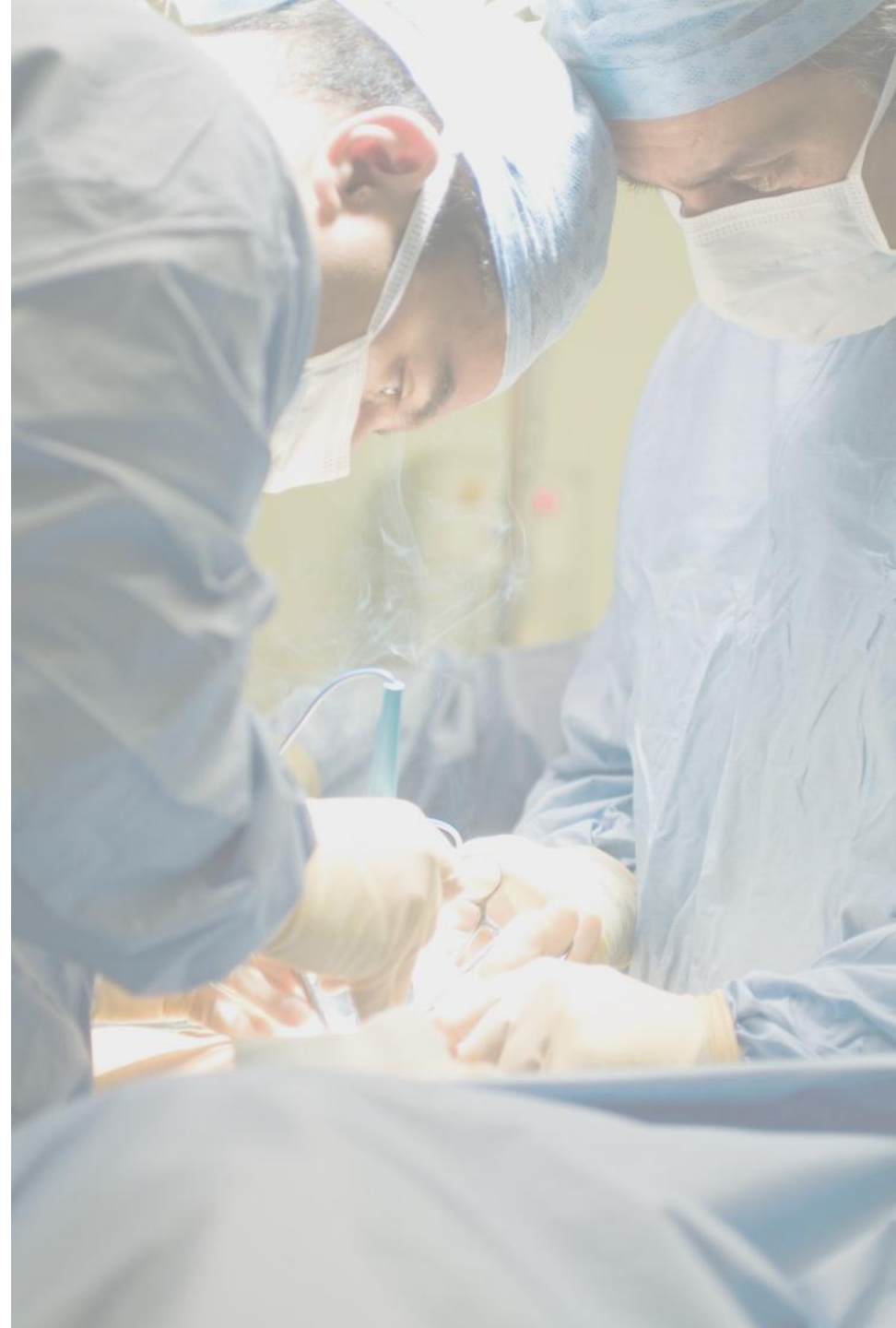


Principles of skills assessment

Dr Nick Sevdalis
Senior Lecturer

Email: n.sevdalis@imperial.ac.uk

Imperial College
London



Learning outcomes for this lecture

1. Define terms used within surgical education
2. Compare and contrast different modes of assessment (technical & nontechnical)
3. Select appropriate tools for training and assessment of surgical skills
4. Define the multidimensional nature of surgical performance
5. Evaluate research studies (in simulation) and their limitations

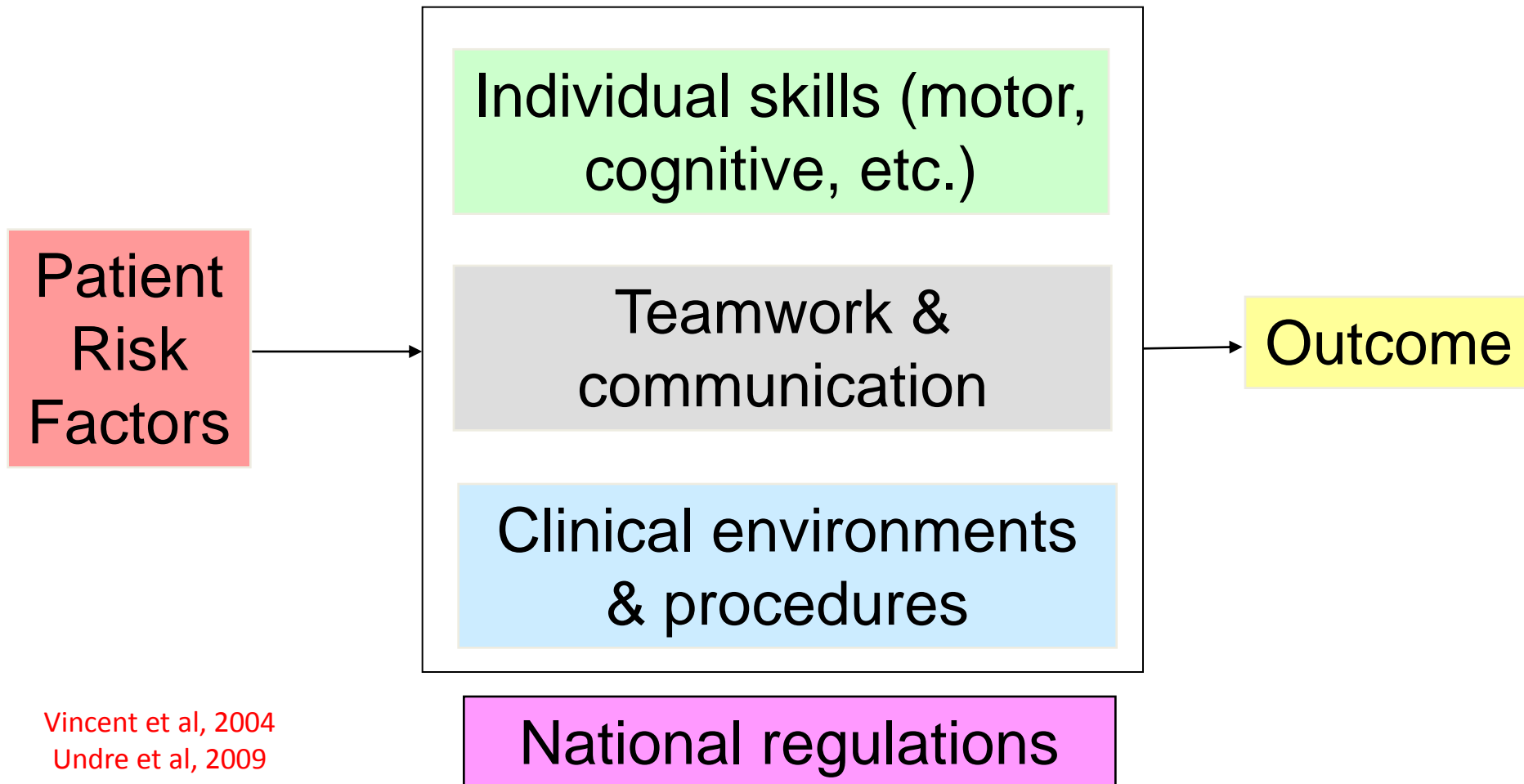


What do YOU think makes a good surgeon?

Small group work (10min) & feedback to class (10mins)



Systems approach to surgical performance





Can Skills and Performance be Measured Scientifically?

“We can only be sure to improve what we can actually measure”

Professor A Darzi
High Quality Care for All
Department of Health (UK) 2008

**What does 'scientific measurement'
consist of?**

Scientific measurement

- **I measure sth = I made an observation and then I assigned a number to my observation**
- Shoe size, intelligence, reaction time, diagnostic accuracy, surgical dexterity

Scientific measurement

- Four levels of measurement:
 - **Nominal**: values are categories (e.g. gender)
 - **Ordinal**: values can be rank-ordered, but we have no idea (or we don't care about) what's going on in the intervals ("class" of mark)
 - **Interval/continuous**: values can be rank-ordered and we can assume that the intervals between them are all equal (scales, 1-5, 1-10, etc).
 - **Ratio/continuous**: values can be rank-ordered; we know the intervals are equal; the scale has a meaningful 0-point (speed, blood loss).

Reliability & validity

- **Scientific measurement ought to be**
 - **Reliable:** reproducible and consistent
 - **Valid:** it should capture what it is intended to capture



Unreliable & Invalid



Unreliable, But Valid



Reliable, Not Valid



Both Reliable & Valid

Key types of reliability

- **Inter-rater:** do 2 or more assessors agree in their assessments of a skill?
- **Internal consistency** (for multi-item scales): if you take the scale as a whole, do different items tend to get scored in the same manner?
 - Cronbach alpha
- **Test-retest:** if you administer an assessment repeatedly, do the results correlate adequately?
 - Correlation coefficients

Key types of validity

- **Face:** does an assessment instrument 'appear' valid to those who will use it?
- **Content:** does an assessment instrument adequately cover all aspects of performance or skill in question?
- **Construct:** this reflect the evidence that we have on the underlying 'construct' (e.g., surgical performance, technical skill, etc) and our scientific understanding of it
 - *Concurrent:* instrument scores correlate with an independent criterion at present
 - *Predictive:* the scores correlate with an independent criterion in the future
 - *Convergent:* 2 different instruments meant to capture similar constructs produce correlated scores
 - *Discriminant:* 2 different instrument meant to capture different constructs produce uncorrelated scores

Key types of validity – more recent

- **Functional:** what are you going to use an assessment instrument for? Is it fit for your purpose?

Take home message:

**from a scientific point of view,
instrument validation never ends**

(we can never know 'enough')



Questions so far?

Reliance on observational methods

- **Technical skills**
 - Psychomotor dexterity and coordination
- **Nontechnical skills**
 - Cognitive and behavioural skills (communication, leadership, etc)
- **Teamworking in operating theatres (elsewhere)**
 - How well a team works together to look after a patient and carry out a procedure
- **Environmental impact on performance**
 - Distractions, interruptions, etc

Surgeons' technical skills

(a) Objective Structured Assessment of Technical Skill (Martin et al, Br J Surg 1998)

General Skill	1	2	3	4	5
Respect For Tissue	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
Time & Motion	Many unnecessary moves		Efficient time/motion but some unnecessary moves		Economy of movement and maximum efficiency
Instrument Handling	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments although occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
Knowledge of Instruments	Frequently asked for the wrong instrument or used an inappropriate instrument		Knew the names of most instruments and used appropriate instrument or the task		Obviously familiar with the instruments required and their names
Use of Assistants	Consistently placed assistants poorly or failed to use assistants		Good use of assistants most of the time		Strategically used assistant to the best advantage at all times
Flow of Operation & Forward Planning	Frequently stopped operating or needed to discuss next move		Demonstrated ability for forward planning with steady progression of operative procedure		Obviously planned course of operation with effortless flow from one move to the next
Knowledge of Specific Procedure	Deficient knowledge. Needed specific instruction at most operative steps		Knew all important aspects of the operation		Demonstrated familiarity with all aspects of the operation

Table 1
Global rating scale component of the intraoperative assessment tool*

Depth perception

1. Constantly overshoots target, wide swings, slow to correct
- 2.
3. Some overshooting or missing of target, but quick to correct
- 4.
5. Accurately directs instruments in the correct plane to target

Bimanual dexterity

1. Uses only one hand, ignores nondominant hand, poor coordination between hands
- 2.
3. Uses both hands, but does not optimize interaction between hands
- 4.
5. Expertly uses both hands in a complimentary manner to provide optimal exposure

Efficiency

1. Uncertain, inefficient efforts; many tentative movements; constantly changing focus or persisting without progress
- 2.
3. Slow, but planned movements are reasonably organized
- 4.
5. Confident, efficient and safe conduct, maintains focus on task until it is better performed by way of an alternative approach

Tissue handling

1. Rough movements, tears tissue, injures adjacent structures, poor grasper control, grasper frequently slips
- 2.
3. Handles tissues reasonably well, minor trauma to adjacent tissue (ie, occasional unnecessary bleeding or slipping of the grasper)
- 4.
5. Handles tissues well, applies appropriate traction, negligible injury to adjacent structures

Autonomy

1. Unable to complete entire task, even with verbal guidance
- 2.
3. Able to complete task safely with moderate guidance
- 4.
5. Able to complete task independently without prompting

Table 2
Task-specific checklist: dissection of the gallbladder from the liver bed

Tasks	Done (1 point)	Not done (0 points)
1. Uses cautery only when all conducting areas are in field of view		
2. Has good control of the instrument, minimizes recoil		
3. Grasps gallbladder near clips to begin dissection		
4. Readjusts tension on gallbladder to optimize exposure		
5. Avoids dissecting into liver causing undue bleeding		
6. Avoids perforation of the gallbladder		
7. Avoids spillage of gallstones		
8. Maximizes useful dissection in 1 area before changing approach		
9. Performs dissection in appropriate plane the majority of the time		
10. Obviates the need for surgeon takeover		
Total		/10

**GLOBAL OPERATIVE ASSESSMENT
OF LAPAROSCOPIC SKILLS
(GOALS)**

Vassiliou et al 2005

Evidence base

The American Journal of Surgery (2011) 202, 469–480

The American
Journal of Surgery

Review

Observational tools for assessment of procedural skills: a systematic review

Kamran Ahmed, M.B.B.S., M.R.C.S., Danilo Miskovic, M.D., F.R.C.S.,
Ara Darzi, M.D., F.R.C.S., F.A.C.S., K.B.E., Thanos Athanasiou, Ph.D., F.E.T.C.S.,
George B. Hanna, Ph.D., F.R.C.S.*

Review

Objective assessment of technical surgical skills

P. D. van Hove^{1,2,4}, G. J. M. Tuijthof^{1,3}, E. G. G. Verdaasdonk², L. P. S. Stassen⁴ and J. Dankelman¹

¹Department of Biomechanical Engineering, Delft University of Technology, and ²Department of Surgery, Reinier de Graaf Group, Delft, ³Department of Orthopaedic Surgery, Amsterdam Medical Centre, Amsterdam, and ⁴Department of Surgery, Maastricht University Medical Centre, Maastricht, The Netherlands

Correspondence to: Dr P. D. van Hove, Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands (e-mail: P.D.vanHove@tudelft.nl)

Surgeons' non-technical skills

Aviation

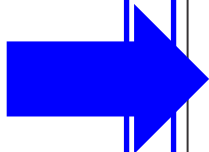
Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR



NLR-TP-98518

NOTECHS:
Non-technical skill evaluation in JAR-FCL

J.A.G. van Avermaete



Categories	Elements	Example Behaviours
COOPERATION	Team building and maintaining	- Establishes atmosphere for open communication and participation
	Considering others	- Takes condition of other crew members into account
	Supporting others	- Helps other crew members in demanding situation
	Conflict solving	- Concentrates on what is right rather than who is right
LEADERSHIP & MANAGERIAL SKILLS	Use of authority and assertiveness	- Takes initiative to ensure involvement and task completion
	Maintaining standards	- Intervenes if task completion deviates from standards
	Planning and co-ordinating	- Clearly states intentions and goals
	Workload management	- Allocates enough time to complete tasks
SITUATION AWARENESS	System awareness	- Monitors and reports changes in system's states
	Environmental awareness	- Collects information about the environment
	Anticipation	- Identifies possible future problems
DECISION MAKING	Problem definition / diagnosis	- Reviews causal factors with other crew members
	Option generation	- States alternative courses of action - Asks other crew member for options
	Risk assessment / Option choice	- Considers and shares risks of alternative courses of action
	Outcome review	- Checks outcome against plan

Very Poor	Poor	Acceptable	Good	Very Good
Observed behaviour directly endangers flight safety	Observed behaviour in other conditions could endanger flight safety	Observed behaviour does not endanger flight safety but needs improvement	Observed behaviour enhances flight safety	Observed behaviour optimally enhances flight safety and could serve as an example for other pilots

van Avermaete, 1998

Non-Technical Skills for Surgeons (NOTSS)

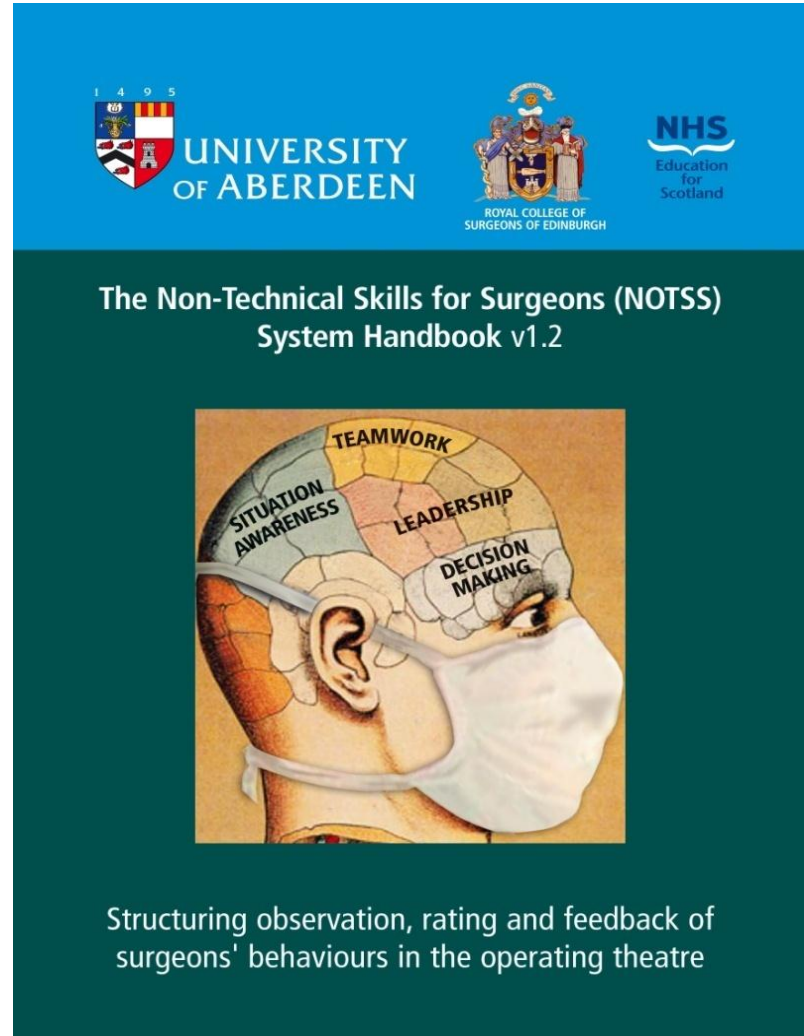
4-Good: Performance was of a consistently high standard, enhancing patient safety

3-Acceptable: Performance was of a satisfactory standard

2-Marginal: Performance indicated cause for concern

1-Poor: Performance endangered or potentially endangered patient safety

Not Applicable: Skill not required/not relevant

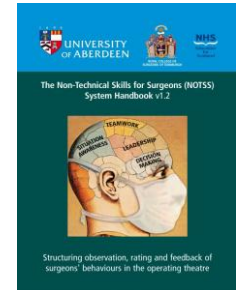


The Non-Technical Skills for Surgeons (NOTSS)
System Handbook v1.2

Structuring observation, rating and feedback of surgeons' behaviours in the operating theatre

NOTSS

Scoring: 1-4



Category	Category rating*	Element	Element rating*	Feedback on performance and debriefing notes
Situation Awareness	3	Gathering information	2	
		Understanding information	4	
		Projecting and anticipating future state	3	
Decision Making	3	Considering options	2	Consider discussing the decision to convert with the anaesthetist next time
		Selecting and communicating option	3	
		Implementing and reviewing decisions	3	
Leadership	2	Setting and maintaining standards	3	
		Supporting others	2	Ensure you delegate tasks appropriately
		Coping with pressure	N/A	
Communication and Teamwork	1	Exchanging information	2	Be more precise when asking for instruments
		Establishing a shared understanding	1	Brief theatre personnel beforehand about the operation and your expectations
		Co-ordinating team activities	4	

**Surgeons as members of the
operating theatre team**

Operating theatre team: OTAS

OTAS® Working Version – Sep 2009

Imperial College
London

OBSERVATIONAL TEAMWORK ASSESSMENT FOR SURGERY®

Surgical Team – Intra-Operative Phase

RATING ANCHORS	BRIEF ANCHOR DEFINITION
6	Exemplary behaviour; very highly effective in enhancing team function
5	Behaviour enhances highly team function
4	Behaviour enhances moderately team function
3	Team function neither hindered nor enhanced by behaviour
2	Slight detriment to team function through lack of/inadequate behaviour
1	Team function compromised through lack of/inadequate behaviour
0	Problematic behaviour; team function severely hindered

BEHAVIOUR	DEFINITION	RATING SCALE						
COMMUNICATION	Quality and quantity of information exchanged among team members	0	1	2	3	4	5	6
COORDINATION	Management and timing of activities and tasks	0	1	2	3	4	5	6
COOPERATION/ BACK UP BEHAVIOUR	Assistance provided among members of the team, supporting others, and correcting errors	0	1	2	3	4	5	6
LEADERSHIP	Provision of directions, assertiveness, and support among members of the team	0	1	2	3	4	5	6
MONITORING/ SITUATIONAL AWARENESS	Team observation and awareness of ongoing processes	0	1	2	3	4	5	6

	EXAMPLE/SAMPLE BEHAVIOURS
COMMUNICATION	Asks team if all prepared to begin the operation Requests and instructions to team communicated clearly and effectively Provides information to whole team on progress Surgeon informs the team of technical difficulties and /or changes of plan
COORDINATION	Gives prior notification of requirements to Scrub Nurse to enhance timing of instrument exchange Surgeons co-ordinate use of equipment, such as camera in minimal access surgery providing adequate view of operating field Contribute to smooth exchange of instruments and provisions with Scrub Nurse
COOPERATION/ BACK UP BEHAVIOUR	Reacts positively to questions and requests from Nursing group Responds to requests or questions from Anaesthetic group Helps with smooth instrument exchange with Scrub Nurse Supports Surgical group assistants and compensates for lack of experience
LEADERSHIP	Instructions and explanations provided to assistants Advises Anaesthetist if unfamiliar with operative techniques (e.g. tube insertion) to call for senior help Supervision provided for staff lacking familiarity with tasks or equipment
MONITORING/ SITUATIONAL AWARENESS	Check table positioning and positions of members Assistants monitor direction of light Checks team condition Aware of patient condition including anaesthesia

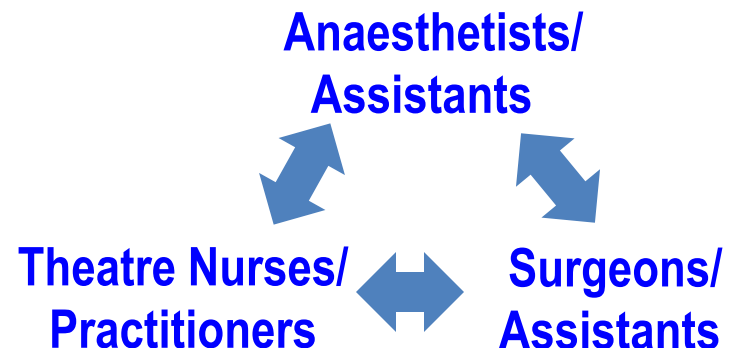
Imperial College
London

CPSSQ
Centre for Patient Safety & Service Quality

Observational Teamwork Assessment for Surgery (OTAS)

USER Training Manual (draft)

February 2011



OTAS team skills

Communication: Quality and quantity of information exchanged among team members

Leadership: Provision of directions, assertiveness, and support among members of the team

Mutual Support/Cooperation: Assistance provided among members of the team, supporting others, and correcting errors

Team monitoring/Situational awareness: Team observation and awareness of ongoing processes

Team coordination: Management and timing of activities and tasks

OTAS rating scale

7-Point Scale



0	1	2	3	4	5	6
Team function severely hindered	Team function compromised	Slight detriment to team function	Team function neither hindered nor enhanced	Team function moderately enhanced	Team function highly enhanced	Exemplary; team function very highly enhanced

COMPROMISED ← - - - - - **PATIENT SAFETY** - - - - - → **ENHANCED**

Evidence base

COLLECTIVE REVIEW

The Impact of Nontechnical Skills on Technical Performance in Surgery: A Systematic Review

Louise Hull, MSc, Sonal Arora, PhD, MRCS, Rajesh Aggarwal, PhD, FRCS, Ara Darzi, MD, FACS, Charles Vincent, PhD, Nick Sevdalis, PhD

-
- BACKGROUND:** Failures in nontechnical and teamwork skills frequently lie at the heart of harm and near-misses in the operating room (OR). The purpose of this systematic review was to assess the impact of nontechnical skills on technical performance in surgery.
- STUDY DESIGN:** MEDLINE, EMBASE, PsycINFO databases were searched, and 2,041 articles were identified. After limits were applied, 341 articles were retrieved for evaluation. Of these, 28 articles were accepted for this review. Data were extracted from the articles regarding sample population, study design and setting, measures of nontechnical skills and technical performance, study findings, and limitations.
- RESULTS:** Of the 28 articles that met inclusion criteria, 21 articles assessed the impact of surgeons' nontechnical skills on their technical performance. The evidence suggests that receiving feedback and effectively coping with stressful events in the OR has a beneficial impact on certain aspects of technical performance. Conversely, increased levels of fatigue are associated with detriments to surgical skill. One article assessed the impact of anesthesiologists' nontechnical skills on anesthetic technical performance, finding a strong positive correlation between the 2 skill sets. Finally, 6 articles assessed the impact of multiple nontechnical skills of the entire OR team on surgical performance. A strong relationship between teamwork failure and technical error was empirically demonstrated in these studies.
- CONCLUSIONS:** Evidence suggests that certain nontechnical aspects of performance can enhance or, if lacking, contribute to deterioration of surgeons' technical performance. The precise extent of this effect remains to be elucidated. (J Am Coll Surg 2012;214:214-230. © 2012 by the American College of Surgeons)
-



Questions so far?

Definition of assessment

5-Stage Process

1. OBSERVATION
2. RECORDING
3. INTERPRETATION
4. EVALUATION
5. FEEDBACK



Observational assessments in clinical settings

PRESENT

- Provide trainees feedback on their skill development
- Identify training needs for individuals/teams/department
- Determine whether an intervention is effective
- Audit
- Research

FUTURE?

High-Stakes

- Revalidation
- Selection



The assessment process

Assessor

Assessment Process

Assessee

**OBSERVES, RECORDS, INTERPRETS, EVALUATES,
and PROVIDES FEEDBACK**



**OBSERVES, RECORDS, INTERPRETS, EVALUATES,
and PROVIDES FEEDBACK**



INFORMED and ENGAGED

Requirements for Assessment

Assessor

Fair
Objective
Unbiased



Trained

Assessment Process

Reliable
Valid
Accurate
Fair



**Robust Assessment
Instruments/Tools**

Assessee

Informed
Engaged

Requirements for Assessment

CAN BE RESOURCE INTENSIVE
TIME, COST, TRAINING



Assessor

Fair
Objective
Unbiased



Trained



Assessment Process

Reliable
Valid
Accurate
Fair



**Robust Assessment
Instruments/Tools**



Assessee

Informed
Engaged

Components of Assessment



Assessment Components

OBSERVATION
of a skill

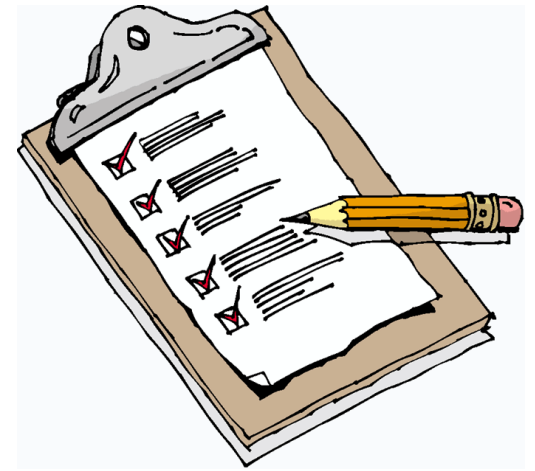
NUMERICAL SCORING
of a skill

Mental Processes:

Detection of behaviour
Perception of behaviour
Recall of behaviour

Mental Processes:

Categorisation of
observations
Mental integration of
observations
Evaluation of observations



Assessors' Biases



Key biases to consider...

- **NON-BLINDED RATINGS**
 - Knowing assesses
 - Previous knowledge of performance/ability
- **HAWTHORNE EFFECT**
 - Knowledge of being assessed → *'angel performance'*

Common rating errors



- **HALO EFFECT**-one particular aspect of behaviour is over-emphasised and enhances the ratings on other behavioural dimensions
- **HORNS EFFECT**-one particular negative aspect of behaviour is overemphasised and reduces the ratings on other behavioural dimensions
- **CENTRAL TENDENCY**-ratings mainly given around the mid-point of the scale

Common rating errors



- **LENIENCY**-tendency to give favourable (higher) ratings
- **SEVERITY**-tendency to give unfavourable (lower) ratings
- **PRIMACY**-remembering better/overweighting behaviours that were observed first
- **RECENCY**-remembering better/overweighting behaviours that were observed last

How do you ensure a valid, reliable, accurate and fair assessment?

(small group work if time allows)



How do the pilots do it?

NASA/TM—2003–212809



A Gold Standards Approach to Training Instructors to Evaluate Crew Performance

*David P. Baker and R. Key Dismukes
Ames Research Center, Moffett Field, California*

Guidelines for assessor training

Guidelines

1. Detailed discussion of skills and behaviours to be evaluated
2. Review the standards of performance associated with each skill or behaviour
3. Include training on how to observe skills and behaviours
4. Provide trainee faculty with opportunities to practice ratings and receive feedback
5. Provide trainee faculty with 'gold standard' rating derived via expert consensus as a benchmark for their own ratings



**Assessing others' skills and performance is a skill in itself –
and thus it requires training**



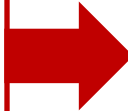
National guidelines for the UK: Faculty training requirements

NASA/TM—2003—212809



A Gold Standards Approach to Training Instructors to Evaluate Crew Performance

David P. Baker and R. Key Dismukes
Ames Research Center, Moffett Field, California



ORIGINAL ARTICLE

Training Faculty in Nontechnical Skill Assessment

National Guidelines on Program Requirements

Louise Hull, MSc, Sonal Arora, PhD, Nicholas R.A. Symons, MSc., Rozh Jalil, MRCS, Ara Darzi, FACS,
Charles Vincent, PhD, and Nick Sevdalis, PhD

Objective: To develop guidelines for the requirements for a faculty training program in nontechnical skill assessment in surgery.

Background: Nontechnical skills in the operating room are critical for patient safety. The successful integration of these skills into workplace-based assessment is dependent upon the availability of faculty who are able to teach and assess them. At present, no guidelines exist regarding the training requirements for such faculty in surgical contexts.

Methods: The development of the guidelines was carried out in several stages: stage 1—a detailed literature review on current training for nontechnical skill assessors; stage 2—semistructured interviews with a multidisciplinary panel (consisting of clinicians and psychologists/human factors specialists) of experts in surgical nontechnical skills; and stage 3—interview findings fed into an Expert Consensus Panel (ECP) Delphi approach to establish consensus regarding training requirements for faculty assessing nontechnical skills in surgery.

Results: The ECP agreed that training in nontechnical skill assessment should be delivered by a multidisciplinary team consisting of clinicians and psychologists/human factors specialists. The ECP reached consensus regarding who should be targeted to be trained as faculty (including proficiency and reevaluation requirements). Consensus was reached on 7 essential training program content elements (including training in providing feedback/debriefing) and 8 essential methods of evaluating the effectiveness of a “train-the-trainers” program.

Conclusions: This study provides evidence-based guidelines that can be used to guide the development and evaluation of programs to educate faculty in the training and assessment of nontechnical skills. Uptake of these guidelines could accelerate the development of surgical expertise required for safe and high-quality patient care.

Keywords: assessment, education, faculty, nontechnical skills, operating room, surgery, teamwork, training, train-the-trainers

(*Ann Surg* 2012;00: 1–6)

have yet to be fully integrated into mainstream surgical education, training, or assessment. The availability of psychometrically robust (ie, with evidence for reliability and validity) assessment tools that measure nontechnical skills, coupled with adequately trained faculty, is essential for this endeavor.

Several tools have been developed to assess nontechnical skills in the operating room, for example, the Observational Teamwork Assessment for Surgery (OTAS) and Anesthetists’ Non-Technical Skills (ANTS).^{6,7} Research has demonstrated the psychometric robustness of such tools, and these are now widely available to the surgical community.^{6,8} Surgical educators, however, remain concerned about the complex faculty training requirements for those who will use these tools to deliver nontechnical skill teaching and assessment.⁹ This is a particular challenge because nontechnical skill assessment tools appear straightforward to use. In contrast, evidence suggests that considerable training is required to use these tools in a valid and reliable manner.^{9,10}

Within surgery, the wide availability of tools to assess nontechnical skills and the lack of recommendations regarding their application is in stark contrast to the stringent regulations enforced by other professions.¹¹ Such regulations are in place to prevent unreliable assessments, which could be unfair and potentially damaging for the individual being assessed.^{8,9} In other high-risk industries, faculty performing behavioral and cognitive assessments receive extensive training, demonstrate a minimum level of proficiency, and hold an accreditation to apply these measures in practice.¹¹ For example, the aviation industry has long recognized the need to training faculty to assess and debrief nontechnical performance as key characteristics of a high-reliability organization.^{12,13} Specific faculty training programs have been developed that focus on teaching novice faculty to identify and assess nontechnical performance in the same way as highly experienced assessors.^{12,13} This principle also applies to modern technical skill assessment in surgery,^{14,15} in which trainers participate in for-

Take home message:

- A biased assessment is one or all of the below:
 - **Unreliable**
 - **Invalid**
 - **Unfair**
 - **Inaccurate**
- A biased assessment is **WORSE** than no assessment at all



Questions?