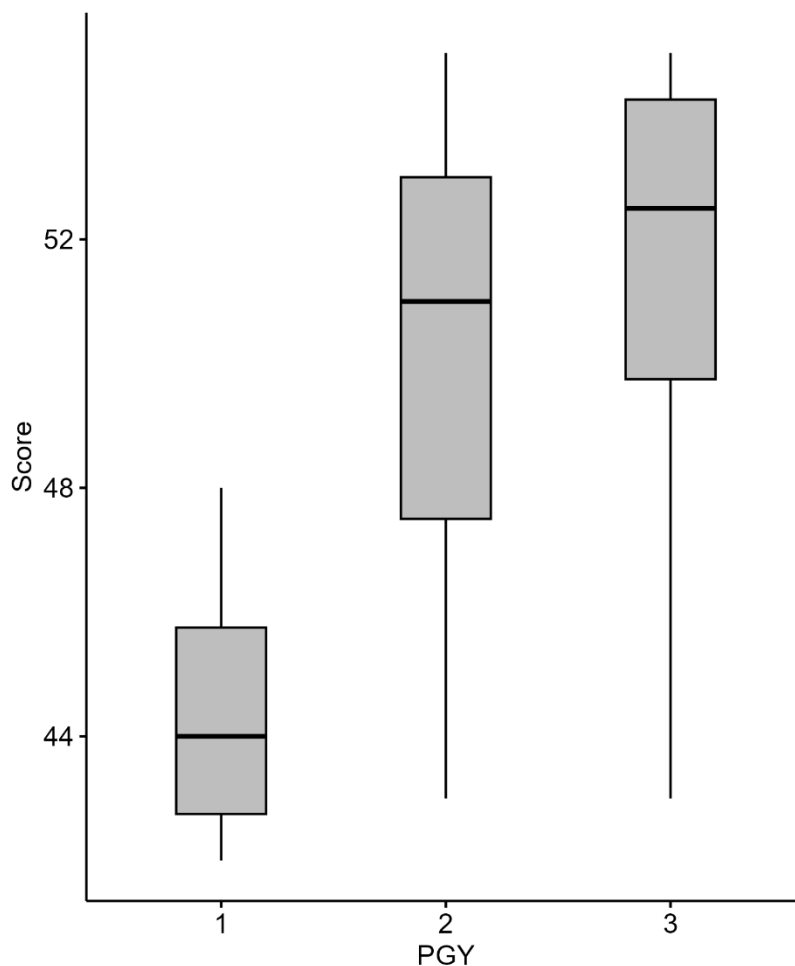


Figure 3.2 Box and whisker plot of total score and PGY 1, 2, and 3 of radiologic residents

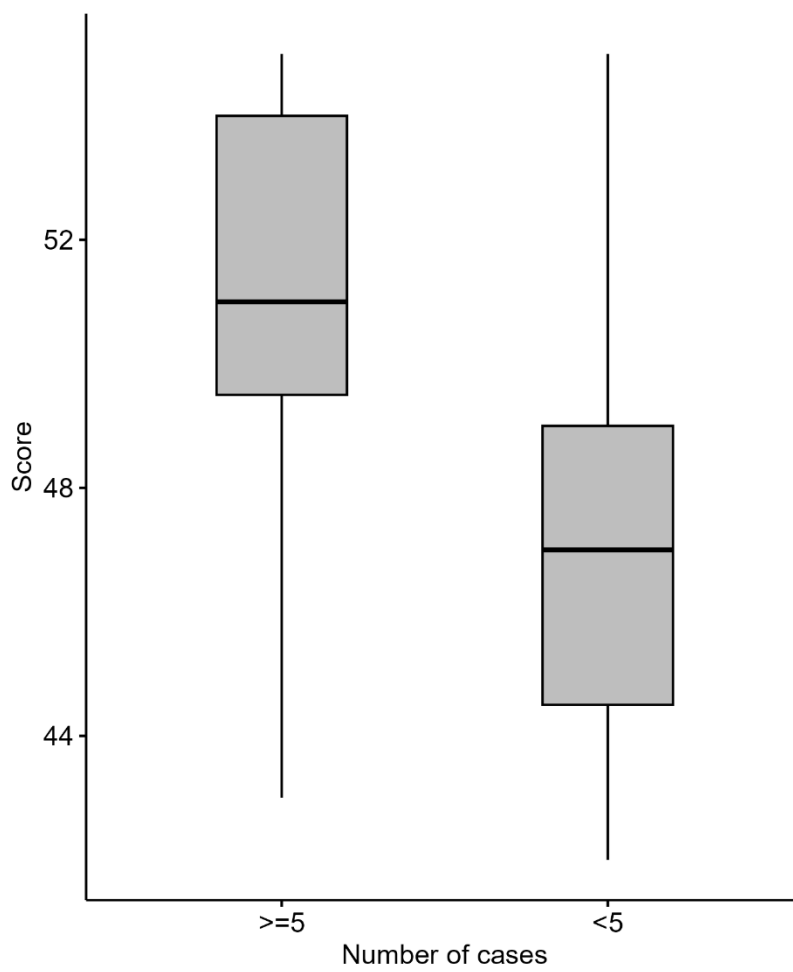


Figure 3.3 Box and whisker plot of total score and experience of cranial US cases of radiologic residents

Overall performance of radiologic residents was also evaluated according to their PGY and experience of cranial US. Overall performance of PGY 1 radiologic residents and PGY 3 radiologic residents demonstrated statistical significance ($p = 0.026$, figure 3.4). Statistically significant between overall performance and experience of cranial US was also found ($p = 0.022$, figure 3.5).

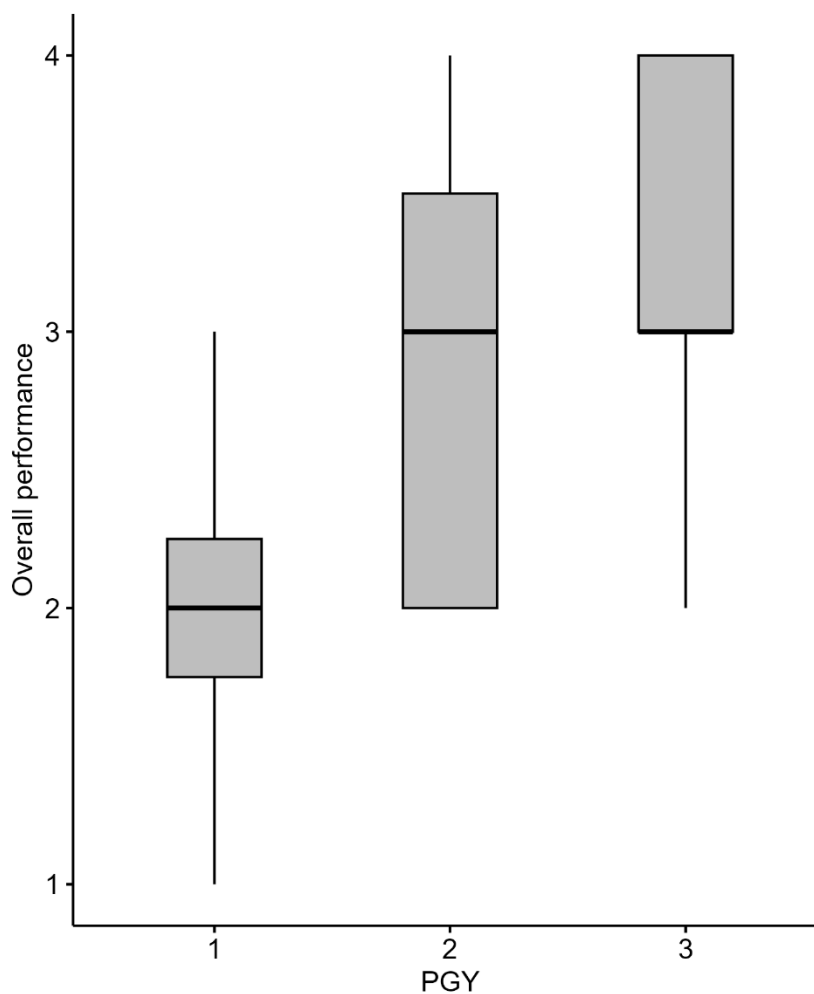


Figure 3.4 Box and whisker plot of overall performance and PGY 1, 2, and 3 of radiologic residents

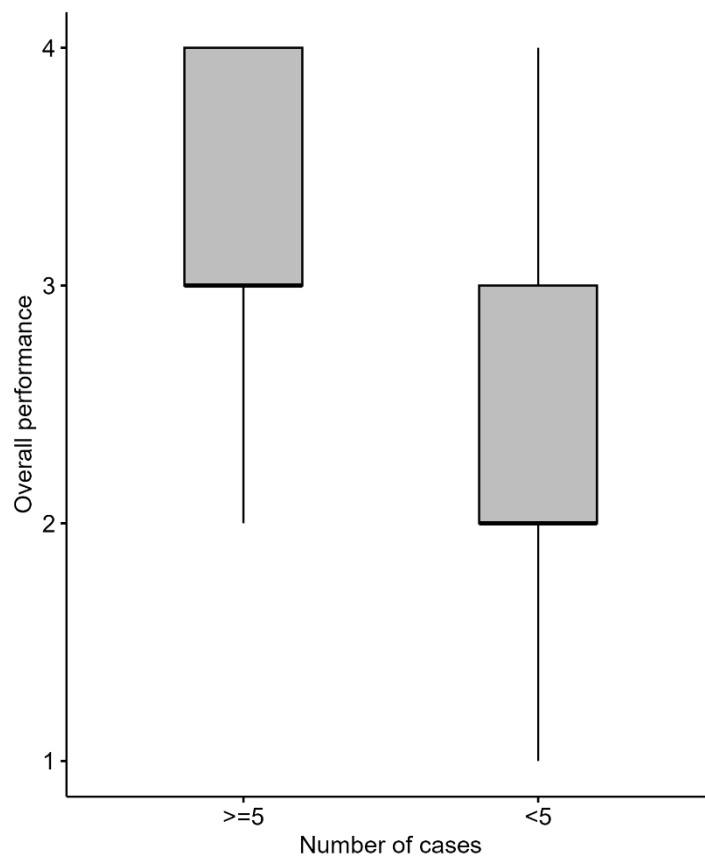


Figure 3.5 Box and whisker plot of overall performance and experience of cranial US cases of radiologic residents

3.2.5 Internal consistency reliability

The internal consistency of the assessment tool was evaluated using Cronbach's alpha. The Cronbach's alpha value of both raters were about 0.856 and 0.891.

Chapter 4

Discussion

In the field of neonatal care, cranial ultrasound is a necessary supportive examination. Early detection of intracranial abnormalities in premature neonates, particularly IVH or PVL is crucial due to their long-term prognosis.^{1,2,6-8} The cranial US is an effective tool to identify these abnormalities, however it's required training to improve practical skills and experience. Like the US of other parts of human bodies, there is no uniform assessment tool for cranial US to evaluate the learner's competency.^{4,5} We, therefore, designed this study to develop the cranial US assessment form and determine the validity and reliability of this tool.

Delphi method and content validity

The Delphi method is an excellent technique to create consensus and produce new ideas. This cranial US assessment form using Delphi process should be regarded as a high-quality assessment tool due to it was created with agreement from almost all pediatric radiologists, expertise on cranial US in Thailand.

Originally, 14 items were created. After Delphi round 1, two additional new items including 'additional US by employing linear probe' and 'identified patient' were added from the expert recommendation. Although there was a number of drop off, 48% of expert panel remained participating in the Delphi round 2 and 3. In Delphi round 2, the expert suggestion was to merge the new item of 'additional US by employing linear probe' to the original item of 'selected appropriate prove', resulting in the final 15 items cranial US assessment form for approval in Delphi round 3. The rubric scale cranial US assessment form was successfully established, approved and verified by the expert panel. All 15 items created by the expert panel reflects the content validity of the assessment form during this Delphi method.

The advantage of this study is that all expert panelists are pediatric radiologists with expertise in cranial US and the items of assessment refer to content validity. Most participant's workplace is the university hospital, however some working at pediatric and general hospital. This could reflect the generalizability of using this assessment

form in variety setting. Moreover, the optimum assessment technique for performance evaluation is the DOPS methodology.^{21,22} The rubric scale's information enables the rater to make a confident, accurate, and reliable judgment about a candidate's level of proficiency. For the aforementioned reasons, it is predicted that this assessment form will be a competent and trustworthy tool to assess cranial US performance.

In order to prevent group thinking based on reputation or authority, all experts in this study were given anonymous. Experts could, therefore, independently comment and decide. However, participant anonymity prevented responses from being connected to professional experience, which is a drawback of this study. Additionally, the participant response rate decreased with each cycle of the Delphi process, which led to an overemphasis of the responses from fewer participants.

Interrater reliability

In this study, 27 radiologic residents from PGY1-3 participated in the reliability test for the new cranial US assessment tool. By using interrater reliability evaluation, almost all items demonstrated moderate to practically perfect agreement (Kappa = 0.42-0.94 and NaN). The NaN with 100% agreement of three items of 'identification of patient', 'selected appropriate probe', and 'cleaning the probe after using' showed the most objective items. Only two items of 'adjusted gain and frequency correctly' and 'displayed cranial US approach systematically' showed fair Kappa value of 0.39, each. Due to the subjective nature of the evaluation of both items and the vastly divergent levels of acceptance of vivid color contrast in US photographs across raters in 'adjusted gain and frequency correctly', these might be explained the relatively low Kappa and percent agreement of those two items. Careful training of raters and modifying the description at each level of mastery of these two items should enhance Kappa's value. Several studies attempted to estimate the reliability of US performance assessment tools, which vary in organs, number of raters, and forms of assessment tool. The previous research of Chuan et al., which evaluated the ICC of the rating scale for ultrasound-guided regional anesthesia, showed poor to moderate agreement (ICC = -0.01 to 0.43)⁷⁶. Another study by Rice et al. demonstrated the reliability of the checklist tool for assessing the placement of US-guided intravenous access that had

an almost perfect agreement (weight Kappa value = 0.93)⁶⁵. Wong et al. demonstrated the ICC of checklists and global rating scales for the assessment of procedural skills in anesthesia. The result of Wong's study presented moderate agreement (ICC =0.44) of the total score⁹⁵. Sultan et al. also studied the reliability of clinical assessment tools for US-guided axillary brachial plexus block. The answers to their study indicated substantial agreement for rating scales and almost perfect agreement for a checklist (ICC = 0.795 and 0.842, respectively)⁷⁹. Overall agreement on assessment tools from previous studies varied from poor to almost perfect agreement. This implies that our study's developed assessment tool was accepted and is a good, reliable assessment tool with moderate to perfect agreement of most items compared to previous studies' results.

Construct validity

Both total score and overall performance were able to differentiate between PGY of radiologic resident and their experience of cranial US. As the results in this research, total score was significantly difference between PGY 1 and PGY 2 and PGY 3 and also difference between experience of cranial US < 5 cases and $\geq$ 5 cases. Overall performance between PGY1 and PGY 3 and experience of cranial US < 5 cases and $\geq$ 5 cases were also statistically difference. These results suggested good construct validity.

Internal consistency reliability

Cronbach's alpha was used to assess the internal consistency. The previous study by Gaudet et al. studied assessment tools to evaluate image acquisition quality for focused critical care echocardiography. The Cronbach's alpha of their assessment tool was 0.95. Another study by Kunz et al. presented Cronbach's alpha of performance checklist for US-guided arthrocentesis, which was about 0.89⁸⁹. The Cronbach's alpha value of our study was a good reliability level (0.856-0.891) that was relatively comparable with the previous studies.

Limitations of the study

Apart from the limitation by the nature of Delphi process mentioned above, our study did not evaluate the construct validity of this new assessment tool. Therefore, the discrimination of resident's ability or the pass/fail score could not be identified and evaluated. The narrow range of cranial US background ability of radiologic residents in our hospital at the beginning, resulting in homogeneous skill performance and less varieties, causes difficulty in this kind of analysis. This new assessment tool for cranial US performance requires user's training to increase familiarity and reliability. Further study of construct validity should be also investigated.

Chapter 5

Conclusions

To the best of our knowledge, this is the first study to develop the objective structured assessment tool for cranial US performance skills. This tool was refined using a Delphi method to achieve consensus from pediatric radiologist expert panel in Thailand. The assessment tool demonstrates a high content validity and moderate to high interrater agreement according to weighted Kappa value. The assessment tool also exhibits construct validity, that can differentiate between the cranial US performers by PGY and cranial US experience. Internal consistency of the assessment tool is also in good level.

The simple-to-use cranial US assessment tool with rubric scale holds all raters to the same standard of evaluation. Moving from a fixed number of examinations to competency-based evaluation in real life workplace situation using this valid and reliable assessment form together with other measurement tools for formative assessment is reasonable and suitable in the era of competency-based curriculum. Realtime reflection and feedback is feasible to improve the learner's performance. Adaptation from formative assessment to summative assessment can be done from this assessment tool by indicate cutoff point that is appropriate for the learner levels.

User's training of this assessment tool and generalization is encouraging to enhance interrater reliability and modify the tool for more optimization.

References

1. Diwakar RK, Khurana O. Cranial sonography in preterm infants with short review of literature. *J Pediatr Neurosci* 2018;13(2):141-9.
2. Volpe JJ. Brain injury in the premature infant. *Clin Perinatol* 1997;24(3):567-87.
3. Stanojevic M. Training of ultrasound in neonatology: global or local? Donald Sch J *Ultrasound Obstet Gynecol* 2013;7(3):338-45.
4. Bould MD, Crabtree NA, Naik VN. Assessment of procedural skills in anaesthesia. *Br J Anaesth* 2009;103(4):472-83.
5. Tolsgaard MG, Ringsted C, Dreisler E, Klemmensen A, Loft A, Sorensen JL, et al. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology: Assessment of ultrasound competence. *Ultrasound Obstet Gynecol* 2014;43(4):437-43.
6. AIUM practice parameter for the performance of neurosonography in neonates and infants. *J Ultrasound Med* 2020;39(5):E57-E61.
7. Guillot M, Chau V, Lemyre B. Routine imaging of the preterm neonatal brain. *Paediatr Child Health* 2020;25(4):249-55.
8. Papile LA, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: A study of infants with birth weights less than 1,500 gm. *J Pediatr* 1978;92(4):529-34.
9. Fitts PM, Posner MI. Human performance. Oxford, England: Brooks/ Cole Pub.Co.;1967.
10. Taylor JA, Ivry RB. The role of strategies in motor learning. *Ann N Y Acad Sci* 2012;1251:1-12.
11. Caro-Domínguez P, Lecacheux C, Hernandez-Herrera C, Llorens-Salvador R. Cranial ultrasound for beginners. *Transl Pediatr.* 2021 Apr;10(4) : 1117- 1137. doi: 10.21037/tp-20-399.
12. Lowe LH, Bailey Z. State-of-the-Art cranial sonography: part 1, modern techniques and image interpretation. *Am J Roentgenol* 2011;196(5):1028-33.

13. Ecury-Goossen GM, Camfferman FA, Leijser LM, Govaert P, Dudink J. State of the art cranial ultrasound imaging in neonates. *J Vis Exp JoVE* 2015;2(96):e52238. doi:10.3791/52238.
14. Özek E, Kersin SG. Intraventricular hemorrhage in preterm babies. *Turk Pediatri Ars* 2020.
15. Lowe LH, Bailey Z. State-of-the-art cranial sonography: part 2, pitfalls and variants. *Am J Roentgenol* 2011;196(5):1034–9.
16. de Vries LS, Eken P, Dubowitz LMS. The spectrum of leukomalacia using cranial ultrasound. *Behav Brain Res* 1992;49(1):1–6.
17. Sood R, Singh T. Assessment in medical education: evolving perspectives and contemporary trends. *Natl Med J India* 2012;25(6):357-64.
18. Elshama SS. How to use and apply assessment tools in medical education? *Iberoamericanjm* 2020;04:351-9.
19. Teaching and learning in higher education [Internet]. [place unknown: publisher unknown] . Assessment strategies [cited 2023 May 4] ; Available from: <https://www.queensu.ca/teachingandlearning/modules/index.html>.
20. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of educational objectives: The classification of educational goals. Vol. Handbook I: Cognitive domain. 1956, New York: David McKay Company.
21. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65(9 suppl):S63-S67.
22. Norcini J and Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher* 2007;29:855–71.
23. Sim JH, Aziz YFA, Mansor A, Vijayananthan A, Foong CC, Vadivelu J. Students' performance in the different clinical skills assessed in OSCE: what does it reveal? *Med Educ Online* 2015;20(1):26185.
24. Syed AT. Assessment methods in medical education. *Int J Health Sci* 2008;2(2):3-7.
25. Epstein RM. Assessment in medical education. *N Engl J Med* 2007;356(4):387-96.

26. McCracken DJ, Laursen CB, Barker G, Gleeson FV, Cullen KM, Rahman NM. Thoracic ultrasound competence for ultrasound-guided pleural procedures. *Eur Respir Rev* 2019;28(154):190090.
27. Al-Wardy NM. Assessment methods in undergraduate medical education. *Sultan Qaboos Univ Med J* 2010;10(2):203-9.
28. Kwong JSW, Chen H, Sun X. Development of evidence-based recommendations: implications for preparing expert consensus statements. *Chin Med J* 2016;129(24):2998-3000.
29. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and nominal group in medical education research. *Med Teach* 2017;39(1):14-9.
30. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm* 2016;38:655-62.
31. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;74(9):979-83.
32. Davoudi M, Colt HG, Osann KE, Lamb CR, Mullon JJ. Endobronchial ultrasound skills and tasks assessment tool: assessing the validity evidence for a test of endobronchial ultrasound-guided transbronchial needle aspiration operator skill. *Am J Respir Crit Care Med* 2012;186(8):773-9.
33. Alotaibi G, Youssef A. Development of an assessment tool to measure students' perceptions of respiratory care education programs: Item generation, item reduction, and preliminary validation. *J Fam Community Med* 2013;20(2):116-22.
34. Tolsgaard MG, Todsen T, Sorensen JL, Ringsted C, Lorentzen T, Ottesen B, et al. International multispecialty consensus on how to evaluate ultrasound competence: A Delphi consensus survey. *PLoS ONE* 2013;8(2).
35. Patrawalla P, Eisen LA, Shiloh A, Shah BJ, Savenkov O, Wise W, et al. Development and validation of an assessment tool for competency in critical care ultrasound. *J Grad Med Educ* 2015;7(4):567-73.
36. Janke K, Kelley K, Sweet B, Kuba SE. A modified Delphi process to define competencies for assessment leads supporting a doctor of pharmacy program. *Am J Pharm Educ* 2016;80(10):167.

37. Millington SJ, Arntfield RT, Guo RJ, Koenig S, Kory P, Noble V, et al. The assessment of competency in thoracic sonography (ACTS) scale: validation of a tool for point-of-care ultrasound. *Crit Ultrasound J* 2017;9(1):25.
38. Damewood SC, Leo M, Bailitz J, Gottlieb M, Liu R, Hoffmann B, et al. Tools for measuring clinical ultrasound competency: recommendations from the ultrasound competency work group. *AEM Educ Train* 2020;4(S1):S106-12.
39. Adamson R, Morris AE, Sun Woan J, Ma IWY, Schnobrich D, Soni NJ. Development of a focused cardiac ultrasound image acquisition assessment tool. *Sch* 2020;1(3):260–77.
40. Kim C-H, Yeo K. Beyond consensus: a review of Delphi research published in Malaysian social science journals. *International Journal of Business and Society* 2018;19:S2:312-23.
41. Turoff M, Linstone HA. The Delphi method: techniques and applications. 18. Addison-Wesley Publishing Company;2002.
42. Rowe G, Wright G, Bolger F. The Delphi technique: a re-evaluation of research and theory. *ORG Groups Teams Top*. 2011.
43. Skulmoski GT, Hartman F, Krahn J. The Delphi method for graduate research. *J Inf Technol Educ Res* 2007;6:001-21.
44. Allen D, Tanner K. Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sci Educ* 2006;5(3):197-203.
45. Assessment and curriculum support center [Internet]. [place unknown: publisher unknown] . Creating and using rubrics [cited 2023 May 5] ; Available from: <https://manoa.hawaii.edu/assessment/resources/creating-and-using-rubrics/>
46. Mertler CA. Designing scoring rubrics for your classroom. *Practical assessment, research, and evaluation*. 2000-2001;7.
47. Grote D. A survival guide for managers. Question and answer book. 2002. AMACOM, NY.
48. Chyung SY, Roberts K, Swanson I, Hankinson A. Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance improvement* 2017;56(10):15-23.

49. Assessment for learning [Internet] . [place unknown: publisher unknown] . Advantages and disadvantages of rubrics in assessment [cited 2023 May 8] ; Available from: <https://educationsummary.com/lesson/advantages-and-disadvantages-of-rubrics-in-assessment/>
50. Heale R, Twycross A. Validity and reliability in quantitative research. *Evid Based Nurs* 2015;18:66-7.
51. Golafshani N. Understanding reliability and validity in qualitative research. *The qualitative report* 2003;8(4):597-607.
52. Sullivan GM. A primer on the validity of assessment instruments. *J Grad Med Educ* 2011.
53. Harvey ND. A simple guide to inter-rater, intra-rater and test-retest reliability for animal behaviour studies. *OSF Preprints 8stpy*, Center for Open Science 2021.
54. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20(1):37-46.
55. Mohamad AB, Nurakmal B. Guidelines of the minimum sample size requirements for Cohen's Kappa. *Epidemiology Biostatistics and Public Health* 2017;14.
56. Dettori JR, Norvell DC. Kappa and beyond: Is there agreement? *Global Spine J* 2020;10(4):499-501.
57. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. *Fam Med* 2005;37(5):360-3.
58. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica* 2012;22(3):276-82.
59. Ishak NM, Bakar AYA. Qualitative data management and analysis using NVivo: An approach used to examine leadership qualities among student leaders. *Education Research Journal* 2012;2(3):94-103.
60. Tolsgaard MG, Ringsted C, Dreisler E, Nørgaard LN, Petersen JH, Madsen ME, et al. Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial. *Ultrasound Obstet Gynecol* 2015;46(3):312-8.
61. Reproductive health access project. Basic ultrasound skills evaluation [Internet]. [Reproductiveaccess.org](https://reproductiveaccess.org);2014 [cited 2020 Apr 6]. Available from:

https://www.reproductiveaccess.org/wp-content/uploads/2014/12/basic_ultrasound_skills_evaluation.pdf

62. Shafqat A, Rafi M, Thanawala V, Bedford NM, Hardman JG, McCahon RA. Validity and reliability of an objective structured assessment tool for performance of ultrasound-guided regional anaesthesia. *Br J Anaesth* 2018;121(4):867-75.
63. Skaarup SH, Laursen CB, Bjerrum AS, Hilberg O. Objective and structured assessment of lung ultrasound competence. A multispecialty Delphi consensus and construct validity study. *Ann Am Thorac Soc* 2017;14(4):555-60.
64. Schmidt JN, Kendall J, Smalley C. Competency assessment in senior emergency medicine residents for core ultrasound skills. *West J Emerg Med* 2015;16(6):923-6.
65. Rice J, Crichlow A, Baker M, Regan L, Dodson A, Hsieh Y-H, et al. An assessment tool for the placement of ultrasound-guided peripheral intravenous access. *J Grad Med Educ* 2016;8(2):202-7.
66. Kahr Rasmussen N, Nayahangan LJ, Carlsen J, Ekberg O, Brabrand K, Albrecht-Beste E, et al. Evaluation of competence in ultrasound-guided procedures—a generic assessment tool developed through the Delphi method. *Eur Radiol* 2020;17.
67. Rebel A, Srour H, DiLorenzo A, Nguyen D, Ferrell S, Dwarakanatli S, et al. Ultrasound skill and application of knowledge assessment using an innovative OSCE competition- based simulation approach. *J Educ Perioper Med JEPM* 2016;18(1):E404.
68. Di Pietro S, Mascolo M, Falaschi F, Brambilla W, Ruzga R, Mongodi S, et al. Lung-ultrasound objective structured assessment of technical skills (LUS-OSAUS): utility in the assessment of lung- ultrasound trained medical undergraduates. *J Ultrasound* 2020;1-9.
69. Gaudet J, Waechter J, McLaughlin K, Ferland A, Godinez T, Bands C, et al. Focused critical care echocardiography: development and evaluation of an image acquisition assessment tool. *Crit Care Med* 2016;44(6):e329.
70. Laurent DAB, Niazi AU, Cunningham MS, Jaeger M, Abbas S, McVicar J, et al. A valid and reliable assessment tool for remote simulation- based ultrasound- guided regional anesthesia: *Reg Anesth Pain Med* 2014;39(6):496-501.

71. Wani S, Keswani RN, Han S, Aagaard EM, Hall M, Simon V, et al. Competence in endoscopic ultrasound and endoscopic retrograde cholangiopancreatography, from training through independent practice. *Gastroenterology* 2018;155(5):1483-94.
72. Ziesmann MT, Park J, Unger BJ, Kirkpatrick AW, Vergis A, Logsetty S, et al. Validation of the quality of ultrasound imaging and competence (QUICK) score as an objective assessment tool for the FAST examination. *J Trauma Acute Care Surg* 2015;78(5):1008-13.
73. Hofer M, Kamper L, Sadlo M, Sievers K, Heussen N. Evaluation of an OSCE assessment tool for abdominal ultrasound courses. *Ultraschall Med- Eur J Ultrasound* 2011;32(02):184–90.
74. Jaffer U, Singh P, Pandey VA, Aslam M, Standfield NJ. Validation of a novel duplex ultrasound objective structured assessment of technical skills (DUOSATS) for arterial stenosis detection. *Heart Lung Vessels* 2014;6(2):92-104.
75. Jaffer U, Normahani P, Lackenby K, Aslam M, Standfield NJ. Validation of a novel venous duplex ultrasound objective structured assessment of technical skills for the assessment of venous reflux. *J Surg Educ* 2015;72(4):754-60.
76. Chuan A, Thillainathan S, Graham PL, Jolly B, Wong DM, Smith N, et al. Reliability of the direct observation of procedural skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesth Intensive Care* 2016;44(2):201-6.
77. Limchareon S, Kongpromsuk S. A single static breast model education of ultrasound skill in final year medical students of Burapha university. *J Med Ultrasound* 2018;26(3):143-6.
78. Black H, Sheppard G, Metcalfe B, Stone-McLean J, McCarthy H, Dubrowski A. Expert facilitated development of an objective assessment tool for point-of-care ultrasound performance in undergraduate medical education. *Cureus* 2016;8(6):e636.
79. Sultan SF, Iohom G, Saunders J, Shorten G. A clinical assessment tool for ultrasound-guided axillary brachial plexus block: assessment tool for axillary block. *Acta Anaesthesiol Scand* 2012;56(5):616-23.

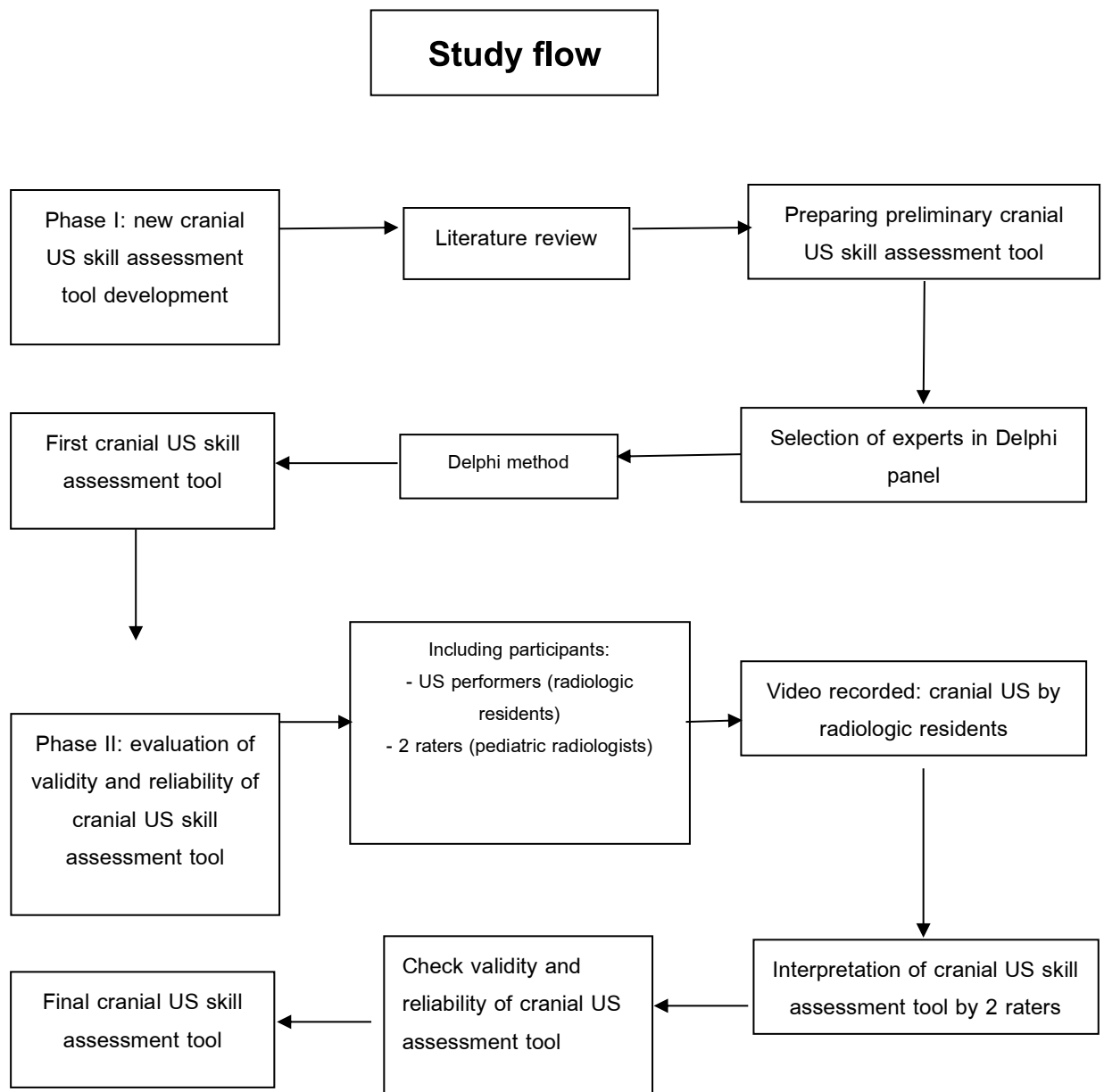
80. Ma IWY, Desy J, Woo MY, Kirkpatrick AW, Noble VE. Consensus-based expert development of critical items for direct observation of point-of-care ultrasound skills. *J Grad Med Educ* 2020;12(2):176-84.
81. Cheung JJH, Chen EW, Darani R, McCartney CJL, Dubrowski A, Awad IT. The creation of an objective assessment tool for ultrasound-guided regional anesthesia using the Delphi method: *Reg Anesth Pain Med* 2012;37(3):329-33.
82. Shahab N, Mahmood E, Huiwen L, Istvan S, Bijan S. Development and validation of an automated assessment tool of echocardiography skills acquired on a neonatal simulator. *Echocardiography* 2020.
83. Le CK, Lewis J, Steinmetz P, Dyachenko A, Oleskevich S. The use of ultrasound simulators to strengthen scanning skills in medical students: a randomized controlled trial. *J Ultrasound Med* 2019;38(5):1249-57.
84. Desy J, Noble VE, Woo MY, Walsh M, Kirkpatrick AW, Ma IWY. Use of critical items in determining point-of-care ultrasound competence. *Eval Health Prof* 2020;163278720975833.
85. Enquobahrie A, Horvath S, Arikatla S, Rosenberg A, Cleary K, Sharma K. Development and face validation of ultrasound-guided renal biopsy virtual trainer. *Healthc Technol Lett* 2019;6(6):210-3.
86. Ng DL, Vuhahula E, Zhang L, Waterhouse EG, White KL, Mushi BP, et al. Efficacy of an intensive, ultrasound-guided fine-needle aspiration biopsy training workshop in Tanzania. *J Glob Oncol* 2018;4:1-9.
87. Todsén T, Melchior J, Charabi B, Henriksen B, Ringsted C, Konge L, et al. Competency-based assessment in surgeon-performed head and neck ultrasonography: a validity study. *The Laryngoscope* 2018;128(6):1346-52.
88. Han C, Nie C, Shen X, Xu T, Liu J, Ding Z, et al. Exploration of an effective training system for the diagnosis of pancreatobiliary diseases with EUS: A prospective study. *Endosc Ultrasound* 2020;9(5):308.
89. Kunz D, Pariyadath M, Wittler M, Askew K, Manthey D, Hartman N. Derivation of a performance checklist for ultrasound-guided arthrocentesis using the modified Delphi method. *J Ultrasound Med* 2017;36(6):1147-52.

90. Chory MK, Schnettler WT, March M, Hacker MR, Modest AM, Rodriguez D. ACES: accurate cervical evaluation with sonography. *J Ultrasound Med* 2016;35(1):25-8.
91. Heinzow HS, Friederichs H, Lenz P, Schmedt A, Becker JC, Hengst K, et al. Teaching ultrasound in a curricular course according to certified EFSUMB standards during undergraduate medical education: a prospective study. *BMC Med Educ* 2013;13:84.
92. Chin KJ, Tse C, Chan V, Tan JS, Lupu CM, Hayter M. Hand motion analysis using the Imperial college surgical assessment device: validation of a novel and objective performance measure in ultrasound-guided peripheral nerve blockade. *Reg Anesth Pain Med* 2011;36(3):213-9.
93. Starkie T, Drake EJ. Assessment of procedural skills training and performance in anesthesia using cumulative sum analysis (cusum). *Can J Anesth* 2013;60(12):1228-39.
94. Mariano ER, Harrison TK, Kim TE, Kan J, Shum C, Gaba DM, et al. Evaluation of a standardized program for training practicing anesthesiologists in ultrasound-guided regional anesthesia skills. *J Ultrasound Med* 2015;34(10):1883-93.
95. Wong DM, Watson MJ, Kluger R, Chuan A, Herrick MD, Ng I, et al. Evaluation of a task-specific checklist and global rating scale for ultrasound-guided regional anesthesia: *Reg Anesth Pain Med* 2014;39(5):399-408.
96. Hsi RS, Dunmire B, Cunitz BW, He X, Sorensen MD, Harper JD, et al. Content and face validation of a curriculum for ultrasonic propulsion of calculi in a human renal model. *J Endourol* 2014;28(4):459-63.
97. Meneses EL, Olivencia JLL, Vázquez-Cano E. The impact of personal interactions on the stress of school administrators: the validation and application of an assessment tool. *Intang Cap* 2017;13(3):499-522.
98. Soleimani F, Torkzahrani S, Rafiey H, Salavati M, Nasiri M. Development and psychometric testing of a scale for the assessment of the quality of developmental care in neonatal intensive care units in Iran. *Electron Physician* 2016;8(1):1686-92.
99. Maria TM, Jose M OM, Eva OM. Educational environments with cultural and religious diversity: psychometric analysis of the cyberbullying scale. *Religions* 2019;10(7):443.

100. De Silva Weliange SH, Fernando D, Gunatilake J. Development and validation of a tool to assess the physical and social environment associated with physical activity among adults in Sri Lanka. *BMC Public Health* 2014;14:423.
101. Klemenc-Ketiš Z, Kersnik J. New virtual case-based assessment method for decision making in undergraduate students: a scale development and validation. *BMC Med Educ* 2013;13:160.
102. Flack VF, Afifi AA, Lachenbruch PA, Schouten HJA. Sample size determinations for the two rater kappa statistic. *Psychometrika* 1988;53:321-325.

Appendices

Appendix A: Study flow



Appendix B: Time schedule of the study

Activity	Oct 2019 to Jan 2020	Feb 2020 to Jun 2020	Jul 2020 to Dec 2020	Jan 2021 to Jan2023
1. Proposal submission for EC				
2. Preparatory phase				
3. Data collection				
4. Data analysis				
5. Manuscript writing				
6. Submit research				

Appendix C: Record forms

Record form for participated experts

Number E _____

Gender	1. Male 2. Female	Egender []
Position in pediatric radiologic view	1. Instructor 2. Associated Professor 3. Assistance Professor 4. Professor	Epos []
Workplace	1. University hospital 2. Pediatric hospital 3. General hospital	Ework []
Years of experience on cranial US	1. 3-5 years 2. 5-10 years 3. > 10 years	Eyear []

Record form for participated radiologic residents

Number R _____

Gender	1. Male 2. Female	Rgender []
Level of training	1. 1 st year radiologic residents (PGY-1) 2. 2 nd year radiologic residents (PGY-2) 3. 3 rd year radiologic residents (PGY-3) 4. Others	Rposition []
Experience in cranial US training	1. No If yes, please identified method of training (can choose more than 1 choice) 2. Yes, lecture 3. Yes, observation 4. Yes, workshop 5. Yes, self-learning from other sources 6. Yes, others _____	Rexptrain []
Experience in performed cranial US	_____ cases 1. 0 case 2. < 5 cases 3. 5-10 cases 4. 11-15 cases 5. 16-20 cases 6. > 20 cases	Rexpcase []

Appendix D: Initial cranial US skill assessment form (5-points rating scale for Delphi method)

Domains and items	Score				
	1	2	3	4	5
Motor skills Applying US equipment 1. Proper positioning of US machine 2. Selecting appropriate probes 3. Correct handling of probe 4. Applying of warm gel to probe 5. Using proper buttons on US machine 6. Adjust gain and frequency, correctly 7. Adjust depth and focus to include whole brain and appropriate area of focus 8. Cleaned probe after US					
Visual-cognitive skills Systemic cranial US examination 9. Correct placement of probe at anterior fontanel to start US 10. Confirmation side of brain orientation 11. Displaying cranial US approach systematically 12. Presenting of relevant brain structures according to standard protocol 13. Identified the positive findings (e.g. IVH, PVL, hydrocephalus, or anomaly) 14. Overall performance					

Appendix E: Initial cranial US skill assessment form with rubric score, example

Domains and items	1 not obtained	2 critical error	3 sub- optimal	4 optimal
Proper positioning of US machine	wrong place		right hand close to the patient	
Selecting appropriate probes	choose other prove		small curvilinear 8.5 MHz probe	
Correct handling of probe				
Applying of warm gel to probe	do not use gel	put gel to head	put gel to prove	put gel to probe with optimal amount
Using proper buttons on US machine	can not operate	operate with some experience	operate with experience but inconsistency	familiar to operate
Adjust gain and frequency, correctly	can not operate	wrong	inappropriate	appropriate
Adjust depth and focus to include whole brain and appropriate area of focus	can not operate	include partial brain	include whole brain but inappropriate depth and focus	optimum FOV, depth, focus
Cleaned probe after US	do not done		done	
Correct placement of probe at anterior fontanel to start US	can not done	place in other area	place in anterior fontanel with poor contact	place in anterior fontanel with good contact
Confirmation side of brain orientation	do not done	wrong side	confirm several times	correct side with confidence
Displaying cranial US approach systematically	can not done	unsystematic display	some systematic display	consistently systematic display
Presenting of relevant brain structures according to standard protocol	can not done	presence of some images	inconsistency	follow standard protocol
Identified the positive findings (e.g. IVH, PVL, hydrocephalus, or anomaly)	can not done	miss positive findings	identified some findings	completely identified positive findings
Image recording per standard protocol	can not done	record some images	complete all images	complete with good quality images
Overall performance	fail		pass	

Appendix F: Dummy table 1: Predictive factors information

Predictive factors	Number (%)	Mean
Places: NICU US room		
Subjects Age (months) Intubation status: Yes No		
Trainee PGY: Year 1 Year 2 Year 3 Number of US brain before video record (cases) Cranial US training: Yes No		
Time: Video duration (minutes) Time to complete assessment (minutes)		

Appendix G: Dummy table 2: Summary DOPS scores and interrater reliability

DOPS items	4-point scale: Score mean (actual)	Kappa	Percent agreement
1. Proper positioning of US machine 2. Selecting appropriate probes (small curvilinear 8.5 MHz probe) 3. Correct handling of probe 4. Applying of warm gel to probe 5. Using proper buttons on US machine 6. Adjust gain and frequency, correctly 7. Adjust depth and focus to include whole brain and appropriate area of focus 8. Correct placement of probe at anterior fontanel to start US 9. Confirmation side of brain orientation 10. Displaying cranial US approach systematically 11. Presenting of relevant brain structures according to standard protocol 12. Identified the positive findings (eg. IVH, PVL, hydrocephalus, or anomaly) 13. Image recording per standard protocol 14. Overall performance (fail/pass)			

Appendix H: Item generation

All 35 articles, details are described in table and examples are shown below.

Articles	Total items	Relevant/Not relevant items	Detail of relevant items
1. EBUS-STAT 10 points assessment tool ³²	10	1/9	- US obtained without artifacts
2. LUS-OSAUS ⁶³	17	4/13	- Correct choice of transducer - Correct depth - Correct gain - Correct handling of transducer
3. POC US exam type and area of evaluation scored as one point each for OSCE ⁶⁴	28	5/23	- Correct transducer selection x5
4. US-Guided Peripheral Intravenous Access Checklist ⁶⁵	10	1/9	- Turn on US
5. IUSE ⁶⁶	12	2/10	- US equipment optimization of images (choice of transducer and pre-settings and use of knobology including Doppler) - Hand-eye coordination
6. US Olympics exam station 3 (TTE anatomy) ⁶⁷	14	10/4	- Selects appropriate orientation x2 - Selects appropriate depth x2 - Selects appropriate gain x2 - Overall picture quality x2 - Identifies target on Screen x2
7. LUS-OSAUS ⁶⁸	17	4/13	Same as number 2.
8. Focused critical care echocardiography image	50	20/30	- Appropriate depth x6 - Appropriate gain x6

Articles	Total items	Relevant/Not relevant items	Detail of relevant items
acquisition assessment tool ⁶⁹			- Appropriate depth and gain x2 - Structures centered in view x6
9. Global rating scale for US-guided regional anesthesia ⁷⁰	9	2/7	- Instrument handling - Overall performance
10. Quality indication in EUS and ERCP ⁷¹	12	0/12	-
11. QUICK score ⁷²	7	1/6	- Sets gain appropriately
12. OSCE: retroperitoneal space with aorta ⁷³	12	5/7	- Orientation - Positioning - Connection - Adequate zoom factors - Overall performance
13. DUOSATS ⁷⁴	11	3/8	- Transducer selection - Ultrasound coupling gel usage - Acquires images in B mode (Appropriate depth of field, TGC, and focus setting, appropriate gain control?)
14. V-DUOSATS ⁷⁵	8	3/5	- Transducer selection - Ultrasound coupling gel usage - Image optimization in B mode-
15. DOPS item ⁷⁶	8	2/6	- Technical ability - Overall performance
16. Technical US skills checklist ⁷⁷	7	6/1	- Selects correct ultrasound probe - Applies gel on probe or on the patient

Articles	Total items	Relevant/Not relevant items	Detail of relevant items
			- Correctly identifies probe orientation - Adjust image depth or contrast to optimize view of lesion - Correctly identifies lesion - Saves and stores the image
17. Global rating scales ⁷⁸	9	6/3	- Preparation for procedure (e.g. machine, machine placement, US gel, towels) - Image optimization (probe choice, optimization of gain and depth) - Probe technique - Image interpretation - Knowledge of procedure (steps, familiarity, and confidence) - Overall performance
18. Generic technical skills global rating scale ⁷⁹	8	5/3	- Time and motion - Instrument handling - Flow of procedure - Knowledge of procedure - Overall performance
19. Items assessment tool ⁸⁰	31	14/17	- Appropriately positions the machine - Appropriately applies basic knobology (e.g. on/off, depth, gain)

Articles	Total items	Relevant/Not relevant items	Detail of relevant items
			- Appropriately uses examination presets - Chooses correct transducer - Appropriately enters patient identifier - Able to store relevant images and clips - Appropriately cleans machine and transducers - Conducts the appropriate types of scans - Conducts scans in the appropriate prioritization/sequence - Scans adequately through the entire area of interest - Able to optimize image appropriately when necessary - Adjusts focal zone appropriately - Scans with efficiency of hand motion - Able to recognize key findings
20. Assessment checklist for ultrasound-guided regional anesthesia and Global rating scale ⁸¹	31	11/20	- Correct placement of ultrasound machine relative to patient to allow easy visualization of both - Choice of correct transducer

Articles	Total items	Relevant/Not relevant items	Detail of relevant items
			- Correct depth, gain, and focal zone choices - Holds the probe appropriately (3 fingers holding the probe and 1 finger touching the patient) - Knowledge or confirmation of screen orientation (ie, which side of probe corresponds to which side of screen) - Scanning of anatomy and proper identification of target - Time and motion - Instrument handling - Flow of procedure - Knowledge of procedure - Overall performance
21. Scoring system for automate evaluation echocardiography method ⁸²	6	0/6	-
22. Practical examination score sheet for heart, lung, IVC US ⁸³	21	0/21	-
23. Technical skill for ultrasound-guided FNA ⁸⁶	11	3/8	- Satisfactory position of hands - Satisfactory position of transducer - Satisfactory position of lesion on screen

Articles	Total items	Relevant/Not relevant items	Detail of relevant items
24. OSAUS items ⁸⁷	5	5/0	- Applied knowledge of US equipment - Image optimization - Systematic examination - Interpretation of images - Documentation of examination
25. The EUS skills assessment tool ⁸⁸	17	1/16	- Identify lesion of interest
26. Ultrasound-guided arthrocentesis checklist ⁸⁹	25	1/24	- Obtains correct equipment
27. Items worksheet for cervical evaluation US ⁹⁰	10	0/10	-
28. DOPS for US ⁹¹	7	3/4	- Identification of organ structures - Investigation of organ structures - Technical ability
29. Task-specific checklist for USG supraclavicular block ⁹²	30	6/24	- Positions US machine appropriately - Choose appropriate probe - Checks probe orientation - Selects appropriate depth - Selects appropriate gain - Handles probe with control
30. Quantitative assessment of procedural performance ⁹⁴	15	4/11	- Equipment preparation - Probe stability - Visual focus - Position of US machine

Articles	Total items	Relevant/Not relevant items	Detail of relevant items
31. Global rating scale for US-guided regional anesthesia ⁹⁵	10	3/7	- Preparation for procedure - Instrument handling - Overall performance
32. Cognitive and technical skill assessment for renal US in human renal model ⁹⁶	14	2/12	- Optimized adjustment of US controls - Correct transducer orientation

1. The Endobronchial Ultrasound Skills and Tasks Assessment Tool³²**EBUS-STAT 10 Point Assessment Tool**

Learner: _____ Year of Training _____
 Faculty _____ Date _____

Educational Item* Items 1-10 each scored separately	Satisfactory Yes/No
1. Able to maneuver the scope through upper airway into trachea, without trauma or difficulty (5 points for single item tested) ☐ Mouth and Vocal cords ☐ ET Tube ☐ Laryngeal mask airway	Yes / No Score ____/5
2. Able to maneuver scope using white light bronchoscopy within tracheobronchial tree without trauma (4 points, no partial points) ☐ Scope centered in airway lumen avoiding airway wall trauma	Yes / No Score ____/4
3. Ultrasound image obtained without artifacts (5 points, no partial points) ☐ Absence of artifacts on image, any target	Yes / No Score ____/5
4. Identify major mediastinal vascular structures (4 points per item) ☐ Aorta ☐ Pulmonary artery ☐ Superior vena cava ☐ Azygos vein ☐ Left atrium	Yes / No Score ____/20
5. Identify lymph node station (Select 3 targets, 5 points each) ☐ 2R ☐ 2L ☐ 4R ☐ 10R ☐ 7 ☐ 4L ☐ 10L ☐ 11L ☐ 11Rs ☐ 11Ri	Yes / No Score ____/15
6. Able to operate EBUS processor (2 points each item) ☐ Gain ☐ Depth ☐ Doppler	Yes / No Score ____/6
7. Performance of EBUS-TBNA (1 point each, target 15 points) ☐ Advance needle through working channel (neutral position) ☐ Secure needle housing by sliding the flange ☐ Release sheath screw ☐ Advance and lock sheath when it touches wall ☐ Release needle screw ☐ Advance needle using jab technique ☐ Visualize needle entering target node ☐ Move stylet in and out a few times ☐ Remove stylet ☐ Attach syringe ☐ Apply suction ☐ Pass needle in and out of node 10-15 times ☐ Release suction ☐ Retract needle into sheath ☐ Unlock and remove needle and sheath	Yes / No Score ____/15
8. Image analysis: CT scans (1 point each, target 10 points) ☐ Image 1 ☐ Image 2 ☐ Image 3 ☐ Image 4 ☐ Image 5 ☐ Image 6 ☐ Image 7 ☐ Image 8 ☐ Image 9 ☐ Image 10	Yes / No Score ____/10
9. Image analysis: EBUS views (1 point each, target 10 points) ☐ Image 1 ☐ Image 2 ☐ Image 3 ☐ Image 4 ☐ Image 5 ☐ Image 6 ☐ Image 7 ☐ Image 8 ☐ Image 9 ☐ Image 10	Yes / No Score ____/10
10. Decision-making tasks: (2 points each, target 10 points) ☐ Image 1 ☐ Image 2 ☐ Image 3 ☐ Image 4 ☐ Image 5	Yes / No Score ____/10

* The combined use of the 10 items tests competencies needed to climb the learning curve from novice to advanced beginner to intermediate to competent bronchoscopist able to independently perform EBUS-TBNA.

FINAL GRADE **PASS** **FAIL** **SCORE** ____/100

2. Lung ultrasound objective structured assessment of technical skills tool in its final form⁶³

	1	2	3	4	5
Indication					
Evaluates the indication for lung ultrasound	None		Some		Sufficient
Suggests focused questions that can be examined by lung ultrasound	None		Some		Sufficient
Systematic lung ultrasound examination					
Performs lung ultrasound systematically	No systematic approach		Some systematic approach		Sufficient systematic approach
Performs lung ultrasound on the basis of focused questions	No correlation between focused questions and scanning		Some correlation between focused questions and scanning		Sufficient correlation between focused questions and scanning
Technical skills					
Correct placement of patient (e.g., supine when scanning for pneumothorax)	Wrong placement (e.g., evaluating pleural effusion with patient in Trendelenburg position)				Optimal placement (e.g., evaluating pleural effusion with patient in seated position)
Correct choice of transducer	Wrong choice of transducer				Optimal choice of transducer
Correct depth	Wrong depth setting				Optimal depth setting
Correct gain	Wrong setting of gain				Optimal setting of gain
Correct handling of transducer	Poor transducer handling				Optimal transducer handling
Findings					
Correct assessment of pleura	Not able to assess correctly		Properly assessed sometimes		Properly assessed every time
Correct assessment of B-lines	Not able to assess correctly		Properly assessed sometimes		Properly assessed every time
Correct assessment of consolidations	Not able to assess correctly		Properly assessed sometimes		Properly assessed every time
Correct assessment of pleural effusion	Not able to assess correctly		Properly assessed sometime		Properly assessed every time
Correct assessment of whether ultrasound-guided thoracentesis is safe	Not able to assess correctly		Proper assessed sometimes		Properly assessed every time
Documentation					
Documents findings in patient's chart	None		Main findings are described		Described sufficiently
Conclusion					
Able to make a diagnosis on the basis of lung ultrasound findings	No diagnosis		Some diagnosis		Correct diagnosis
Able to integrate lung ultrasound findings with patient's history	No integration		Some integration		Optimal integration

*Operators were scored 1–5 for each element, and the final rating is the sum of all 17 elements.

3. Point-of-care ultrasound exam type and area of evaluation scored as one point each for objective structured clinical examination⁶⁴

Exam type	Area of evaluation
ECHO (15 points)	Correct transducer selection Identifies parasternal long axis view with RV, LA, LV Measures aortic outflow tract Identifies parasternal short axis view with RV, LV Identifies apical four-chamber view with RV, LV, RA, LA
Aorta (18 points)	Correct transducer selection Identifies target and inferior vena cava by compression and/or doppler Obtains trans. proximal, middle, distal and bifurcation views Obtains longitudinal aorta view Performs correct measurement of aorta Identifies vertebral body
FAST (11 points)	Correct transducer selection Subxiphoid view Identifies right upper quadrant with Morisons, tip of liver, inf. pole kidney Identifies left upper quadrant view Identifies the splenorenal recess Visualizes the inferior pole of the left kidney Identifies pelvic view in transverse and sagittal planes
Pelvic (12 points)	Correct transducer selection
Trans-abdominal	Obtains uterus view in long and short axis Scans through bilateral ovaries in two planes Explains how to calculate the fetal heart rate with M-mode
Trans-vaginal	Obtains coronal and sagittal uterus view
Central line (7 points)	Correct transducer selection Explains probe positioning/marker orientation Identifies target and associated artery Measures depth of vein Appropriately demonstrates needle entry/ angle of insertion

ECHO, echocardiogram; *RV*, right ventricle; *LA*, left atrium; *LV*, left ventricle; *RA*, right atrium; *FAST*, focused assessment with sonography for trauma

4. Ultrasound-Guided Peripheral Intravenous Access Checklist⁶⁵

Step	Attempt 1	Attempt 2	Attempt 3
Wash hands, apply gloves			
Tourniquet placed			
Turn on ultrasound			
Locate vessel with linear transducer			
Demonstrate vessel is venous			
Disinfect area			
Insert IV and follow needle tip to vessel			
Advance catheter and demonstrate placement with draw back 1 cc			
Flush 3 to 5 cc saline			
Secure IV			
Score			

5. Interventional Ultrasound Skills Evaluation (IUSE) tool⁶⁶

	1	2	3	4	5
1. Indications Assessment of the indications for the procedure based on patient history, referral information, and prior imaging	Not able to explain the indications		Partially explains the indications		Clearly explains the indications
2. Contraindications Handling of existing relative and absolute contraindications (allergies, medication, and blood tests) in accordance with guidelines	Does not check contraindications OR does not act accordingly		Checks and handles contraindications suboptimally but without compromising patient safety		Checks and clearly explains all contraindications and acts accordingly
3. Preparation Planning of the procedural strategy based on available prior imaging (e.g., CT scan, MRI or ultrasound)	Plans an unsafe strategy OR does not use available prior imaging		Plans a suboptimal but safe strategy based on available prior imaging		Plans and clearly explains an optimal strategy based on available prior imaging
4. Analgesia Administration of local anesthesia	Administers insufficient local anesthesia using incorrect AND/OR non-sterile technique		Administers local anesthesia in a suboptimal way		Administers sufficient local anesthesia using the correct technique
5. US equipment Optimization of images (choice of transducer and pre-settings and use of knobology including Doppler)	Acquires inadequate images thereby compromising safety or yield		Acquires suboptimal images without compromising safety or yield		Acquires optimal images consistently
6. Procedural instruments Choice and handling of instruments (e.g., steer, draping, biopsy needle, catheter)	Chooses wrong instruments OR uses them incorrectly, e.g. does not maintain sterility		Chooses correct instruments but handles them insecurely		Chooses correct instruments and handles them optimally

	1	2	3	4	5
7. Route Choice of procedural route	Chooses suboptimal route that compromises safety		Chooses suboptimal route without compromising safety		Chooses optimal route and deliberately avoids critical structures
8. Hand-eye coordi- nation Demonstration of hand-eye coordi- nation and targeting during the procedure	Demonstrates poor hand-eye coordination and targeting		Demonstrates suboptimal hand-eye coordi- nation but suffi- cient targeting		Demonstrates superior hand-eye coordina- tion and precise targeting
9. Procedural outcome Assessment of procedural outcome (e.g., placement of tube, biopsy material, or sufficient drainage of abscess)	Assesses procedural outcome incorrectly OR does not act accordingly		Assesses and acts on procedural outcome suboptimally but without compromising yield		Assesses the procedural outcome optimally and acts accordingly
10. Monitoring Monitoring of the patient during and immediately after the procedure (e.g., blood pressure, ultrasound) and handling of potential complications	Fails to monitor the patient OR handles complications incorrectly		Monitors the patient suboptimally without compromising safety		Monitors the patient optimally and correctly handles complications
11. Patient communication Communication with the patient throughout the procedure (e.g., instruction on breathing and body positioning)	Fails to provide instructions and communi- cates poorly with the patient		Provides insufficient instructions OR communicates with the patient suboptimally		Provides optimal instructions and maintains good communication
12. Teamwork Communication with members of the team	Communicates poorly with the team		Communicates suboptimally with the team		Communicates optimally with the team

6. US Olympics exam station 3 (TTE anatomy)⁶⁷

Ultrasound Olympics Exam Station 3 (TTE Anatomy)

Resident Name:

Evaluated by:

Date:

Training level: ☐ CA1 ☐ CA2 ☐ CA3**Task: Identify Aortic valve in Parasternal short axis view**

US Task		Grade (GRS)	comments
Ultrasound Basic Setup	Selects appropriate orientation		
	Selects appropriate depth		
	Selects appropriate gain		
Ultrasound scanning	Picture shows parasternal short axis view		
	Picture shows left ventricular outflow tract		
	Overall picture quality		
Target Identification	Identifies target on Screen (marked with arrow) Arrow identifies Aortic valve		☐ Success ☐ Fail
	Total points	/35	

Task: Identify Left Ventricle in apical view

US Task		Grade (GRS)	comments
Ultrasound Basic Setup	Selects appropriate orientation		
	Selects appropriate depth		
	Selects appropriate gain		
Ultrasound scanning	Picture shows apical view of the heart		
	Picture shows ventricular structures (right or left)		
	Overall picture quality		
Target Identification	Identifies target on Screen (marked with arrow) Arrow identifies Left ventricle		☐ Success ☐ Fail
	Total points	/35	

Comments:

7. Lung-ultrasound objective structured assessment of technical skills (LUS-OSAUS): same assessment tool to number 2.⁶⁸

8. Focused critical care echocardiography image acquisition assessment tool⁶⁹

SECTION 1: REQUIRED VIEWS		
A. Parasternal Long Axis View		
Structure(s) To Be Imaged	Score	
A1. Aortic valve, LV outflow tract, aortic root?		
A2. Left atrium, mitral valve and leaflets?		
A3. RV and interventricular septum?		
A4. LV endocardial borders and papillary muscles?		
A5. Posterior LV wall and pericardium?		
A6. Structures centered in view?		
A7. Appropriate depth?		
A8. Appropriate gain?		
A9. Section total (12 points)		
B. Parasternal Short Axis View		
Level	Structure(s) To Be Imaged	Score
Aortic Valve	B1. Aortic valve and leaflets?	
	B2. Structures centered in view?	
	B3. Appropriate depth?	
	B4. Appropriate gain?	
Mitral Valve	B5. Mitral valve, leaflets and LV cavity?	
	B6. Interventricular septum?	
	B7. Axis	
	B8. Structures centered in view?	
	B9. Appropriate depth?	
	B10. Appropriate gain?	
Mid Papillary	B11. Papillary muscles	
	B12. Axis	
	B13. Structures centered in view?	
	B14. Appropriate depth?	
	B15. Appropriate gain?	
LV Apex	B16. Ventricular apex?	
	B17. Appropriate depth and gain?	
	B18. Section total (18 points)	
SECTION 2: OVERALL IMAGE DIAGNOSTIC QUALITY		
Images of diagnostic quality for clinical questions?		
DQ1. LV systolic function		
DQ2. LV size		
DQ3. RV systolic function		
DQ4. RV size		
DQ5. Volume status (reflected by IVC measurement)		
DQ6. Presence of significant pericardial effusion		
DQ7. Section total (12 points)		
SCORING DEFINITIONS AND POINT VALUES		
Section 1: Required Views		
Inadequate: image of structures not obtained or images are inadequate for assessment (0 points); Adequate: images obtained are adequate for assessment (1 point); Optimal: images obtained are optimized for assessment (2 points); Centered: structures centered in view (1 point); Depth: appropriate (0.5 points), not appropriate (0 points); Gain: appropriate (0.5 points) not appropriate (0 points); Axis: On (1 point), Off (0 points).		
Section 2: Overall Image Diagnostic Quality		
Non-diagnostic: images are inadequate to permit diagnostic assessment (0 points); Borderline: images are of borderline quality to permit diagnostic assessment (1 point); Diagnostic: images are adequate to permit diagnostic assessment (2 points).		

Figure 1. Focused critical care echocardiography image acquisition assessment tool. LV = left ventricle, RV = right ventricle, IVC = inferior vena cava.

9. Global rating scale for US-guided regional anesthesia⁷⁰

Item	Score			
	1	2	3	5
Preparation for procedure (eg, monitors, intravenous access, ultrasound machine)	Did not organize equipment well. Has to stop procedure frequently to prepare equipment.		Equipment generally organized. Occasionally has to stop to prepare items.	All equipment neatly organized, prepared, and ready for use.
Patient interaction	Little to no rapport established; patient is unaware of procedures. No sedation is provided.		Rapport is generally established; patient is aware and informed of most procedures. Patient anxiety is alleviated adequately with sedative(s).	Strong rapport is established and maintained throughout the procedure. Patient is well informed of all relevant procedures. Patient anxiety is alleviated through sedation and verbal comforting.
Asepsis (eg, use of sterile gloves, proper patient draping, probe sterility, cleansing of skin before infiltration, use of op site)	Practice of proper aseptic technique not generally apparent. Many errors in aseptic technique made throughout the procedure.		Generally practices proper aseptic technique. Occasional errors in aseptic technique made during the procedure.	Excellent demonstrates proper aseptic technique. Few or no errors in aseptic technique made during the procedure.
Respect for tissue	Frequently uses unnecessary force on tissue or causes damage.		Carefully handles tissue but occasionally causes unintentional damage.	Consistently handles tissues appropriately with minimal damage.
Time and motion	Many unnecessary movements.		Efficient time/motion but some unnecessary movements.	Clear economy of movements. Maximum efficiency.
Instrument handling	Repeatedly makes tentative and awkward movements.		Competent with instruments but occasionally makes awkward or stiff movements.	Fluid movements with instruments and no awkwardness.
Flow of procedure	Frequently stops procedure and seems unsure of next move.		Demonstrates some forward planning with reasonable progression of the procedure.	Obviously planned course of the procedure, with effortless flow from one move to the next.
Knowledge of procedure	Deficient knowledge.		Knows all important steps of the procedure.	Demonstrates familiarity with all aspects of the procedure.
Overall performance	Very poor		Competent	Clearly superior

10. Performance of advanced endoscopy trainees in first year of independent practice based on ASGE and American College of Gastroenterology established quality indicators in EUS and ERCP⁷¹

Performance of Advanced Endoscopy Trainees in First Year of Independent Practice Based on ASGE and American College of Gastroenterology Established Quality Indicators in EUS and ERCP (Phase 2)

QI ^d (measure type and performance target)	Procedures, n	Overall AET performance		AETs reaching performance target, n (%)
		n (%)	Range, %	
EUS				
Diagnostic rate of adequate sample in all solid lesions undergoing EUS-FNA (outcome ≥ 85%)	1255	1185 (94.4)	77.1–100	19 (90.5)
Diagnostic rates for malignancy in patients undergoing EUS-FNA of pancreatic masses (outcome ≥70%; priority indicator)	519	435 (83.8)	45–100	17(81)
Incidence of adverse events after EUS-FNA				
Acute pancreatitis (outcome < 2%)	3258	13 (0.4)		NA
Perforation (outcome < 0.5%)	3258	2 (0.06)		NA
Clinically significant bleeding (outcome < 1%)	3258	8(0.25)		NA
ERCP				
Frequency with which deep cannulation of ducts of interest is achieved (process NA)	2668	2532 (94.9)	84–100	
Frequency with which deep cannulation of ducts of interest in patients with native papillae is achieved (process > 90%; priority indicator)	1552	1445 (93.1)	76.5–100	17 (77.3)
Frequency with which common bile duct stones < 1 cm are extracted successfully (outcome ≥ 90%)	1141	1068 (93.6)	62.1–100	18 (81.8)
Frequency with which stent placement for biliary obstruction is successfully achieved (outcome ≥ 90%)	1325	1244 (93.9)	80–100	15 (68.2)
Adverse events				
Rate of post-ERCP pancreatitis (outcome NA; priority indicator)	2673	67 (2.51)		
Rate of perforation (outcome ≤ 0.2%)	2673	9 (0.34)		
Rate of clinically significant hemorrhage (outcome ≤ 1%)	2673	22 (0.82)		

NA, not applicable.

^aBased on the ASGE and American College of Gastroenterology QIs in EUS and ERCP.

11. Example of TSC contents, hepatorenal space scoring⁷²

TABLE 1. Example of TSC Contents, Hepatorenal Space Scoring

Hepatorenal Space	Accomplished 1 = yes 0 = no
Orients image with liver to the left and kidney to the right	
Adjusts depth so the image ends just below the kidney	
Sets gain appropriately	
Visualizes the interface between the liver and kidney clearly	
Visualizes the interface between the liver and kidney in entirety by sweeping through the entire kidney	
Visualizes the caudal tip of the liver clearly	
Total score for hepatorenal space	____/6

12. OSCE: retroperitoneal space with aorta⁷³

Examination Protocol Sono OSCE date:	Examinee: (last name, first name): Examiner:
--	---

Station 10 (Retroperitoneal space with aorta)

A patient presents with suspected lymphoma: Please 1.) Examine the entire retroperitoneal space to measure or exclude enlarged lymph nodes (LN) and 2.) Determine the diameter of the aortic lumen 3.) Please identify five hypoechoic/anechoic egg-shaped structures in a frozen image, which might be changed by mistake with LN. Systematic scanning of RP space from L paraaortic to R paracaval sections, following the aorta to its bifurcation, AO-2, shows 5 dark/black egg-shaped DDs.

Initial handling of the transducer

Orientation:

- Correct, initial check by disconnecting the cranial part of the transducer, done by himself/herself 1
- Initial problems or forgotten, corrected after examiner's reminder (see: guideline) 0
- Needs assessor's help to find the adequate orientation

Positioning:

- Correct, puts transducer in place right away 2
- Initial problems, but can adjust to examiner's reminder (see: guideline) 1
- Needs assessor's help to position the transducer at the adequate place 0

Connection:

- Connects the transducer adequately to the skin, adjusts pressure if necessary 2
- Initial problems, but can adjust to examiner's reminder (see: guideline) 1
- Insufficient pressure, does not succeed in connecting the transducer to the skin without help 0

Adequate zoom factors:

- Adjusts the zoom factor by himself/herself immediately if necessary for the present task 2
- Initial problems, but can adjust to examiner's reminder (see: guideline) 1
- Does not achieve an adequate zoom factor even after examiner's feedback / 2° to time limits 0

Communication with the patient/model

Breathing commands:

- Correct: "Please take a deep breath – and hold it." 4
- Incomplete/initial problems (forgotten)/reminder from examiner necessary (see guideline) 2
- Even after reminder incomplete or forgets several times to ask the patient to inhale 0
- Immediately asks the patient to breathe again after freezing the image 2

Scanning performance

Scanning/following the blood vessels:

- RP space completely scanned – even with left paraaortic and right paracaval spaces 8
- RP space scanned down to the bifurcation without scanning the paravascular areas 4
- Scanning of the RP space only possible with examiner's help (see guideline) 2
- Even with examiner's support no adequate scanning possible 0

Measurement:

- Correct: perpendicular to the long axis of the vessel, vessel walls not included 5
- Wrong endpoints of the diameter/corrective feedback from examiner necessary (see guideline) 0
- Wrong or missing measurement – even after examiner's feedback / 2° to time limits

Explanation of frozen image

- Identifies five hypoechoic/anechoic egg-shaped structures (as DD to LN) correctly:
 Esophagus (1) Crura of the diaphragm (1) Confluens PV (1) Left renal vein (1) Duodenum (1) 0 - 5

Overall performance (global rating scale)

Outstanding 8 - 7 - 6 - 5 - 4 - 3 - 2 - 1 Poor 0 - 8

Theoretical background

- Lymph nodes are nodular structures – how can you differentiate them from tubular structures? Please explain two options. 0 - 4
Continuous scanning (1) => LN: sudden (dis-)appearance (1); rotation of the transducer by 90° around its cable (1) LN: keeps round shape and does not become tubular (1)
- Please use a diagram to explain the increased risk for rupture of a partially thrombosed aneurysm with eccentric lumen in contrast to an aneurysm with concentric lumen: 0 - 6
Diagram of aortic aneurysm with concentric (1) versus eccentric (1) lumina; protective thrombotic ring in case of concentric lumina (2) / correct predilection site for rupture in eccentric cases (2)

Station score (max. 50):

Version 2010

13. Duplex Ultrasound Objective Structured Assessment of Technical Skills (DUOSATS)⁷⁴

Duplex Ultrasound Objective Assessment of Technical Skill (DUOSATS)				
	SCORING			
DOMAIN	1	2	3	4
Patient positioning	Does not position patient appropriately	Positions patient appropriately		
Transducer selection	Incorrect transducer	Correct transducer with prompting	Correct transducer without prompting	Correct transducer without prompting with optimum frequency selected
Ultrasound coupling gel usage	None/too much	Correct amount		
Acquires images in B mode	No attempt made	Attempt made at B mode imaging		Appropriate depth of field, TGC and focus setting Appropriate gain control?
Evaluation of stenosis in colour	No	Yes		Good colour, filling with no colour bleed. Appropriate velocity range selected with no colour aliasing colour aliasing
Spectral Doppler		Appropriate gate position and size	Appropriate gate position and size. Appropriate velocity range, baseline and inversion settings	Appropriate gate position and size. Appropriate velocity range, baseline and inversion settings. No aliasing seen
Doppler Angle	>60 degree angle Doppler angle to direction of flow		< 60 degree Doppler angle to direction of flow in centre stream	< 60 degree Doppler angle to direction of flow in centre stream
Assessment of PSVR	No pre and post PSV measurements	Pre/post stenosis PSV measurement too close to stenosis		PSV measurement taken in stenosis jet and 4 diameter lengths away
Calculation of diameter reduction at stenosis	Colour Doppler used to assess lumen with appropriate beam steering	Colour Duplex used to assess lumen in longitudinal section with appropriate beam steering	Colour Duplex used to assess lumen in longitudinal section with appropriate beam steering at appropriate point in cardiac cycle	Colour Duplex used to assess lumen in cross section with appropriate beam steering at appropriate point in cardiac cycle. Good attempt made to make an un-diseased and actual diameter measurement of vessel
Calculation of degree of stenosis	No attempt made		With 20% error of correct stenosis	Correct decile calculated for degree of stenosis
Reporting	Poor legibility, comments not relevant	Legible, comments not relevant	Legible report with appropriate comments on PSVR	Legible report with appropriate comments on PSVR, diameter reduction and conclusions

14. Venous Duplex Ultrasound Objective Structured Assessment of Technical Skills (V-DUOSATS)⁷⁵

V- DUOSATS for simulation		1	2	3	4
x	Patient positioning	Doesn't position appropriately	Positions patient appropriately with prompting		Positions patient appropriately (standing or sitting on edge of couch), with examined leg externally rotated and in front, resting weight on contra-lateral leg with no prompting
✓	Transducer selection	Incorrect transducer	Correct transducer with prompting	Correct transducer without prompting	Correct transducer without prompting with optimum frequency and image preset selected
✓	Ultrasound coupling gel usage	None/ too much	Correct amount		
✓	Image optimisation in B mode	No attempt made	Attempt made at B mode imaging		Appropriate depth of field, TGC and focus setting Appropriate gain control?
✓	Acquisition of SFJ image	No attempt	Partial image of LSV or CFV		Good longitudinal image of LSV entering CFV. Measurement of LSV diameter taken.
✓	Evaluation of reflux in colour	No	Yes		Good colour, filling with no colour bleed. Appropriate velocity range selected .
✓	Reflux time assessment using spectral Doppler		Appropriate gate position and size.	Appropriate gate position and size. Appropriate velocity range, baseline and inversion settings.	Appropriate gate position and size. Appropriate velocity range, baseline and inversion settings. No aliasing seen
x	Reporting	Poor legibility, comments not relevant	Legible, comments not relevant	Legible report with appropriate comments on junctional incompetence	Legible report with appropriate comments on junctional incompetence, reflux time, LSV diameter and conclusions

15. Direct observation of procedural skills assessment tool for ultrasound-guided regional anesthesia⁷⁶

Item (scoring scale) *	Video 1		Video 2	
	Min-Max	Median (IQR)	Min-Max	Median (IQR)
1. Clinical knowledge (1–9)	4–9	7 (6–7)	3–8	6 (4.75–7)
2. Consent (1–9)	4–9	6 (5–7)	2–9	5 (4–5)
3. Preparation (1–9)	2–8	6 (4–7)	2–9	6 (4–7)
4. Vigilance (1–9)	1–8	6 (4–7)	3–9	6 (4–7)
5. Infection control (1–9)	1–8	5 (3–7)	1–8	5 (3–7)
6. Technical ability (1–9)	2–9	7 (6–8)	2–8	6 (3–7)
7. Patient interaction (1–9)	1–9	7 (5–7)	2–9	5 (4–6)
Overall Performance for this Procedure (1–9)	1–8	6 (4–7)	1–8	5 (3–6)
Was the procedure completed satisfactorily? (Yes/No)	Yes: 76%		Yes: 55%	

16. A single static breast model education education of ultrasound skill in final year medical students of Burapha university⁷⁷

Table 1: Technical skills checklist

Skill	Yes	No
1. Selects correct ultrasound probe		
2. Applies gel on probe or on patient		
3. Correctly identifies probe orientation		
4. Adjust image depth or contrast to optimize view of lesion		
5. Correctly identifies lesion		
6. Measures the lesion		
7. Saves and stores the image		

17. Global rating scales⁷⁸

Global Rating Scale				
	Score			
Item	1	2	3	4
Preparation for procedure (e.g., machine, machine placement, ultrasound gel, towels)	Was not prepared or organized, stops frequently to organize equipment		Equipment organized, some forward thinking evident, occasionally stops to organize equipment	Equipment well organized, performs procedure with efficiency
Patient interaction (rapport, patient comfort)	Little or no rapport with patient, does not consider patient comfort		Some rapport with patient, occasionally considers patient comfort	Excellent rapport with patient, consistently considers patient comfort
Image optimization (probe choice, optimization of gain and depth)	Does not choose appropriate probe, depth of field or gain, obviously unsure of how to use machine		Usually chooses correct probe, appropriate depth of field and gain.	Consistently chooses correct probe, efficient choice of depth of field, gain and use of machine
Probe technique	Many unnecessary movements, inefficient, lengthy or no image generation		Some unnecessary movements, frequently generates accurate image	Purposeful movements, maximum efficiency, proficient image generation
Image interpretation	Cannot identify abnormalities		Identifies abnormalities	Identifies abnormalities and understands the clinical context
Use of sterile technique (If Applicable)	No or poor sterile technique		Occasional errors in sterile technique made	No errors in sterile technique, obvious comfort with sterile technique
Knowledge of procedure (If Applicable)	Deficient in knowledge		Knows most important steps of procedure	Knows all important steps of procedure, demonstrates familiarity and confidence in performing steps of procedure
Troubleshooting (adjusts approach as necessary)	Does not recognize problem or that an alternate approach is necessary		Recognizes problem and occasionally attempts to take an alternate approach/ troubleshoot	Recognizes problem and always attempts an alternate approach/troubleshoot
Overall performance	Poor		Competent	Superior aptitude

18. Generic technical skills global rating scale⁷⁹

Appendix 2: Generic technical skills global rating scale^{7,9,10}

Respect for tissue	1 Frequently used unnecessary force on tissue or caused damage	2	3 Careful handling of tissue but occasionally caused inadvertent damage	4	5 Consistently handled tissue appropriately with minimal damage
Time and motion	1 Many unnecessary moves	2	3 Efficient time/motion but some unnecessary moves	4	5 Clear economy of movement and maximum efficiency
Instrument handling	1 Repeatedly makes tentative or awkward moves with instruments by inappropriate use of instruments	2	3 Competent use of instruments but occasionally appeared stiff or awkward	4	5 Fluid moves with instruments and no awkwardness
Knowledge of instrument	1 Frequently asked for wrong instruments or used inappropriate instrument	2	3 Knew names of most instruments and used appropriate instruments	4	5 Obviously familiar with the instruments and their names
Flow of procedure	1 Frequently stopped procedure and seemed unsure of next move	2	3 Demonstrated some forward planning with reasonable progression of procedure	4	5 Obviously planned course of procedure with effortless flow from one move to the next
Use of assistants	1 Consistently placed assistants poorly or failed to use	2	3 Appropriate use of assistants most of the times	4	5 Strategically used assistants to the best advantage at all times
Knowledge of procedure	1 Deficient knowledge	2	3 Knew all important steps of operation	4	5 Demonstrated familiarity with all aspects of operation/procedure
Overall performance	1 Very poor	2	3 Competent	4	5 Clearly superior

☐ Pass; ☐ Fail

19. Items assessment tool⁸⁰

Item	Critical Items ^a	Expert Estimate % ^b (SD)
Introduction		
1. Introduces self where applicable (ie, if not already known to patient, patient not critically ill)		72.8 (20.4)
2. Explains procedure (explains ultrasound, its role, and images) where applicable (ie, patient not critically ill)		74.2 (16.1)
3. Washes hands		49.0 (17.8)
4. Ensures patient appropriately and discreetly exposed		55.3 (22.1)
5. Explains ultrasound findings appropriately (even if unsure of results), where applicable	Yes (1)	74.6 (18.1)
Appropriate use of the machine		
6. Appropriately positions the machine		54.3 (19.6)
7. Appropriately applies basic knobology (eg, on/off, depth, gain)	Yes (1)	86.7 (14.8)
8. Appropriately uses examination presets		52.5 (24.8)
9. Chooses correct transducer	Yes (1)	90.0 (14.1)
10. Appropriately enters patient identifier		43.2 (15.7)
11. Able to store relevant images and clips	Yes (1)	61.3 (21.5)
12. Appropriately cleans machine and transducers	Yes (2)	42.1 (16.3)
13. Able to ensure safety of transducers (eg, not dropping transducers)		42.8 (24.1)
Choice of scans based on clinical relevance		
14. Conducts the appropriate types of scans	Yes (1)	80.8 (14.0)
15. Conducts scans in the appropriate prioritization/sequence		64.1 (23.2)
16. Applies appropriate clinical reasoning in choice of scans	Yes (1)	70.1 (10.2)
Image acquisition		
17. Attains minimal criteria	Yes (1)	84.2 (16.1)
18. Positions patient appropriately for specific scans		60.1 (18.6)
19. Scans with adequate transducer pressure		56.5 (19.0)
20. Scans adequately through the entire area of interest	Yes (1)	78.8 (19.8)
21. Able to optimize image appropriately when necessary		42.1 (17.6)
22. Adjusts focal zone appropriately (where relevant and available)		32.5 (18.0)
23. Scans with efficiency of hand motion		37.8 (20.6)
Image interpretation		
24. Able to recognize key findings	Yes (1)	88.3 (11.1)
25. Able to recognize when images are inadequate/insufficient for a given indication	Yes (1)	87.1 (20.5)
26. Recognizes relevant artifacts	Yes (1)	68.3 (19.1)
Scan integration/clinical decision making		
27. Able to determine when and what additional imaging studies/investigations are necessary	Yes (1)	82.2 (17.4)
28. Able to appropriately determine patient disposition based on ultrasound findings	Yes (1)	79.2 (16.9)
29. Able to appropriately incorporate test characteristics (eg, sensitivity/specificity/likelihood ratios) into clinical decision making		60.0 (17.5)
30. Able to appropriately manage unexpected or unknown findings on bedside ultrasound	Yes (3)	67.9 (17.5)
31. Overall, able to determine appropriate next clinical steps	Yes (1)	83.3 (12.1)
Final cut score for the 31-item tool		65.2 (17.0)
Final cut score for the 16 critical item tool		76.5 (12.4)

^a Critical items are those that the experts indicated that a learner should fail the competency-based assessment if the item was not performed satisfactorily; the numbers in parentheses indicate the round in which consensus for the critical item was achieved.

^b Expert estimate % refers to the expert estimated percentage of borderline learners who would successfully complete the item.

20. Assessment checklist for ultrasound-guided regional anesthesia and Global rating scale⁸¹

TABLE 1. Assessment Checklist for Ultrasound-Guided Regional Anesthesia

Task	Not Performed	Performed Poorly	Performed Well
1. Proper positioning of patient			
2. Correct placement of ultrasound machine relative to patient to allow easy visualization of both			
3. Choice of correct transducer			
4. Correct depth, gain, and focal zone choices			
5. Holds the probe appropriately (3 fingers holding the probe and 1 finger touching the patient)			
6. Knowledge or confirmation of screen orientation (ie, which side of probe corresponds to which side of screen)			
7. Scanning of anatomy and proper identification of target			
8. Use of Doppler to rule out vascular structures (if applicable)			
9. Appropriate needle alignment			
10. Maintenance of needle tip image during advancement of needle			
11. Efficiency of regaining needle tip position image (PART maneuver)			
12. Recognition of proper nerve stimulation at appropriate levels (if nerve stimulation used)			
13. Ensure that current is not <0.2 mA (if nerve stimulation is used)			
14. Ask for initial aspiration to rule out intravascular injection			
15. Visualization of needle tip before injection			
16. Ask for 1- to 2-mL initial injection to rule out intraneural and intravascular injection			
17. Ask patient or at least look for signs of pain/discomfort			
18. Ask for proper aspiration every 5-mL increments injection			
19. Recognition of proper needle tip position			
20. Perform appropriate needle tip adjustments			
21. Assessment of ease of injection (high pressure)			
22. Recognition of correct local anesthetic spread in relation to nerve			

TABLE 2. Global Rating Scale Generated After the Second Iteration of Feedback

Item	Score				
	1	2	3	4	5
Preparation for procedure (eg, monitors, intravenous access, ultrasound machine)	Did not organize equipment well. Has to stop procedure frequently to prepare equipment.		Equipment generally organized. Occasionally has to stop and prepare items.		All equipment neatly organized, prepared, and ready for use.
Patient interaction	Little to no rapport established; patient is unaware of procedures. No sedation is provided.		Rapport is generally established; patient is aware and informed of most procedures. Patient anxiety is alleviated adequately with sedative(s).		Strong rapport is established and maintained throughout procedure. Patient is well informed of all relevant procedures. Patient anxiety alleviated through sedation and verbal comforting.
Asepsis (eg, use of sterile gloves, proper patient draping, probe sterility, cleansing of skin before infiltration, use of op site)	Practice of proper aseptic technique not generally apparent. Many errors in aseptic technique made throughout procedure.		Generally practices proper aseptic technique. Occasional errors in aseptic technique made during procedure.		Excellent demonstrates proper aseptic technique. Few or no errors in aseptic technique made during procedure.
Respect for tissue	Frequently uses unnecessary force on issue or causes damage.		Carefully handles tissue but occasionally causes unintentional damage.		Consistently handles tissues appropriately with minimal damage
Time and motion	Many unnecessary movements.		Efficient time/motion but some unnecessary movements.		Clear economy of movements. Maximum efficiency.
Instrument handling	Repeatedly makes tentative and awkward movements.		Competent with instruments but occasionally makes awkward or stiff movements.		Fluid movements with instruments and no awkwardness.
Flow of procedure	Frequently stops procedure and seems unsure of next move.		Demonstrates some forward planning with reasonable progression of procedure.		Obviously planned course of procedure, with effortless flow from one move to the next.
Knowledge of procedure	Deficient knowledge.		Knows all important steps of procedure.		Demonstrates familiarity with all aspects of procedure.
Overall performance	Very poor		Competent		Clearly superior
Overall, should the candidate: pass/fail					

21. Scoring system for automate evaluation echocardiography method⁸²

Evaluate placement and orientation of probe match with machine.

23. Technical skills for ultrasound-guided FNA⁸⁶

Technical Skill	Preworkshop (n = 23)	Postworkshop* (n = 21)		P
		No	Yes	
Satisfactory position of hands				.013
No	15 (65)	5 (38)	8 (62)	
Yes	8 (35)	0 (0)	7 (100)	
Satisfactory position of transducer				.001
No	17 (74)	3 (20)	12 (80)	
Yes	6 (26)	0 (0)	5 (100)	
Satisfactory position of lesion on screen				.013
No	15 (65)	5 (38)	8 (62)	
Yes	8 (35)	0 (0)	7 (100)	
Placed needle in appropriate amount of time				.070
No	17 (74)	6 (46)	9 (54)	
Yes	6 (26)	2 (40)	3 (60)	
Satisfactory repositioning of needle				.004
No	18 (28)	5 (33)	10 (67)	
Yes	5 (72)	0 (0)	5 (100)	
Performed adequate excursions				< .001
No	19 (83)	2 (13)	14 (87)	
Yes	4 (17)	0 (0)	4 (100)	
Visualized lesion in maximum dimension				.008
No	13 (57)	2 (18)	9 (82)	
Yes	10 (43)	0 (0)	9 (100)	
Measured lesion in three dimensions				< .001
No	20 (87)	3 (18)	14 (82)	
Yes	3 (13)	0 (0)	3 (100)	
Visualized needle tip in lesion				< .001
No	22 (96)	3 (16)	16 (82)	
Yes	1 (4)	0 (0)	1 (100)	
Targeted sample with parallel method				.001
No	20 (87)	5 (29)	12 (71)	
Yes	3 (13)	0 (0)	3 (100)	
Targeted sample with perpendicular method				< .001
No	19 (83)	3 (19)	13 (81)	
Yes	4 (17)	0 (0)	4 (100)	

*Total number may sum to < 21 if a specific skill was omitted from the postworkshop assessment by participants.

24. Objective structure assessment of ultrasound skill (OSAUS)⁸⁷

Likert scale	1	2	3	4	5
1. Applied knowledge of ultrasound equipment: Familiarity with the equipment and its functions.	Unable to operate the equipment		Operates the equipment with some experience		Familiar with operating the equipment
2. Image optimization: Consistently ensuring optimal image quality by adjusting gain, depth, focus, frequency.	Fails to optimize images		Competent image optimization but not done consistently		Consistent optimization of images
3. Systematic examination: Consistently displaying systematic approach to the examination and presentation of relevant structures according to guidelines.	Unsystematic approach		Displays some systematic approach		Consistently displays a systematic approach
4. Interpretation of images: Recognition of image pattern and interpretation of findings.	Unable to interpret any findings		Does not consistently interpret findings correctly		Consistently interprets findings correctly
5. Documentation of examination: Image recording and focused verbal/written documentation.	Does not document any images		Documents most relevant images		Consistently documents relevant images

25. The EUS skills assessment tool⁸⁸

Affix AET sticker here

Affix patient sticker here

1=no assistance
2=minimal assistance (one instruction)
3=moderate assistance (multiple instructions)
4=significant assistance (attg manipulates scope to optimize visualization)
5=unable to achieve

☐ Radial ☐ Linear

Indication ☐ Panc mass ☐ Panc cyst ☐ Bil dil ☐ PD dil
☐ Lymphadenopathy ☐ Lumina GI cancer ☐ Mediastinal mass ☐ possible submucosal lesion
☐ Chronic Pancreatitis ☐ Other _____

Date ____/____/____ Case No. for fellow ____/Attending initials____/Fellow initials____

EUS stations: Allow about 1 minute per station before assistance (please specify reason for NA)

AP window	1	2	3	4	5	NA _____
Celliac axis	1	2	3	4	5	NA _____
Body of pancreas	1	2	3	4	5	NA _____
Tail of pancreas	1	2	3	4	5	NA _____
Portosplenic confluence	1	2	3	4	5	NA _____
Head/neck of pancreas	1	2	3	4	5	NA _____
CBD/CHD	1	2	3	4	5	NA _____
Gallbladder	1	2	3	4	5	NA _____
Uncinate	1	2	3	4	5	NA _____
Ampulla	1	2	3	4	5	NA _____

Characterization/Sampling of Lesion

Identify lesion of interest	1	2	3	4	5	NA _____
Appropriate TNM stage	1	2	3	4	5	NA _____
Characterize subepithelial lesion (wall layers)	1	2	3	4	5	NA _____
Achieved FNA	1	2	3	4	5	NA _____

Procedural complication: _____

Supplement Figure 1. The EUS Skills Assessment Tool was to assess EUS skills throughout the duration of training

26. Ultrasound-guided arthrocentesis checklist⁸⁹**Table 1.** Checklist

Criterion Number	Delphi Round 1		Delphi Round 2		Final Version of Item
	Mean; Standard Deviation	Modification	Mean; Standard Deviation	Modification	
1	8.2; 0.92	None	8.1; 1.85	None	1. Identifies need for arthrocentesis by history and physical exam.
2	8.5; 0.85	Combined items	8.3; 0.67	None	2. Explains to patient the indications and alternatives for procedure and obtains informed consent.
3	8.4; 1.07	None	8.2; 0.79	None	3. Ensures no contraindications for arthrocentesis.
4	7.7; 1.64	Wording change	8; 0.94	None	4. Using ultrasound, identifies area of largest effusion and most accessible needle entry point. Identifies any structures that must be avoided (nerves, vessels, tendons).
5	8.1; 1.29	Wording change	8.2; 0.79	None	5. Obtains correct equipment.
6	8.3; 1.34	None	8.4; 0.70	None	6. Washes hands prior to procedure.
7	7.7; 1.49	None	8; 0.82	None	7. Places joint in proper position.
8	7.4; 1.65	Wording change	8.3; 0.82	None	8. Cleans area surrounding aspiration site with betadine/chlorhexidine so as to create sterile field.
9	5.3; 2.63	Wording change	8.1; 1.10	None	9. Personal protective equipment items, such as sterile gloves, are used.
10	5.9; 2.23	Wording change	7.4; 1.26	None	10. Creates sterile field with drapes/towels if needed.
11	6.3; 2.45	Wording change	7.2; 1.81	None	11. Uses local anesthetic as indicated.
12	7.6; 1.84	Wording change	7.7; 1.34	None	12. Places probe cover on ultrasound.
13	7.8; 1.31	Wording change	7.9; 1.10	None	13. With ultrasound, once again identifies area with largest fluid collection, appropriate needle entry site, and any structures that should be avoided.
14	4.3; 3.82	Dropped	0; 0	None	14. Applies traction to joint as appropriate.
15	8; 1.41	None	8.2; 0.63	None	15. Properly sized needle on syringe is inserted into subcutaneous tissue and identified by ultrasound.
16	8.7; 0.48	None	8.3; 0.67	None	16. Under ultrasound guidance, needle is advanced into fluid pocket and fluid aspirated.
17	7.8; 1.55	None	7.9; 1.20	None	17. Sufficient sample is obtained in appropriate sized syringe.
18	7; 2.36	Wording change	8; 0.82	None	18. Needle is carefully removed from joint, removed from syringe, and safely discarded in sharps container. A sterile cap is then placed on the syringe.
19	4.4; 1.96	Dropped	0; 0	None	19. Gentle pressure over needle insertion site applied for approximately 1 min.
20	4; 2.05	Dropped	0; 0	None	20. Insertion site cleansed with anti-septic solution.
21	6.9; 1.52	None	7.7; 1.06	None	21. Bandage applied over needle insertion site.
22	7.3; 2.50	Wording change	8.1; 1.00	None	22. If fluid analysis is indicated, specimen properly labeled with patient identification by person who obtained fluid.
23	7.5; 2.22	Wording change	8.1; 1.00	None	23. Culture/gram stain, cell count, and crystal analysis are ordered as indicated.
24	7.6; 1.26	None	7.4; 1.26	None	24. Reassess the patient to ensure that no immediate complications, including pain/bleeding, are present.
25	7.2; 2.04	Combined items	8.3; 0.95	None	25. Discusses results and return precautions, including hemarthrosis and infection, with patient.

27. Items worksheet for cervical evaluation US⁹⁰**Table 1.** Ten-Item Worksheet

Item: Each With Yes or No Response
1. Positioning: was the patient placed in the dorsal lithotomy position with the head of the table slightly raised no more than 30° from horizontal? 2. Was the image sector adjusted to encompass at least a 120° field of view to facilitate simultaneous visualization of the required landmarks without excessive probe manipulation? 3. Was the image "trapezoid," kept in a vertical orientation to maintain a sagittal view of the cervix and lower uterine segment? 4. Was the image depth adjusted to display the entire cervical length and anterior and posterior lower uterine segments? 5. Landmarks: was the external os "triangle" clearly visualized? 6. Landmarks: was the entirety of the endocervical canal clearly visualized? 7. Landmarks: was the internal os clearly visualized? 8. Were the calipers placed correctly? 9. Was the use of moderate fundal pressure (insufficient to cause any patient discomfort) applied for approximately 15 seconds along the uterine axis to provoke dynamic change? 10. Was color Doppler imaging applied briefly over the lower uterine segment and internal cervical os to identify and exclude vasa previa or to assist with placental margin identification?

28. Direct observation of procedural skills (DOPS)⁹¹

Direct Observation of Procedural Skills (DOPS)		Sonographiekurs, Med B			
<i>Student's Surname:</i> <i>Forename:</i> <i>Date of birth:</i> <i>Clinical setting:</i> out patient department, study hospital <i>Procedure Number:</i> thyroid (1), portal area of the liver (2), pancreas (3), liver (4), spleen (5), kidney (6), liver left (8), retroperitoneum II (9) <i>Assessor's position:</i> consultant (1), doctor in advanced education (2) <i>DEGUM-Stage:</i> 1,2,3 <i>Number of previous DOPS observed by assessor with any trainee:</i> 0, 1,2,3,4,5-9,>9 <i>Number of times Procedure performed by trainee:</i> 0, 1-4, 5-9, >10 <i>Difficulty of procedure:</i> low (1), average (2), high (3)					
Please grade the following areas using the scale below	Below expectation for completion (0)	Borderline expectation for completion (1)	Meets expectations for completion (2)	above expectation for completion (3)	U/C*
1.) Identification of organ structures by US					
2.) Investigation of organ structures by US					
3.) Measurement of organ structures and/or blood flow by US					
4.) Description of the procedure					
5.) Background information (DD and normal values)					
6.) Communication skills					
7.) Technical ability					
* U/C: Please mark this if you have not observed the behavior and therefore feel unable to comment. Please use this space to record areas of strength or any suggestions for development: Have you had training in the use of this assessment tool ? (no=0, yes=1) Time taken for observation: 10min Time taken for feedback: 5min Assessor's Signature: Assessor's Name:					

Figure 1 Direct observation of procedural skills assessment form.

29. Task-specific checklist for USG supraclavicular block⁹²

APPENDIX 1
TASK-SPECIFIC CHECKLIST FOR USG SUPRACLAVICULAR BLOCK

Tasks	Task Performance		
	Not Done	Poorly Done/Assistance and Prompting Required	Well Done/Independent
1. Adjusts bed to appropriate height			
2. Positions US machine appropriately			
3. Positions patient appropriately			
4. Sedates patient appropriately			
5. Shows awareness of adequate monitoring and vital signs			
6. Chooses appropriate probe			
7. Prepares probe in sterile fashion (not applicable if prepared by assistant)			
8. Prepares skin (application of skin prep to an adequate area—ear to below clavicle)			
9. Checks that nerve stimulator is attached			
10. Checks that block needle is appropriate and primed			
11. Checks probe orientation			
12. Selects appropriate depth			
13. Selects appropriate gain			
14. Handles probe with control			
15. Identifies subclavian artery and any other vessels in the area			
16. Identifies first rib			
17. Identifies pleura			
18. Identifies plexus and best location for injection			
19. Warns patient of local anesthetic (LA) skin infiltration			
20. Warns patient of block needle insertion			
21. Visualizes needle tip before advancing toward plexus, artery, and pleura			
22. Warns patient about nerve stimulation			
23. Instructs assistant clearly regarding adjustment of nerve stimulator			
24. Instructs assistant to perform negative aspiration test			
25. Warns patient about discomfort on injection			
26. Instructs assistant on volume to be injected			
27. LA is injected in small increments (≤ 5 mL) with previous aspiration			
28. Assesses LA spread and repositions needle appropriately			
29. Cleans site after needle removal			
30. Assesses block success appropriately			

30. Task-specific checklist for USG supraclavicular block⁹²

APPENDIX 1
TASK-SPECIFIC CHECKLIST FOR USG SUPRACLAVICULAR BLOCK

Tasks	Task Performance		
	Not Done	Poorly Done/Assistance and Prompting Required	Well Done/Independent
1. Adjusts bed to appropriate height			
2. Positions US machine appropriately			
3. Positions patient appropriately			
4. Sedates patient appropriately			
5. Shows awareness of adequate monitoring and vital signs			
6. Chooses appropriate probe			
7. Prepares probe in sterile fashion (not applicable if prepared by assistant)			
8. Prepares skin (application of skin prep to an adequate area—ear to below clavicle)			
9. Checks that nerve stimulator is attached			
10. Checks that block needle is appropriate and primed			
11. Checks probe orientation			
12. Selects appropriate depth			
13. Selects appropriate gain			
14. Handles probe with control			
15. Identifies subclavian artery and any other vessels in the area			
16. Identifies first rib			
17. Identifies pleura			
18. Identifies plexus and best location for injection			
19. Warns patient of local anesthetic (LA) skin infiltration			
20. Warns patient of block needle insertion			
21. Visualizes needle tip before advancing toward plexus, artery, and pleura			
22. Warns patient about nerve stimulation			
23. Instructs assistant clearly regarding adjustment of nerve stimulator			
24. Instructs assistant to perform negative aspiration test			
25. Warns patient about discomfort on injection			
26. Instructs assistant on volume to be injected			
27. LA is injected in small increments (≤ 5 mL) with previous aspiration			
28. Assesses LA spread and repositions needle appropriately			
29. Cleans site after needle removal			
30. Assesses block success appropriately			

31. Quantitative assessment of procedural performance⁹⁴

Table 1. Quantitative Assessment of Procedural Performance^{1,13}

Category	Description	Points
Procedural time, min	Starting when ultrasound probe touches skin and ending when placement needle is removed	2 (≤5 min) 1 (6–10 min) 0 (>10 min)
Needle passes, n	Withdrawal of placement needle >1 cm with readvancement	2 (1) 1 (2) 0 (>2)
Procedural performance	Needle visualization during advancement	2 (all of the time) 1 (part of the time) 0 (none of the time)
	Equipment preparation (eg, probe selection, machine settings)	2 (excellent) 1 (good) 0 (poor)
	Target positioning (eg, able to see target and feasible needle trajectory)	2 (all of the time) 1 (part of the time) 0 (none of the time)
	Probe stability (eg, no unintentional movement)	2 (all of the time) 1 (part of the time) 0 (none of the time)
	Needle manipulation (eg, comfortable grip on needle and catheter)	2 (all of the time) 1 (part of the time) 0 (none of the time)
	Visual focus (eg, appropriately focused on machine and not hands during procedure)	2 (all of the time) 1 (part of the time) 0 (none of the time)
	Confirmation of proper injectate spread	2 (excellent) 1 (good) 0 (poor)
	Confirmation of proper catheter tip position	2 (excellent) 1 (good) 0 (poor)
	Positioning of ultrasound machine	2 (excellent) 1 (good) 0 (poor)
	No thoracolumbar flexion (≥45°)	2 (all of the time) 1 (part of the time) 0 (none of the time)
Ergonomic factors	No head/neck rotation (≥45°)	2 (all of the time) 1 (part of the time) 0 (none of the time)
	No lateral shoulder tilt (≥30°)	2 (all of the time) 1 (part of the time) 0 (none of the time)
	No crossing sterile field to nondominant side	2 (all of the time) 1 (part of the time) 0 (none of the time)

32. Global rating scale for US-guided regional anesthesia⁹⁵

ITEM	SCORE				
	1	2	3	4	5
Preparation for procedure (eg, monitors, intravenous access, ultrasound machine)	Did not organize equipment well. Has to stop procedure frequently to prepare equipment		Equipment generally organized. Occasionally has to stop and prepare items.		All equipment neatly organized, prepared, and ready for use
Patient interaction	Little to no rapport established; patient is unaware of procedures. No sedation is provided.		Rapport is generally established; patient is aware and informed of most procedures. Patient anxiety is alleviated adequately with sedative(s).		Strong rapport is established and maintained throughout procedure. Patient is well informed of all relevant procedures. Patient anxiety alleviated through sedation and verbal comforting.
Asepsis (eg, use of sterile gloves, proper patient draping, probe sterility, cleansing of skin before infiltration, use of op site)	Practice of proper aseptic technique not generally apparent. Many errors in aseptic technique made throughout procedure.		Generally practices proper aseptic technique. Occasional errors in aseptic technique made during procedure.		Excellent demonstration of proper aseptic technique. Few or no errors in aseptic technique made during procedure.
Respect for tissue	Frequently uses unnecessary force on tissue or causes damage.		Carefully handles tissue but occasionally causes unintentional damage.		Consistently handles tissues appropriately with minimal damage.
Time and motion	Many unnecessary movements.		Efficient time/ motion but some unnecessary movements.		Clear economy of movements. Maximum efficiency.
Instrument handling	Repeatedly makes tentative and awkward movements.		Competent with instruments but occasionally makes awkward or stiff movements.		Fluid movements with instruments and no awkwardness.
Flow of procedure	Frequently stops procedure and seems unsure of next move.		Demonstrates some forward planning with reasonable progression of procedure.		Obviously planned course of procedures, with effortless flow from one move to the next.
Knowledge of procedure	Deficient knowledge		Knows all important steps of procedure.		Demonstrates familiarity with all aspects of procedure.
Overall performance	Very poor		Competent		Clearly superior
Overall, should the candidate pass/fail			Pass/ Fail		

33. Cognitive and technical skill assessment for renal US in human renal model⁹⁶

<i>Task</i>	
Cognitive assessment, ^a mean $\pm$ SD (%) correct	
Pretest	56% $\pm$ 17%
Post-test	91% $\pm$ 11%
B-mode imaging	
Optimized adjustment of ultrasound controls	100%
Correct transducer orientation	100%
Obtained transverse and longitudinal views	100%
Identified kidney stone and location	100%
Translucent maze	
Time (minutes), mean $\pm$ SD	4.6 $\pm$ 2.2
Total pushes delivered, mean $\pm$ SD	12.5 $\pm$ 3.5
% Effective pushes delivered, mean $\pm$ SD (%)	71% $\pm$ 21%
Successful task completion	100%
Human phantom	
Reposition lower pole calyceal stone	
To infundibulum	100%
To renal pelvis (goal)	90%
Time (minutes), mean $\pm$ SD	4.5 $\pm$ 3.1
Total pushes delivered, mean $\pm$ SD	15.7 $\pm$ 13.3
% Effective pushes delivered, mean $\pm$ SD (%)	80% $\pm$ 15%

^a $p < 0.0001$.

From total 133 items of all 32 references, we deleted the repetitive or same items. The final items (which were inserted into the initial assessment form with modification as initiated items) were 21 items as described below;

1. US obtained without artifacts/picture quality/image optimization
2. Correct choice of transducer
3. Correct depth
4. Correct gain
5. Correct handling of transducer
6. Turn on US
7. Hand-eye coordination/applied knowledge of US equipment
8. Appropriate probe orientation
9. Identified target on screen/structure centered in view
10. Overall performance
11. Positioning of machine
12. US coupling of gel usage
13. Technical ability/technical scan
14. Applies gel on probe
15. Identified lesion/image interpretation
16. Saves and stores images
17. Step of procedure/systematic examination
18. Appropriately enters patient identifier
19. Clean machine and transducer
20. Scans adequately through the entire area of interest
21. Adjust focus

Appendix I: The ethics committee approval

26/5/66 10:09

RMIS :: pm



คณะกรรมการจริยธรรมการวิจัยในมนุษย์
คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์

หนังสือรับรองฉบับนี้ให้ไว้เพื่อแสดงว่า

รหัสโครงการ : REC.62-342-7-1

ชื่อโครงการ :

ความถูกต้องของเครื่องมือที่ใช้ประเมินผลประสิทธิภาพในการทำอัลตราซาวด์สมองของทารก (Validation of new assessment tool for neonatal cranial ultrasound(US) performance)

ผู้วิจัยหลัก: ปัทมา ธนอนันต์ศรีภักดิ์ : ภาควิชารังสีวิทยา คณะแพทยศาสตร์

มหาวิทยาลัยสงขลานครินทร์

ผู้ร่วมวิจัย : ภาสุรี แสงสุภาวนิชภักดิ์ : ภาควิชากุมารเวชศาสตร์ คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์

เอกสารที่รับรอง:

1. โครงการวิจัยฉบับสมบูรณ์ เวอร์ชัน 2.0 ฉบับวันที่ 17 ธันวาคม 2562
2. เอกสารชี้แจงอาสาสมัคร เวอร์ชัน 2.0 ฉบับวันที่ 17 ธันวาคม 2562
3. เอกสารแสดงเจตนายินยอมของอาสาสมัคร เวอร์ชัน 2.0 ฉบับวันที่ 17 ธันวาคม 2562
4. แบบบันทึกข้อมูล เวอร์ชัน 1.0 ฉบับวันที่ 23 กันยายน 2562
5. ใบขอความยินยอมถ่ายภาพเพื่อการตีพิมพ์ เวอร์ชัน 1.0 ฉบับวันที่ 17 ธันวาคม 2562
6. ประวัติผู้วิจัยและหลักฐานการบอจริยธรรมการวิจัย

ได้ผ่านการพิจารณาและรับรองจากคณะกรรมการจริยธรรมการวิจัยในมนุษย์คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์ โดยยึดหลักจริยธรรมของประกาศเฮลซิงกิ (Declaration of Helsinki) และแนวทางการปฏิบัติการวิจัยทางคลินิกที่ดี (The International Conference on Harmonization in Good Clinical Practice)

โดยบรรจุไว้ในการประชุมคณะกรรมการจริยธรรมการวิจัยในมนุษย์ ครั้งที่ 1/2563 วาระที่ 3.4

ขอให้นักวิจัยรายงานความก้าวหน้าโครงการวิจัย ทุก 12 เดือน และยื่นต่ออายุก่อนถึงวันหมดอายุอย่างน้อย 30 วัน

งชื่อ

(รศ.นพ.บุญสิน ตั้งตระกูลนิช)

ประธานคณะกรรมการจริยธรรมการวิจัยในมนุษย์

คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์

วันที่รับรอง : 25 ธันวาคม พ.ศ. 2562

หมดอายุ : 24 ธันวาคม พ.ศ. 2563

คณะกรรมการจริยธรรมการวิจัยในมนุษย์

คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์

15 ถนนกาญจนาภิเษย์ อำเภอหาดใหญ่ จังหวัดสงขลา 90110

โทรศัพท์ 074451149, 074451157

โทรสาร 074212900

Ref no. 3K6ho-ogy7n-Cl3qx-1n9lu-8coZy

มอ 351.7.2/ec.1310

Cranial ultrasound performance assessment tool: development and validation

Pattama Tanaanantarak¹, Supika Kritsaneepaiboon¹, Vorawan Charoonratana¹, Chomanad Chittchang², Kanokporn Chutiwongthanaphat², Preeyacha Pacharn³, Kriengkrai Iemsawatdikul³, Ramida Amornsitthiwat³, Nichanan Ruangwattanapaisarn⁴, Suwadee Eng-Chuan⁴, Thitiporn Junhasavasdikul⁴, Atchara Mahayosnond⁵, Nipaporn Tewattananarat⁶, Piyaporn Kitpraphasiri⁷, Teerasak Phewplung⁸, Thipsumon Tangsiwong⁹, Pasuree Sangsupawanich¹⁰

¹Department of Radiology, Faculty of Medicine, Prince of Songkla University, Songkhla, ²Department of Radiology, Queen Sirikit National Institute of Child Health, Bangkok, ³Department of Radiology, Siriraj Hospital, Faculty of Medicine, Mahidol University, Bangkok, ⁴Department of Diagnostic and Therapeutic Radiology, Ramathibodi Hospital, Faculty of Medicine, Mahidol University, Bangkok, ⁵Department of Radiology, Faculty of Medicine, Chulalongkorn University, Bangkok, ⁶Department of Radiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, ⁷Department of Diagnostic and Interventional Radiology, Chulabhorn Hospital, Bangkok, ⁸Department of Radiology, Faculty of Medicine, King Chulalongkorn Memorial Hospital and Chulalongkorn University, Bangkok, ⁹Department of Radiology, Buddhachinaraj Phitsanulok Hospital, Phitsanulok, ¹⁰Department of Pediatrics, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand

Abstract

Aim: A standard assessment tool for direct evaluation of procedural skills to ensure proficiency of trainees is necessary for cranial ultrasound (US) in clinical practice. This study created and validated an assessment tool for cranial US performance by radiologists. **Material and methods:** An initial evaluation tool for cranial US using criteria was developed based on existing literature. The assessment form was modified using a three-round Delphi process by an expert panel, conducted between January 2021 and April 2021. Rubric scales for grading were added once consensus regarding generated items was reached. Experts confirmed the final assessment tool using a rubric scale. Two raters evaluated cranial US performance of 27 residents in video clips using the tool. Reliability and percent agreement were assessed. **Results:** Seventeen pediatric radiologists working in different settings participated in the expert panel. The content validation of the proposed evaluation tool was enabled by expert pediatric radiologists. Following three rounds of the Delphi process, the initial 14-item assessment form became a final 15-item form. A three-part rubric scale was used in the final form (preparation, US machine operation, and cranial US performance). Interrater reliability was evaluated with Cohen's Kappa. The Kappa value and percent interrater agreement for most items was moderate to almost perfect (0.42–0.93 and 77.8–100%, respectively). The Cronbach's alpha values for both raters were 0.856 and 0.891. **Conclusions:** This study produced the first validated cranial US assessment tool using a modified Delphi method. The final assessment form is a simple and reliable tool.

Keywords: cranial ultrasound; Delphi method; pediatric radiology; validation

Received 22.07.2023 Accepted 07.10.2023

Med Ultrason

2023;0 Online first, 1-8

Corresponding author: Pasuree Sangsupawanich, MD
Department of Pediatrics, Faculty of Medicine,
Prince of Songkla University,
15 Kanjanavanich Rd, Hat Yai, Songkla,
Thailand, 90110
Phone/fax: +66815404777/+6674451251
Email: pasurees@gmail.com

Introduction

The rapid evolution in obstetrics and neonatal care has increased the survival rate of premature babies. However, premature babies may present serious brain conditions, such as intraventricular hemorrhage and periventricular leukomalacia, the effects of which can range in severity from minor cognitive disorders to severe men-

tal retardation. Both abnormalities can be demonstrated clearly on cranial ultrasound (US) [1,2].

Cranial US is the standard method to evaluate brain pathology in neonates via a widened anterior fontanel. Nevertheless, cranial US experts are limited [2]. Moreover, cranial US training methods comprise predominantly workplace-based learning with unstandardized assessment methods using a logbook [3]. However, logbooks have limitations and are inappropriate for assessing competency due to their low reliability because of the self-assessment process involved and have low validity given the lack of guarantee of their accuracy [4]. Furthermore, only evaluating the number of procedures an individual has performed does not reflect performance and competency in those procedures [3-5].

Assessment methods based on competency are appropriate for evaluating skill performance, including the direct observation of procedural skills (DOPS) [6-10]. Using a standard and appropriate assessment tool is critical in the DOPS process. DOPS takes two forms: without and with criteria. Compared to DOPS without criteria, DOPS with criteria is better due to its excellent reliability [4]. The Delphi method is the proper method to develop a DOPS with criteria as an assessment tool. The Delphi method is a consensus development process recommended when creating lists or items in an assessment tool. It has many advantages, such as the inclusion of a large number of participants, the use of the email questionnaire method, low cost, and the avoidance of abusive dominance [11-13]. Several previous studies have used the Delphi method to develop an assessment tool for evaluating US competency in various organs [14-17].

The development of a standard and appropriate assessment tool to apply the DOPS method with criteria allowing proficient trainees to develop an independent cranial US practice is necessary. This study aimed to develop an assessment tool for cranial US skill performance. The reliability of the developed tool was also assessed.

Materials and methods

Phase 1. Assessment tool development

Settings and participants

We used the modified Delphi method to establish a new cranial US performance assessment tool [15,18]. From January 2021 to April 2021, three steps of the procedure were conducted, including the development of the initial cranial US evaluation form, selecting a group of experts for the expert panel, and shaping and verifying the final assessment form using the modified Delphi method.

Generating items for the initial assessment tool

After reviewing the literature related to US assessment forms to determine the performance required for a US user, the primary researcher developed an initial assessment form as part of the modified Delphi method [14,15,19]. The literature review was conducted by searching PubMed for articles published from 2011 to 2020 using the keywords “ultrasound,” “skill,” “assessment,” and “tool.” Overall, 300 related articles were identified; 265 were excluded due to them being unrelated or repetitive articles and 35 articles were reviewed [19-53]. All items regarding general US skills that were significantly correlated with the cranial US were pooled in the initial assessment form.

Selecting a group of experts for the expert panel

The specialists invited to the expert panel had to be practicing pediatric radiologists with at least 3 years of cranial US experience. Each expert had to complete a Google Form to participate in this study while maintaining anonymity.

Delphi process

After the primary researcher developed the initial assessment tool following the literature review, it was sent to the participating experts as a Google Form via email during the first round of the Delphi process. The participating experts were requested to score each item using a 5-point rating scale, with scores ranging from 1 to 5, where 1 meant that the item was not relevant, 2 was defined as the item being relevant but not important, 3-5 meant that the items were relevant and important, and 5 was equivalent to the item being “most important” (3 = “a little important,” 4 = “moderately important,” and 5 = “absolutely important”). The experts were also asked to provide comments or opinions to eliminate items, add other items, or modify the initial items. The experts returned the responses to the researcher and the mean scores and their distribution were evaluated. All items with mean scores equal to or more than 4 were retained, and those with lower mean scores were eliminated. In each round, we allotted a total response time of 1 month, with email reminders sent to non-responders after 2 weeks.

During the second and third rounds of the Delphi process, the cranial US assessment tool, with the items with lower mean scores eliminated and new items added based on expert opinions, was sent again to all experts and the mean scores and distribution were re-evaluated. The experts were requested to re-rank their agreements and provide opinions, with an opportunity to change their scores after viewing the group’s reply to the previous round. These repeated processes were stopped after

the third round due to consensus being achieved with no further suggestions.

After completing the items to identify performance criteria for cranial US, the next step was setting the performance levels. Four performance levels were chosen: distinguished, proficient, emergent, and insufficient. We chose these four levels for formative and summative decision-making with greater reliability in scoring performance than that of fewer levels. Rubrics were used to develop performance descriptions for each level of each performance criterion; they provide information to guide rater judgements. Finally, all items in the third round of the Delphi process were divided into three parts: the preparation phase, US machine operation, and cranial US performance, with their summation as the overall performance score. The final cranial US skill assessment tool with a rubric scale was sent to all experts for verification [54-57].

Phase 2. Assessment tool performance

The reliability of the final cranial US skill assessment tool was determined. The tool was used to evaluate cranial US skill performance of radiology residents recorded in video footage by two raters who were pediatric radiologists. The video footages of cranial US procedure were recorded after the information and objective of the study was given to radiology residents, and the neonate's parents participated in the study after providing written informed consent. Two raters were assigned to use the assessment tool and completed the form after standardization with the instruction sheets. All radiology residents had different levels of experience with cranial US lectures and skills.

The video footage was recorded using two mobile video cameras. The first device recorded the hand movements of the US performers during cranial US and the second device recorded the US screen monitor and the ability to use the US console. Two video recorded footages of the same neonate were merged into one footage, synchronously, with high definition (HD) resolution.

Two raters completed all three parts of the cranial US assessment tool and determined the overall performance of the US performer. Inter-rater reliability was measured using Cohen's kappa. Total percent agreement was also calculated.

Statistical analysis

The characteristics of experts were reported as percentage values. During the Delphi process, the mean score of each item was evaluated. Reliability of each item in the assessment tool were calculated using Cohen's Kappa. Cronbach alpha coefficients were calculated using individual item scores completed by both raters. Percent agreement between two raters was also measured. All data were analyzed using R software version 3.6.1.

Ethical statement

The study protocol was approved by the Human Research Ethics Committee (HREC) of XXX. This study was performed based on the ethical principles of the Declaration of Helsinki and The International Conference on Harmonization in Good Clinical Practice.

Results

Phase 1. Assessment tool development

Overall, 28 pediatric radiologists were invited to participate as expert panelists in this study. Invitation forms were sent via an online group of pediatric radiologists. Seventeen (60.71%) pediatric radiologists agreed to join in the Delphi process. The characteristics of the experts are shown in Table I.

Three rounds of the Delphi process progressed as described below. The response rates of experts to round 1 and rounds 2-3 were 17 of 28 (60.71%) and 13 of 17 (76.47%), respectively.

Round 1 of the Delphi process

After the primary researcher conducted the literature review, 14 relevant items were created and inserted into the initial assessment form. This form was sent to all participating experts in Delphi round 1. Before sending the results back to the researcher, all experts were asked to provide the relevant score for each item and add, eliminate, or modify the items, as necessary. Table II presents the initial items with the mean scores and their distributions from Delphi round 1. The suggested new items from experts in round 1 included "additional US by using linear probe" and "identified patient."

Table I. Characteristics of experts in the expert panel

Characteristic	Number (%)
Sex	
Male	3 (17.65)
Female	14 (82.35)
Position	
Instructor	9 (52.94)
Assistant Professor	3 (17.65)
Associate Professor	3 (17.65)
Professor	2 (11.76)
Workplace	
University Hospital	12 (70.59)
Pediatric Hospital	2 (11.76)
General Hospital	2 (11.76)
Other	1 (5.88)
Experience in cranial ultrasound	
3-5 years	8 (47.06)
5-10 years	4 (23.53)
>10 years	5 (29.41)

Round 2 of the Delphi process

The revised assessment form, including all initial items that had a mean score more than 4 and the newly added suggested items, was resent to the experts in Delphi round 2 as part of the iterative process. The comment “additional US by using linear probe” was proposed to be included in the rubric scale for the item “selected appropriate probe.” In round 2, there were no additional sug-

gested or modified items, and all items had a mean score more than 4, as shown in Table III.

Round 3 of the Delphi process

All 15 finalized items with a mean score more than 4 on the rubric scale were accepted. Subsequently, the components were divided into four levels. The rubric scale in the assessment form was developed by the primary researcher according to the properties for each item. The

Table II. Initial items with mean scores and distributions in round 1 of the Delphi process

Items	Distribution (n = 17)					Mean score
	5	4	3	2	1	
1. Positioned the US machine appropriately	7	9	1	0	0	4.35
2. Selected the appropriate probe	16	1	0	0	0	4.94
3. Corrected probe handling	14	3	0	0	0	4.82
4. Applied gel on the probe	10	7	0	0	0	4.59
5. Used proper functions of the US machine	13	4	0	0	0	4.76
6. Adjusted appropriate gain and frequency	12	5	0	0	0	4.71
7. Adjusted depth and focus to include the whole brain and appropriate area of focus	17	0	0	0	0	5
8. Cleansed the probe after use	8	8	1	0	0	4.41
9. Corrected placement of probe at anterior fontanel to start US	15	2	0	0	0	4.88
10. Corrected probe orientation	16	1	0	0	0	4.94
11. Displayed cranial US approach systematically	11	6	0	0	0	4.65
12. Presented the relevant brain structures according to standard protocol	13	4	0	0	0	4.76
13. Identified positive findings (e.g., IVH, PVL, hydrocephalus, and anomaly)	17	0	0	0	0	5
14. Overall performance	13	4	0	0	0	4.76

IVH, intraventricular hemorrhage; PVL, periventricular leukomalacia; US, ultrasound

Table III. Initial items with mean scores and distributions in round 2 of the Delphi process

Items	Distribution (n = 13)					Mean score
	5	4	3	2	1	
1. Positioned the US machine appropriately	5	7	1	0	0	4.31
2. Selected the appropriate probe	12	1	0	0	0	4.92
3. Corrected probe handling	11	2	0	0	0	4.85
4. Applied gel on the probe	10	3	0	0	0	4.77
5. Used proper functions of the US machine	10	3	0	0	0	4.77
6. Adjusted the appropriate gain and frequency	11	2	0	0	0	4.85
7. Adjusted the depth and focus to include the whole brain and appropriate area of focus	11	2	0	0	0	4.85
8. Cleansed the probe after use	7	5	1	0	0	4.46
9. Correctly placed the probe at the anterior fontanel to start US	10	3	0	0	0	4.77
10. Corrected probe orientation	13	0	0	0	0	5
11. Displayed cranial US approach systematically	8	4	1	0	0	4.53
12. Presented the relevant brain structures according to standard protocol	9	4	0	0	0	4.69
13. Identified positive findings (e.g., IVH, PVL, hydrocephalus, and anomaly)	13	0	0	0	0	5
14. Overall performance	8	5	0	0	0	4.62
15. Identified the patient	10	3	0	0	0	4.77
16. Additional US using a linear probe	7	5	1	0	0	4.46

IVH, intraventricular hemorrhage; PVL, periventricular leukomalacia; US, ultrasound

Table IV. Final cranial US performance assessment form with rubric

Cranial US performance assessment form with rubric				
	4 (Distinguished)	3 (Proficient)	2 (Emergent)	1 (Insufficient)
I. Preparation phase				
1. Identified the patient	Done and confirmed	Done	Not done	Wrong patient
2. Properly positioned the US machine	Located to the right, close to the patient	Located to the right of the patient	Any side but tries to do with the dominant hand	Any side and operates with the non-dominant hand
II. US machine operation				
1. Selected the appropriate probe	Small curvilinear 8-5 MHz probe	Linear high-frequency probe	Large curvilinear low-frequency probe	Hockey stick probe
2. Corrected probe handling	Complete	Inadequate	Improper	Incorrect
3. Applied warm gel onto the probe	Applied the optimal amount of gel onto the probe	Applied the gel onto the probe	Applied the gel on the patient's head	Did not use gel
4. Used proper buttons on the US machine	Familiar with operating	Operated with experience but was inconsistent	Operated with inexperience	Unable to operate
5. Adjusted gain and frequency correctly	Appropriate	Inappropriate	Incorrect	Unable to operate
6. Adjusted the depth and focus to include the whole brain and appropriate area of focus	Optimum field of view, depth, and focus	Included the whole brain but with inappropriate depth and focus	Included partial brain	Unable to operate
7. Cleansed the probe after use	Done			Not done
III. Cranial US performance				
1. Corrected placement of the probe at the anterior fontanel	Placed on the anterior fontanel with good contact	Placed on the anterior fontanel with poor contact	Place on another fontanel	Placed on other areas
2. Confirmed the side of brain orientation	Correct side with confidence	Confirmed several times	Not done	Wrong side
3. Displayed cranial US approach systematically	Consistently systematic display	Some systematic display	Unsystematic display	Could not display
4. Presented the relevant brain structures according to standard protocol	Followed standard protocol	Inconsistent	Presence of some images	Could not do
5. Identified positive findings	Completely identified positive findings	Identified some findings	Missed positive findings	Could not do
Overall performance	Excellent	Good	Borderline	Poor

final assessment form, including the 15 adjusted items and the four levels of competence assigned to each cell of the rubric scale, was divided into three sections: the preparation phase, US machine operation, and cranial US performance. The form is user-friendly and is shown in Table IV. The final assessment form with the rubric scale was sent to the experts for approval. In round 3, consensus and verification by experts were achieved.

Phase 2. Assessment tool performance

Validity

Content validity of a tool refers to whether the tool that can accurately measure all aspects of content.

All experts in the Delphi panel were pediatric radiologists who were content experts in cranial US. Good content validity was confirmed during the Delphi process.

Reliability

The Kappa value for each item in the assessment tool ranged from 0.15 to 0.93 and “NA”. “NA” indicated that the Kappa value could not be calculated due to lack of variation in the data. The percent agreement of all items ranged from 37–100%. The internal consistency of the assessment tool was assessed using Cronbach’s alpha, which were 0.856 and 0.891 for both raters (table V).

Discussion

Cranial US is a supporting assessment required in the field of neonatal care. A standardized assessment process to determine individual proficiency in clinical practice is required. Thus, new and valid cranial US assessment forms need to be developed.

The cranial US assessment form developed in this study should be considered a quality assessment tool because it was developed by a panel of 17 experts in cranial US. It reflects the content validity of the assessment form developed using the Delphi process. The rubric scale in each cell provides the details of the rater’s assessment [58]. The information provided by the rubric scale allows the rater to assess competency confidently, precisely, and reliably [58].

All experts in this study were anonymized to prevent group guidance due to authority or reputation. Therefore, experts could comment and make decisions independently [59]. Nevertheless, the anonymization of participants

prevented the responses from being associated with professional experience, which is a limitation of this study. Moreover, each cycle of the Delphi process was associated with a reduction in the participant response rate, which caused the responses from fewer participants to be overemphasized.

After evaluating the reliability of the assessment tool, all items had moderate to almost perfect agreement (Kappa = 0.42–0.93 and NA), except only two items, “adjusted gain and frequency correctly” and “displayed cranial US approach systematically.” Although the item “displayed cranial US approach systematically” had a Kappa value of 0.27, the percent agreement was slightly higher at 63%. The item “adjusted gain and frequency correctly” had lower values for both Kappa and percent agreement, 0.15 and 37%, respectively. The low Kappa and percent agreement of this item was likely owing to subjective nature of the evaluation due to the widely variable acceptance of bright color contrast of US images between raters. Moreover, the internal consistency of the assessment tool was good, as indicated by the Cronbach’s alpha values of 0.856 and 0.891 for the two raters.

A limitation of this study was that all cranial US performers evaluated were radiology residents with some experience in general US and familiarity with the US machine due to the vulnerability of the neonatal population in the study. This limitation led to a complicated analysis, skill homogeneity, and lack of high variability of US performers. An inexperienced individual was not eligible as

Table V. Interrater reliability of the assessment tool

Items	Kappa (95% CI)	% Agreement
1. Identification of the patient	NA	100
2. Properly positioned the US machine	0.93 (0.80–1)	96.3
3. Selected appropriate probe	NA	100
4. Corrected probe handling	0.69 (0.45–0.92)	81.5
5. Applied warm gel onto the probe	0.42 (0.03–0.81)	77.8
6. Used proper buttons on the US machine	0.44 (0.14–0.74)	70.4
7. Adjusted gain and frequency correctly	0.15 (–0.003–0.3)	37.0
8. Adjusted the depth and focus to include the whole brain and appropriate area of focus	0.55 (0.26–0.84)	77.8
9. Cleansed the probe after use	NA	100
10. Corrected placement of the probe at the anterior fontanel	0.55 (0.21–0.89)	81.5
11. Confirmed the side of brain orientation	0.57 (0.29–0.86)	77.8
12. Displayed cranial US approach systematically	0.27 (–0.002–0.55)	63.0
13. Presented the relevant brain structures according to standard protocol	0.82 (0.64–1)	88.9
14. Identified positive findings (e.g., IVH, PVL, hydrocephalus, and anomaly)	0.65 (0.38–0.93)	85.2
15. Overall performance	0.84 (0.66–1)	88.9

IVH, intraventricular hemorrhage; PVL, periventricular leukomalacia; US, ultrasound; CI, confident interval; NA, Kappa could not be measured due to lack of variation in the data.

a cranial US performer in this study for ethical reasons and this contributed to the limited range in cranial US ability evaluated in this study.

Conclusion

The cranial US assessment form and rubric scale developed in this study is straightforward to use, and it aligns all raters to the same standard of evaluation. It was confirmed to have a good preliminary validity and reliability.

Conflict of interest: none

References

- Diwakar RK, Khurana O. Cranial sonography in preterm infants with short review of literature. *J Pediatr Neurosci* 2018;13:141–149.
- Volpe JJ. Brain injury in the premature infant. Neuropathology, clinical aspects, pathogenesis, and prevention. *Clin Perinatol* 1997;24:567–587.
- Davis PJ, Cox RM, Brooks J. Training in neonatal cranial ultrasound: a questionnaire survey. *Br J Radiol* 2005;78:55–56.
- Bould MD, Crabtree NA, Naik VN. Assessment of procedural skills in anaesthesia. *Br J Anaesth* 2009;103:472–483.
- Stanojevic M. Training of ultrasound in neonatology: global or local? *Donald Sch J Ultrasound Obstet Gynecol* 2013;7:338–345.
- McCracken DJ, Laursen CB, Barker G, Gleeson FV, Cullen KM, Rahman NM. Thoracic ultrasound competence for ultrasound-guided pleural procedures. *Eur Respir Rev* 2019;28:190090.
- Al-Wardy NM. Assessment methods in undergraduate medical education. *Sultan Qaboos Univ Med J* 2010;10:203–209.
- Sood R, Singh T. Assessment in medical education: evolving perspectives and contemporary trends. *Natl Med J India* 2012;25:357–364.
- Tabish SA. Assessment methods in medical education. *Int J Health Sci (Qassim)* 2008;2:3–7.
- Sim JH, Abdul Aziz YFA, Mansor A, Vijayananthan A, Foong CC, Vadivelu J. Students' performance in the different clinical skills assessed in OSCE: what does it reveal? *Med Educ Online* 2015;20:26185.
- Patrawalla P, Eisen LA, Shiloh A, et al. Development and validation of an assessment tool for competency in critical care ultrasound. *J Grad Med Educ* 2015;7:567–573.
- Janke KK, Kelley KA, Sweet BV, Kuba SE. A modified Delphi process to define competencies for assessment leads supporting a Doctor of Pharmacy program. *Am J Pharm Educ* 2016;80:167.
- Adamson R, Morris AE, Sun Woan J, Ma IWY, Schnobrich D, Soni NJ. Development of a focused cardiac ultrasound image acquisition assessment tool. *ATS Sch* 2020;1:260–277.
- Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. *Med Teach* 2017;39:14–19.
- McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm* 2016;38:655–662.
- Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;74:979–983.
- Tolsgaard MG, Todsén T, Sørensen JL, et al. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. *PLoS ONE* 2013;8:e57687.
- Morgan PJ, Lam-McCulloch J, Herold-McIlroy J, Tarshis J. Simulation performance checklist generation using the Delphi technique. *Can J Anaesth* 2007;54:992–997.
- Skaarup SH, Laursen CB, Bjerrum AS, Hilberg O. Objective and structured assessment of lung ultrasound competence. a multispecialty Delphi consensus and construct validity study. *Ann Am Thorac Soc* 2017;14:555–560.
- Davoudi M, Colt HG, Osann KE, Lamb CR, Mullon JJ. Endobronchial ultrasound skills and tasks assessment tool: assessing the validity evidence for a test of endobronchial ultrasound-guided transbronchial needle aspiration operator skill. *Am J Respir Crit Care Med* 2012;186:773–779.
- Schmidt JN, Kendall J, Smalley C. Competency assessment in senior emergency medicine residents for core ultrasound skills. *West J Emerg Med* 2015;16:923–926.
- Rice J, Crichlow A, Baker M, et al. An assessment tool for the placement of ultrasound-guided peripheral intravenous access. *J Grad Med Educ* 2016;8:202–207.
- Kahr Rasmussen N, Nayahangan LJ, Carlsen J, et al. Evaluation of competence in ultrasound-guided procedures—a generic assessment tool developed through the Delphi method. *Eur Radiol* 2021;31:4203–4211.
- Rebel A, Srouf H, DiLorenzo A, et al. Ultrasound skill and application of knowledge assessment using an innovative OSCE competition-based simulation approach. *J Educ Perioper Med* 2016;18:E404.
- Di Pietro S, Mascolo M, Falaschi F, et al. Lung-ultrasound objective structured assessment of technical skills (LUS-OSAUS): utility in the assessment of lung-ultrasound trained medical undergraduates. *J Ultrasound* 2021;24:57–65.
- Gaudet J, Waechter J, McLaughlin K, et al. Focused critical care echocardiography: development and evaluation of an image acquisition assessment tool. *Crit Care Med* 2016;44:e329–e335.
- Burckett-St Laurent DA, Niazi AU, Cunningham MS, et al. A valid and reliable assessment tool for remote simulation-based ultrasound-guided regional anesthesia. *Reg Anesth Pain Med* 2014;39:496–501.
- Wani S, Keswani RN, Han S, et al. Competence in endoscopic ultrasound and endoscopic retrograde cholangiopancreatography, from training through independent practice. *Gastroenterology* 2018;155:1483–1494.e7.

29. Ziesmann MT, Park J, Unger BJ, et al. Validation of the quality of ultrasound imaging and competence (QUICK) score as an objective assessment tool for the FAST examination. *J Trauma Acute Care Surg* 2015;78:1008–1013.
30. Hofer M, Kamper L, Sadlo M, Sievers K, Heussen N. Evaluation of an OSCE assessment tool for abdominal ultrasound courses. *Ultraschall Med* 2011;32:184–190.
31. Jaffer U, Singh P, Pandey VA, Aslam M, Standfield NJ. Validation of a novel duplex ultrasound objective structured assessment of technical skills (DUOSATS) for arterial stenosis detection. *Heart Lung Vessel* 2014;6:92–104.
32. Jaffer U, Normahani P, Lackenby K, Aslam M, Standfield NJ. Validation of a Novel venous duplex ultrasound objective structured assessment of technical skills for the assessment of venous reflux. *J Surg Educ* 2015;72:754–760.
33. Chuan A, Thillainathan S, Graham PL, et al. Reliability of the direct observation of procedural skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesth Intensive Care* 2016;44:201–209.
34. Limchareon S, Kongpromsuk S. A single static breast model education of ultrasound skill in final year medical students of Burapha University. *J Med Ultrasound* 2018;26:143–146.
35. Black H, Sheppard G, Metcalfe B, Stone-McLean J, McCarthy H, Dubrowski A. Expert facilitated development of an objective assessment tool for point-of-care ultrasound performance in undergraduate medical education. *Cureus* 2016;8:e636.
36. Sultan SF, Iohom G, Saunders J, Shorten G. A clinical assessment tool for ultrasound-guided axillary brachial plexus block. *Acta Anaesthesiol Scand* 2012;56:616–623.
37. Ma IYW, Desy J, Woo MY, Kirkpatrick AW, Noble VE. Consensus-based expert development of critical items for direct observation of point-of-care ultrasound skills. *J Grad Med Educ* 2020;12:176–184.
38. Cheung JJH, Chen EW, Darani R, McCartney CJL, Dubrowski A, Awad IT. The creation of an objective assessment tool for ultrasound-guided regional anesthesia using the Delphi method. *Reg Anesth Pain Med* 2012;37:29–333.
39. Noori S, Ebrahimi M, Luo H, Seri I, Siassi B. Development and validation of an automated assessment tool of echocardiography skills acquired on a neonatal simulator. *Echocardiography* 2021;38:217–221.
40. Le CK, Lewis J, Steinmetz P, Dyachenko A, Oleskevich S. The use of ultrasound simulators to strengthen scanning skills in medical students: a randomized controlled trial. *J Ultrasound Med* 2019;38:1249–1257.
41. Desy J, Noble VE, Woo MY, Walsh M, Kirkpatrick AW, Ma IYW. Use of critical items in determining point-of-care ultrasound competence. *Eval Health Prof* 2021;44:220–225.
42. Enquobahrie A, Horvath S, Arikatla S, Rosenberg A, Cleary K, Sharma K. Development and face validation of ultrasound-guided renal biopsy virtual trainer. *Health Technol Lett* 2019;6:210–213.
43. Ng DL, Vuhahula E, Zhang L, et al. Efficacy of an intensive, ultrasound-guided fine-needle aspiration biopsy training workshop in Tanzania. *J Glob Oncol* 2018;4:1–9.
44. Todsen T, Melchioris J, Charabi B, et al. Competency-based assessment in surgeon-performed head and neck ultrasonography: A validity study. *Laryngoscope* 2018;128:1346–1352.
45. Han C, Nie C, Shen X, et al. Exploration of an effective training system for the diagnosis of pancreatobiliary diseases with EUS: A prospective study. *Endosc Ultrasound* 2020;9:308–318.
46. Kunz D, Pariyadath M, Wittler M, Askew K, Manthey D, Hartman N. Derivation of a performance checklist for ultrasound-guided arthrocentesis using the modified Delphi method. *J Ultrasound Med* 2017;36:1147–1152.
47. Chory MK, Schnettler WT, March M, Hacker MR, Modest AM, Rodriguez D. ACES: accurate cervical evaluation with sonography. *J Ultrasound Med* 2016;35:25–28.
48. Heinzow HS, Friederichs H, Lenz P, et al. Teaching ultrasound in a curricular course according to certified EF-SUMB standards during undergraduate medical education: a prospective study. *BMC Med Educ* 2013;13:84.
49. Chin KJ, Tse C, Chan V, Tan JS, Lupu CM, Hayter M. Hand motion analysis using the Imperial College London Surgical Assessment Device: validation of a novel and objective performance measure in ultrasound-guided peripheral nerve blockade. *Reg Anesth Pain Med* 2011;36:213–219.
50. Starkie T, Drake EJ. Assessment of procedural skills training and performance in anesthesia using cumulative sum analysis (cusum). *Can J Anaesth* 2013;60:1228–1239.
51. Mariano ER, Harrison TK, Kim TE, et al. Evaluation of a standardized program for training practicing anesthesiologists in ultrasound-guided regional anesthesia skills. *J Ultrasound Med* 2015;34:1883–1893.
52. Wong DM, Watson MJ, Kluger R, et al. Evaluation of a task-specific checklist and global rating scale for ultrasound-guided regional anesthesia. *Reg Anesth Pain Med* 2014;39:399–408.
53. Hsi RS, Dunmire B, Cunitz BW, et al. Content and face validation of a curriculum for ultrasonic propulsion of calculi in a human renal model. *J Endourol* 2014;28:459–463.
54. Allen D, Tanner K. Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sci Educ* 2006;5:197–203.
55. Ciocan I, Pabst WL, Wolf RW. Transfontanelar ultrasound in preterm infants – a survival guide for the young radiologist on call. Poster: ECR 2017/C-1668. 2017. doi:10.1594/ecr2017/C-1668
56. Caro-Dominguez P, Lecacheux C, Hernandez-Herrera C, Llorens-Salvador R. Cranial ultrasound for beginners. *Transl Pediatr* 2021;10:1117–1137.
57. Ecury-Goossens GM, Camfferman FA, Leijser LM, Govaert P, Dudink J. State of the art cranial ultrasound imaging in neonates. *J Vis Exp* 2015;96:e52238.
58. Wolfe K, Stevens E. The role of rubrics in advancing and assessing student learning. *J Effect Teach* 2007;7:3–14.
59. Rowe G, Wright G, Bolger F. Delphi: A reevaluation of research and theory. *Technol Forecast Soc Change* 1991;39:235–251.

VITAE**Name** Pattama Tanaanantarak**Student ID** 6110330023**Educational Attainment**

Degree	Name of Institution	Year of Graduation
Bachelor of Medicine	Prince of Songkla University	2005
Board of Diagnostic Radiology	Royal College of Radiology	2008

Scholarship Awards during Enrolment

None

Work - Position and Address

Radiologist at Pediatric radiology section, Department of radiology, Prince of Songkla University, Hat Yai, Songkhla, Thailand