



**University of Dundee**

**Validity and reliability of metrics for translation of regional anaesthesia performance from cadavers to patients**

McLeod, Graeme; McKendrick, Mel; Taylor, Alasdair; Lynch, Joanna; Ker, Jean; Sadler, Amy

*Published in:*  
British Journal of Anaesthesia

*DOI:*  
[10.1016/j.bja.2019.04.060](https://doi.org/10.1016/j.bja.2019.04.060)

*Publication date:*  
2019

*Licence:*  
CC BY-NC-ND

*Document Version*  
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

*Citation for published version (APA):*

McLeod, G., McKendrick, M., Taylor, A., Lynch, J., Ker, J., Sadler, A., Halcrow, J., McKendrick, G., Mustafa, A., Seeley, J., Raju, P., & Corner, G. (2019). Validity and reliability of metrics for translation of regional anaesthesia performance from cadavers to patients. *British Journal of Anaesthesia*, 123(3), 368-377. <https://doi.org/10.1016/j.bja.2019.04.060>

**General rights**

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1  
2  
3 **Validity and reliability of metrics for translation of regional anaesthesia performance from**  
4  
5 **cadavers to patients**  
6  
7  
8  
9

10 Graeme McLeod  
11 Consultant Anaesthetist, NHS Tayside  
12 Honorary Professor of Anaesthesia  
13 Institute of Academic Anesthesia  
14 Ninewells Hospital & University of Dundee, UK.  
15 g.a.mcleod@dundee.ac.uk  
16

17 Mel McKendrick  
18 CEO & Research Psychologist  
19 Optomize Ltd, Glasgow, UK  
20 mel@optomize.solutions  
21 Assistant Professor  
22 Heriot-Watt University  
23 M.McKendrick@hw.ac.uk  
24  
25

26 Alasdair Taylor  
27 Anaesthesia SpR  
28 Ninewells Hospital, Dundee, UK  
29 alasdairtaylor1@nhs.net  
30  
31

32 Joanna Lynch  
33 Consultant Anesthetist  
34 Ninewells Hospital, Dundee, UK  
35 Joanna.lynch@nhs.net  
36  
37

38 Jean Ker  
39 Emeritus Professor of Medical Education  
40 NHS Education for Scotland  
41 Postgraduate Medical Office  
42 Ninewells Hospital and Medical School, Dundee, DD1 9SY  
43 jean.ker@nes.scot.nhs.uk  
44

45 Amy Sadler  
46 Anaesthesia SpR  
47 Ninewells Hospital, Dundee, UK  
48 amysadler1@nhs.net  
49  
50

51 Jayne Halcrow  
52 Anaesthesia SpR  
53 Ninewells Hospital, Dundee, UK  
54 jaynehalcrow@nhs.net  
55  
56  
57  
58  
59  
60

1  
2  
3 Gary McKendrick  
4 Researcher  
5 Optomize Ltd, Glasgow, UK  
6 gary@optomize.solutions  
7  
8  
9

10 Ayman Mustafa  
11 Consultant Anaesthetist  
12 Ninewells Hospital, Dundee, UK  
13 amustafa1@nhs.net  
14

15 Jonathan Seeley  
16 Consultant Anaesthetist  
17 Ninewells Hospital, Dundee, UK  
18 jonathan.seeley@nhs.net  
19  
20

21 Pavan Raju  
22 Consultant Anesthetist  
23 Ninewells Hospital, Dundee, UK  
24 pbangalore@nhs.net  
25

26 George Corner  
27 Honorary Professor of Bioengineering  
28 University of Dundee, Dundee, UK  
29 g.a.corner@dundee.ac.uk  
30  
31  
32  
33  
34  
35

36 Correspondence to g.a.mcleod@dundee.ac.uk  
37

38 **Key words**

39  
40 Simulation  
41 Eye tracking  
42 Regional anaesthesia  
43 Metrics  
44

45 Running head: Validity and reliability of performance metrics  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Abstract**

Background: We wish to develop metrics that quantify translation of performance from cadavers to patients. Our primary objective was to develop steps and error checklists developed from a Delphi questionnaire. Our second objective was to show that our test scores were valid and reliable.

Methods: Sixteen UK experts identified 15 steps conducive to good performance and 15 errors to be avoided during interscalene block on the soft embalmed cadaver and patients. Thereafter, 6 experts and 6 novices were trained then tested. Training consisted of: psychometric assessment; an anatomy tutorial; volunteer scanning; ultrasound guided needle insertion on a pork phantom; and on a soft embalmed cadaver. For testing, participants conducted a single interscalene block on a dedicated soft embalmed cadaver while wearing eye tracking glasses.

Results: We developed a 15 step checklist and a 15 error checklist. The internal consistency of our steps measures was 0.83 (95%CI: 0.78 – 0.89) and 0.90 (95%CI: 0.87 – 0.93) for our error measures. Experts completed more steps, mean difference 3.2 (95%CI: 1.5 - 4.8),  $P < 0.001$ ; had less errors, mean difference 4.9 (95%CI: 3.5 – 6.3)  $P < 0.001$ ; better global ratings scores, mean difference 6.8 (95%CI: 3.6 – 10.0),  $P < 0.001$  and more eye gaze fixations, median of differences 128 (95% CI: 0 – 288),  $P = 0.048$ . Fixation count correlated negatively with steps ( $r = -0.60$ ,  $P = 0.04$ ) and with errors ( $r = 0.64$ ,  $P = 0.03$ ).

Conclusions: Our tests were valid and reliable.

1  
2  
3  
4  
5 High quality regional anaesthesia, performed safely, has driven changes in patient  
6  
7  
8 experience and postoperative pain relief after surgery. The Royal College of Anaesthetists  
9  
10 (RCOA) oversees training standards in the UK, and its 2010 curriculum states that after  
11  
12 completion of higher training, trainees should be able to perform plexus and regional blocks  
13  
14 with distant supervision<sup>1</sup>. However, exposure to regional anaesthesia is sporadic and many  
15  
16 trainees complete training with limited regional anaesthesia aptitudes and skills. Our  
17  
18 experience is that only fellows in regional anaesthesia tend to achieve the desired  
19  
20 competencies. Many trainees lack confidence<sup>2</sup>, find interpretation of ultrasound  
21  
22 challenging<sup>3</sup>, display a wide intra and inter subject variability in performance<sup>4</sup>, and may  
23  
24 expose patients to repeated attempts, pain and harm<sup>5</sup>. For example, in a study of axillary  
25  
26 brachial plexus block, trainees committed three times more errors than experts<sup>6</sup>. One out of  
27  
28 six errors were sentinel errors – events that represent a serious deviation from optimal  
29  
30 performance, jeopardize outcome, or harm patients<sup>5</sup>.  
31  
32  
33  
34  
35

36  
37 Our wish is to train novice anaesthetists on a simulator before patient exposure, and  
38  
39 incorporate simulator training into routine medical activity. As yet little evidence exists to  
40  
41 justify this approach. A recent systematic review comparing simulation versus non-  
42  
43 simulation training for regional anaesthesia showed large variability for participants, mode  
44  
45 of simulation and outcome<sup>7</sup>.  
46  
47

48 Our view is that in order to show any benefits of simulation we need to develop valid and  
49  
50 reliable metrics <sup>8,9</sup> that are simple to use and specifically measure translation of skills from  
51  
52 the cadaver to the operating room.  
53  
54

55  
56 However, current tools such as procedure duration and cumulative sum charts (CUSUM)<sup>4</sup> do  
57  
58 not measure the quality of clinical intervention, direct observation of procedural skills  
59  
60

(DOPS) lacks reliability for formative assessment<sup>10</sup>, and dynamic confidence scores remain unvalidated<sup>11</sup>. The Global Rating Scale (GRS)<sup>12, 13 14</sup> is summative and descriptive. Checklists identify key steps and errors before and during nerve block<sup>5, 12, 14, 15</sup>, but are limited to patient use, and may be generic or block specific<sup>6</sup>. Eye gaze tracking quantitative measure of technical performance that maps fixations and saccades (movement of the pupil from one fixation to another), and offers an insight into cognitive intention during ultrasound guided regional anaesthesia (UGRA)<sup>16-18</sup>.

We hypothesised was that our steps and error checklist scores were valid and reliable, and were better with experts compared to novices when tested on the soft embalmed Thiel cadaver, a highly durable and reliable simulator used for regional anaesthesia training<sup>19</sup>. Thus, the primary objective of this study was to develop steps and error checklists. Our secondary objectives were to show: validity and internal and external reliability of our checklist scores; and validity of eye tracking metrics, global rating score and self-rating questionnaires.

## Methods

### *Development of steps and errors*

After University of Dundee non-clinical Ethics Committee approval, an SpR in Anaesthesia, with a psychology degree and research experience, interviewed 7 local UGRA experts. Each interview was tape recorded and used the Consolidated Criteria for Reporting Qualitative research (COREQ), a 32-item checklist for interviews and focus groups<sup>20</sup>. Interviews lasted between 20 and 40 minutes. Recordings were analysed using methodological framework analysis<sup>21</sup> and informed a questionnaire (Online Surveys, Bristol, UK) that sought to identify generic steps and errors critical to the performance of nerve block during both cadaver simulation training and clinical practice (Table 1).

Each question had three sub-questions:

- (i) How important was the item? (answered using a 5-point categorical scale - nil, minor, moderate, major, extreme importance).
- (ii) What was the likelihood of causing patient harm? (answered using a 5-point categorical scale - not very likely, not likely, equally likely, likely, very likely).
- (iii) What was the potential consequence of such harm? (answered using a 4-point categorical score – negligible, minor, serious, catastrophic).

The questionnaire was refined using the Delphi method, an iterative process that looks for consensus amongst experts. We asked UK consultant anaesthetists to participate who had previously been a member of faculty at our annual regional anaesthesia mastery training course on Thiel cadavers in Dundee. All were tutors or graduates of the MSc in regional anaesthesia run from the University of East Anglia, Norwich. Responses were analysed and questions rejected if inter-rater agreement was  $< 0.80^6$ . The remaining items formed a second, smaller questionnaire that was re-distributed to the same consultants. Items achieving  $\geq 80\%$  agreement were included in the final steps and error checklists.

We defined psychometric tests according to the American Association of Psychologists Standards for Educational and Psychological Testing, 2014<sup>22</sup>. They include tests of validity, reliability and fairness. The definitions are as follows:

- Test: A device or a procedure in which a sample of an examinee's behaviour in a specified domain is obtained and subsequently evaluated and scored using a standardised process.
- Assessment: A broader term than test, commonly referring to a process that integrates test information with information from other sources

- **Validity:** A unitary concept which describes the degree to which accumulated evidence and theory support the interpretation of test scores for proposed uses of tests.
- **Validation:** The process involving accumulation of relevant evidence to provide a sound scientific basis for the proposed test interpretations. It is the interpretations of test scores for proposed uses that are evaluated, not the test itself.
- **Construct:** The concept or characteristic the test is designed to measure
- **Reliability/precision:** A general notion of consistency of the scores across instances of the testing procedure
- **Reliability coefficient:** Reliability coefficient of classical test theory, i.e. the correlation between scores on two equivalent forms of the test. Three broad categories exist: (i) alternate form: (ii) test-retest and (iii) internal consistency.
- **Generalizability coefficient:** Ratio of universe score variance to observed score variance. Provides separate measures of components of variance
- **Item response function:** A model representing the increasing proportion of correct responses to an item at increasing levels of the ability or trait being measured.
- **Fairness:** Responsiveness to individual characteristics and testing contexts so that test scores will yield valid interpretations for intended uses.

Study methods entailed rater training and testing, subject training and testing and rater assessment.

#### *Rater video training*

One month before the study, the principal investigator conducted interscalene blocks on the soft embalmed cadaver under ultrasound guidance. Thirty eight videos and ultrasound movies were recorded and stored on computer hard disc. Five raters attended a training session, viewed all recordings, and were taught to recognise all steps and errors including



1  
2  
3 sentinel errors, errors that may lead to patient harm. Thereafter, raters were tested on two  
4  
5 videos and two ultrasound recordings, and completed the 15-item steps checklist and the  
6  
7 15-item error checklist. Successful training was defined as 80% agreement between raters.  
8  
9

### 10 11 12 *Subject training*

13  
14 We conducted the study at the Centre for Anatomy and Human Identification (CAHID). The  
15  
16 sponsor was NHS Tayside, R&D Number 2017AN04, REC Number18/WS/0082, IRAS No  
17  
18 243797, ISRCTN14180589. Twelve anaesthetists (6 novice and 6 experts) voluntarily opted-in  
19  
20 and took part in the study. Novices and experts were defined according to criteria set out by  
21  
22 Dreyfus and Dreyfus<sup>23</sup>. Novices saw uncomplicated actions as a series of steps, had some  
23  
24 working knowledge of key aspects of practice, were able to achieve some steps using their  
25  
26 own judgement, but needed guidance or supervision. In contrast experts regarded  
27  
28 themselves as individuals who had a deep understanding of regional anaesthesia, were able  
29  
30 to take full responsibility for their own work, routinely achieved acceptable standards, saw  
31  
32 the full picture, and modified movement patterns to address difficulties.  
33  
34  
35  
36  
37  
38

39 At the start of the study, we collected baseline demographic characteristics, including sex,  
40  
41 handedness, age, years in anaesthesia, grade, number of supervised and unsupervised  
42  
43 interscalene blocks performed, (0-5, 6-10, 11-30, >30), and details of undergraduate and  
44  
45 postgraduate training. Two independent anaesthetists acted as trainers and delivered four  
46  
47 training sessions over two consecutive days. Experts and novices were taught in groups of  
48  
49 three, either in the morning or afternoon, over 2 days. Participants were rotated through  
50  
51 four training stations, each lasting 20 minutes. Training stations comprised: (i) psychometric  
52  
53 testing; (ii) a lecture describing the anatomy and performance of interscalene nerve block  
54  
55 using ultrasound; (iii) scanning of the interscalene nerve ventral rami on a volunteer; (iv)  
56  
57 needle insertion practice on a pork belly with embedded tendon (WetLab-MedMeat,  
58  
59  
60

1  
2  
3 Warwick, UK); and (v) repetitive performance of interscalene block on a soft embalmed  
4 Thiel cadaver. The subjects received the same tutorial that is given at our Ultrasound for  
5  
6 Novice Anaesthetics Trainees course  
7  
8 [https://dihs.dundee.ac.uk/courses/anaesthetics/ultrasound-novice-anaesthetics-trainees.](https://dihs.dundee.ac.uk/courses/anaesthetics/ultrasound-novice-anaesthetics-trainees)  
9  
10  
11  
12 Our collective experience is that novices' most difficult task is visualisation of the needle at  
13  
14 all times. This is best achieved for novices using an in-plane technique and keeping the  
15  
16 needle as parallel as possible to the transducer elements<sup>24</sup>.  
17

18  
19 Each subject was informed of all the necessary steps to perform and errors essential to avoid,  
20  
21 before and during the block. A trainer was present at all stations in order to demonstrate  
22  
23 procedures and offer guidance similar to that provided at standard UGRA teaching courses.  
24  
25

26  
27  
28 The soft embalmed Thiel cadaver is soaked in a salt and acid solution for 6 months before  
29  
30 use<sup>25</sup>. It is soft, flexible and durable<sup>26</sup> and exhibits similar elasticity and strain displacement  
31  
32 as patients<sup>19</sup>. Bolus injection of embalming fluid results in perineural spread similar to  
33  
34 patients followed by tissue relaxation and fluid dispersion. These properties have allowed  
35  
36 our research group to investigate the fundamental mechanisms underpinning differences in  
37  
38 performance<sup>3</sup>, and the application of mastery training to UGRA training<sup>18</sup>. The cadaver is  
39  
40 highly durable, leaves no needle tracks and withstands hundreds of injections with minimal  
41  
42 change in anatomy<sup>19</sup>. The legal requirements governing cadaver use are set out in the  
43  
44 Human Tissue (Scotland) Act 2006.  
45

46  
47  
48 Psychometric testing used a range of validated instruments (Inquisit 5 Lab, Millisecond  
49  
50 Software, Seattle, WA) to measure mood<sup>27</sup>, sleepiness<sup>28</sup>, handedness<sup>29</sup>, wakefulness<sup>30, 31</sup>,  
51  
52 visuo-spatial skills<sup>32</sup>, dexterity, sustained attention<sup>33 34</sup> and visual scanning skills<sup>35</sup> (Table 1).  
53  
54  
55

56  
57 *Subject testing*  
58  
59  
60

1  
2  
3 For testing, subjects were tasked to perform an interscalene block on a second cadaver.  
4  
5 Subjects first scanned over the airway then moved the transducer laterally to sit posterior to  
6  
7 the right sternocleidomastoid muscle. Once, the ventral nerve roots of C5 and C6 were  
8  
9 identified, a 21g B.Braun needle was inserted in the plane of the ultrasound elements and  
10  
11 directed, through scalenus medius, towards the junction between C5 and C6. Once the  
12  
13 needle tip was as close to, but not touching the ventral nerve roots, 0.25 to 0.5ml of Thiel  
14  
15 embalming solution was injected. This volume is sufficient to observe hydrolocation. No time  
16  
17 limit was used for testing and no instruction was given. Each subject wore SMI ETG 2w  
18  
19 wireless eye-tracking glasses (SensoMotoric Instruments, Teltow, Berlin, Germany) that were  
20  
21 calibrated by psychologists before use. We videoed performance using a fixed camera  
22  
23 focused on the transducer. Ultrasound and eye tracking metrics were recorded continuously.  
24  
25 At the end of the study, participants completed a 15-item self-rating questionnaire with an  
26  
27 11-point anchored scale where 0 represented the worst performance and 10 represented  
28  
29 the best performance possible. Each question was based on the list of steps selected by the  
30  
31 Delphi process.  
32  
33  
34  
35  
36  
37  
38

#### 39 *Rater video analysis*

40  
41 All test video and ultrasound recordings from all participants were examined by raters using  
42  
43 the validated steps and errors checklists developed from our Delphi questionnaire. Our  
44  
45 primary end point was the summation of the number of correctly performed steps and  
46  
47 errors. Raters also completed an 8-point global rating score (GRS) with 5 anchored  
48  
49 categories. Parameters included: preparation; asepsis; respect for tissue; time and motion;  
50  
51 instrument handling; flow of procedure; knowledge; and overall performance. Thus, our  
52  
53 secondary end-points were: (i) psychometric scores; (ii) GRS; (iii) eye gaze metrics; (iv) pre-  
54  
55 procedural scanning and procedural needling time; and (v) participant self-rating of  
56  
57 performance.  
58  
59  
60

## Statistics

Normality of data was assessed by the Shapiro-Wilk test. Steps and error scores and GRS scores were regarded as dependant and total scores assessed using the paired t-test.

Analysis of non-parametric continuous data such as self-report measures, cognitive tests and eye gaze results used the Mann-Whitney U test and are presented as median [interquartile range]. Significant results are presented as the median (95%CI:) of the differences using the Hodges-Lehmann estimate (Graph Pad Prism 7, La Jolla, CA.) We analysed 2x2 contingency tables using the Fisher Exact Probability Test and 4x2 contingency tables with the Freeman-Halton extension of the Fisher exact probability test. Internal reliability of our steps and errors used Cronbach's alpha. External reliability was calculated using intraclass correlation (2, 5) using a 2-way random effects for absolute agreement (SPSS, v25.0.0.1, Chicago, IL). Eye tracking data was analysed using BeGaze 3.7 software (SensoMotoric Instruments). We also used generalizability statistics<sup>36</sup> to model the variance of items (steps and errors), raters, and subjects. (gtheory package (Rstudio, Version 1.1.456 – © 2009-2018 RStudio, Inc. Boston, MA) We used the "ltm" package (Rstudio) to develop a rasch model of the probability of success or failure for specific steps and errors. The mathematical function of the item characteristic curve for binary data is:

$$p = P(X = 1) = \frac{\exp(\theta - \delta)}{1 + \exp(\theta - \delta)}$$

where X is a random variable indicating success or failure on the item. For steps, X = 1 indicated successful performance and X = 0 indicated failure. Inherent ability is indicated by  $\delta$  on the latent variable scale, and  $\theta$  is an item-parameter, generally called the item difficulty, on the same scale. In order to analyse errors, maintain the format of the item characteristic curve, and take account of the wording of the error questionnaire, we altered code entry. Errors were highlighted as X = 0. Thus, for the error item characteristic curve

1  
2  
3 probability of success should be interpreted as “success in not failing to...” or “successful  
4 avoidance of...”. Best fit of the rasch model used bootstrap goodness-of-fit Pearson chi-  
5 squared analysis. Correlation between tests was evaluated using the Spearman correlation.  
6  
7

8  
9 Kappa for multiple raters was calculated using the Online Kappa Calculator.<sup>37</sup>

10  
11 In all analyses, a two-tailed P-value less than 0.05 indicated statistical significance. We used  
12 the Inquisit Lab 5 platform (www.millisecond.com). (https://www.psychtoolkit.org)<sup>8</sup> to analyse  
13 psychometric tests and *Standards for Educational and Psychological Testing*<sup>22</sup>  
14  
15

16  
17 Teltow, Berlin, Germany).

18  
19  
20  
21  
22  
23 **Power analysis:** Our experience was that a difference of at least 3 errors represented a  
24 clinically meaningful difference between experts and novices. Our pilot work suggested that  
25 the standard deviation was 3 errors. Given  $\alpha=0.05$  and power=0.8, effect size = 1, we  
26 recruited 12 participants in order to measure the difference between two dependant means  
27 (G\*power 3.1, Dusseldorf).  
28  
29  
30  
31  
32  
33  
34  
35  
36

## 37 Results

### 38 Checklist development

39  
40 Fifty five questions were identified using methodological framework analysis and divided into  
41 eight groups (Table 2): (i) position and preparation, n = 7; (ii) pre-procedural steps, n = 9; (iii)  
42 pre-procedural scanning, n = 7; (iv) needle, n = 12; (v) needle tip, n = 11; (vi) needle tip  
43 feedback, n = 3 and (viii) local anaesthetic injection, n = 6.  
44  
45  
46  
47  
48

49  
50 The Delphi process identified 20 steps conducive to good performance and 18 errors. Five  
51 steps and 3 errors were considered not relevant to translation of performance from cadavers  
52 to patients, giving a total of 15 steps and 15 errors for testing. The reliability or internal  
53 consistency (95%CI) of our steps checklist was 0.83 (95%CI: 0.78 – 0.89) and 0.90 (95%CI: 0.87  
54  
55  
56  
57 – 0.93) for our errors checklist.  
58  
59  
60

### *Rater video testing*

Five experts were tested on two independent videos. Experts absolute agreement was 79.8% for steps and 80.5% for errors. Reliability (Kappa) was 0.62 (95%CI: 0.31 – 0.88) and 0.61 (95%CI: 0.36 – 0.86) respectively.

### *Subject characteristics*

Experts had more experience than novices, 14.5 [8.5 – 16.8]yr vs 3.5 [3.0 – 4.8]yr, median (95%CI) of differences, 9.5 (95%CI: 2 – 16)yr,  $P < 0.001$ . All novices had performed  $< 5$  unsupervised blocks, and all experts had performed  $> 30$  unsupervised blocks ( $P = 0.002$ ). There was no difference in age ( $P = 0.13$ ), sex ( $p = 0.24$ ), hand dominance ( $P = 1.00$ ) or premedical qualifications ( $P = 1.00$ ). Psychological, visuo-spatial and motor tests were similar in both groups (Table 3).

### *Subject testing – steps and errors*

Nine hundred assessments of steps and 900 assessments of errors were made by 5 raters on 12 participants (6 experts and 6 novices). The external reliability of rater assessment (intraclass correlation, (95%CI)) was 0.85 (95%CI: 0.69 – 0.94),  $P < 0.001$  for steps and 0.94 (95%CI: 0.87 – 0.98) for errors,  $P < 0.001$ .

Experts performed more steps, 8.9 (3.4) vs 5.7 (3.2), mean (95%CI) difference 3.2 (95%CI: 1.5 – 4.8),  $P < 0.001$  and had less errors, 4.1 (3.9) vs 9.0 (3.4) mean (95%CI) difference 4.9 (95%CI: 3.5 – 6.3),  $P < 0.004$ . The variance of steps, subjects, grade and raters and their combinations are shown in Table 4. Subjects, steps and their combination contributed most to measurable variance. Rasch models of the probability of success or failure for specific steps and errors are shown in Fig 1. The easiest steps (Fig 1 A, Nos. 4, 5, 8) and errors (Fig 1 B, Nos. 1, 2, 3) were associated with preprocedural scanning and identification of the nerve. The most

1  
2  
3 difficult steps (Fig 1 A, 10, 13, 15) and errors (Fig 1 B, 9, 13, 15) were associated with keeping  
4 the needle tip in view at all times and recognition of needle/nerve contact and intraneural  
5 injection.  
6  
7  
8  
9

#### 10 11 12 *Subject testing - global rating scores*

13  
14 Experts had better mean (SD) GRS scores (Table 5) than novices, 26.0 (8.8) vs 19.0 (5.4),  
15 mean (95%CI) difference 6.8 (95%CI: 3.6 – 10.0),  $P < 0.001$ . Internal consistency (Cronbach's  
16 alpha) of GRS was 0.62 (95%CI: 0.16 - 0.90),  $P < 0.001$ .  
17  
18  
19

#### 20 21 22 *Subject testing - eye tracking*

23  
24 Experts completed steps quicker, median (95% CI:) of differences 113.9 (25.9 – 417.8) s,  $P =$   
25 0.04. During the scanning phase, experts glanced less often towards the monitor, median  
26 (95% CI:) of differences 4 (95% CI: 0 - 9),  $P = 0.04$  (Table 5).  
27  
28  
29  
30  
31

32  
33 In the needling phase, expert fixation duration was longer, median (95% CI:) of differences  
34 525 (95% CI: 83 - 993) ms,  $P = 0.03$ , and fixation count was greater, median (95% CI:) of  
35 differences 128 (95% CI: 0 – 288),  $P = 0.048$ . Expert self-report scores were better than novices,  
36 mean (95% CI:) difference 3.9 (95%CI: 3.4 - 4.5),  $P < 0.001$ .  
37  
38  
39  
40  
41  
42  
43  
44  
45

#### 46 47 *Subject self-report measures*

48 Experts self-report scores were better than novices. The mean (95% CI:) difference was 3.9  
49 (95%CI: 3.4 - 4.5),  $P < 0.001$ . All questions showed differences between experts and novices  
50 (all  $P < 0.001$ ).  
51  
52  
53  
54

#### 55 56 57 *Metric correlation*

Steps and errors were negatively correlated ( $r = -0.94$ ,  $P < 0.001$ ). Errors correlated with procedure duration ( $r = 0.70$ ,  $p = 0.01$ ) and negatively with self-report scores ( $r = -0.59$ ,  $P = 0.045$ ). Self report scores correlated negatively with scanning dwell time ( $r = -0.63$ ,  $P = 0.03$ ); and glances ( $r = -0.68$ ,  $P = 0.02$ ). Eye gaze fixation count during needling correlated with the number of steps ( $r = -0.60$ ,  $P = 0.04$ ) and the number of errors ( $r = 0.64$ ,  $P = 0.03$ ).

#### *Cadaver durability*

Scanning and needling images from subject 1 (expert) and subject 12 (novice) are shown in Fig. 2. No difference in anatomy is seen despite injection of embalming fluid between C5 and C6 ventral nerve roots. Subject 1 aligns the needle tip precisely whereas subject 12 inserts the needle too superficially. The needle tip is not seen.

#### **Discussion**

We developed steps and error tests that were valid and reliable. Steps and error metrics, eye gaze tracking metrics, global rating scores and participant self-reporting of performance showed differences between novices and experts. These results will enable us to use our metrics to quantify the translation of training performance on the soft embalmed Thiel cadaver to clinical performance on patients.

#### *Strengths and weaknesses of the study*

The principal strengths of our study relate to the process of checklist development, quality of rater training, and depth and range of subject assessment.

Checklist development was comprehensive. We applied COREQ<sup>20</sup> and methodological framework analysis<sup>21</sup> and identified 15 steps and 15 errors from 16 UK experts. The Delphi questionnaire was designed in order that we constructed a generic questionnaire that could



1  
2  
3 be applied to translation of performance from cadaver to patients. Our results suggest that  
4  
5 rather than recommending a particular block technique, that observation of the needle and  
6  
7 avoidance of nerve contact are most important.  
8  
9

10 Our checklists differed from others<sup>5, 14, 15</sup> with respect to size and content: they were up to  
11  
12 four-times smaller than other published checklists, and tended to focus on the details of  
13  
14 needle nerve interaction, similar to that of Cheung et al<sup>14</sup>. We can account for the disparity  
15  
16 between our checklist and those of Ahmed<sup>5</sup> and Sultan<sup>15</sup> because our checklists are applicable  
17  
18 to all blocks and, uniquely, are transferable from soft cadavers to patients, and easy to use.

19  
20 The consensus of UK experts on the technical details of needle tip position are reflected by  
21  
22 our rasch models which show the odds of skills attainment or risk of error relative to  
23  
24 performance. Easiest steps and errors were associated with preprocedural scanning and  
25  
26 identification of the nerve, whereas the most difficult steps were associated with keeping the  
27  
28 needle tip in view at all times and recognition of needle/nerve contact and intraneural  
29  
30 injection. These results echo our teaching experience and approach, and further validate our  
31  
32 checklists.  
33  
34

35  
36 Differences between experts and novices were clinically relevant. The mean number of steps  
37  
38 differed by over three and steps by almost 5, and justified our small number of participants.  
39  
40 Validity studies<sup>6, 13, 15, 38</sup> comparing the performance of experts and novices do not need large  
41  
42 numbers because the differences between groups is large, the effect size is ~1, and data is  
43  
44 dependant because the same test cadaver is being used. However, comparison of subject  
45  
46 performance between studies is difficult due to variable checklist composition, number of  
47  
48 experts used, testing, size, and statistical analysis.  
49  
50  
51

52  
53 Global rating differences were in accord with those recorded previously<sup>12, 14</sup> and our self-rating  
54  
55 questionnaire results were in keeping with this difference. Psychomotor skills did not differ  
56  
57 between experts and novices, in contrast to the work of Shafqat et al<sup>39</sup> who demonstrated a  
58  
59  
60

1  
2  
3 negative correlation between mental rotation tests and novice needle performance on a  
4 turkey phantom. However, our sample size was five-fold smaller and our range of values was  
5 large. We will require larger samples in order to judge whether psychomotor tests impact on  
6 either scanning or needling skills during UGRA on the soft embalmed cadaver.  
7  
8  
9  
10

11  
12  
13  
14 Our objective eye tracking results agree with our previous work on the soft embalmed cadaver  
15 that showed that expertise was associated with more focussed attention<sup>11</sup>. We showed that  
16 reductions in pre-procedural glances towards the screen and procedural eye fixation time  
17 were consistent with expert performance, and that eye tracking metrics discriminated  
18 between different skill levels during the scanning and needling phases of interscalene nerve  
19 block. It is likely that a lower fixation count is driven in part by reduced switches between the  
20 screen and tools and our results reflect patterns shown in laparoscopy<sup>36-38</sup>.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

32 Our findings show validity of eye gaze tracking: fixations and glances differentiated between  
33 experts and novices and correlated with steps and errors. Longer fixation durations evident  
34 in the expert group have been associated with more local processing and higher memory  
35 load<sup>39</sup>. This was more evident in the needling phase where focus was concentrated around  
36 the needle as the trajectory of the needle was followed. The lack of inherent group differences  
37 in cognitive tests indicate that steps, error and eye movement data are due to skill acquisition.  
38  
39  
40  
41  
42  
43  
44

45 Only 3 studies have incorporate eye tracking into assessment of regional anaesthesia  
46 performance. Two UGRA eye tracking studies<sup>17, 40</sup> recruited 6 trainees and 6 experts. The  
47 first<sup>17</sup> demonstrated qualitative eye gaze “heat-maps” when injecting into a gel phantom,  
48 while the second quantified trainees’ interpretation of five ultrasound images<sup>40</sup>. A third,  
49 more extensive study, created individual learning curves for eye gaze fixation and eye  
50 glance, albeit with a wide intra-and inter-subject variation in the rate of skill acquisition, and  
51 discriminated trainee performance based on the slope and variance of the learning curve<sup>18</sup>.  
52  
53  
54  
55  
56  
57  
58  
59  
60

### *Wider impact of research*

Our study is the first to validate steps and errors for the measurement of translational performance from cadavers to patients, validates eye tracking technology as a quantitative measure of regional anaesthesia performance, and informs the debate surrounding the best measure of performance reliability<sup>41</sup>.

Our metrics will enable us to measure the translation of performance from simulators to patients and ask important questions about the impact of simulation in regional anaesthesia on clinical performance. This study is the first of three sequential studies. We will use the results of this study to power two RCTs that will (i) endeavour to compare high fidelity cadaver training vs low fidelity pork specimen training, and (ii) compare mastery vs standard teaching. The latter two studies will train subjects on cadavers and test subjects' performance on patients. Valid objective metrics will allow accurate measurement of performance to a predefined benchmark at basic, intermediate, higher and advanced levels, and offer the opportunity to standardise performance globally. Limited training resource may be more efficiently used to target areas of weakness or to optimise performance in a simulated setting prior to clinical practice.

Widespread application of eye tracking to UGRA is presently limited by the cost of equipment, and the need for an experienced psychologist to input data by hand and interpret results. Automation of eye tracking data analysis is being developed by our research group. Our intention is to enhance training experience with real time objective performance feedback in order to accelerate learning and answer key questions in UGRA robustly. The cost of eye tracking technology is falling rapidly making this affordable to an increasing number of educational institutions worldwide.

1  
2  
3 We used generalizability (G) theory<sup>36</sup> because it calculates the variance associated with all  
4 factors impacting on performance. It is the current recommended statistical approach used  
5 in medical educational studies. Our work showed much variance between subjects, items  
6 and their combination in both steps and error models. For comparison purposes, we  
7 purposefully measured inter-rater agreement during rater training, and calculated kappa  
8 and type 2 intraclass correlation for subject testing as recommended by others<sup>5, 12</sup>. We  
9 obtained good overall agreement ~ 80% during rater training, and during subject testing,  
10 kappa > 0.60 and intraclass correlation, 0.85 to 0.94. These results hide the variance  
11 attributable to subjects and checklist items as well as a large unaccountable error variance.  
12 Our approach exposes the limitations of inter-rater reliability and ICC, and suggests the  
13 need for a validated objective measurement of UGRA performance such as eye tracking which  
14 is not reliant on subjective assessment.

15 We are not aware of any application of rasch modelling to high stakes anaesthetic and  
16 surgical training. The logistic function modelled the probability of success and probability of  
17 failure to spot errors. Thus, we present, for the first time, a graphical means of predicting  
18 item difficulty according to skill levels. The rasch model lends itself to binary measurement  
19 of surgical and anaesthetic skills assessment and we hope this approach will prove useful to  
20 trainers and trainees alike.

21 In conclusion, our steps and error metrics. were valid and reliable. Eye tracking, GRS scores  
22 and self-report metrics were valid and correlated with step and error metrics. We intend to  
23 use our metrics to study the effectiveness of translation of interscalene block from the  
24 cadaver simulator to the clinical setting.

## References

1. RCoA. CCT in Anaesthetics - Higher Level Training (Annex D). 2 version 1.8 ed. London, UK: Royal College of Anaesthetists, 2011; August 2010 | Version 1.8.
2. McIndoe AK. Modern anaesthesia training: is it good enough? *Br J Anaesth* 2012; **109**: 16-20.
3. Munirama S, Zealley K, Schwab A, *et al.* Trainee anaesthetist diagnosis of intraneural injection-a study comparing B-mode ultrasound with the fusion of B-mode and elastography in the soft embalmed Thiel cadaver model. *Br J Anaesth* 2016; **117**: 792-800.
4. Barrington MJ, Wong DM, Slater B, Ivanusic JJ, Ovens M. Ultrasound-guided regional anesthesia: how much practice do novices require before achieving competency in ultrasound needle visualization using a cadaver model. *Reg Anesth Pain Med* 2012; **37**: 334-9.
5. Ahmed OM, O'Donnell BD, Gallagher AG, Shorten GD. Development of performance and error metrics for ultrasound-guided axillary brachial plexus block. *Adv Med Educ Pract* 2017; **8**: 257-63.
6. Ahmed OM, O'Donnell BD, Gallagher AG, Breslin DS, Nix CM, Shorten GD. Construct validity of a novel assessment tool for ultrasound-guided axillary brachial plexus block. *Anaesthesia* 2016; **71**: 1324-31.
7. Chen XX, Trivedi V, AlSaflan AA, *et al.* Ultrasound-Guided Regional Anesthesia Simulation Training: A Systematic Review. *Reg Anesth Pain Med* 2017.
8. Gallagher AG. Metric-based simulation training to proficiency in medical education:- what it is and how to do it. *Ulster Med J* 2012; **81**: 107-13.
9. Barsuk JH, Cohen ER, McGaghie WC, Wayne DB. Long-term retention of central venous catheter insertion skills after simulation-based mastery learning. *Acad Med* 2010; **85**: S9-12.
10. 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**: 201-9.
11. Byrne AJ BM, McDougall SJ. . Dynamic confidence during simulated clinical tasks. . *Postgraduate Medical Journal* 2005; **81**: 785-8.
12. Chuan A, Graham PL, Wong DM, *et al.* Design and validation of the Regional Anaesthesia Procedural Skills Assessment Tool. *Anaesthesia* 2015; **70**: 1401-11.
13. 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.
14. Cheung JJ, Chen EW, Darani R, McCartney CJ, 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**: 329-33.
15. Sultan SF, Iohom G, Saunders J, Shorten G. A clinical assessment tool for ultrasound-guided axillary brachial plexus block. *Acta Anaesthesiol Scand* 2012; **56**: 616-23.
16. Borg LK, Harrison TK, Kou A, *et al.* Preliminary Experience Using Eye-Tracking Technology to Differentiate Novice and Expert Image Interpretation for Ultrasound-Guided Regional Anesthesia. *J Ultrasound Med* 2018; **37**: 329-36.
17. Harrison TK, Kim TE, Kou A, *et al.* Feasibility of eye-tracking technology to quantify expertise in ultrasound-guided regional anesthesia. *J Anesth* 2016; **30**: 530-3.
18. McKendrick M, Corner G, Tafili T, *et al.* Development of a progressive clinical expertise indicator for utrasound-guided regional anaesthesia. *Br J Anaesth* 2016; **116**: e938.
19. Munirama S, Eisma R, Columb M, Corner GA, McLeod GA. Physical properties and functional alignment of soft-embalmed Thiel human cadaver when used as a simulator for ultrasound-guided regional anaesthesia. *Br J Anaesth* 2016; **116**: 699-707.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
20. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007; **19**: 349-57.
  21. Richie J, Spencer L. Qualitative data analysis for applied policy research. In: Burgess Ba, ed. *Analysing Qualitative Data*. London, 1994; 173-94.
  22. Standards for Educational and Psychological Testing American Psychological Association, & National Council on Measurement in Education, 2014.
  23. Dreyfus HL, S.E. D. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* New York: The Free Press (Macmillan), 1986.
  24. Munirama S, Joy J, Columb M, *et al*. A randomised, single-blind technical study comparing the ultrasonic visibility of smooth-surfaced and textured needles in a soft embalmed cadaver model. *Anaesthesia* 2015; **70**: 537-42.
  25. Thiel W. [Supplement to the conservation of an entire cadaver according to W. Thiel]. *Ann Anat* 2002; **184**: 267-9.
  26. Benkhadra M, Faust A, Ladoire S, *et al*. Comparison of fresh and Thiel's embalmed cadavers according to the suitability for ultrasound-guided regional anesthesia of the cervical region. *Surg Radiol Anat* 2009; **31**: 531-5.
  27. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther* 1995; **33**: 335-43.
  28. Akerstedt T, Gillberg M. Subjective and objective sleepiness in the active individual. *Int J Neurosci* 1990; **52**: 29-37.
  29. Veale JF. Edinburgh Handedness Inventory - Short Form: a revised version based on confirmatory factor analysis. *Laterality* 2014; **19**: 164-77.
  30. Thomann J, Baumann CR, Landolt HP, Werth E. Psychomotor vigilance task demonstrates impaired vigilance in disorders with excessive daytime sleepiness. *J Clin Sleep Med* 2014; **10**: 1019-24.
  31. Lee IS, Bardwell WA, Ancoli-Israel S, Dimsdale JE. Number of lapses during the psychomotor vigilance task as an objective measure of fatigue. *J Clin Sleep Med* 2010; **6**: 163-8.
  32. Shepard RN, Metzler J. Mental rotation of three-dimensional objects. *Science* 1971; **171**: 701-3.
  33. Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J. 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* 1997; **35**: 747-58.
  34. Wilson KM, Finkbeiner KM, de Joux NR, Russell PN, Helton WS. Go-stimuli proportion influences response strategy in a sustained attention to response task. *Exp Brain Res* 2016; **234**: 2989-98.
  35. Neisser U. Decision time without reaction time: experiments in visual scanning. *Am J Psychology* 1963; **76**: d376 - 85.
  36. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach* 2012; **34**: 960-92.
  37. Randolph JJ. Online Kappa Calculator [Computer software]. <http://justus.randolph.name/kappa>. 2008.
  38. Naik VN, Perlas A, Chandra DB, Chung DY, Chan VW. An assessment tool for brachial plexus regional anesthesia performance: establishing construct validity and reliability. *Reg Anesth Pain Med* 2007; **32**: 41-5.
  39. Shafqat A, Ferguson E, Thanawala V, Bedfordth NM, Hardman JG, McCahon RA. Visuospatial Ability as a Predictor of Novice Performance in Ultrasound-guided Regional Anesthesia. *Anesthesiology* 2015; **123**: 1188-97.

- 1  
2  
3 40. Borg LK, Harrison TK, Kou A, *et al.* Preliminary Experience Using Eye-Tracking  
4 Technology to Differentiate Novice and Expert Image Interpretation for Ultrasound-Guided  
5 Regional Anesthesia. *J Ultrasound Med* 2017.  
6  
7 41. Shafqat A, Rafi M, Thanawala V, Bedforth NM, Hardman JG, McCahon RA. Validity  
8 and reliability of an objective structured assessment tool for performance of ultrasound-  
9 guided regional anaesthesia. *Br J Anaesth* 2018; **121**: 867-75.  
10  
11 36. Law, B., Atkins, M.S., Kirkpatrick, A.E., and Lomax, A.J. Eye gaze patterns  
12 differentiate novice and experts in a virtual laparoscopic surgery training environment. in:  
13 Proceedings of the 2004 symposium on eye tracking research & applications. ; 2004: 41–48  
14 View in Article | Crossref | Google Scholar  
15  
16 37. Wilson, M.R., McGrath, J.S., Vine, S.J., Brewer, J., Defriend, D., and Masters, R.S.  
17 Perceptual impairment and psychomotor control in virtual laparoscopic surgery. *Surg*  
18 *Endosc.* 2011; 25: 2268–2274  
19 View in Article | Crossref | PubMed | Scopus (27) | Google Scholar  
20  
21 38. Wilson, M.R., Vine, S.J., Bright, E., Masters, R.S., Defriend, D., and McGrath, J.S. Gaze  
22 training enhances laparoscopic technical skill acquisition and multi-stepping performance: a  
23 randomized, controlled study. *Surg Endosc.* 2011; 25: 3731–3739  
24 <https://doi.org/10.1007/s00464-011-1802-2>  
25  
26 39. Meghanathan, R.N, van Leeuwen, C. and Nikolaev, A.R. Fixation duration surpasses  
27 pupil size as a measure of memory load in free viewing. 2014; 8: 1063.

### 28 Acknowledgments

29  
30 We wish to thank the NIAA for award of the BJA/RCoA project grant (WKRO-2017-0060), and  
31  
32 the RA-UK research grant (WKRO-2018-0016) to fund this study. We also wish to thank the  
33  
34 staff of the Centre for Anatomy and Human Identification (CAHID) at the University of  
35  
36 Dundee.

37  
38  
39  
40  
41 GM, MM, JK designed the study and wrote the paper

42  
43 JL developed the checklist

44  
45 AT coordinated the cadaver study

46  
47 GM and GC provided technical support

48  
49 AS, JH, AM, JS, and PR assessed videos  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21 Table 1 Psychometric and cognitive tests  
22  
23  
24

25 Table 2 Checklist development. Fifty five steps (A) and errors (B) identified and categorised  
26 into seven groups. Inter-rater agreement of iterations 1 and 2 shown in fourth and fifth  
27 columns. Only items with inter-rater reliability  $> 0.80$  indicated consensus amongst experts.  
28 Items indicated by asterisk \* chosen for final steps and error checklists. NA indicates not  
29 accepted for final checklist as not appropriate for translation from cadavers to patients.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40 Table 3. Psychological, visuo-spatial and motor tests. No differences between groups.  
41  
42  
43

44 Table 4. Variance calculated from generalizability theory. Checklist items (steps and errors),  
45 subjects and their interaction comprise account for most measurable variance. Subject garde  
46 associated with more variance around errors. Raters make very small contribution to  
47 variance. Residual variance is substantial for both checklists.  
48  
49  
50  
51  
52  
53

54 Table 5. Eye tracking results of experts and novices. Experts took less time, glanced less  
55 often, fixated for longer on the needle on the monitor.  
56  
57  
58  
59  
60



1  
2  
3  
4  
5  
6 Fig 1. Rasch item characteristic curves. A steps, B errors. The x-axis represents student  
7 ability. Zero represents an average student; points to the right of the zero represent a  
8 comparatively better student and to the left, a worse student. The y-axis represents the  
9 probability that the subject conducts a steps or fails to notice an error. Curves to the left  
10 suggest easier steps and errors more likely to be noticed, and vice versa. For example,  
11 identifying a nerve is relatively easy (steps 4) whereas visualising the needle tip at all times is  
12 difficult (steps10). With regard to errors, the probability of failing to alter image depth,  
13 handle the transducer and to identify the nerve epineurium on scanning is low, whereas the  
14 probability of failing to recognise needle nerve contact (error 13), spread of fluid (error 14)  
15 or enter the nerve is high (error 15). The steeper the slope, the higher the item  
16 discrimination. Slope angles indicate that errors are more discriminatory than steps.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

31 Fig 2.

32  
33 Scanning and needling phases of interscalene block on soft embalmed Thiel cadaver  
34 performed by subject no. 1 (expert) and subject no. 2 (novice). Ventral nerve roots of C5 and  
35 C6 visible. No difference in anatomical features between images. No fluid accumulation or  
36 needling tracking. The needle shaft is well aligned and tip positioned well by expert. In  
37 contrast, needle poorly aligned by novice and traversing C6 ventral nerve root.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Function	Scale/Steps	Description
Mood	Depression Anxiety Stress Scales (DASS) <sup>23</sup>	Three self-report scales <ul style="list-style-type: none"> <li>• Depression (dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest and inertia).</li> <li>• Anxiety (autonomic arousal, skeletal muscle effects, situational anxiety).</li> <li>• Stress (difficulty relaxing, nervous arousal, and being easily upset/agitated, irritability and impatience).</li> </ul>
Sleepiness	The Karolinska Sleep Scale <sup>24</sup>	9-item scale from 1 (extremely alert) to 9 (extremely sleepy).
Handedness	The Edinburgh Handedness Scale <sup>25</sup>	4-item scale cored in units of 50 from -100 (always left) to +100 (always right)
Wakefulness	Psychomotor Vigilance test <sup>26, 27</sup>	Sensitive to sleep loss. Reaction time to appearance of red stopwatch on screen.
Visuo-spatial ability	Mental Rotation Test <sup>28</sup>	Mentally rotate two-dimensional images to match each other.
Dexterity	Pursuit Motor steps	Follow moving dot round circle. Performance based on time on and off target.
Sustained attention	Sustained Attention to Response Steps (SART) <sup>29, 30</sup>	Presented with a single digit from 1 to 9 on screen in varying font sizes. Asked to press spacebar if any digit other than 3 was seen (Go trial) and to withhold the response if digit 3 was seen (inappropriate response to NoGo).
Visual scanning	Visual scanning <sup>31</sup>	Participants presented with a letter matrix consisting of 25 rows of 5 letters. Stepped with finding the target letter 'K' within 10 seconds and typing in row number within 4 seconds. Responses measured on percentage of correct responses and reaction time.

A

No.	Category	Items – steps	Delphi 1	Delphi 2	Final checklist items*
1	Positioning & Preparation	How important do you consider discussing the block with trainees beforehand?	0.75		
2		How important do you consider the pre-block positioning of the cadaver joints and limbs?	1.00	1.00	*
3		How important do you consider the positioning of the ultrasound machine	0.75		
4		How important do you consider making sure the ultrasound cables are not on the floor?	0.50		
5		How important do you consider covering the transducer with a sheath?	0.75		
6		How important do you consider sterile technique?	0.63		
7		How important do you consider making the trainee sit comfortably?	0.75		
8	Pre-procedural tasks	How important do you consider stopping before you block?	0.63		
9		How important do you consider palpation of anatomical landmarks before undertaking a block?	0.25		
10		How important do you consider flushing the needle before skin insertion?	0.63		
11		How important do you consider matching needle length to the type of block?	0.63		
12		How important do you consider the application of gel to the transducer?	0.63		
13		How important do you consider the choice of transducer?	1.00	1.00	NA
14		How important do you consider altering image depth and gain on the ultrasound machine in order to optimise the image?	1.00	1.00	*
15		How important do you consider how the transducer is handled during scanning (rotation, tilt, pressure)?	1.00	1.00	*
16		How important do you consider alignment of the transducer to the screen image?	0.75		
17		How important do you consider scanning and identifying the nerve?	1.00	1.00	*
18	Pre-procedural scanning	How important do you consider scanning and identifying the nerve epineurium?	0.63		
19		How important do you consider scanning and identifying blood vessels?	1.00	1.00	NR
20		How important do you consider scanning and identifying muscles?	0.38		
21		How important do you consider scanning and identifying muscles fascial planes?	0.63		
22		How important do you consider the level of trainee focus while conducting the pre-procedural scan?	1.00	1.00	NR
23		How important do you consider scanning proximally and distally before conducting the block?	0.88	1.00	*
24		Needle	How important do you consider the choice of needle insertion site?	0.75	
25	How important do you consider trainees looking at their hands when conducting the block?		0.75		
26	How important do you consider aligning the needle to the transducer?		1.00	1.00	*
27	How important do you consider checking the needle trajectory?		0.75	1.00	NR
28	How important do you consider aligning the needle at a tangent to the nerve?		1.00	0.82	*
29	How important do you consider optimising the visibility of the target nerve?		1.00	0.94	*
30	How important do you consider keeping the transducer still?		0.63		
31	How important do you consider identifying the entire length of the needle?		0.50		
32	In order to improve visibility of the needle, how important do you consider moving the transducer rather than the needle?		0.63		
33	How important do you consider keeping the target nerve in the middle of the screen?		0.00		
34	In order to improve visibility of the needle, how important do you consider checking the orientation of the needle with the transducer at the skin surface?		0.88	0.69	
35	Needle tip	How important do you consider being aware of the rate of needle insertion?	0.38		
36		How important do you consider identifying the needle tip before advancing the needle?	1.00	1.00	*
37		How important do you consider visualising the needle tip at all times?	0.88	0.88	*
38		How important do you consider being able to adjust the position of the needle tip?	1.00	1.00	*
39		How important do you consider needle tip migration?			
40		How important do you consider quickly regaining needle tip position when tip visibility is lost?	0.88		
41		How important do you consider being able to identify the needle tip before injection?	1.00	1.00	*
42		How important do you consider needing one needle pass?	0.00		
43		How important do you consider using hydrolocation if the needle tip is not seen?	0.63		
44		How important do you consider injecting at the best possible anatomical site?	0.50		
45		How important do you consider being ambidextrous?	0.13		
46	How important do you consider being able to inject as close to but not touching the epineurium?	0.50			
47	Needle tip feedback	How important do you consider feeling fascial pops?	0.38		
48		How important do you consider recognition of needle nerve contact?	0.88	0.94	*
49		How important do you consider assessment of injection pressure?	0.38		
50	Fluid injection	How important do you consider injection of a 0.5ml to 1ml hydrolocation bolus of Thiel embalming fluid in order to confirm needle tip position	0.63		
51		How important do you consider circumferential local anaesthetic spread?	0.50		
52		How important do you consider knowing how much local anaesthetic has been injected?	0.88	0.69	
53		How important do you consider knowing where local anaesthetic has spread?	0.88	1.00	*
54		How important do you consider recognition of intraneural injection?	1.00	0.94	*
55		How important do you consider trainees communicating that they are out of their depth?	1.00	1.00	NR

## B

No.	Category	Items – errors including sentinel errors	Delphi 1	Delphi 2	Final checklist items*
1		What would be the likelihood of causing patient harm by failing to discuss the block beforehand?	0.50	0.31	
		What would be the potential consequence of such harm?	0.88	0.88	NA
2		What would be the likelihood of causing patient harm by failing to position the joints and limbs appropriately?	0.50		
		What would be the potential consequence of such harm?	0.75		
3		What would be the likelihood of causing patient harm by failing to see the ultrasound machine and the cadaver at the same time?	0.50		
		What would be the potential consequence of such harm?	0.75		
4	Positioning & Preparation	What would be the likelihood of causing patient harm by failing to remove ultrasound cables on the floor?	0.13		
		What would be the potential consequence of such harm?	0.63		
		What would be the likelihood of causing patient harm by failing to cover the transducer with a sheath?	0.50		
		What would be the potential consequence of such harm?	0.38		
		What would be the likelihood of causing patient harm by failing to use sterile technique?	0.50		
		What would be the potential consequence of such harm?	0.50		
		What would be the likelihood of causing patient harm by failing to get the trainee sitting comfortably?	0.13		
		What would be the potential consequence of such harm?	0.38		
9	Pre-procedural tasks	What would be the likelihood of causing patient harm by failing to stop before you block	1.00		
		What would be the potential consequence of such harm?	0.25		
		What would be the likelihood of causing patient harm by failing to palpate anatomical landmarks before undertaking a block	0.00		
		What would be the potential consequence of such harm?	0.00		
		What would be the likelihood of causing patient harm by failing to flush the needle?	0.25		
		What would be the potential consequence of such harm?	0.50		
		What would be the likelihood of causing patient harm by failing to match the needle length to the type of block?	0.25		
		What would be the potential consequence of such harm?	0.50		
		What would be the likelihood of causing patient harm by failing to apply gel to the transducer?	0.63		
		What would be the potential consequence of such harm?	0.63		
		What would be the likelihood of causing patient harm by failing to choose the appropriate transducer?	0.38		
		What would be the potential consequence of such harm?	0.38		
		What would be the likelihood of causing patient harm by failing to alter image depth and gain?	0.50	1.00	*
		What would be the potential consequence of such harm?	0.88	0.94	
		What would be the likelihood of causing patient harm by failing to handle the transducer appropriately during scanning?	0.63	0.69	*
		What would be the potential consequence of such harm?	1.00	0.94	
17	Pre-procedural scanning	What would be the likelihood of causing patient harm by failing to align the transducer to the screen image?	0.50		
		What would be the potential consequence of such harm?	1.00		
		What would be the likelihood of causing patient harm by failing to identify the nerve on scanning?	1.00		
		What would be the potential consequence of such harm?	1.00		
		What would be the likelihood of causing patient harm by failing to identify the nerve epineurium on scanning?	0.50	0.56	*
		What would be the potential consequence of such harm?	0.88	1.00	
		What would be the likelihood of causing patient harm by failing to identify the blood vessels on scanning?	0.88	0.81	NA
		What would be the potential consequence of such harm?	1.00	1.00	
		What would be the likelihood of causing patient harm by failing to identify the muscles on scanning?	0.13		
		What would be the potential consequence of such harm?	0.38		
		What would be the likelihood of causing patient harm by failing to identify the fascial planes on scanning?	0.13		
What would be the potential consequence of such harm?	0.38				
22	Needle	What would be the likelihood of causing patient harm by failing to focus during the pre-procedural scan?	0.88	0.69	
		What would be the potential consequence of such harm?	1.00	0.81	
		What would be the likelihood of causing patient harm by failing to scan proximally and distally before conducting the block?	0.63		
		What would be the potential consequence of such harm?	0.63		
		What would be the likelihood of causing patient harm by failing to identify the most appropriate needle insertion site?	0.38		
		What would be the potential consequence of such harm?	0.50		
		What would be the likelihood of causing patient harm by the trainee failing to not look at his/her hands?	0.38		
		What would be the potential consequence of such harm?	0.63		
		What would be the likelihood of causing patient harm by failing to align the needle to the transducer?	1.00	1.00	*
		What would be the potential consequence of such harm?	1.00	1.00	
		What would be the likelihood of causing patient harm by failing to check the needle trajectory?	0.50		
What would be the potential consequence of such harm?	0.63				
What would be the likelihood of causing patient harm by failing to align the needle at a tangent to the nerve?	0.63	0.50	*		
What would be the potential consequence of such harm?	1.00	0.94			
What would be the likelihood of causing patient harm by failing to optimise the visibility of the target nerve?	0.88	1.00	*		
What would be the potential consequence of such harm?	1.00	1.00			
What would be the likelihood of causing patient harm by failing to keep the transducer still?	0.38				
What would be the potential consequence of such harm?	0.63				

31		What would be the likelihood of causing patient harm by failing to identify the entire length of the needle?	0.50	0.63	*
		What would be the potential consequence of such harm?	0.88	1.00	
32		What would be the likelihood of causing patient harm by failing to move the transducer rather than the needle?	0.50	0.38	
		What would be the potential consequence of such harm?	0.75	0.75	
33		What would be the likelihood of causing patient harm by failing to keep the target nerve in the middle of the screen?	0.13		
		What would be the potential consequence of such harm?	0.25		
34		What would be the likelihood of causing patient harm by failing to check the orientation of the needle with the transducer at the skin surface?	0.50		
		What would be the potential consequence of such harm?	0.50		
35		What would be the likelihood of causing patient harm by failing to check the rate of needle insertion?	0.38		
		What would be the potential consequence of such harm?	0.50		
36	Needle tip	What would be the likelihood of causing patient harm by failing to identify the needle tip before advancing the needle?	0.88	1.00	*
		What would be the potential consequence of such harm?	1.00	1.00	
37		What would be the likelihood of causing patient harm by failing to identify the needle tip at all times?	0.63	0.81	*
		What would be the potential consequence of such harm?	0.88	1.00	
38		What would be the likelihood of causing patient harm by failing to identify the fascial planes?	0.50		
		What would be the potential consequence of such harm?	0.63		
39		What would be the likelihood of causing patient harm by failing to recognise needle tip migration?	1.00	0.94	*
		What would be the potential consequence of such harm?	1.00	1.00	
40		What would be the likelihood of causing patient harm by failing to quickly regaining needle tip position when tip visibility was lost?	0.88	0.82	*
		What would be the potential consequence of such harm?	0.88	1.00	
41		What would be the likelihood of causing patient harm by failing to identify the needle tip before injection?	1.00	0.94	*
		What would be the potential consequence of such harm?	1.00	1.00	
42		What would be the likelihood of causing patient harm by needing more than one needle pass?	0.00		
		What would be the potential consequence of such harm?	0.13		
43		What would be the likelihood of causing patient harm using hydrolocation if the needle tip was not seen?	0.13		
		What would be the potential consequence of such harm?	0.38		
44		What would be the likelihood of causing patient harm by failing to inject at the best possible anatomical site?	0.13		
		What would be the potential consequence of such harm?	0.13		
45		What would be the likelihood of causing patient harm by not being ambidextrous?	0.00		
		What would be the potential consequence of such harm?	0.00		
46		What would be the likelihood of causing patient harm by failing to inject as close to but not touching the epineurium?	0.00		
		What would be the potential consequence of such harm?	0.13		
47		Needle tip feedback	What would be the likelihood of causing patient harm by failing to feel fascial pops?	0.13	
	What would be the potential consequence of such harm?		0.13		
48	What would be the likelihood of causing patient harm by failing to identify needle nerve contact?		0.88	0.88	*
	What would be the potential consequence of such harm?	1.00	0.94		
49	What would be the likelihood of causing patient harm by failing to assess injection pressure?	0.50	0.38		
	What would be the potential consequence of such harm?	1.00	1.00	*	
50	Fluid injection	What would be the likelihood of causing patient harm by failing to identify the hydrolocation bolus?	0.38		
		What would be the potential consequence of such harm?	0.25		
51		What would be the likelihood of causing patient harm by failing to provide circumferential local anaesthetic spread?	0.13		
		What would be the potential consequence of such harm?	0.13		
52		What would be the likelihood of causing patient harm by failing to note how much local anaesthetic had been injected?	0.50	0.63	NA
		What would be the potential consequence of such harm?	0.75	1.00	
53		What would be the likelihood of causing patient harm by failing to know where local anaesthetic has spread?	0.50		
		What would be the potential consequence of such harm?	0.63		
54		What would be the likelihood of causing patient harm by failing to identify intraneural injection?	1.00	1.00	*
		What would be the potential consequence of such harm?	1.00	1.00	
55	What would be the likelihood of causing patient harm by failing to communicate they are out of their depth?	1.00	1.00	NA	
	What would be the potential consequence of such harm?	0.88	1.00		

	<b>Novice</b>	<b>Expert</b>	<b>p</b>
Total DASS	9 [6 - 44]	12 [2 - 36]	0.94
Depression	2 [0 - 14]	4 [0 - 28]	0.56
Anxiety	2 [0 - 6]	2 [0 - 2]	0.69
Stress	7 [2 - 26]	6 [0 - 26]	0.68
Handedness	94 [-75 - 100]	94 [-75 - 100]	1.00
Subjective sleepiness	3 [2 - 4]	2 [1 - 5]	0.18
Wakefulness	0 [0 - 1]	1 [0 - 2]	0.32
Mental rotation (% correct)	80 [70 - 90]	75 [60 - 100]	0.68
Mental rotation time (s)	4.1 [2.4 - 35.2]	8.0 [5.4 - 27.5]	0.13
Pursuit rotor. Time on target (%)	99 [96 - 100]	99 [98 - 100]	0.24
Sustained attention Time on Go Trials (s)	3.44 [3.13 - 4.83]	3.25 [2.80 - 4.28]	0.49
Sustained attention. Commission errors No-Go trials (n)	10 [4 - 15]	15 [4 - 23]	0.17
Generic visual scanning skills (%)	52 [31 - 81]	36 [0 - 64]	0.31
Generic visual scanning skills (s)	56.5 [52.0 - 69.0]	63.0 [56.9 - 75.2]	0.24

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

	<b>Steps</b>		<b>Errors</b>	
	<b>Variance</b>	<b>% of total</b>	<b>Variance</b>	<b>% of total</b>
		<b>variance</b>		<b>variance</b>
<b>Item</b>	0.049	18.5	0.052	19.0
<b>Subject</b>	0.020	7.4	0.033	11.8
<b>Grade</b>	0.016	6.2	0.044	15.9
<b>Rater</b>	0.003	1.3	0.003	1.3
<b>Subject:Item</b>	0.017	6.3	0.030	11.0
<b>Subject:Grade</b>	0.010	3.8	0.012	4.3
<b>Rater:Item</b>	0.030	11.4	0.006	2.1
<b>Residual</b>	0.120	45.2	0.095	34.5

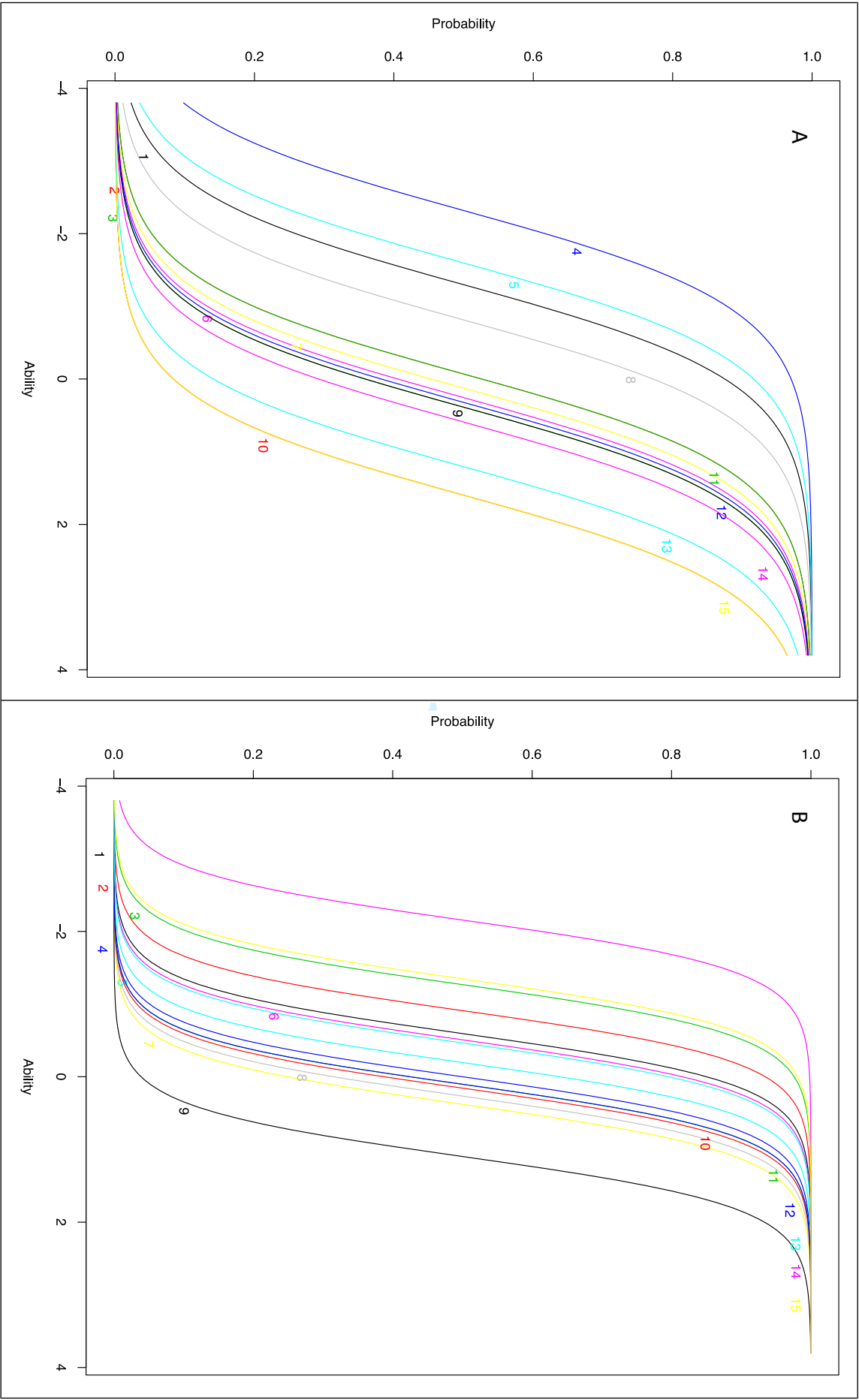
For Peer Review

		Novice Median [Range]	Expert Median [Range]	P-value
<b>Duration of task(s)</b>				
	Pre-Procedural	98.4 [27.3 - 282.5]	55.3 [29.2 - 87.9]	0.09
	Procedural	175.4 [103.7 - 367.7]	71.3 [39.4 - 171.6]	0.04
	Total	274.3 [152.9 - 650.2]	126.1 [96.4 - 232.4]	0.03
<b>Dwell time (s)</b>				
	Pre-procedural			
	Monitor	70.0 [13.7 - 94.1]	48.4 [26.7 - 99.4]	0.18
	Tools	5.1 [0.1 - 80.3]	1.6 [0.5 - 8.7]	0.06
	Other	2.2 [0 - 9.7]	0.5 [0 - 2.5]	0.19
	Total	83.8 [14.5 - 160.4]	49.4 [28.7 - 73.3]	0.18
	Procedural			
	Monitor	53.4 [17.3 - 168.3]	59.8 [36.2 - 141.1]	0.39
	Tools	11.9 [0.4 - 43.5]	9.8 [2.0 - 58.9]	0.06
	Other	0.7 [0.0 - 7.2]	0.0 [0.0 - 0.5]	0.04
	Total	69.2 [18.2 - 212.6]	85.0 [38.2 - 152.5]	0.06
<b>Glance (n)</b>				
	Scanning			
	Monitor	8.5 [4 - 15]	3.5 [3 - 9]	0.04
	Tools	5.5 [1 - 9]	2.5 [1 - 6]	0.32
	Other	3.5 [0 - 11]	0.5 [0 - 3]	0.08
	Total	17 [8 - 31]	7 [5 - 18]	0.02
	Needling			
	Monitor	12.5 [3 - 27]	4 [1 - 12]	0.06
	Tools	6 [3 - 24]	4 [1 - 8]	0.19
	Other	3 [0 - 18]	0 [0 - 8]	0.12
	Total	29 [6 - 56]	8 [2 - 28]	0.07
<b>Fixation (ms)</b>				
	Scanning			
	Monitor	351.3 [194.1 - 801.8]	471.4 [401.2 - 754.9]	0.49
	Tools	215.0 [133.1 - 248.9]	223.1 [166.4 - 371.6]	0.70
	Other	192.0 [0.0 - 276.4]	115.7 [0 - 632.4]	0.95
	Needling			
	Monitor	334.7 [187 - 923]	974.9 [710 - 1336.6]	0.03
	Tools	306.5 [217 - 522]	258.8 [133.1 - 990.0]	0.60
	Other	161.0 [0 - 182]	0 [0 - 228.4]	0.17
<b>Fixation count (n)</b>				
	Scanning			
	Monitor	127 [41 - 249]	85 [38 - 136]	0.18
	Tools	18 [1 - 27]	5.5 [3 - 33]	0.79
	Other	8 [0-31]	1.5 [0-10]	0.13
	Total	168 [46 - 293]	90 [44 - 179]	0.18
	Needling			
	Monitor	160 [75-394]	60 [28-170]	0.07
	Tools	42 [16 - 131]	23 [2 - 36]	0.07
	Other	4 [0 - 35]	0 [0 - 36]	0.18
	Total	222 [91 - 530]	81 [30 - 242]	0.048


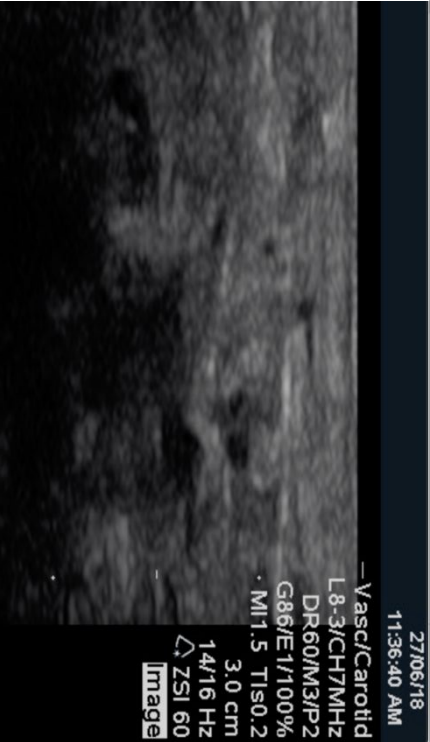




1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

	Scanning phase	Needling phase	
Expert. Participant No. 1			
Novice Participant No. 12			

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46