

Development and Evaluation of a New Formative Assessment of Surgical Performance

Ronnie M. Davies, MB/BChir,* Lisa Hadfield-Law, MSc,^{†,1} and Philip G. Turner, MB, ChB^{‡,2}

*Department of Orthopaedics, Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom; [†]British Orthopaedic Association, Royal College of Surgeons, London, United Kingdom; and [‡]Department of Orthopaedics, Stepping Hill Hospital, Stockport, United Kingdom

BACKGROUND: Formative assessment of operative performance is a mandatory part of surgical training. Engagement with assessment is limited in part by the time-consuming nature and the high perceived stakes of current assessment tools.

OBJECTIVES: Our aims were to develop and collect validity evidence for a new operative assessment tool that addresses barriers to assessment that current trainers and trainees experience.

METHODS: We developed the Generic Operative Supervised Learning Event (GOSLE). Orthopedic trainees were invited to complete GOSLEs with their trainers after surgical procedures. Experienced consultants assessed videotaped operations performed by trainees using the GOSLE. Validity evidence for content, relationships to other scores, internal structure, response process, and consequences of testing were evaluated.

RESULTS: A total of 250 GOSLEs were completed. A strong correlation was found between the GOSLE scores and the Procedure-Based Assessment ratings ($r = 0.87$, $p < 0.001$). Rasch analysis confirmed satisfactory internal structure of the rating scale, with sequential increases in rating as performance improved. The reproducibility coefficient was 0.88, with 10 assessments of the same trainee who has to achieve a reliability coefficient of 0.8. Over 90% of users found the GOSLE easy to use, with most preferring it to other assessment methods. Feedback quality was higher using the GOSLE than with current assessments.

CONCLUSION: We have collected validity evidence across multiple domains in support of the GOSLE. Its psychometric performance is comparable to that of current assessments. It is preferred by trainers and trainees over existing assessments. It stimulates high-quality, actionable feedback which better supports formative assessment. By addressing issues experienced with existing assessments, we expect engagement among users to be high. (J Surg Ed ■■■■■. © 2018 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: educational measurement, clinical competence, psychometrics, formative feedback, education

COMPETENCIES: Practice-Based Learning and Improvement

INTRODUCTION

Over the past 10 years, surgical training and assessment has shifted from an apprenticeship model to a competency-based model.¹ To facilitate the introduction of this training model, a raft of workplace-based assessment (WBA) tools was developed to monitor, both formatively and summatively, trainees' progress. These assessments are used to demonstrate competency, providing much-needed quality assurance of both a trainee's abilities and the training program itself.

The mainstay of assessment of operative performance in the United Kingdom is the Procedure-based Assessment (PBA). It is a lengthy assessment, comprising 40 to 50 tick boxes and 12 free-text areas. It was introduced prior to any formal validation, beyond a Delphic process of agreement among experts, and that it was a valid assessment method.² Several years following its introduction, a large validation study was undertaken that focused on the reliability of its rating scales.³ There is little doubt that they demonstrated

Role of funding source: No funding was received for this study.

Correspondence: Inquiries to Ronnie M. Davies, MB/BChir, Department of Orthopaedics, Central Manchester University Hospitals NHS Foundation Trust, Oxford Road, Manchester M13 9WL, United Kingdom; e-mail: ronnie@etailers.co.uk

¹ Surgical Educationalist, British Orthopaedic Association, 35-43 Lincoln's Inn Fields, London WC2A 3PE, United Kingdom.

² Department of Orthopaedics, Stepping Hill Hospital, Stockport SK2 7JE, United Kingdom.

acceptable reliability and construct validity of the PBA. Using traditional validation frameworks, subsequent validation studies confirm that the PBA is a valid and reliable measure of performance⁴ and responsive to change.⁵ This has made it the current gold standard in assessment of operative performance.

Unfortunately, engagement with the PBA has been poor. The initial validation study reported “mixed, but predominantly positive” feedback relating to user satisfaction with the PBA. This is reflected in subsequent studies, confirming that trainees and trainers viewed the PBA as a mere form-filling exercise.⁶ Further work has clarified the reasons behind poor engagement with the PBAs. This includes the time taken to complete assessments, inability to demonstrate progression due to a coarse rating scale, and general assessment fatigue.⁷ In spite of issues with PBAs and WBA in general, there has been little change in their structure since their introduction.

The aim of this study was to design a new assessment tool that addresses issues with current assessments, and to collect evidence from multiple sources to assess its validity. After describing the rationale behind the structure of the new assessment, its systematic validation will be presented.

METHODS

The new assessment was designed by the authors, who represent the key stakeholders in operative assessment: medical education experts, a senior surgical trainer, and a senior surgical trainee, with over 50 years’ cumulative experience of surgical assessment. A design brief was produced for a new assessment based on published best practices and issues with the PBA. The overarching principles were to develop an assessment that

- is quick to complete,
- has a rating scale that demonstrates progression,
- makes clear the distinction between formative and summative assessment,
- requires minimal training in its use,
- is applicable to any surgical procedure, and
- has educational impact.

To address the dichotomy between formative and summative assessment, it was designed to be a solely formative assessment coinciding with guidance regarding assessment published by the General Medical Council.⁸ It was named the Generic Operative Supervised Learning Event (GOSLE) and produced as an online form that can be completed for any procedure listed in eLogbook (Appendix 1). The electronic version can be accessed at <http://www.gosle.co.uk>.

The rating scale was designed from scratch to discriminate between performance levels and eliminate the floor and ceiling effect currently observed. Previous 4-point rating scales were too coarse to demonstrate progression. Eight levels were chosen, conforming to psychometric norms (Table 1).⁹ Ratings chosen were construct-aligned descriptive levels, which are more reliable than an ordinal scale in an assessment of performance.¹⁰ The scale was based on the Zwisch model of skill acquisition¹¹ because it closely follows the gradual withdrawal of expert guidance from the trainer. A task-specific checklist was not used, as it increases the time to complete an assessment and is less reliable than global ratings.¹²

Feedback is invited on the GOSLE in 3 separate boxes, 2 of which are compulsory. One focuses on reinforcing areas of good practice and another on areas for improvement. This feedback structure is familiar to the surgical workforce. All usage instructions are provided in-line with the GOSLE so that it may be used without any formal training. The assessment was piloted on various surgical training courses, and feedback from users was sought to incrementally improve the assessment until a final version was established.

Validation of the GOSLE followed a contemporary framework. This defines validity as the “appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores.”¹³ It is now considered to be a unitary construct. Validation is defined as the “hypothesis-driven process of accumulating evidence to support such inferences.” Using this contemporary framework, validity evidence is collected from multiple sources, including relationships to other variables, content, internal structure, response process, and consequences of testing. In contrast, traditional validation frameworks collect evidence for different types of validity, for example, face validity, construct validity, and predictive validity.

TABLE 1. GOSLE Rating Scale, with Comparison to Zwisch Model

Level	Description	Zwisch Model
1	Able to assist with guidance (was not familiar with all steps of procedure)	Show and tell
2	Able to assist without guidance (knew all steps of procedure and anticipated next move)	
3	Guidance required for most/all of the procedure (or part performed)	
4	Guidance or intervention required for key steps only	Smart help
5	Procedure performed with minimal guidance or intervention (needed occasional help)	
6	Procedure performed competently without guidance or intervention but rather hesitant	Dumb help
7	Procedure performed confidently and fluently without any guidance or intervention	
8	As above and was able to anticipate, avoid, and/or deal with common problems/complications	No help

Download English Version:

<https://daneshyari.com/en/article/10222796>

Download Persian Version:

<https://daneshyari.com/article/10222796>

[Daneshyari.com](https://daneshyari.com)