BJA

Competency-based assessment tools for regional anaesthesia: a narrative review

A. Chuan[1,2,*], A. S. Wan[2], C. F. Royse[3,4] and K. Forrest[5]

[1]South Western Sydney Clinical School, University of New South Wales, Sydney, Australia, [2]Department of Anaesthesia, Liverpool Hospital, Sydney, Australia, [3]Department of Surgery, University of Melbourne, Melbourne, Australia, [4]Department of Anaesthesia, Royal Melbourne Hospital, Melbourne, Australia and [5]Faculty of Health Sciences and Medicine, Bond University, Queensland, Australia

*Corresponding author. PO Box 3438, Putney, New South Wales 2112, Australia. E-mail: dr.chuan@iinet.net.au

Abstract

Competency-based assessment tools are used in regional anaesthesia to measure the performance of study participants, trainees, and consultants. This narrative review was performed to appraise currently published assessment tools for regional anaesthesia. A literature search found 397 citations of which 28 peer-reviewed studies met the inclusion criteria of primary psychometric evaluation of assessment tools for regional anaesthesia. The included studies were diverse in the type of assessment and the skill set being assessed. The types of assessments included multiple-choice questions, hand-motion analysis, cumulative sum, visuospatial and psychomotor screening, checklists, and global rating scales. The skill sets that were assessed included holistic regional anaesthesia technical and non-technical performance observed at the bedside, to isolated part-tasks, such as needle tip visualisation under ultrasound. To evaluate validity and reliability, we compared the studies against published medical education consensus statements on ideal assessment tools. We discuss the relative merits of different tools when used to assess regional anaesthesia, the importance of psychometrically robust assessment tools in competency-based anaesthesia education, and directions for future education research in regional anaesthesia.

Key words: anaesthesia; conduction; credentialing; psychometrics; reproducibility of results

The assessment of complex anaesthesia procedures, such as regional anaesthesia, requires evaluation of disparate knowledge, skill sets, and professional attributes. Practically, this is achieved by the use of assessment tools.[1] With the international shift towards competency-based postgraduate medical education, these tools assume a central role in current anaesthesia curricula by assessing the trainees' performances against predefined learning objectives, standards of performance, and behaviour.[1–6] The performances can range from the entire nerve block procedure on patients in a clinical setting, to part-tasks relevant to regional anaesthesia, such as anatomy knowledge and ultrasound-guided needle skills in simulation settings.

The design, structure, and quantitative characteristics of an assessment tool are defined as its psychometric properties. Evaluating a tool's psychometrics ('assessing the assessment tool') is necessary for acceptance by all stakeholders. These properties include validity, reliability, feasibility, educational stimulus, and acceptability. For clinicians, validity and reliability are required to demonstrate robustness and confidence

in assessment results. In research studies, validated and reliable tools are needed to measure performance changes after training interventions and for serial measurement of performance over time.[7]

Until relatively recently, the principles of assessment and the methods of psychometric evaluation were not well disseminated beyond the medical education literature. To improve the consistency and robustness of studies that sought to design and evaluate assessment tools, the 2010 Ottawa Conference for Competence in Medicine and Healthcare published a consensus statement on ideal assessment tool properties,[8] describing a framework of reporting standards and statistical techniques used in psychometric evaluation.[9] We adopted this framework to appraise the published regional anaesthesia assessment tools. This narrative review discusses the type and purpose of current tools, and categorizes their psychometric properties, implications for clinical adoption, and priorities for research.

Methods

Criteria for ideal assessment tools

A utility index of psychometric properties of the ideal competency-based assessment tool was first proposed by Van der Vleuten[10] in 1996. This was officially adopted as a consensus statement at the 2010 Ottawa Conference as the document 'Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference'.[8] The properties that defined a good assessment tool included validity, reliability, feasibility, educational stimulus, and acceptability of a tool by stakeholders. Table 1 summarizes these properties with a specific focus on regional anaesthesia.

In practice, every assessment tool is a compromise between competing properties. This interdependent balance is most pertinent for validity and reliability. At a minimum, this requires evidence that the important skill sets of that task are examined (face and content validity), whether the tool differentiate trainees of different abilities (construct validity) and reliability (reproducibility of results).[12,13] Of the different types of reliability, external reliability is more important in a clinical context, as it concerns consistency of scoring among different faculty members who assess trainees.

Content validity is typically assured by anchoring with expert consensus guidelines. Examples include the regional anaesthesia learning objectives in the 2015 Accreditation Council for Graduate Medical Education Milestone project.[14] For ultrasound-guided techniques, the joint committees of the American Society of Regional Anesthesia and Pain Medicine and the European Society of Regional Anaesthesia and Pain Therapy describe minimum competencies.[15]

Framework for psychometric evaluation

Psychometric evaluation is an examination of these assessment tool properties against how well it conforms to purpose. These often use quantitative statistical analyses. In research studies designing new assessment tools, a primary aim is to evaluate reliability. However, previous reviews have noted limitations in the statistical analyses of reliability used by published studies.[4,16] These limitations include heterogeneity and incorrect use of tests, precluding any meta-analyses of tools. In an effort to improve the quality of medical education research, an additional 2010 Ottawa Conference statement on

Table 1 Psychometric properties of an ideal competency-based regional anaesthesia assessment tool. Adapted from the 2010 Ottawa Conference for Competence in Medicine and Healthcare 'Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference' by Norcini and colleagues.[8] Further definitions from Bould and colleagues,[1] Ahmed and colleagues,[4] Van der Vleuten,[10] and Gallagher and colleagues[11]

Property	Description and comments
Validity	The tool measures what it purports to measure. A body of evidence, built up over multiple studies, supports validity. Must also demonstrate reliability
Face validity	Suitability of the tool to the real-life regional anaesthesia clinical reality
Content validity	Criterion-based regional anaesthesia knowledge, skill set, or behaviour being tested
Construct validity	Test scores differentiate inexperienced and experienced proceduralists.
Discriminate validity	Test scores differentiate novice *vs* expert proceduralists.
Concurrent validity	Test scores are consistent with an external and objective outcome. Examples have included measures of block performance, such as success, time to onset, number of needle passes, and complication rates.
Predictive validity	Test scores are predictive of clinical practice. An example is *in vitro* simulation scores predict *in vivo* block performance.
Reliability	Reproducibility of test scores if repeated under similar circumstances External reliability is more important than internal reliability. In practice, trainees are scored by different assessors and over multiple occasions throughout training; thus, demonstration of external reliability is essential for validity and acceptability.
Internal reliability	Different trainees scoring consistently in the same section of the tool
External reliability	Different assessors scoring consistently for the same trainee performance Also called between-assessor or inter-rater agreement
Test–retest reliability	Consistency of scores over different times for the same performance
Feasibility	Minimal time, logistics, financial costs, and assessor training required
Educational catalyst	Allows structured feedback, deliberate practice, and reflection (formative assessment)
Acceptability	Acceptability dependent on the stakeholder and purpose of tool For example, medical registration boards accept tools if there is evidence of high reliability and validity. Feasibility and educational effect are not as important. Clinical tutors will accept a tool for its educational benefit and high feasibility. Other properties are not as critical.