The complexity of the operating room requires a surgeon to have both technical ability and an array of non-technical skills. Multiple modalities were identified including high fidelity simulation, low fidelity simulation, didactic teaching and crisis resource management. The most valuable teaching modalities were identified as high and low fidelity simulation, which should be integrated into surgical training curricula.
Non-technical Skills in Surgery: A Systematic Review of Current Training

Modalities

Authors:
Esther Ounounou BSc (Hons)
Abdullatif Aydin BSc (Hons), MBBS
Oliver Brunckhorst BSc (Hons), MBBS
M. Shamim Khan OBE, FRCS Urol, FEBU
Prokar Dasgupta MSc, MD, FRCS Urol, FEBU
Kamran Ahmed PhD, FRCS Urol

Affiliations:
MRC Centre for Transplantation, King's College London
Department of Urology, Guy's and St. Thomas’ NHS Foundation Trust
King's Health Partners, London, United Kingdom

Correspondence:
Dr. Abdullatif Aydin
MRC Centre for Transplantation,
5\textsuperscript{th} Floor Southwark Wing, Guy's Hospital,
King's College London, London SE1 9RT
Ph: +44 (0)20 7188 8580
Fax: +44 (0)20 3312 6787
Email: abdullatif.aydin@kcl.ac.uk
Mini-ABSTRACT

The complexity of the operating room requires a surgeon to have both technical ability and an array of non-technical skills. This study aims to identify current methods used to teach non-technical skills and status of their validity, evidence and role in training. The most valuable teaching modalities were identified as high and low fidelity simulation, which should be integrated into surgical training curricula.
ABSTRACT

Background: The complexity of the operating room requires a surgeon to have both technical ability and an array of non-technical skills. The emphasis on technical skills during surgical training is well established, however it is deficiencies in non-technical skills that have been identified as the main cause of errors in the operating room.

Objective: This systematic review aims to identify current methods used to teach non-technical skills and how these methods are assessed to determine their validity, evidence and role in training.

Methods: MEDLINE and Embase databases were searched for English language articles between 2000 and 2017 for non-technical surgical skills training. Original research articles were included if they described non-technical surgical skills training modalities and their assessment. Results were assessed for the level of evidence and each modality was awarded a level of recommendation, using a modified educational Oxford Centre for Evidence-Based Medicine classification, as adapted by the European Association of Endoscopic Surgery.

Results: A total of 19 studies were identified pertaining to high fidelity simulation (n=8), low fidelity simulation (n=6), didactic teaching (n=2) and crisis resource management (n=3). Of the included studies 1 was classified Level 1b, 1 level 2b, 7 level 2b, 2 level 2c and 8 level 3.

Conclusion: With the importance of non-technical skills being increasingly recognised, it is essential for surgeons to receive adequate training in non-technical skills. Therefore the most valuable teaching modalities such as high and low fidelity simulation needs to be implemented into surgical training curricula.
INTRODUCTION

It is commonly assumed that the marker of a surgeon’s ability is his or her technical ability to perform a given procedure and therefore the emphasis on technical skills during surgical training is well established. However, given the complexity of the operating room, a surgeon requires both technical ability and an array of non-technical skills [1]. Non-technical skills may be divided into 3 distinct categories including social (communication, teamwork and leadership) and cognitive skills (decision-making and situational awareness) as well as the personal resource factors (ability to cope with stress and fatigue) [2].

Deficiencies in non-technical skills have been identified as the main cause of errors in the operating room [3]. Analysis of adverse events in surgery have shown that the underlying causes often originate from failures in non-technical skills as opposed to technical performance [4]. Thus, it is reasonable to suggest that improved non-technical skills of surgeons would result in improved surgical outcomes. Despite this, a relatively small amount of research has been conducted to combine non-technical skills with technical skills teaching [1]. It is desirable to provide surgical trainees with a basic skill set and knowledge-base of non-technical competencies needed for the operating room (OR), at the start of training, similar to the training provided for technical skills [5]. There has been an increased push towards using simulation-based teaching for both technical and non-technical skills [6]. This may be due to the fact that simulation-based education has been shown to directly impact physicians’ clinical behavior and change outcomes [7]. Other teaching methods include traditional didactic teaching and in theatre coaching however there is a lack of studies that look at the
effectiveness of training interventions to teach non-technical skills to surgical trainees [8].

The aim of this study is to identify the current training methods used to teach non-technical surgical skills and how these methods are assessed to determine their validity, level of evidence and role in training. This study also aims to evaluate the level of recommendation for each teaching method.
METHODS

This study was performed using the guidelines set out by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Figure 1) [9]

Search methods
An initial broad search was performed on MEDLINE and Embase databases for English language articles between Jan 2000 and Jan 2017. Search terms used included “non technical surgical skills” “NoTSS” and “surgical training” which allowed for the majority of the articles to be found. A further search using terms such “simulation training “technical surgical skills ” was also performed. Key search terms were then combined to narrow down the results.

Study eligibility criteria
Original research articles were included if they described non-technical surgical skills training methods and assessment of the non-technical skills. Articles describing both non-technical and technical skills training were included. Furthermore, articles describing solely technical surgical skills training or articles not evaluating the training methods were excluded.

Study selection and data collection
Initially titles and abstracts were screened and after meeting the inclusion and exclusion criteria, articles were retrieved in full for further examination. Full text review revealed articles, which were not, focused on non-technical surgical skills training and assessment or validation studies. Following reference reviews, relevant articles found
cited in the included articles were also retrieved and examined under the inclusion and exclusion criteria.

**Data analysis**

Results were tabulated and studies describing each training method were grouped together. The studies were classified into whether they described high fidelity simulation in a fully equipped operating room, low fidelity simulation i.e. full immersion/distributed simulation (using a simulator rather than an operating room) and didactic teaching. Definitions of validity were based on the work of McDougall [10] and van Nortwick et al. [11]. Table 1 presents a summary of all the included studies looking at the demographics, intervention and assessment of non-technical skills teaching. Each study where possible, has been awarded a level of evidence (LoE) and a level of recommendation (LoR) was awarded to each study, using a modified educational Oxford Centre for Evidence-Based Medicine (OCEBM) LoE and LoR classification system, as adapted by the European Association of Endoscopic Surgery [12, 13].
RESULTS

Out of 9719 articles that were initially retrieved, including duplicates, 28 articles met the inclusion criteria, all describing different non-technical surgical skills training modalities. In terms of the different types of training method, the included studies described high fidelity simulation (n=15), low fidelity simulation (n=6), didactic teaching (n=2) and crisis resource management (n=6). The different methods used to assess non-technical skills included in the studies were the NoTSS scoring system (n=9), the NoTECHS scale (n=11), OSANTS (n=1) and the LOSA checklist (n=1). The NoTSS scoring system comprises of four important domains required in the OR for effective performance including situational awareness, decision making, communication and teamwork and leadership [14]. The NoTECHS scale is a surgically adapted assessment tool of observable behaviors first used in aviation. It assess 22 questions that a surgeon is expected to perform [14]. The LOSA checklist was originally developed to assess non-technical skills in aviation but a few elements, which have been considered to be relevant to surgery are now being used to assess non-technical surgical skills [15]. Out of 28 studies, 20 validated their method of non-technical skills training in at least one parameter. Studies that did not demonstrate validity were marked by a strike through the text.

High fidelity simulation

Fifteen studies [4, 6, 7, 14-25] were identified describing high fidelity OR simulation (HFORS) (see figure 2). All of these studies used a simulated operating room (SOR) in order to assess non-technical surgical skills. Abdelshehid et al. [23] conducted a prospective cohort study in which urology residents undertook a laparoscopic partial
nephrectomy simulation-based team-training scenario, using validated simulator models. The scenario was video recorded and were reviewed by 2 trained reviewers using the NoTSS assessment tool. Following the simulation, a debriefing session was held with all participants. They found that the level of urology resident training did significantly affect non-technical performance using the NoTSS score, thus showing construct validity. Briggs et al. [24] carried out a retrospective cohort study where 20 surgical teams were assessed using two high fidelity trauma scenarios; they were also assessed using the NoTSS system. Interestingly, they found that NTS scored across the course of the simulated scenarios decreased and they showed a significant correlation between the team leader’s cognitive skills (situational awareness and decision making) and critical task completion. Face or construct validity was not described and there was no mention of how feedback was given to the participants. Lee et al. [6] conducted a study similar to Abdelshehid et al. [23] whereby urology residents undertook a simulation- based scenario for a laparoscopic nephrectomy but alongside anaesthesiology residents. 94% of participating residents thought that the session was useful and should be included in residency training, demonstrating face validity. After each scenario, participants received a debriefing session.
Moorthy et al. [16] compared the non-technical skills of junior trainees and senior trainees performing a saphenofemoral junction (SFJ) high-tie procedure on a synthetic bench model. The scenario was recorded and the NoTECHS scale was used to assess the skills, they found large variability within the group but there was an absence of construct validity. All participants found the SOR environment to be realistic proving face validity. The participants received an objective, video-based criteria-referenced feedback of their performance [16]. A previous study by Moorthy et al. [15] compared 3 groups (junior, middle level and senior trainees) performing a SFJ procedure, here they used the LOSA checklist, which was developed for the assessment of non-technical skills in aviation considered to be related to surgery. A human factors expert subsequently gave the participants feedback on their performance. They were able to show face validity using a 4-statement questionnaire however non-technical skills assessment showed no differences between the 3 groups except for leadership skills.

Nicksa et al. [7] conducted a study to see whether HFORS of high-risk clinical scenarios made a difference in junior residents’ non-technical skills when done in the first half of the academic year compared to the second half, they assessed these skills using the NoTECHS scale. All sessions were followed by 30-minute de-briefings with real-time feedback. A significant difference was highlighted in terms of differences in ability between resident level of training and time of the year showing construct validity. Furthermore, face validity was also highlighted as 89% of participants found the sessions useful.
Saleh et al. [14] observed trainee and consultant ophthalmologists perform steps of cataract surgery and intravitreous injections in a SOR; their non-technical skills were assessed using the NoTSS and NoTECHS scale. Each scenario was followed by a detailed debriefing and video playback. Participants found the simulation useful and realistic thus demonstrating face validity. Rao et al. [25] observed 15 residents perform an open gastrojejunostomy (GJ) in a SOR as a pretest, this pretest was followed by teamwork training tasks as the intervention, which was then followed by the performance of another open GJ in a SOR. The NoTSS scale was used to assess their skills and feedback was given in person on their performance, immediately after the posttest. They found that mean global NoTSS score improved following intervention (p= 0.01) and a positive survey response from participants proving face validity.

Undre et al. [17] observed 20 surgical teams in a virtual operating theatre environment, non technical skills were evaluated using the NoTECHS scale. Following the simulated scenario, participants were given individual feedback on their non-technical performance. Most of the team skills were scored above 4 on a number of 6-point Likert scale and they found that the SOR represented a useful training environment.

In a study conducted by Sevdalis et al. [20], junior surgical teams underwent 1 and a half days of simulation training to complete a saphenofemoral junction ligation procedure involving crises; their non-technical skills were assessed using a revised NOTECHS scale. Between the training days the surgeons received an intervention covering safety and non-technical skills, they surgeons did not receive any feedback and neither face nor construct validity was demonstrated. The aim of the study was to report on the reliability of the NoTECHS scale, they found adequate internal consistency of all 5 subscales of the revised scale.
Arora et al. [22] observed 25 general surgeons complete a laparoscopic cholecystectomy in a SOR; non-technical skills were assessed using the NoTSS tool. After the simulation, the surgeons completed a self-assessment and had a debriefing session. There was no significant correlation between self-assessment and faculty member’s assessment [22] and neither face nor construct validity were established. In the study conducted by Nguyen et al. [4], general surgical residents completed 2 identical simulation sessions (2 months apart) and a content expert debriefed them after each session. Non-technical skills were assessed using an intraoperative checklist, residents showed a significant improvement in all non-technical skills competencies (p< 0.05) and overall the residents reported that the simulated OR closely resembles that of a typical OR and they felt it was a valuable teaching and training tool [4] therefore showing content and face validity. Paige et al. [18] looked at interdisciplinary team training using simulation scenarios, after completion of the scenario the team had a video-facilitated debriefing discussion. Non-technical skills were not formally assessed, all participants completed a questionnaire and each one found that the sessions were effective to very effective for improving teamwork and communication.

In a later study by Paige et al. [21], 6 general surgeons performed a cholecystectomy in a SOR, the scenarios were videotaped and the participants received feedback. Participants completed a pre and post session questionnaire measuring perceived self-efficacy for performing targeting teamwork competencies in the actual OR setting [21] however non-technical skills were not assessed. In a study by Gettman et al. [19] 17 urology residents participated in a prospective simulation study. Significant
improvement was noted on validated teamwork instruments between scenarios based on resident (pretest 24, posttest 27 p= 0.01) and expert evaluation using a validated teamwork scale. The participants were debriefed using recorded videos of the scenarios they took part in and post scenario Likert questionnaire results revealed face and content validity. HFORS was awarded a LoR of 2.

Low fidelity simulation

Six studies [1-3, 26-28] described the use of low fidelity simulation and all described the use of full immersion/distributed simulation (FIDS). The “Igloo” distributed simulator (DS) (see figure 3) is a 360° inflatable and mobile shell filled with operative equipment to create a realistic environment [1]. Brewin et al. [3] conducted a study where 10 trainee and 10 experienced urologists performed a transurethral resection of the prostrate (TURP) in the DS environment. After their simulation performance they were debriefed and received feedback. The outcomes of this study included face validity, content validity and construct validity. Face and content validity were evaluated using qualitative questionnaires and the results confirmed both face (mean Likert score 4.6/5 SD 0.51) and content (mean Likert score 4.4/5 SD 0.67) validity. The non-technical skills were assessed using the NoTECHS scale and the skills of the experienced urologists were significantly better than those of the trainees, establishing construct validity.

The study by Brunckhorst et al. [1] involved 32 novices who took part in a comparative study of simulation vs non-simulation training (knowledge only). Half of the medical students had technical and non-technical skills training within a simulation based rigid-ureteroscopy curriculum, and the other half only received a short didactic introductory
session. This study aimed to find a relationship between technical and non-technical skills, all technical skill parameters analysed demonstrated a significant correlation with the NoTSS rating scale. The trained students demonstrated a significant negative relationship between time to completion and communication and teamwork, situational awareness and decision making (p < 0.05) but little difference was found between the cohorts. Furthermore, there was no mention of how feedback was given to participants.

A randomized controlled trial by Brunckhorst et al. [2] aimed to develop a simulation-based ureteroscopy training curriculum integrating technical and non-technical skills. Thirty-two novices took part and the intervention arm underwent full ureteroscopy training via the developed curriculum, the non-technical skills module of the designed curriculum involved FIDS and didactic teaching. They used quantitative surveys to confirm content validity however there was no mention of feedback to participants. 86% of experts agreed that the developed curriculum would be effective in teaching ureteroscopy to novices, demonstrating content validity. The curriculum trained group had higher NoTSS scores (p < 0.0001) than the control group also it was found that previous training within the DS significantly improved all NOTSS components (all p < 0.05).

In the study conducted by Kassab et al. [26] 10 novice and 10 expert surgeons performed a laparoscopic cholecystectomy (LC) in the DS and on a box trainer, participants did not receive feedback on their performance. Face and content validity were strongly demonstrated using a 6-point Likert questionnaire and experts performed significantly better than novices in the DS demonstrating construct validity.
Non-technical skills were not quantitatively assessed but surgeons agreed FIDS would be useful for non-technical skills training. In a later study by Kassab et al. [27] 11 surgical residents were recruited to perform a small bowel anastomosis in the DS, the surgeons did not receive feedback on their performance. Face and content validity was analysed and demonstrated both quantitatively (overall mean face and content validity rated as 5.0 (SD 0.57) and qualitatively through interview, surgeons generally perceived the contextualized simulation to be realistic and representative of real-world practice. However this study did not formally assess technical or non-technical skills.

Harris et al recruited 5 senior and 5 junior surgeons to perform a LC in the DS but off-site, non-technical skills were not formally assessed and no feedback on technical or non-technical skills performance was provided to the participants. They were primarily looking at the feasibility of an off-site (not in a hospital environment) DS, off-site face and content validity were assessed with positive ratings from both senior and junior surgeons. The surgeons also supported the use of DS in the training and assessment of non-technical skills. The modality of FIDS was awarded a LoR of 2.

**Didactic and simulation based teaching**

Two studies [5, 8] described the use of a combination of methods, didactic teaching and simulation-based teaching. This combinatory approach involved a structured curriculum that can be incorporated into surgical training [5]. The curriculum consists of lectures and interactive sessions covering non-technical skills as well as putting the knowledge into practice using simulation scenarios. Dedy et al. [5] conducted a randomized controlled trial to evaluate the effectiveness of structured curricular training for non-technical skills. 23 residents were randomized and the intervention
group received a 5-day course of non-technical skills, which included simulations and didactic teaching. The NoTSS rating system and OSANTS rating scale were used; participants were not given any feedback on their non-technical skills. Between-group comparisons showed the intervention group had higher scores on both NoTSS and OSANTS but the results were not statistically significant. Participants completed post-course critique and the majority agreed that skills learnt on the course could be implemented in the OR, demonstrating face validity.

Pena et al. [8] looked at the effectiveness of a simulation based-training plus interactive workshop on non-technical skills performance assessed in a simulated environment. After each scenario, participants took part in a one-on-one debriefing session. Overall, 92% of participants gave positive responses regarding the realism of the simulation scenario. There was no difference found between the groups (simulation vs simulation plus workshop group) for any of the non-technical skills. Didactic teaching was awarded a LoR of 3.

**Crisis resource management**

Six studies [29-34] described the use of crisis or crew resource management (CRM) as a modality for non-technical skills training. CRM is a training course that has been adapted from cockpit or crew resource management used in aviation training; these skills have been adapted for acute medical care [32]. The five basic CRM principles taught through lectures are leadership, communication, situational awareness, teamwork and resource use [32]. McCulloch et al. [29] studied 83 surgeons performing laparoscopic cholecystectomy and carotid endarterectomy before and after receiving a non-technical skills course based on CRM. Following training feedback was provided
during a 3-month “bedding in” period, where the surgeons were supported in the OR. The NoTECHS scale was used and they found that team non-technical skills significantly increased after training (p= 0.021) however no validity was demonstrated.

Mishra et al. [30] evaluated the validity of the NoTECHS system; they used an aviation-style safety training intervention for operating teams performing laparoscopic cholecystectomy. There was a significant improvement (p=0.005) in scores after the team-training programme however no validity was established and the participants did not receive feedback.

Shamim Khan et al. [31] studied 33 urology specialist registrars of different grades participate in CRM sessions as part of seven full-day programmes. The scenarios consisted of the trainees in a HFORS environment followed by a structured debrief led by faculty. Non-technical skills were not formally assessed but face and construct validity were established from an overwhelming positive response from the interviews and questionnaires conducted. In the study by Ziesmann et al. [32] 20 surgical residents participated HFORS trauma simulations before and after undertaking a CRM course, each scenario was followed by a 45-minute debriefing session. Similarly to Shamim Khan et al. [31] Ziesmann et al. [32] did not assess non-technical skills but the study did demonstrate face validity.

Morgan et al. [33] conducted a controlled interrupted time series experiment; the intervention was a combination of standard operating procedures and CRM training. The team training intervention consisted of a 1-day interactive lecture-based training course. They used NoTECHS II scale to measure team non-technical skills. The mean NoTECHS II score statistically improved more than in the intervention group compared
to control group ($p=0.067$). The participants were not debriefed and Morgan et al. [33] failed to demonstrate face or construct validity.

Robertson et al. [34] also conducted a controlled interrupted time series experiment in an orthopaedic and reconstructive hospital. The intervention for non-technical skills was teamwork training based on aviation style CRM in one morning and two evening sessions. The effect of the intervention was assessed using the Oxford NoTECHS II scale, the mean score improved more than in the intervention group than the control group but this improvement was of borderline significance ($p=0.058$) [34]. There was no mention of debriefing and there was no validity established. This modality has been awarded a LoR of 3.
DISCUSSION

Over the past few years, there has been more emphasis placed on non-technical skills training in surgery. It is evident that the quality of a surgeon’s non-technical skills is vital for patient safety [1], yet a standardized approach to non-technical skills training is lacking across all surgical specialties and there is a paucity of relevant literature. This review presents the current modalities and concepts for non-technical skills training in surgery. The current modalities identified in this review include HFORS, FIDS, didactic teaching and CRM.

Amongst the validated non-technical assessment scales found in this review, the NoTSS system was used more commonly; this may be due to the fact that this system is useful in assessing individuals whereas the NoTECHS scale is mainly used to evaluate whole teams [30]. Wood et al. [35] have given levels of recommendation to non-technical assessment scales, it has been suggested that the use of NoTSS is both effective and of high quality and it has previously been awarded a LoR of 2 (Table 2). The NoTECHS scale is effective in highlighting the strengths and weakness of team-based non-technical practice and has also been awarded a LoR of 2 (Table 2) [35].

High-fidelity simulation is stated as the ‘gold standard’ for immersive surgical training [3]. It appears that HFORS can be essential in the development and practice of non-technical skills [23] and trainees agree that it resembles real life as most studies demonstrated face validity (Table 2). As well as being an effective training method, simulation is also useful for the assessment of non-technical skills. Briggs et al highlighted that high fidelity simulation is an ideal environment for analyzing and
studying the performance of non-technical skills in controlled setting [24]. Although high-fidelity training is seemingly popular and it has received a LoR of 2, recently low fidelity training in the form of distributed simulation has been shown to be a preferable method due to its portability and cost-effectiveness [3].

The use of distributed simulation for technical skills training is becoming more fashionable. This review identified 6 studies using distributed simulation, which included the only 1b study [2] of this entire review. It’s increased use in surgical training may be well founded as the study done by Brewin et al showed face, content and construct validity [3]. There may be some value in practicing both technical and non-technical skills in the DS however this practice may not be transferable to the OR [2], nevertheless FIDS received a LoR of 2.

A combination of a more traditional approach using didactic teaching alongside HFORS as a means to teach non-technical skills has been highlighted by two studies in this review [5, 8]. To their knowledge Dedy et al. [5] were the first to conduct a randomized controlled trial evaluating the impact of a structured curricular training programme for non-technical skills [5]. Both studies demonstrating this approach established face validity but due to the lack of the literature assessing a combined approach, it is difficult to strongly recommend this method, in fact Pena et al. [8] found no advantage in using a combined educational strategy. Unsurprisingly, this modality has been awarded a LoR of 3.

It has been stated that improvements in non-technical skills especially teamwork within theatre teams could reduce errors and adverse outcomes in surgery [29]. CRM has
been shown to be beneficial as it combines focused teamwork training and HFORS and it has received a positive response from a multidisciplinary audience [32]. However, due to the lack of evidence found describing and assessing this modality for non-technical skills, CRM has been awarded a LoR of 3 in the current study.

In light of this review, it is clear that a framework is needed for non-technical skills training particularly one that combines the most useful and effective modalities. Figure 4 demonstrates a potential framework that could form a basis for non-technical skills training. Trainers and institutions are recommended to include high or low-fidelity training modalities for nontechnical skills within training programs and/or curricula. Assessment of individuals and teams are recommended with the most comprehensively evaluated assessment scales, the NoTSS or NoTECHS, respectively.

This review has a number of limitations; first of all despite using the largest and most relevant databases for the data search, relevant studies may have been missed. There were few RCT’s identified in the literature pertaining specifically to non-technical skills training and there were no level 1a studies included. Considering that this is an emerging field of research, more randomized controlled trials are necessary in order to compare the different teaching modalities; this will allow for level 1a research to take place. Furthermore, most studies demonstrated at least face validity however there were eight studies that did not demonstrate any validity. The review does not represent all surgical specialties as the majority of articles found were describing urology or general surgery training, therefore it appears that these two specialties have made the most progress in establishing non-technical skills training.
CONCLUSION

With the importance of non-technical skills in surgery being recognised, establishing effective training methods have become more pertinent. It is essential for surgeons to be trained in non-technical skills, therefore the most valuable teaching modalities need to be identified and implemented within curricula. It is hoped that non-technical skills training will be standardized and implemented into curricula across surgical specialties. There is a good basis to recommend the use of high fidelity and low fidelity simulation for non-technical skills training. Future studies should focus on developing a validated training method for non-technical skills that can be implemented into surgical training across all surgical specialties.
References


## TABLES

Table 1: Summary of Validation studies for methods of non-technical skills training. Abbreviations: CRM- crisis resource management, FIDS- full immersion distributed simulation, HFORS- high fidelity operating room simulation, LoE- Level of Evidence, LoR- Level of Recommendation, RCT- Randomised Controlled Trial.

<table>
<thead>
<tr>
<th>Modality</th>
<th>Study</th>
<th>Design</th>
<th>Participants</th>
<th>Assessment</th>
<th>Validation</th>
<th>LoE</th>
<th>LoR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Moorthy et al (2006) [16]</td>
<td>Prospective cohort study</td>
<td>10 junior &amp; 10 senior vascular surgeons</td>
<td>Modified NoTECHS</td>
<td>Face</td>
<td>2b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undre et al (2007) [17]</td>
<td>Prospective cohort study</td>
<td>20 surgical teams (including 80 surgeons, anaesthetists, scrub nurses and operational department practitioners)</td>
<td>NoTECHS</td>
<td>Face</td>
<td>2c</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paige et al (2007) [18]</td>
<td>Prospective cohort study</td>
<td>3 senior surgical residents</td>
<td>N/A</td>
<td>Face, content</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gettman et al (2008) [19]</td>
<td>Prospective cohort study</td>
<td>19 urology residents</td>
<td>N/A</td>
<td>Face, content</td>
<td>2c</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sevdalis (2008) [20]</td>
<td>Prospective cohort study</td>
<td>Surgical teams (trainee surgeons, scrub nurses and operational department practitioners)</td>
<td>NoTECHS</td>
<td>Face, construct</td>
<td>2b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paige et al (2009) [21]</td>
<td>Prospective cohort study</td>
<td>6 general surgery residents</td>
<td>N/A</td>
<td>Face, construct</td>
<td>2c</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arora et al (2011) [22]</td>
<td>Prospective cohort study</td>
<td>25 junior and senior general surgery residents</td>
<td>NoTSS</td>
<td>Face, construct</td>
<td>2b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lee et al (2012) [6]</td>
<td>Prospective cohort study</td>
<td>8 urology &amp; 8 anaesthetic residents</td>
<td>NoTSS</td>
<td>Face</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abdelshehid et al (2013) [23]</td>
<td>Single blind, prospective comparative trial</td>
<td>9 urology residents</td>
<td>NoTSS</td>
<td>Construct</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Briggs et al (2015) [24]</td>
<td>Retrospective cohort study</td>
<td>20 trauma teams (including emergency medicine residents, surgery residents, emergency department nurses and emergency service assistants)</td>
<td>NoTSS (individuals), T-NOTECHS (entire team)</td>
<td>Face, construct</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nguyen et al (2015) [4]</td>
<td>Prospective cohort study</td>
<td>11 general surgery residents</td>
<td>N/A</td>
<td>Face, content</td>
<td>2c</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nicksa et al (2015) [7]</td>
<td>Prospective cohort study</td>
<td>43 surgical residents</td>
<td>Modified NoTECHS</td>
<td>Face, construct</td>
<td>2c</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Saleh et al (2016) [14]</td>
<td>Prospective cohort study</td>
<td>20 trainee &amp; consultant ophthalmologists</td>
<td>NOTSS and NoTECHS</td>
<td>Face</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rao et al (2016) [25]</td>
<td>Prospective cohort study</td>
<td>15 surgical residents</td>
<td>NoTSS</td>
<td>Face</td>
<td>2b</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Design Type</td>
<td>Methodology</td>
<td>Group Description</td>
<td>Outcome Measures</td>
<td>Study Quality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------</td>
<td>------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>--------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kassab et al (2011) [26]</td>
<td>Prospective cohort</td>
<td>10 novice &amp; 10 expert general surgeons</td>
<td>N/A</td>
<td>Face, content</td>
<td>2b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kassab et al (2012) [27]</td>
<td>Prospective cohort</td>
<td>11 surgical residents</td>
<td>N/A</td>
<td>Face, content</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harris et al (2013) [28]</td>
<td>Prospective cohort</td>
<td>5 senior &amp; 5 junior general surgeons</td>
<td>N/A</td>
<td>Face, content</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brunckhorst et al (2015) [2]</td>
<td>RCT</td>
<td>32 medical students</td>
<td>NoTSS</td>
<td>Face, Content, Construct</td>
<td>1b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dedy et al (2016) [5]</td>
<td>RCT</td>
<td>22 postgraduate general surgery residents</td>
<td>NoTSS</td>
<td>Face</td>
<td>2a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCulloch et al (2009) [29]</td>
<td>Historical prospective cohort study</td>
<td>Surgical teams (83 surgeons, anaesthetists, nurses)</td>
<td>NoTECHS (outcome measured in OR style setting)</td>
<td>Face, construct</td>
<td>2c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mishra et al (2009) [30]</td>
<td>Prospective cohort</td>
<td>Surgical teams (65 surgeons, anaesthetists, nurses)</td>
<td>NoTECHS</td>
<td>Face, construct</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shamim Khan et al (2012) [31]</td>
<td>Prospective cohort</td>
<td>33 urology trainees</td>
<td>N/A</td>
<td>Face, construct</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ziesmann et al (2013) [32]</td>
<td>Prospective cohort</td>
<td>20 surgical residents</td>
<td>N/A</td>
<td>Face</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morgan et al (2014) [33]</td>
<td>Controlled interrupted time series</td>
<td>Surgical teams (participant numbers not disclosed)</td>
<td>NoTECHS</td>
<td>Face, construct</td>
<td>2c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robertson et al (2015) [34]</td>
<td>Controlled interrupted time series</td>
<td>Surgical teams (participant numbers not disclosed)</td>
<td>NoTECHS</td>
<td>Face, construct</td>
<td>2c</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: OCEBM level of Recommendation for Training Tools [35] Added with permission from Wood et al.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoTSS</td>
<td>Level 2</td>
</tr>
<tr>
<td>NoTSSdk</td>
<td>Level 3</td>
</tr>
<tr>
<td>OTAS</td>
<td>Level 3</td>
</tr>
<tr>
<td>Oxford NoTECHS</td>
<td>Level 2</td>
</tr>
<tr>
<td>Oxford NoTECHS II</td>
<td>Level 3</td>
</tr>
<tr>
<td>Surgeon’s leadership Inventory</td>
<td>Level 3</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1: Study selection according to the PRISMA statement

Figure 2: High fidelity operating room simulation

Figure 3: Full immersion/distributed simulation

Figure 4: Recommended framework for non-technical skills training in surgery. Abbreviations: CRM- crisis resource management, FIDS- full immersion distributed simulation, HFORS- high fidelity operating room simulation.
Records identified through database searching (n = 9719)

Additional records identified through other sources (n = 21)

Records after duplicates removed (n = 4865)

Records screened (n = 337)

Records excluded (n = 4528)

Full-text articles assessed for eligibility (n = 71)

Full-text articles excluded, with reasons (n = 42)

Studies included in qualitative synthesis (n = 29)


For more information, visit www.prisma-statement.org.
Didactic lectures or e-learning teaching non-technical skills principles

FIDS or HFORS scenario for common procedures according to surgical specialty

CRM and HFORS scenarios complex and emergency procedures according to surgical specialty
Supplementary Tables

Supplementary Table 1: Modified levels of evidence classification for validation studies, adapted from Oxford Centre for Evidence-Based Medicine classification by the European Association of Endoscopic Surgeons (Carter et al) LoE- Level of Evidence

<table>
<thead>
<tr>
<th>LoE</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>Systematic reviews (meta-analysis) containing at least some trial of level 1b evidence, in which results of separate, independently controlled trials are consistent</td>
</tr>
<tr>
<td>1b</td>
<td>Randomised controlled trial of good quality and of adequate sample size (power calculations)</td>
</tr>
<tr>
<td>2a</td>
<td>Randomised trials of reasonable quality and/or of inadequate sample size</td>
</tr>
<tr>
<td>2b</td>
<td>Nonrandomised trials, comparative research (parallel cohort)</td>
</tr>
<tr>
<td>2c</td>
<td>Nonrandomised trials, comparative research (historical cohort, literature controls)</td>
</tr>
<tr>
<td>3</td>
<td>Nonrandomised, non-comparative trials, descriptive research</td>
</tr>
<tr>
<td>4</td>
<td>Expert opinions, including the opinion of Work Group members</td>
</tr>
</tbody>
</table>

Supplementary Table 2: Levels of recommendation for training models, adapted from Oxford Centre for Evidence-Based Medicine classification by the European Association of Endoscopic Surgeons (Carter et al) LoR- level of recommendation

<table>
<thead>
<tr>
<th>LoR</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Based on one systematic review (1a) or at least two independently conducted research projects classified as 1b</td>
</tr>
<tr>
<td>2</td>
<td>Based on at least two independently conducted research projects classified as level 2a or 2b, within concordance</td>
</tr>
<tr>
<td>3</td>
<td>Based one independently conducted research project level 2b or at least two trials of level 3, within concordance</td>
</tr>
<tr>
<td>4</td>
<td>Based on one trial at level 3 or multiple expert opinions, including the opinions of Work Group members (e.g. level 4)</td>
</tr>
</tbody>
</table>