



**Methodological shortcomings in the literature evaluating the role and applications of 3D training for surgical trainees**

Journal:	<i>Medical Teacher</i>
Manuscript ID	Draft
Manuscript Categories:	Articles
Date Submitted by the Author:	n/a
Complete List of Authors:	Kostusiak, Milo; Addenbrooke's Hospital, Division of Neurosurgery; Hart, Michael; Addenbrooke's Hospital, Division of Neurosurgery Barone, Damiano; Addenbrooke's Hospital, Division of Neurosurgery Kirollos, Ramez ; Addenbrooke's Hospital, Division of Neurosurgery Santarius, Thomas; Addenbrooke's Hospital, Division of Neurosurgery Trivedi, Rikin; Addenbrooke's Hospital, Division of Neurosurgery
Keywords:	Surgery < Discipline, Simulation < Teaching & Learning, Postgraduate < Phase of education, Medical education research < Management, Practical procedures < Learning outcomes

SCHOLARONE™  
Manuscripts

1  
2  
3 **Methodological shortcomings in the literature evaluating the role and applications of 3D**  
4  
5 **training for surgical trainees**  
6  
7  
8  
9

10  
11 Milosz Kostusiak, Michael Hart, Damiano Giuseppe Barone, Ramez Kirolos, Tom Santarius  
12  
13 & Rikin Trivedi  
14

15  
16  
17  
18  
19 Division of Neurosurgery, Addenbrooke's Hospital, University of Cambridge, Cambridge UK  
20

21  
22  
23  
24 Corresponding author: Milosz Kostusiak, milokostusiak@doctors.org.uk, Neuroscience  
25  
26 Department, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation  
27  
28 Trust, Hills Road Cambridge CB2 0QQ  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11 **Methodological shortcomings in the literature evaluating the role and applications of 3D**  
12 **training for surgical trainees**  
13  
14

15  
16  
17  
18  
19 Milosz Kostusiak, Michael Hart, Damiano Giuseppe Barone, Ramez Kirolos, Tom Santarius  
20  
21 & Rikin Trivedi  
22  
23

24  
25  
26  
27 Division of Neurosurgery, Addenbrooke's Hospital, University of Cambridge, Cambridge UK  
28  
29

30  
31  
32 Corresponding author: Milosz Kostusiak, milokostusiak@doctors.org.uk, Neuroscience  
33

34 Department, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation  
35

36  
37 Trust, Hills Road Cambridge CB2 0QQ  
38  
39

40  
41  
42 Running title: Shortcomings of 3D surgical training  
43  
44

45  
46 Keywords: 3D, three-dimensional, surgery, surgical training, shortcomings, methodology  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Abstract**

**Aim:** There has been an increased interest in the use of three-dimensional (3D) technology in surgical training. We wish to appraise the methodological rigor applied to evaluating the role and applications of 3D technology in surgical training, in particular, on the validity of these models and assessment methods in simulated surgical training.

**Methods:** Literature search was performed using MEDLINE with the following terms: “3D”; “surgery”; and “training”. Only studies evaluating the role of 3D technology in surgical training were eligible for inclusion and assessed for the level of evidence, validity of the simulation model, and assessment method used.

**Results:** A total of 93 studies were analysed, and majority of reviewed articles focused on 3D displays (36) and 3D printing (35). Most of these studies were case series, the most common assessment was subjective (69), with objective assessment used by 57 studies. Very few studies provided evidence for validity of the model or the assessment methods used.

**Conclusion:** 3D technology has a great potential in simulated surgical training. However, the validity of this technology and strong evidence for its beneficial effects in surgical training is lacking. Further work on validation of 3D technology and assessment tools is needed.

## Introduction

Apprenticeship has been a traditional surgical training model since establishment of surgery as a medical profession, where the learner follows a more experienced and skilled mentor to attain knowledge, judgement, and a range of surgical techniques (Satava et al. 2003). This model still holds true in current surgical training, however, the introduction of the European working time directive (EWTD) and relative increase of surgical trainees to core surgical procedures have led to significant reduction in the hours spend in the operating theatre (Breen et al. 2013; Toll and Davis 2010). Consequently, these changes have necessitated the search for alternative training methods, mainly in the form of simulation.

Simulation is a crucial part of surgical training as it emulates a real life scenario that allows a trainee to practice procedural skills in a safe environment (Sutherland et al. 2006). It has been widely adopted throughout the health care profession and can range from low-fidelity (e.g. glove model for suturing) to high-fidelity simulation models (e.g. pig model for laparoscopic cholecystectomy) (Van Bruwaene et al. 2015). Currently, there is a plethora of simulation models available for the trainees, and they can be categorised into physical, cadaveric, animal, harvested tissue, computer simulation and hybrid models (Samia et al. 2013). Human cadavers are incorporated into surgical training throughout the UK and other countries, however, the evidence for effectiveness of cadaveric workshops in higher surgical training is limited (Gilbody et al. 2011). Furthermore, due to a limited supply of cadavers and the cost associated with it, there has been a trend towards using more affordable and sustainable models, for example three-dimensional synthetic models (Hochman et al. 2014). Three dimensional (3D) technology has improved immensely over the recent years and led to the development of 3D virtual reality (VR), printing, and interactive mobile applications, all of which are slowly being integrated into current surgical training (Lewis et al. 2014; Olivieri et

1  
2  
3 al. 2016). However, the question remains whether 3D technology in simulation training has  
4  
5 any beneficial effect on surgical skill acquisition.  
6  
7

8 Surgical performance, competence, and skill acquisition can be measured through various  
9  
10 assessment methods, broadly categorised into formative and summative, where former is  
11  
12 aimed at providing feedback on the performance and latter is used to formally assess the  
13  
14 acquisition of the skills required for completion of the training (Evgeniou 2013). The  
15  
16 majority of assessment tools in surgical training are subjective and most of these tools are  
17  
18 checklists (Alkhayal et al. 2012), for example the Structured Technical Skills Assessment  
19  
20 Form (STSAF) (Winckel et al. 1994). These tools require standardised training of the  
21  
22 assessors in order to provide reliable results and reduce possibility of a rater bias. On the  
23  
24 contrary, objective assessment tools do not involve subjectivity of the assessor and thus are  
25  
26 more reliable. Examples of such tools include end product assessment, virtual reality (VR),  
27  
28 hand motion analysis (HMA) and objective structured assessment of technical skills  
29  
30 (OSATS) (Moorthy et al. 2003). VR and HMA generate objective results, such as task  
31  
32 completion times, or numbers of inaccurate or corrective manoeuvres, which can be used to  
33  
34 assess the performance by comparing the results against 'expert controls' (Chin et al. 2011;  
35  
36 Perrenot et al. 2012). However, OSATS, which uses checklists with performance being  
37  
38 assessed by expert surgeons, has shown high reliability and construct validity, which suggests  
39  
40 that with extensive validation processes, these checklists can be used as objective assessment  
41  
42 tools (Reznick et al. 1997). OSATS still remains the most popular assessment tool for open  
43  
44 and laparoscopic skills, but unfortunately the transition from bench or cadaveric simulation  
45  
46 models to assessment in operating theatres is seldom seen (Shaharan and Neary 2014).  
47  
48  
49  
50  
51

52  
53 This review focuses on the methodological shortcomings in evaluating the role and  
54  
55 applications of 3D technology in surgical training in particular, on the validity of these  
56  
57 models and assessment methods in simulated surgical training.  
58  
59  
60

## Methods

An extensive literature search was performed using the MEDLINE search engine. The search was performed using the following terms: “3D”, “surgery” and “training”. Only the studies evaluating the role of 3D technology in surgical training, and including both simulation and real-time operations, were accepted. Selected studies were then assessed for the level of evidence, setting and training modality, whether the model was validated for the purpose of the surgical training, assessment methods used, and the overall outcome.

## Results

The literature search identified 761 unique studies that had evaluated the use of 3D in simulated surgical training. After careful screening of titles and abstracts, 93 studies met the inclusion criteria and the full texts of these were then analysed. The studies were screened for the type of research, setting and training modality, assessment method and overall outcome. The vast majority of the 3D related studies included 3D displays (36, 38.7%) and 3D printing (35, 37.6%), the remaining studies used VR, 3D images and videos as well as 3D navigation (Table 1). 3D displays were used with laparoscopic and endoscopic equipment. 89 studies (95.6%) were carried out in a non-clinical setting, whereas the rest were performed in the operating theatre.

Levels of evidence of the studies were assessed according to the widely cited and published criteria developed by the Oxford Centre for Evidence-based Medicine. The level of evidence of these studies ranged from 2b (e.g. poorly designed randomised clinical trials) to 5 (e.g. an expert opinion). The majority of the studies were case series (level 4 evidence) (Table 1).

Validity is a concept defined as “the property of being true, correct, and in conformity with reality”, and includes face, content, construct, concurrent, discriminative and predictive

1  
2  
3 validity (Gallagher et al. 2003). These can evaluate different aspects of the intervention, for  
4  
5 example, while concurrent validity relates to correlation to the gold standard simulator, the  
6  
7 construct validity indicates that the simulator is able to distinguish between different levels of  
8  
9 competence (McDougall 2007). Regrettably, very few studies attempted providing evidence  
10  
11 for validity of the model. 15 studies had reported face validity, 8 had reported content  
12  
13 validity, 3 reported construct validity, and 67 did not report any validation for their models.  
14  
15 However, it is important to note that some of the studies used established 3D models or  
16  
17 cadaveric set-ups that had been previously validated.  
18  
19

20  
21 Assessment methods were divided into objective and subjective assessment categories.  
22  
23 Several studies utilised both methods, however only 57 studies (62%) have used objective  
24  
25 assessment tools. These tools varied considerably, ranging from an anatomy quiz to  
26  
27 measuring performance and quality of the end product. Subjective assessment tools were  
28  
29 used by only 69 studies (74.1%) and included questionnaires or rating scales. The results  
30  
31 showed that 3D techniques were found to be superior to 2D in 90 studies (96.7%), and  
32  
33 participants from all of the studies perceived 3D as useful or very useful. Unfortunately, only  
34  
35 a small fraction of the 3D studies had used validated assessment methods. In one study,  
36  
37 laparoscopic performance was measured using a validated Fundamentals of Laparoscopic  
38  
39 Surgery (FLS) skills set, which measures time to completion, number of attempts required to  
40  
41 achieve proficiency, and the number of errors. Participants in 3D had performed better and  
42  
43 completed the task quicker with fewer errors, and required fewer repetitions to achieve  
44  
45 proficiency (Tanagho et al. 2012). Similar results were obtained by using Imperial College  
46  
47 Surgical Assessment Device (ICSAD) to measure the precision of 3D endoscopic  
48  
49 performance (Taffinder et al. 1999). In another study, 3D printed models improved  
50  
51 anatomical knowledge as evidenced by a post-intervention anatomy quiz, which is also an  
52  
53 objective measure (Ng et al. 2015). [Table 1 near here]  
54  
55  
56  
57  
58  
59  
60



## Discussion

This review provided an overview of the differing methodologies and methodological strategies being used to evaluate the effectiveness or value of 3D technology in simulated surgical training. Before the introduction of the EWTD, a trainee would be exposed to an average of 32,000 hours of experience, compared to almost half of that time after the introduction of the EWTD (Cresswell et al. 2010), thus making simulation an invaluable tool for surgeons in training. However, the effectiveness of simulated training depends on several aspects, including curriculum, validity of the simulation model and assessment methods. This review has presented potential shortcomings when evaluating the role and application of 3D technology in surgical training.

## *Quality of the evidence*

As with any medical research, well designed and performed studies would produce better quality results and thus have higher impact. This is also true for surgical education research, but arguably it is much harder to perform well-designed studies. This review has shown that the evidence in the literature assessing 3D intervention in surgical training is of low to moderate quality. The highest level of evidence was reached by only few studies that have used blinded randomisation, and several other studies have performed randomised trials, but without using a blinded assessor. Majority of the studies were case series and in some studies the methodology was not clear enough to determine the class of evidence, and in these cases they were automatically assigned as level 4 evidence. There are several possible limiting factors that affect the design of the study and these include difficulties in randomisation, time constraints, limited availability of the simulation models and lack of validation of simulation

1  
2  
3 models or assessment methods, and all of these should be addressed to produce high quality  
4  
5 evidence.  
6  
7

### 8 *Validation of simulated models*

9

10  
11 In the reviewed literature, only a handful of studies had reported face or content validity for  
12  
13 their models. Due to the fact that most of the participants in those studies were either students  
14  
15 or junior trainees with little previous exposure to clinical environment, it is unwise to assess  
16  
17 their performances and the usability of the models without prior validation. Inexperienced  
18  
19 participants are more likely to give higher scores to 3D printing models as they would not be  
20  
21 able to compare the model to the real life anatomy. Therefore, simulation models must  
22  
23 resemble the real scenario as close as possible for the developed skills to be transferable to  
24  
25 the clinical world, which can be achieved through the process of validation, and thus can be  
26  
27 used effectively for the skill assessment purposes.  
28  
29  
30  
31  
32  
33

### 34 *Assessment methods in surgical simulation*

35

36  
37 Since simulation has become an integral part of the surgical training, a plethora of assessment  
38  
39 methods were created to provide the trainee with formative and summative feedback. These  
40  
41 assessment tools must be valid, reliable, feasible, in order to have an educational impact. In  
42  
43 addition to 'model validation', assessment tools should ideally contain the essential elements  
44  
45 of validity, including construct, content, and predictive validity. It is also crucial that these  
46  
47 assessment tools provide the participants with accurate feedback so they can monitor skill  
48  
49 acquisition and improve their surgical abilities. To date, this has been best achieved through  
50  
51 objective assessment methods, however, it is also evident that subjective assessment tools  
52  
53 have been successfully employed for that reason. Studies referenced in this review have  
54  
55 utilised an extensive array of assessment tools, ranging from subjective questionnaires to  
56  
57  
58  
59  
60

1  
2  
3 objective quantitative data collection. In nearly all studies, the participants have rated the 3D  
4  
5 training models as 'useful' for surgical training in post-intervention questionnaire. User's  
6  
7 satisfaction is an important aspect of model evaluation, but are subjective in nature, and  
8  
9 expert panels or consensus groups that determine the characteristics of good or 'model'  
10  
11 simulations are potentially liable to similar criticisms. Therefore, many have focussed on  
12  
13 obtaining quantitative data, such as the time to complete task, error rate or number of  
14  
15 corrective manoeuvres required. These parameters are very generic and can be used to assess  
16  
17 any surgical procedure, and thus were the most commonly used assessment tools to obtain  
18  
19 quantitative data. Fundamentals of Laparoscopic Surgery (FLS), McGill Inanimate System  
20  
21 for Training and Evaluation of Laparoscopic Skill (MISTELS) (Fried et al. 2004), or task  
22  
23 specific checklists and global rating scales (GRS) (Ezra et al. 2009) are some of the examples  
24  
25 of validated assessment tools used only by handful of reviewed studies.  
26  
27  
28  
29  
30  
31  
32

### 33 ***Study design and its limitations***

34  
35  
36 The study design is single most important step when setting up a study to evaluate a new  
37  
38 intervention. However, the highest class of evidence is not always achievable or feasible, thus  
39  
40 one must carefully chose the most suitable methodology. When assessing the effectiveness of  
41  
42 3D simulation models in surgical training, the authors would need to consider several aspects,  
43  
44 including, but not limited to, characteristics of the participants, study environment, validation  
45  
46 of the simulation model and assessment methods.  
47  
48

49  
50 The majority of studies in surgical education employs medical students, core and higher  
51  
52 surgical trainees, and consultants as participants. This inhomogeneity among participants  
53  
54 allows for the intervention to be tested at different stages of surgical training and might be  
55  
56 useful at assessing construct validity. However, varying experience levels within the cohort  
57  
58  
59  
60

1  
2  
3 may prevent the study to be conducted as a well-designed randomised controlled trial (RCT)  
4  
5 due to possible selection bias leading to baseline imbalances (Roberts and Torgerson 1999).  
6  
7 Furthermore, trainees may refuse randomisation as they would not want to miss critical  
8  
9 educational experiences, or due to clinical rotation differences the trainees may experience  
10  
11 intervention at different time points (Sullivan 2011b). In addition, the control group within  
12  
13 RCT that evaluate 3D technology would be exposed to the current gold standard, 2D  
14  
15 equivalent, and thus blinding of participants in this example would not be possible. This may  
16  
17 lead to comparison choice bias whereby the control group pre-emptively believes 3D to be  
18  
19 better and rating 2D lower (Jadad and Enkin). One of the ways to reduce the risk of selection  
20  
21 and control biases would be to apply a single subject design, where participants are equally  
22  
23 exposed to all intervention groups, therefore the performance in each intervention can be  
24  
25 compared within subject. This study design removes the need to randomise participants, each  
26  
27 participant serves as its own control, and fewer participants may be required. However, as  
28  
29 with all crossover-design studies, there is a risk of carry-over or order effects, where the  
30  
31 effect of the first intervention is carried onto the next. The way to minimise this risk, is to  
32  
33 incorporate a lengthy washout period, i.e. the time between the interventions. However, this  
34  
35 would not necessarily negate the impact of the sequence of interventions. This type of design  
36  
37 is the most commonly reported in submissions to medical education journals (Sullivan  
38  
39 2011a).

40  
41  
42  
43  
44  
45 Another important aspect of study design is the selection of the best assessment (or  
46  
47 evaluation) method(s). The choice of the assessment tool must be governed by the research  
48  
49 question, study design, type of intervention and participant characteristics. Nonetheless, one  
50  
51 must critically consider the feasibility of the assessment tool, for example, assessing temporal  
52  
53 bone surgery skills using virtual reality or 3D printed models is more practical and safe  
54  
55 compared to real-life surgery (Linke et al. 2013; Mowry et al. 2015). The matter becomes  
56  
57  
58  
59  
60

1  
2  
3 more complicated, however, when evaluating the intervention that is being tested on a large  
4  
5 group. In the case of assessing 3D imaging in surgical anatomy lectures, where the whole  
6  
7 class may participate in the study, self-assessment may offer a feasible alternative. Self-  
8  
9 assessment of surgical skills has been shown to have good reliability and validity  
10  
11 (MacDonald et al. 2003; Mandel et al. 2005), but its role in assessing anatomy knowledge is  
12  
13 less clear (Anthony 1986). Confidence-based marking (CBM) and self-assessment were  
14  
15 shown to have a strong correlation with test performance (Barr and Burke 2013; Schiekirka et  
16  
17 al. 2013), while other forms of self-assessment had weaker correlations (Vivekananda-  
18  
19 Schmidt et al. 2007). These pre- and post-intervention questionnaires, such as the Likert  
20  
21 scales, are very common in surgical education research, but can be subjective. Thus, a more  
22  
23 suitable and objective tool to evaluate the effectiveness of 3D imaging in anatomy teaching  
24  
25 would be pre- and post-intervention tests (Muller-Stich et al. 2013).  
26  
27  
28  
29

### 30 31 32 ***Future Directions*** 33

34  
35 There is limited evidence for the effectiveness of 3D teaching of neuroanatomy in surgical  
36  
37 education. To this end, there is a need to construct a study that aims to assess whether 3D  
38  
39 visualisation of neuroanatomy (3D lectures or cadaveric dissection) is an effective method of  
40  
41 enhancing the knowledge and confidence of neurosurgeons, and how it compares to  
42  
43 traditional training.  
44

45  
46 Accepting that any study proposal will have some limitations, we propose a study that would  
47  
48 employ a three-armed experimental design where participants are exposed to either standard  
49  
50 training (2D lecture based visualisation and cadaveric training) or one of two novel  
51  
52 interventions: 3D lecture based visualisation alone or 3D lecture based visualisation  
53  
54 combined with cadaveric training. This design would allow a comparison of several different  
55  
56 interventions at the same time. The study group should include participants of varying levels  
57  
58  
59  
60

1  
2  
3 of experience, including undergraduate medical students, surgical trainees and consultants.  
4  
5 However, the majority of the participants are likely to be neurosurgical trainees as this  
6  
7 intervention is primarily aimed at them. Furthermore, both medical students, especially at the  
8  
9 very early stages of education, and consultants will serve as negative and positive controls,  
10  
11 which would allow a measure of construct validity of the tested model.  
12  
13

14 The assessment tool(s) will include both objective and subjective methods. Baseline  
15  
16 assessment (pre-intervention) and final assessment (post-intervention) will be employed and  
17  
18 these may include pre- and post-intervention quiz, combined with self-assessment. This  
19  
20 would allow for the participants to be act as their own controls. For cadaveric training and  
21  
22 assessment, the use of a validated task specific checklist would be used, and an experienced  
23  
24 surgeon should carry this assessment out. Furthermore, a comparative survey may be used if  
25  
26 the participants have been exposed to both interventions (e.g. 2D vs 3D lectures). The overall  
27  
28 effectiveness or value of these models will be assessed using comparative statistical  
29  
30 evaluations of the self-reporting rating scales and observer checklists, between groups and  
31  
32 within groups with pre- and post- comparisons. The reliability of self-reporting compared  
33  
34 with independent observer checklists will be determined by correlation analyses.  
35  
36  
37  
38  
39  
40  
41  
42

### 43 **Conclusion**

44 3D technology has a great potential in simulated surgical training as the majority of trainees  
45  
46 find the concept very useful and believe their surgical skills improve after exposure to such  
47  
48 models. However, there are a variety of methodologies used to determine the effectiveness of  
49  
50 3D simulation, and the question still remains whether these models have a beneficial effect  
51  
52 on either skills acquisition or surgical performance. Lack of model validation, poor quality of  
53  
54 study design and inadequate assessment methods are the main methodological shortcomings  
55  
56 and whilst not all of these shortcomings are likely to be addressed in any single study, the  
57  
58  
59  
60

1  
2  
3 proposed study design offers a pragmatic way of implementing a comparative assessment  
4  
5 paradigm into existing 3D assisted training courses.  
6  
7

8 **Practice points:**  
9

- 10 1. Simulation training forms an integral part of surgical trainee's pathway.
- 11 2. 3D technology can be used effectively in simulated surgical training.
- 12 3. Studies assessing 3D technology in surgical training must carefully consider  
13 appropriate study design with validated assessment methods.  
14  
15  
16  
17  
18  
19

20  
21  
22 **Notes on contributors**  
23

24 All authors are part of Neuroanatomy Education Group at Division of Neurosurgery,  
25 Neurosciences Department, Addenbrooke's Hospital, Cambridge University Hospitals NHS  
26 Foundation Trust, Cambridge, UK.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

37 **Declaration of interest:** None  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

- Alaraimi B, El Bakbak W, Sarker S, Makkiyah S, Al-Marzouq A, Goriparthi R, Bouhelal A, Quan V, Patel B. 2014. A randomized prospective study comparing acquisition of laparoscopic skills in three-dimensional (3d) vs. Two-dimensional (2d) laparoscopy. *World J Surg.* 38(11):2746-2752.
- Alkhayal A, Aldhukair S, Alselaime N, Aldekhayel S, Alhabdan S, Altaweel W, Magzoub ME, Zamakhshary M. 2012. Toward an objective assessment of technical skills: A national survey of surgical program directors in Saudi Arabia. *Adv med educ pract.* p. 97-104.
- Anthony TR. 1986. A discrepancy in objective and subjective measures of knowledge: Do some medical students with learning problems delude themselves? *Med Educ.* 20(1):17-22.
- Barr DA, Burke JR. 2013. Using confidence-based marking in a laboratory setting: A tool for student self-assessment and learning. *J Chiropr Educ.* 27(1):21-26.
- Breen KJ, Hogan AM, Mealy K. 2013. The detrimental impact of the implementation of the European Working Time Directive (EWTD) on surgical senior house officer (SHO) operative experience. *Irish Journal of Medical Science.* 182(3):383-387.
- Chalasan V, Cool DW, Sherebrin S, Fenster A, Chin J, Izawa JI. 2011. Development and validation of a virtual reality transrectal ultrasound guided prostatic biopsy simulator. *Can Urol Assoc J.* 5(1):19-26.
- Chin KJ, Tse C, Chan V, Tan JS, Lupu CM, Hayter M. 2011. Hand motion analysis using the Imperial College Surgical Assessment Device validation of a novel and objective performance measure in ultrasound-guided peripheral nerve blockade. *Regional Anesthesia and Pain Medicine.* 36(3):213-219.
- Cresswell B, Marron C, Hawkins W, Harrison E, Fitzgerald E, von Roon A. 2010. Optimising working hours to provide quality in training and patient safety. *The Association of Surgeons in Training Year Book 2009-2010.* The Association of Surgeons in Training Year Book 2009-2010. p. 30-34.
- Curro G, La Malfa G, Lazzara S, Caizzone A, Fortugno A, Navarra G. 2015. Three-dimensional versus two-dimensional laparoscopic cholecystectomy: Is surgeon experience relevant? *J Laparoendosc Adv Surg Tech A.* 25(7):566-570.
- Donnelly L, Patten D, White P, Finn G. 2009. Virtual human dissector as a learning tool for studying cross-sectional anatomy. *Med Teach.* 31(6):553-555.
- Drapkin ZA, Lindgren KA, Lopez MJ, Stabio ME. 2015. Development and assessment of a new 3d neuroanatomy teaching tool for MRI training. *Anat Sci Educ.* 8(6):502-509.
- Engel DC, Ferrari A, Tasman AJ, Schmid R, Schindel R, Haile SR, Mariani L, Fournier JY. 2015. A basic model for training of microscopic and endoscopic transsphenoidal pituitary surgery: The egghead. *Acta Neurochir (Wien).* 157(10):1771-1777; discussion 1777.
- Evgeniou E. 2013. Assessment methods in surgical training in the United Kingdom.
- Ezra DG, Aggarwal R, Michaelides M, Okhravi N, Verma S, Benjamin L, Bloom P, Darzi A, Sullivan P. 2009. Skills acquisition and assessment after a microsurgical skills course for ophthalmology residents. *Ophthalmology.* 116(2):257-262.
- Felisati G, Pipolo C, Maccari A, Cardia A, Revay M, Lasio GB. 2013. Transnasal 3d endoscopic skull base surgery: Questionnaire-based analysis of the learning curve in 52 procedures. *Eur Arch Otorhinolaryngol.* 270(8):2249-2253.



- 1  
2  
3 Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Ghitulescu G, Andrew CG. 2004.  
4 Proving the value of simulation in laparoscopic surgery. *Ann surg.* p. 518-528.
- 5 Gallagher AG, Ritter EM, Satava RM. 2003. Fundamental principles of validation, and reliability:  
6 Rigorous science for the assessment of surgical education and training. *Surg Endosc.*  
7 17(10):1525-1529.
- 8 Gilbody J, Prasthofer AW, Ho K, Costa ML. 2011. The use and effectiveness of cadaveric workshops in  
9 higher surgical training: A systematic review. *Annals of the Royal College of Surgeons of*  
10 *England.* 93(5):347-352.
- 11 Hochman JB, Kraut J, Kazmerik K, Unger BJ. 2014. Generation of a 3d printed temporal bone model  
12 with internal fidelity and validation of the mechanical construct. *Otolaryngol Head Neck*  
13 *Surg.* 150(3):448-454.
- 14 Jadad AR, Enkin MW. Bias in randomized controlled trials.
- 15 Lewis TL, Burnett B, Tunstall RG, Abrahams PH. 2014. Complementing anatomy education using  
16 three-dimensional anatomy mobile software applications on tablet computers. *Clin Anat.*  
17 27(3):313-320.
- 18 Lin Y, Wang X, Wu F, Chen X, Wang C, Shen G. 2014. Development and validation of a surgical  
19 training simulator with haptic feedback for learning bone-sawing skill. *J Biomed Inform.*  
20 48:122-129.
- 21 Linke R, Leichtle A, Sheikh F, Schmidt C, Frenzel H, Graefe H, Wollenberg B, Meyer JE. 2013.  
22 Assessment of skills using a virtual reality temporal bone surgery simulator. *Acta*  
23 *Otorhinolaryngol Ital.* 33(4):273-281.
- 24 MacDonald J, Williams RG, Rogers DA. 2003. Self-assessment in simulation-based surgical skills  
25 training. *Am J Surg.* 185(4):319-322.
- 26 Mandel LS, Goff BA, Lentz GM. 2005. Self-assessment of resident surgical skills: Is it feasible? *Am J*  
27 *Obstet Gynecol.* 193(5):1817-1822.
- 28 McDougall EM. 2007. Validation of surgical simulators. *J Endourol.* 21(3):244-247.
- 29 Metzler R, Stein D, Tetzlaff R, Bruckner T, Meinzer HP, Buchler MW, Kadmon M, Muller-Stich BP,  
30 Fischer L. 2012. Teaching on three-dimensional presentation does not improve the  
31 understanding of according ct images: A randomized controlled study. *Teach Learn Med.*  
32 24(2):140-148.
- 33 Moorthy K, Munz Y, Sarker SK, Darzi A. 2003. Objective assessment of technical skills in surgery. *Bmj.*  
34 p. 1032-1037.
- 35 Mowry SE, Jammal H, Myer Ct, Solares CA, Weinberger P. 2015. A novel temporal bone simulation  
36 model using 3d printing techniques. *Otol Neurotol.* 36(9):1562-1565.
- 37 Muller-Stich BP, Lob N, Wald D, Bruckner T, Meinzer HP, Kadmon M, Buchler MW, Fischer L. 2013.  
38 Regular three-dimensional presentations improve in the identification of surgical liver  
39 anatomy - a randomized study. *BMC Med Educ.* 13:131.
- 40 Ng CL, Liu X, Chee SC, Ngo RY. 2015. An innovative 3-dimensional model of the epitympanum for  
41 teaching of middle ear anatomy. *Otolaryngol Head Neck Surg.* 153(5):832-837.
- 42 Olivieri LJ, Su L, Hynes CF, Krieger A, Alfares FA, Ramakrishnan K, Zurakowski D, Marshall MB, Kim PC,  
43 Jonas RA et al. 2016. "Just-in-time" Simulation training using 3-d printed cardiac models  
44 after congenital cardiac surgery. *World J Pediatr Congenit Heart Surg.* 7(2):164-168.
- 45 Perrenot C, Perez M, Tran N, Jehl JP, Felblinger J, Bresler L, Hubert J. 2012. The virtual reality  
46 simulator dv-trainer((r)) is a valid assessment tool for robotic surgical skills. *Surg Endosc.*  
47 26(9):2587-2593.
- 48 Prinz A, Bolz M, Findl O. 2005. Advantage of three dimensional animated teaching over traditional  
49 surgical videos for teaching ophthalmic surgery: A randomised study. *Br J Ophthalmol.*  
50 89(11):1495-1499.
- 51 Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. 1997. Testing technical skill via an innovative  
52 "Bench station" Examination. *Am J Surg.* 173(3):226-230.
- 53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 Roberts C, Torgerson DJ. 1999. Baseline imbalance in randomised controlled trials. *Bmj*.  
4 319(7203):185.
- 5 Samia H, Khan S, Lawrence J, Delaney CP. 2013. Simulation and its role in training. *Clinics in Colon*  
6 *and Rectal Surgery*. 26(1):47-55.
- 7 Satava RM, Gallagher AG, Pellegrini CA. 2003. Surgical competence and surgical proficiency:  
8 Definitions, taxonomy, and metrics. *Journal of the American College of Surgeons*.  
9 196(6):933-937.
- 10 Schiekirka S, Reinhardt D, Beissbarth T, Anders S, Pukrop T, Raupach T. 2013. Estimating learning  
11 outcomes from pre- and posttest student self-assessments: A longitudinal study. *Acad Med*.  
12 88(3):369-375.
- 13 Shaharan S, Neary P. 2014. Evaluation of surgical training in the era of simulation. *World J*  
14 *Gastrointest Endosc*. 6(9):436-447.
- 15 Sullivan GM. 2011a. Deconstructing quality in education research. *J grad med educ*. Suite 2000, 515  
16 North State Street, Chicago, IL 60654. p. 121-124.
- 17 Sullivan GM. 2011b. Getting off the "gold standard": Randomized controlled trials and education  
18 research. *J grad med educ*. Suite 2000, 515 North State Street, Chicago, IL 60654. p. 285-289.
- 19 Sutherland LM, Middleton PF, Anthony A, Hamdorf J, Cregan P, Scott D, Maddern GJ. 2006. Surgical  
20 simulation - a systematic review. *Annals of Surgery*. 243(3):291-300.
- 21 Taffinder N, Smith SG, Huber J, Russell RC, Darzi A. 1999. The effect of a second-generation 3d  
22 endoscope on the laparoscopic precision of novices and experienced surgeons. *Surg Endosc*.  
23 13(11):1087-1092.
- 24 Tanagho YS, Andriole GL, Paradis AG, Madison KM, Sandhu GS, Varela JE, Benway BM. 2012. 2d  
25 versus 3d visualization: Impact on laparoscopic proficiency using the fundamentals of  
26 laparoscopic surgery skill set. *J Laparoendosc Adv Surg Tech A*. 22(9):865-870.
- 27 Toll E, Davis C. 2010. More trainees and less operative exposure: A quantitative analysis of training  
28 opportunities for junior surgical trainees. *Ann R Coll Surg Engl (Suppl)*. p. 170-173.
- 29 Van Bruwaene S, Schijven MP, Napolitano D, De Win G, Miserez M. 2015. Porcine cadaver organ or  
30 virtual-reality simulation training for laparoscopic cholecystectomy: A randomized,  
31 controlled trial. *Journal of Surgical Education*. 72(3):483-490.
- 32 Vivekananda-Schmidt P, Lewis M, Hassell AB, Coady D, Walker D, Kay L, McLean MJ, Haq I, Rahman  
33 A. 2007. Validation of msat: An instrument to measure medical students' self-assessed  
34 confidence in musculoskeletal examination skills. *Med Educ*. 41(4):402-410.
- 35 Winckel CP, Reznick RK, Cohen R, Taylor B. 1994. Reliability and construct-validity of a structured  
36 technical skills assessment form. *American Journal of Surgery*. 167(4):423-427.
- 37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1.** Twelve most important studies identified in the literature review

Ref.	Modality	Setting	n	Objective Evaluation	Subjective Evaluation	Evidence Level	Methodology validation	Outcome
(Alaraimi et al. 2014)	3D Display	Lab	50	Time and error rate	N/A	2b	Face validation	Improved accuracy, but no effect on global performance
(Tanagho et al. 2012)	3D Display	Lab	33	Time, number of repetitions and error rate	Survey	4	Validated assessment method (FLS)	Improved performance and preferred by participants
(Curro et al. 2015)	3D Display	Theatre	80	Time	Survey	2b	Yes (real patients)	3D not superior to 2D equivalent
(Felisati et al. 2013)	3D Display	Theatre	8	Post-operative follow up	Survey	4	Yes (real patients)	Improved performance and depth perception
(Engel et al. 2015)	3D Printing	Lab	19	Time and end product	N/A	4	Validated model and checklist	Validation of the model
(Lin et al. 2014)	3D Printing	Lab	25	Time, the maximal acceleration of drill	Survey	3b	Construct, face and transfer validity of the model	Validation of the model
(Ng et al. 2015)	3D Printing	Lab	72	Quiz	Survey	2b	No	Higher quiz scores and preferred by participants
(Chalas)	3D VR	Lab	26	Time and	Survey	4	Face, content	Validation

ani et al. 2011)				Accuracy			and construct validity	of the model
(Donnelly et al. 2009)	3D VR	Lab	89	Quiz	N/A	3b	No	3D not superior to 2D equivalent
(Metzler et al. 2012)	3D Images	Lab	73	Quiz	N/A	2b	No	3D not superior to 2D equivalent
(Drapkin et al. 2015)	3D Images	Lab	73	Quiz	Survey	2b	No	Higher quiz scores and preferred by participants
(Prinz et al. 2005)	3D Video	Lab	172	Quiz	Survey	2b	No	Higher quiz scores and preferred by participants