



<b>Title</b>	<b>Assessing the Surgical Skills of Trainees in the Operating Theatre: A Prospective Observational Study of the Methodology</b>
<b>Agency</b>	NETSCC, HTA, NIHR Evaluation and Trials Coordinating Centre Alpha House, University of Southampton Science Park, Southampton, SO16 7NS, United Kingdom; Tel: +44 2380 595 586, Fax: +44 2380 595 639; hta@soton.ac.uk, www.hta.ac.uk
<b>Reference</b>	Volume 15.01. ISSN 1366-5278. <a href="http://www.hta.ac.uk/project/1626.asp">www.hta.ac.uk/project/1626.asp</a>

## Aim

To compare user satisfaction and acceptability, reliability, and validity of 3 different methods of assessing the surgical skills of trainees by direct observation in the operating theatre across a range of surgical specialties and index procedures.

## Conclusions and results

Of 558 patients, 437 (78%) cases were included in the study, and 51 consultant clinical supervisors, 56 anaesthetists, 39 nurses, 2 surgical care practitioners, and 4 independent assessors provided 1635 assessments on 85 trainees undertaking the 437 cases. In total, 749 Procedure Based Assessments (PBAs), 695 Non Technical Skills for Surgeons (NOTSS), and 191 Objective Structured Assessment of Technical Skills (OSATS) were performed. Non obstetrics and gynecology (O&G) clinical supervisors and trainees provided mixed, but predominantly positive, responses on a range of applications of PBA. Most felt that PBA was important in surgical education, would use it again, and did not feel that it added time to the operating list. Overall satisfaction of O&G clinical supervisors and trainees with OSATS was not as high, and most those who used both preferred PBA. Most anaesthetists and nurses felt that NOTSS allowed them to rate interpersonal skills more easily than cognitive skills, that it had formative value, and that it was a valuable adjunct in assessing technical skills. PBA demonstrated high reliability ( $G > 0.8$  for only 3 assessor judgments on the same index procedure). OSATS had lower reliability ( $G > 0.8$  for 5 assessor judgments on the same index procedure). Both were less reliable on a mix of procedures because of strong procedure-specific factors. A direct comparison of PBA between O&G and non O&G cases showed a striking difference in reliability. Within O&G, a good level of reliability ( $G > 0.8$ ) could not be obtained using a feasible number of assessments. Conversely, the reliability within non O&G cases was exceptionally high, with only 2 assessor judgments being required. The reasons for this difference probably include the more summative purpose of

assessment in O&G and the much higher proportion of O&G trainees in this study with training concerns. The reliability of NOTSS was lower than that for PBA. Reliability for the same procedure ( $G > 0.8$ ) required 6 assessor judgments. However, as procedure-specific factors exerted a lesser influence on NOTSS, reliability on a mix of procedures could be achieved using only 8 assessor judgments. PBA and NOTSS showed better construct validity than OSATS, the year of training and the number of recent index procedures performed being significant independent predictors of performance. We found little variation in scoring between different procedures or different designations of assessor. The results suggest that PBA reliable, valid, and acceptable in assessing technical skills of surgical trainees.

## Recommendations

Specialties that use OSATS may wish to consider changing the design or switching to PBA. NOTSS should be considered for the assessment of nontechnical skills.

## Methods

The 3 methods selected were PBA, OSATS, and NOTSS. PBA is used routinely to assess the technical skills of surgical trainees, OSATS is used in the same way for obstetric and gynecology (O&G) trainees, and NOTSS is a newly developed tool for assessing nontechnical skills.

## Further research/reviews required

Whatever workplace-based assessment method is used, the purpose, timing, and frequency of assessment require detailed guidance.