Article Type: Correspondence

Regional anaesthesia assessment tools – a reply

We thank Ahmed et al. for their interest in our paper and for seeking further information on the educational concepts underlying the design of the Regional Anaesthesia Procedural Skills (RAPS) tool [1]. With regards to the descriptors for the checklist items, it is important to not confuse subjectivity with context. Items were designed to encompass the behaviour that was being assessed, while each institution further applies the behaviour to their local context.

As an example, 'sets up equipment properly' is a necessary step in regional anaesthesia performance. However, this will have different meanings for an institution that uses a specially-designed block room with fixed locations for equipment, using an ultrasound machine from Company A, compared to another institution without a procedure bay using a machine from Company B. One can write a tool with step-by-step descriptors, but being excessively prescriptive limits cross-institutional applicability, as well as creating an unwieldy tool that defeats clinical feasibility.

Instead, we (like others [2]) encourage users of assessment tools to define the behaviour in context, and this was the approach taken in our previous work in evaluating assessment tools [3, 4]. We also point out that, despite no training in the use of RAPS, we found that the checklist was highly reliable amongst the ten experts (representing eight different institutions across five Australian cities) during the test-retest validation phase of our study, suggesting that users can successfully interpret and apply the items included in it.

The assertion that Likert-based global rating scales are less reliable than dichotomous checklists is not supported by the evidence. Both types share similar characteristics [5-7]. Indeed, we found for the RAPS tool that both the checklist and global rating scale had identical intraclass correlation coefficients (ICC) of 0.80 during the clinical validation phase. A recent systematic review of 45 studies has confirmed this, finding that global rating scales are actually more likely to discriminate expert behaviours, and slightly more reliable than checklists [8]. We agree with Bould et al's editorial that 'the current best evidence for a gold standard for assessment of procedural skills in anaesthesia consists of a combination of

previously validated checklists and global rating scales, used prospectively by a trained observer, for a procedure performed in an actual patient' [9].

It is not entirely correct that ICC does not measure agreement. We advise that ICC has six different forms, and it is important to articulate which form is used for analysis [10, 11]. In our study methods, we specified the ICC (2A, *k*) model, which measures both absolute agreement and correlation between assessors' score differences. This type of ICC is actually a very stringent test of reliability.

The consultant anaesthetist and anaesthesia assistant were advised not to prepare, comment, or intervene during filming of trainees performing the procedure. They were permitted to intervene if they believed patient safety was about to be compromised. In the videos, this intervention was almost always in the form of a verbal prompt to assess the trainee's cognitive process. If the trainee's response still showed a lack of understanding, the supervising anaesthetist provided a corrective statement. This automatically constituted an unsatisfactory performance in that aspect of the block.

Lastly, there is no dispute that different regional anaesthesia techniques activate different cognitive abilities and psychomotor skills. Acknowledgement of these different levels of complexity already exists by defining blocks as basic, intermediate, and advanced. However, we do not believe that each block deserves it's own assessment tool. Instead, we have shown in our study that the RAPS tool can encompass these different skillsets, and achieve successful assessment of performance with good reliability, validity, and feasibility.

A. Chuan
P. L. Graham
K. Forrest
Macquarie University,
Sydney, Australia.
M. J. Barrington
C. F. Royse
University of Melbourne,
Melbourne, Australia.
D. M. Wong
St Vincent's Hospital,
Melbourne, Australia.
A. J. D. Cameron
Middlemore Hospital,

Auckland, New Zealand.

Y. C. Lim

Changi General Hospital,

Singapore.

D. B. Auyong

Virginia Mason Medical Centre,

Seattle, WA, USA.

Email: alwin.chuan@mq.edu.au

AC received financial support from the National Health and Medical Research Council (APP1056280) and the Australian Society of Anaesthetists (PhD Support Grant) to assist this work. No other conflicts of interest declared. Previously posted on the *Anaesthesia* correspondence website: www.anaesthesiacorrespondence.com

References

1. Chuan A, Graham PL, Wong DM, et al. Design and validation of the Regional Anaesthesia Procedural Skills Assessment Tool. *Anaesthesia* 2015; **70**: 1401-11.
2. Hill F, Kendall K, Galbraith K, Crossley J. Implementing the undergraduate mini-CEX: a tailored approach at Southampton University. *Medical Education* 2009; **43**: 326-34.
3. Wong DM, Watson MJ, Kluger R, et al. Evaluation of a task-specific checklist and global rating scale for ultrasound-guided regional anesthesia. *Regional Anesthesia and Pain Medicine* 2014; **39**: 399-408.
4. Watson MJ, Wong DM, Kluger R, et al. Psychometric evaluation of a direct observation of procedural skills assessment tool for ultrasound-guided regional anaesthesia. *Anaesthesia* 2014; **69**: 604-12.
5. Cunnington JPW, Neville AJ, Norman GR. The risks of thoroughness: Reliability and validity of global ratings and checklists in an OSCE. *Advances in Health Sciences Education* 1996; **1**: 227-33.
6. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine* 1998; **73**: 993-7.
7. Morgan PJ, Cleave-Hogg D, Guest CB. A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Academic Medicine* 2001; **76**: 1053-5.
8. Ilgen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for

checklists versus global rating scales in simulation-based assessment. *Medical Education* 2015; **49**: 161-73.

9. Bould MD, Crabtree NA, Naik VN. Assessment of procedural skills in anaesthesia. *British Journal of Anaesthesia* 2009; **103**: 472-83.
10. Bartko JJ. On various intraclass correlation reliability coefficients. *Psychological Bulletin* 1976; **83**: 762-5.
11. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 1979; **86**: 420-8.