Proficiency-Based Progression (PBP) Training- the future model for dental operative skills training?: A systematic review and meta-analysis of existing literature

A Masters thesis submitted to Cork University Dental School & Hospital, National University of Ireland, Cork

for the qualification of
MDS

by

Dr. Elaine Kehily

January 2023
Head of School: Prof. Paul Brady
Supervisors: Prof. Anthony Roberts, Prof. Finbarr Allen, Prof. Anthony Gallagher

Table of contents

Abstract

Objective: To evaluate the effectiveness of proficiency-based progression (PBP) operative training using validated performance metrics, by comparing this to standard, conventional training methods.

Data: This systematic review was conducted in accordance with the PRISMA guidelines for the Transparent Reporting of Systematic Reviews and Meta-Analyses. Study quality was assessed using the MERSQI tool and the Cochrane Risk of Bias tool. Results were pooled using biased corrected standardized mean difference and ratio-of-means (ROM). Summary effects were evaluated using a series of fixed and random effects models. The primary outcome was the number of procedural errors performed comparing PBP and non-PBP-based training pathways. In quantitative synthesis testing for procedural errors, a pooled meta-analysis on 87 trainees was conducted using random-effects models. In a ROM analysis, PBP was estimated to reduce the mean rate of errors by 62%, when compared to standard training (ROM 0.38, 95% CI: 0.25; 0.58; $p < 0.001$)

Sources: The electronic databases of PubMed, Embase, Web of Science, MEDLINE and Cochrane library's CENTRAL were searched from inception to 8/11/2021. Filters activated were Randomized Controlled trials, clinical trial.

Study selection: 13 studies were included for review with 11 included in the quantitative synthesis from 174 potentially relevant publications identified by the search strategy. Main inclusion criteria were studies comparing standard surgical/operative training with proficiency-based simulation training using validated metrics based on expert performance.

Conclusions: Our meta-analysis found that PBP training improved trainees' performances, by decreasing procedural errors. There is sufficient evidence to explore PBP training for use in dental skills training.

Clinical significance

PBP training was estimated to reduce the mean rate of operative errors by 62%, when compared to standard training. Given that there is a direct correlation between operative skill and patient outcomes, these data suggest that there is sufficient evidence to explore PBP training for use in dental skills training.

Key points

Question: Does proficiency-based simulation training using validated performance metrics improve the surgical/operative skills of trainees compared with standard/conventional training methods without a requirement to achieve proficiency standards?

Findings: In this systematic review and meta-analysis that included 13 studies (11 were included for quantitative analysis), Proficiency Based Progression training was estimated to significantly reduce the mean rate of operative errors by 62%, when compared to standard training.

Meaning: There is sufficient evidence to recommend Proficiency-Based simulation training with validated performance metrics prior to operating on a live patient

Acknowledgments

To my supervisors Professor Roberts, Professor Allen and Professor Gallagher- thank you for all of your support and guidance throughout this process and for your patience and flexibility as the design and execution of this project changed.

To my many work colleagues who helped me along the way and gave me invaluable advice based on their own personal research experience. Who have inspired me and driven me to keep pursuing my career in academia.

To the Health Research Board who funded this research and continue to fund my research and career development- for their understanding and willingness to accommodate changes to the project along the way.

And finally to my husband Barry who always supports me in life and my two wonderful daughters who I love endlessly and who have taught me more than anyone else ever could.

Originality declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources and clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Index of tables

Index of figures

1. Introduction

The acquisition of practical skills for clinicians providing treatment for their patients is clearly an important training consideration. Nowhere is this more evident than the training requirements for operative procedures by doctors, dentists and other health care workers with a direct correlation between operative skill and patient outcomes[1]. Operative procedures in this respect are hands-on medical or dental procedures that typically involve a high level of manual operative skill. Examples in dentistry include: a filling, a root canal treatment, oral surgical procedures such as wisdom tooth removal etc.

Proficiency-Based Progression (PBP) Training

Simulation-based training has a growing role to play in skill acquisition and re-validation[2] with Virtual Reality being first proposed by Satava almost 30-years ago for the acquisition of surgical skills[3]. More specifically, Proficiency-Based Progression (PBP) simulation training has come to the forefront of surgical skill development and maintenance over recent years[4]. PBP is a robust methodology where the operative procedure in question is subject to a task analysis to identify performance metrics essential to the completion of the task. These include steps, errors and critical errors:

- *Steps* are stages of a procedure that are unambiguously defined (including a strictly defined start and end point) so that they can be observed (and/or scored) in a binary fashion[2].
- *Errors* are procedure actions which deviate from optimal practice and are not necessarily bad but potentially unsafe[2].
- *Critical errors* in contrast are procedure actions which are unsafe but may not always lead to a bad outcome[2].

These performance metrics are then subject to a validation procedure (face and content, construct, concurrent and predictive validity) and can be used to establish an objectively pre-defined proficiency benchmark performance to which trainees are trained:

- Face validity is the degree to which the performance metrics will measure what they are designed to measure i.e. operator performance
- Concurrent validation: do operators who score highly according to metric scoring also perform well on similar or related procedures/tasks?
- Construct validation refers to the performance metrics ability to distinguish between expert and novice operator performance i.e. does a more experienced operator score more highly than a novice or inexperienced individual?[2]
- Predictive validation: does the individuals metric scoring predict future skilled performance?[2]

The proficiency benchmark performance is typically based on the mean performance of experienced practitioners.

The performance characteristics of a proficient operator are defined in Table 1. A study by Seymour et al[5] demonstrated that trainees who underwent a virtual reality (VR)-based simulation training pathway performed significantly better than traditionally trained surgeons. It was the first study to introduce the ‘‘Proficiency-Based Progression’’ (PBP) training methodology as an evidence-based alternative to more traditional training.

It was also the first prospective, randomized, double-blind clinical study of simulation-based training for the operating room which demonstrated that surgical residents trained to a proficiency benchmark on a VR simulator made significantly fewer objectively assessed intra-operative errors when compared to the control group. These results have been replicated in other studies and there is increasing confidence that simulation-based training produces a superior skill set when compared to traditional training methods.

STAGE	PERFORMANCE CHARACTERISTICS
Expert	• Source of knowledge and information for others • Continually looks for better methods • Work primarily from intuition • Being forced to follow rules degrades performance
Proficient	• Seeks to understand larger context • Frustrated by over-simplification • Can self-correct performance • Can learn from experience of others
Competent	• Can troubleshoot problems on his/her own • Seeks out expert user advice • Develops conceptual models
Advanced Beginner	• Starts trying tasks on his/her own • Has difficulty troubleshooting • Begins to formulate principles, but without holistic understanding
Novice	• Has little or no previous experience • Doesn't know how to respond to mistakes • Needs rules to function

Table 1. Five stage model of adult skill acquisition- Dreyfus & Dreyfus[6]

Current training and assessment methodologies in dentistry

Unlike surgery and other medical disciplines, dental practitioners are expected to be competent to operate independently when they graduate. Traditionally in dentistry, undergraduates are trained to levels of competence with benchmarking based on a performance level reached by consensus amongst senior staff members. Competences are typically shaped around guidelines such as the ADEE (Association for Dental Education in Europe) 'The Graduating European Dentist' framework[7], which focuses on 4 domains of dental education: Professionalism, Safe and Effective Clinical Practice, Patient-Centred Care, and Dentistry in Society. Contrary to this, metric-based training to proficiency, or PBP, is benchmarked on the actual mean performance level of experienced or proficient practitioners.

An important goal for dental educators is to establish evidence-based validated quality assured training programmes across dentistry. Our methods of assessment in clinical and pre-clinical training should exhibit high validity and reliability:

- *Validity* refers to the extent that an assessment measure what it is supposed to measure
- *Reliability* refers to the extent to which an assessment tool consistently and accurately measures performance

Our assessment tools for dental undergraduate and postgraduate performance should exhibit construct validity (do more skilled individuals perform better on their assessments than less skilled or experienced individuals?) and predictive validity (do our assessments predict future skilled performance?)[2] and define a skill level that is transparent, objective and fair. Metric-based training to a pre-defined benchmark, such as proficiency, demonstrates all of these characteristics.

Conventionally skills in dentistry have been acquired through repeated practice in the clinical or simulation environment with summative feedback at the end of the procedure or when the trainee requests it to help facilitate learning. Using a feedback approach with metrics which are strictly defined and validated allows the trainee to know if they are performing a procedure correctly or incorrectly or in the right or wrong order. In PBP training, trainees should receive this feedback proximate (i.e. immediately after) to a performance error[2] with this approach being termed deliberate practice.

Currently, undergraduates are trained to competency in Cork University Dental School and Hospital (CUDSH), pre-clinically and clinically. Pre-clinical skills training takes place in the Operative Techniques Laboratory (OTL) using dental simulator mannequins with simulation model maxillary and mandibular teeth with plastic typodonts. (Figures 1 and 2).

Figure 1. A dental simulator mannequin used in CUDSH for pre-clinical skills training

Figure 2. Maxillary simulation model with plastic typodonts used for pre-clinical skills training in CUDSH

Clinical skills are taught in a traditional manner through repeated practice with summative feedback as described above. At the beginning of the academic year, undergraduates are provided with a 'schedule' of operative procedures they are required to complete by the end of the year (Figure 3) with the level of complexity increasing as they progress.

Adhesive Restoration Schedule

For each restoration you are required the teaching staff each of the following stages (where applicable);

1. Cavity preparation
2. Matrix placement
3. Completed Restoration

Fissure sealants: 38________________ 48_________________

Student

Schedule Commenced

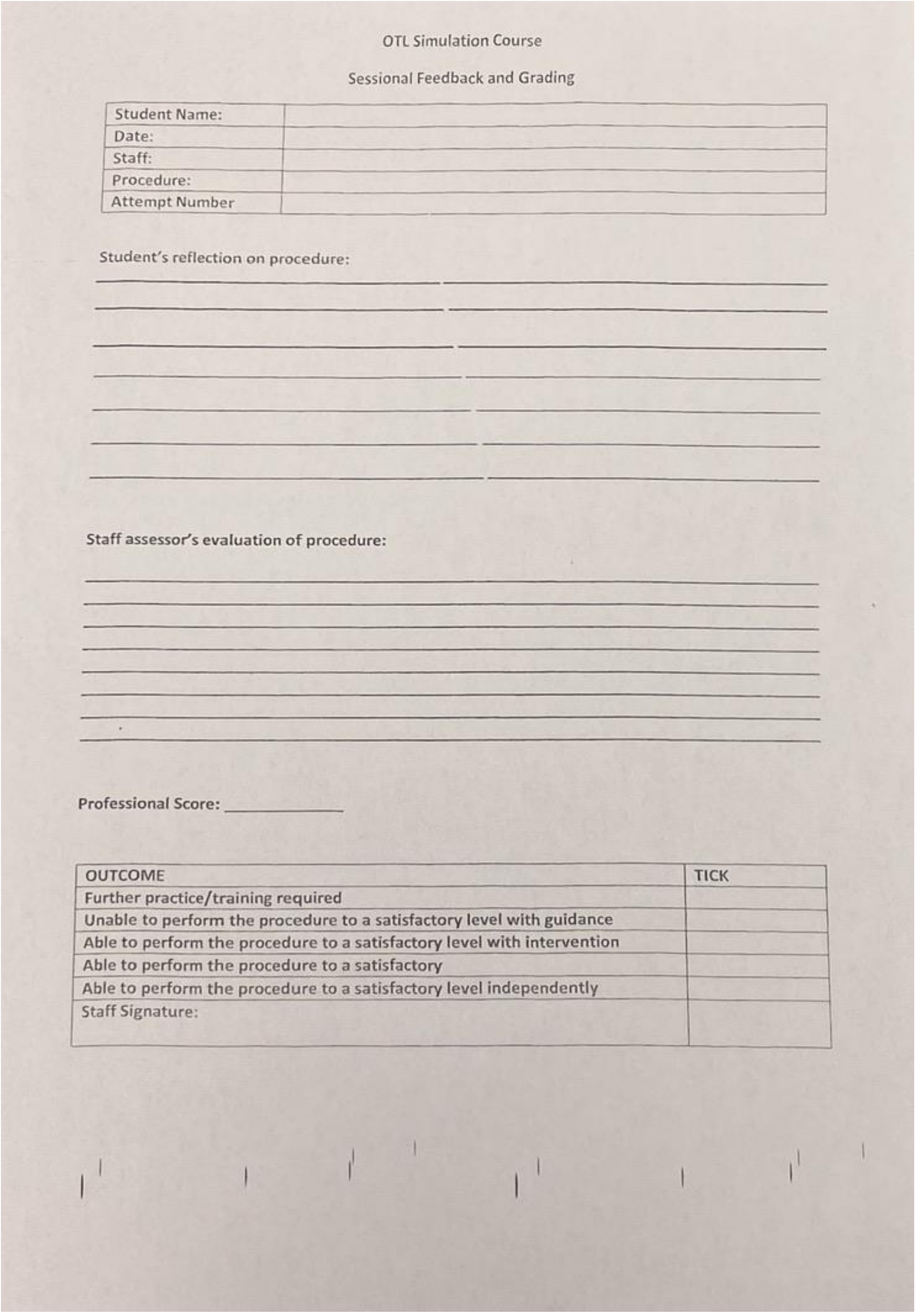
Restoration	Start date	Cavity prep	Matrix	Restoration and polish	Finish date
45^{o}					
26^{o}					
12^{m}					
22^{d}					
41^{d}					
42^{m}					
11^{di}					
21^{mi}					
$35^{\text{mo (slot)}}$					
46^{mo}					
$14^{\text{mo(slot)}}$					
16^{mo}					

Schedule completion date:

Staff signature:

Figure 3. An example of an operative schedule used in CUDSH pre-clinical skills training

Undergraduates are typically permitted three attempts at each procedure, with a ‘pass’ mark required before being allowed to continue to the next procedure. Throughout the session, the student can request feedback as they perform the procedure and are given either verbal or procedural intervention. At the end of each clinical session, which typically lasts three hours, the student is provided with summative feedback where they are encouraged to reflect on their own performance (Figure 4) and given a professional score depending on the outcome of the session as determined by the clinical tutor. Clinical scoring is guided by a scoring rubric used by clinical tutors. This consists of specific ‘criteria’ used to score the operative procedure being performed (Figure 5). These criteria are however not strictly defined and are ambiguous in nature leading to the possibility of subjective and unfair assessment.



OTL Simulation Course

Sessional Feedback and Grading

Student Name:	
Date:	
Staff:	
Procedure:	
Attempt Number	

Student's reflection on procedure:

Staff assessor's evaluation of procedure:

Professional Score: ____________

OUTCOME	TICK
Further practice/training required	
Unable to perform the procedure to a satisfactory level with guidance	
Able to perform the procedure to a satisfactory level with intervention	
Able to perform the procedure to a satisfactory	
Able to perform the procedure to a satisfactory level independently	
Staff Signature:	

Figure 4. Sessional feedback and grading form used in CUDSH for pre-clinical skills training

Quality Evaluation Criteria for Prepared Composite Resin Cavities

Rating	Grade	Finish of Walls and Margins – Cavity Definition	Retention	External Outline	Internal Outline
Satisfactory	4	Walls and margins smooth, cavity well defined		Minimum extension to remove caries lesion only. Convenience form to allow visual access to ADJ and allow material to be placed in cavity.	Extended into dentine No excessive tissue loss
	3	Slight roughness of cavity walls or margins Slight lack of cavity definition		External outline form slightly underextended External outline form slightly overextended	Pulpal or axial walls shallow. Pulpal or axial walls deeper than required
Unsatisfactory	2	Walls or margins rough		Cavity extended beyond margins of lesion. Unsupported enamel removed Supporting or adjacent tooth damaged	Pulpal or axial walls extended beyond caries lesion (too deep) Pulpal floor, axial wall entirely in enamel
	1	Enamel margins grossly overextended		External outline grossly underextended External outline grossly overextended Supporting or adjacent tooth mutilated	Mechanical pulp exposure Frank caries remaining

Figure 5. An example of a scoring rubric for an operative procedure taught in the pre-clinical skills course in CUDSH

This method of assessment displays low levels of validity and reliability due to its subjective nature. Employing an assessment tool with a pre-defined benchmark of performance and tightly-defined metrics (as in PBP) would give us more information on how the student's performance compares to the 'gold standard' or expert performance (construct validity) and how we can expect them to perform in the future (predictive validity) so we could therefore target our training in a more efficient and effective way.

Current recommendations for dental undergraduate training

A recent systematic review of digital undergraduate education in dentistry[8] found that the use of Virtual Reality (VR) technologies for motor skills training is increasing. The review concluded that these technologies (including web-based knowledge transfer/e-learning, digital surface mapping and VR- based simulators) are valuable in dental undergraduate and postgraduate education but also highlighted the need for a better evidence base for the utility of VR in dental education. This coincides with reported shortages of suitable patients available for undergraduate students to treat[9] and a corresponding reduction in clinical experience. In a scoping review of methods and trends in undergraduate clinical skills teaching in Ireland and the UK, McGleenon et al[9] highlight that the use of clinical competencies without minimum experience requirements had the effect of "reduced confidence and perceived preparedness for practice". This evidence further underlines the importance of VR and simulation (including phantom head, computer-supported and haptic-enhanced VR simulators) in the future of dental skills training. However, without formalised performance metrics and effective feedback it's pedagogical effectiveness remains unclear. The systematic review by Zitzmann et al[8] found that operative performance on VR units, such as haptic simulators, with continuous feedback from clinical instructors led to better quality of tooth preparation than any real-time feedback from the VR simulator itself.

In a recent ADEE (Association for Dental Education in Europe) consensus paper[10] a special interest group focusing on the teaching of pre-operative clinical skills in dentistry made the following recommendations when planning operative skills courses:

- "educators should employ and co-create resources with students that break down difficult tasks", and
- "educators should consider whether to plan skills sessions in relation to deconstructed skills development rather than clinical presentation and complexity"

Proficiency-Based-Progression training has the potential to fulfil both of these requirements when planning dental operative skills training.

As PBP training has yet to be utilised in dentistry, we can only look at how this training methodology is employed in surgical operative skills training and the effect that this has on both operator performance and patient outcomes. The simulators utilised (mannequins, box trainers, high fidelity simulators) in these surgical training programmes (Table 2) bear a close resemblance to the phantom head, computer-supported and haptic-enhanced VR simulators currently used in dental training globally and so the authors believe is comparable to dental undergraduate and postgraduate training.

Proficiency-Based Progression Training in Dentistry

A recent systematic review and meta-analysis carried out by Mazzone et al[11] showed that PBP training reduced the number of performance errors by 60% and procedural time by 15%. These data demonstrate significant advantages of PBP training, however their review did not focus specifically on the use of proficiency-based simulation training with validated performance metrics and did not include a Risk of Bias analysis or evaluate the impact of PBP training on Likert scale scoring.

Given the paucity of data on Proficiency-Based-Progression training in Dentistry, the aim of this systematic review was to evaluate the effectiveness of proficiency-based surgical/operative training (the closest comparator for dental clinical skills training) using validated performance metrics, by comparing this to standard conventional training methods. Uniquely, this study considers important aspects of the quality and risk of bias of the studies and, if the results are compelling, could justify the exploration of PBP training in dentistry.

2. <u>Materials and methods</u>

This systematic review was conducted in accordance with the guidelines of the Preferred Reporting of Systematic Reviews and Meta-Analyses (PRISMA)[12].

The primary aim of this systematic review was to answer the following PICO question: Does proficiency-based simulation training using validated performance metrics improve the surgical/operative skills of trainees compared with standard/conventional training methods without a requirement to achieve proficiency standards

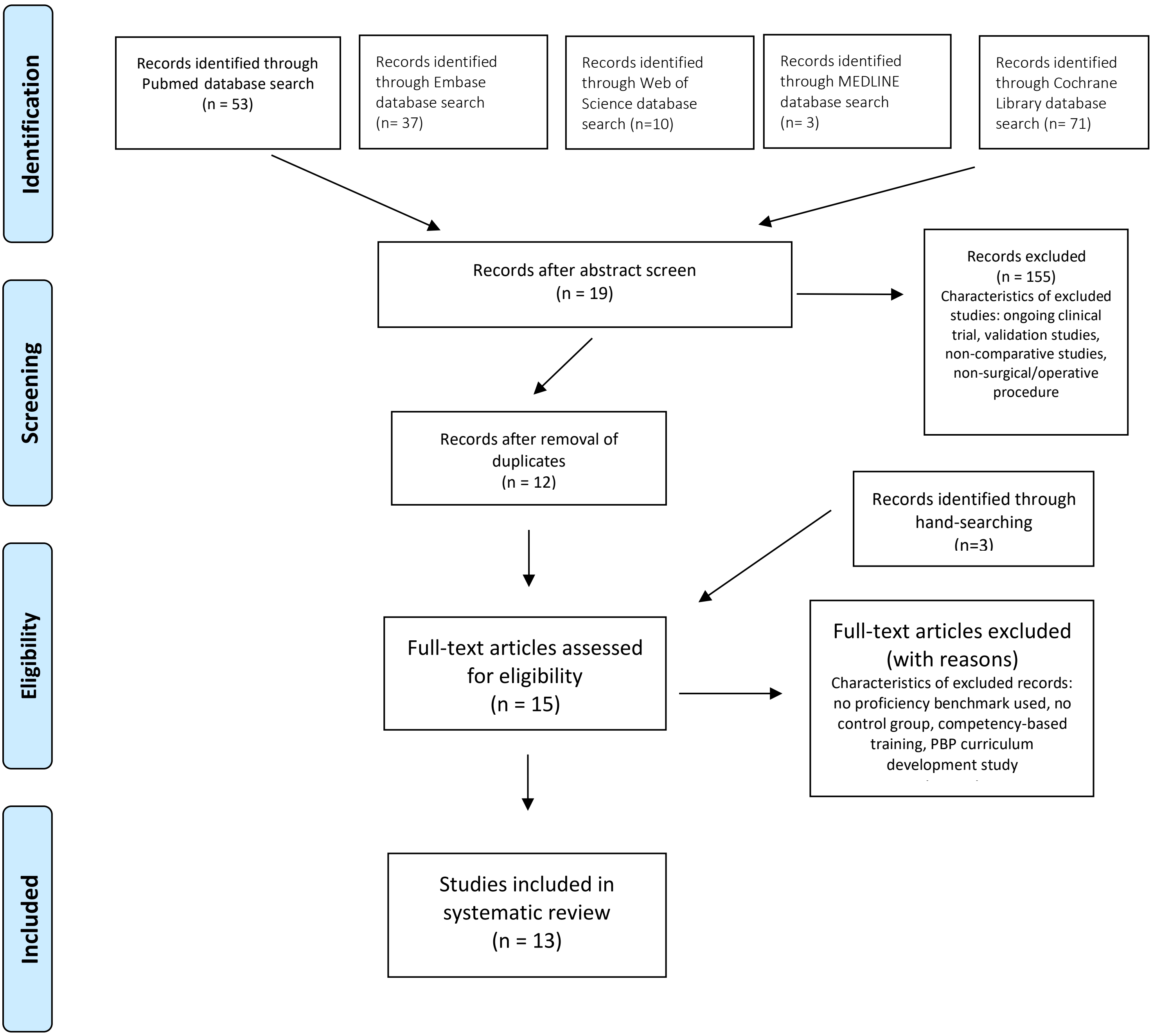
2.1 Search strategy and selection criteria

The electronic databases of PubMed, Embase, Web of Science, MEDLINE and the Cochrane library were searched (last search date 08/11/2021). The search terms entered were validat* AND ((performance) AND metrics) AND surgi* AND proficiency. The filters activated were Randomized Controlled trials, clinical trial. The databases were searched with no publication date limits and no language restriction. Two reviewers conducted the search independently (EK and NC), and any conflict was resolved with discussion. The references quoted in the full text articles were hand searched for any further eligible studies. At the end of the process, 13 and 11 studies have been included for, respectively, the qualitative synthesis and the quantitative meta-analysis.

2.2 Eligibility criteria for included studies

The main inclusion criteria were studies comparing standard surgical/operative training with proficiency-based simulation training using validated metrics based on expert performance were included. All Randomised Controlled Trials (RCTs), quasi-randomised controlled trials and non-Randomised Controlled Trials (CCTs) were considered in the search. Only studies using a validated surgical simulator were included.

Studies involving the training of novice or inexperienced operators in a specific procedure (e.g. surgical residents, specialist trainees) were included, where the proficiency benchmark for training was based on expert (experienced surgeons) performance. A PRISMA flow diagram describing the search process is shown in Figure 6.



PBP: Proficiency-Based-Progression; RCT: Randomized controlled trial

Figure 6. Flow-chart of studies through the screening process according to the PRISMA methodology

The primary outcomes reported were metrics such as time, errors, the number of procedural steps achieved, and whether the proficiency benchmark was reached. Secondary outcome measures included baseline abilities (psychomotor skills), efficiency/economy of motion, the number of training repetitions required to achieve proficiency, economy of movement, tissue handling, instrument control, tool manipulation, tool path length, skill retention (at 6 months, 12 months, 18 months and 24 months), and the number of 'consultant' (senior clinician) takeovers. The characteristics of each study included in this study are presented in Table 2.

Study	Methods	Subjects N; Type	Comparison arm	Task/Procedure	Primary outcomes measured	Other scale used	Intraoperative patient performance	MERSQI score[h]
Ahlberg et al[33]	RCT	13; Residents (PGY 1-2)	Swedish National Surgical Residency Training Programme	Laparoscopic cholecystectomy	Errors, time	_____	Yes	17
Angelo et al[27]	RCT	44; Residents (PGY4-5)	ACGME approved Orthopedic Residency & Arthroscopy Association of North America Shoulder Course	Arthroscopic Bankart Procedure	Errors, steps, time	_____	Yes	15.5
Cates et al[22]	RCT	12; interventional cardiologists	Industry sponsored CASES education and training system	Carotid artery angiography	Errors, time	______	Yes	15
Fried et al[31]	Prosp. cohort	14; residents (PGY1-3)	ACGME approved Otolaryngology Surgery Residency Programme	Endoscopic sinus surgery (ES3)	Errors, time	______	Yes	12.5
Gerull et al[24]	Prosp. cohort	31; residents	ACGME approved General Surgery, Urology & Obstetrics and Gynecology Residency Programmes	RALS[a] case	________	RO-SCORE[b]; NTLX[c]	Yes	11.5
Kurashima et al[29]	RCT	16; residents (PGY 2-5)	General Surgery Residency Training Programme McGill University Canada	Inguinal hernia repair	________	GOALS-GH[d]	Yes	12.5
Lendvay et al[32]	RCT	51; residents (PGY 1-6)	ACGME accredited University of Washington Medical Center and Madigan Army Medical Center Residency Programmes	Laparoscopic procedure	Time, errors	______	Yes	13
Maertens et al[25]	RCT	32; residents (PGY 1-6)	Master of Medicine in Specialist Medicine (Surgery) training Programme, Ghent University	Endovascular procedure	Time, steps	OSATs[e]	Yes	16
Palter et al[30]	RCT	25; residents (PGY1-2)	ACGME approved General Surgery Residency Training Programme	Laparoscopic Right Colectomy	Time, errors, steps	OSATs	Yes	15
Seymour et al[5]	RCT	16; residents (PGY 1-4)	ACGME approved General Surgery Residency Training Program	Laparoscopic cholecystectomy	Time, errors, steps	______	Yes	14
Srinivasan et al[23]	RCT	17; residents	Irish National Anesthesia Training Program	Epidural analgesia	Errors	GRS[f], TSCL[g]	Yes	15
Sroka et al[28]	RCT	17; residents (PGY 1-3)	General Surgery Residency Training Programme McGill University Canada	Laparoscopic cholecystectomy (FLS training programme)	________	GOALS	Yes	15
Van Sickle et al[26]	RCT	22; residents (PGY 3,5,6)	ACGME approved General Surgery Residency Training Program	Nissen fundoplication	Errors, time	______	Yes	15

2Operative Assessment of Laparoscopic Skills-Groin Hernia); e. OSATs (The modified Ottawa Surgical Competency Operating Room Evaluation for surgical robotic cases); f. GRS (Global Rating Scales); g. TSCL, Task-Specific Checklists; h. MERSQI (Medical Education Research Study Quality Instrument);

Table 2. General characteristics of 13 randomized clinical trials studies included in the final qualitative analysis of the systematic review.

Operative metric validation studies were excluded from this review. Studies that used standard/manufacturer simulator metrics to measure outcomes were also excluded (metrics needed to be proficiency-based i.e. based on expert performance, and validated). Ongoing clinical trials registered on the Cochrane Central Register of Controlled Trials (CENTRAL) database were also excluded.

2.3 Data extraction and quality assessment

Study methods, participants, interventions and outcomes were extracted and recorded independently by two reviewer authors (EK and NC). These were compared, and a third reviewer (AR) was available for resolution of any differences.

The quality of each trial was assessed using the Medical Education Research Study Quality Instrument (MERSQI) tool (Figure 7) and the Cochrane Collaboration's 'Risk of Bias' tool (Figures 8 and 9) (http://handbook.cochrane.org). Items evaluated to generate this score included randomisation, sequence generation, blinding, allocation concealment and incomplete outcome data.

Domain	**MERSQI Item**	**Score**	**Max Score**
Study design	Single group cross-sectional or single group posttest only	1	3
	Single group pretest & posttest	1.5	
	Nonrandomized, 2 groups	2	
	Randomized controlled trial	3	
Sampling	*Institutions studied*:		3
	1	0.5	
	2	1	
	3	1.5	
	Response rate, %:		
	Not applicable		
	<50 or not reported	0.5	
	50-74	1	
	>75	1.5	
Type of data	Assessment by participants	1	3
	Objective measurement	3	
Validity of evaluation instrument	*Internal structure:*		3
	Not applicable		
	Not reported	0	
	Reported	1	
	Content:		
	Not applicable		
	Not reported	0	
	Reported	1	
	Relationships to other variables:		
	Not applicable		
	Not reported	0	
	Reported	1	
Data analysis	*Appropriateness of analysis:*		3
	Inappropriate for study design or type of data	0	
	Appropriate for study design, type of data	1	
	Complexity of analysis:		
	Descriptive analysis only	1	
	Beyond descriptive analysis	2	
Outcomes	Satisfaction, attitudes, perceptions, opinions, general facts	1	3
	Knowledge, skills	1.5	
	Behaviors	2	
	Patient/health care outcome	3	
Total possible score*			**18**

Figure 7. Medical Education Research Study Quality Instrument (MERSQI) tool

2.4 Data synthesis and statistical analysis

Data not suitable for meta-analytic evaluation was presented in narrative fashion (qualitative analysis). Meta-analysis was deemed appropriate as 11 included studies met the following criteria (as recommended by Cochrane Consumers and Communication Group[13]):

- outcomes are comparable (i.e. errors, time, steps, Likert scales) and could be pooled meaningfully
- interventions and comparators are similar enough to be combined meaningfully
- correct data are available for the included studies.

In quantitative synthesis, the reported results for continuous outcomes were pooled using biased corrected standardized mean difference (SMD) (Hedges' g effect size) according to previous established methodology.[14,15] Moreover, as previously described, ratio of means (ROM) was applied to provide an estimation of the pooled effect of PBP on the considered outcomes. [16-18] All results were reported with 95% confidence intervals.

Heterogeneity between studies was measured using the I2 statistic [19] and the between-study variance (t2) from the random-effect analyses. I2 values >50% indicate large inconsistency. Unless otherwise indicated all models have allowed for different effect sizes (random effects). In case of large heterogeneity, random effect models (using the DerSimonian and Laird approach [20]) were used.

For the assessment of small study effects and publication bias, values of the SMD or ROM were plotted against their standard error in a contour-enhanced funnel plot. The latter bias represents the error in connection with whether a study is published or not depending on the characteristics and result of individual studies.[21] This error is caused because statistically significant study results generally have a higher likelihood of being published. Statistical significance for all analysis was defined as two-sided $p < 0.05$. Statistical analysis was performed with the R software (version 3.6.3; http://www.r-project.org/).

3. Results

The search strategy identified 174 potentially relevant publications; 53 from PubMed, 37 from Embase, 10 from Web of Science, 3 from Medline and 71 from the Cochrane Library. Following screening of these abstracts and application of the inclusion and exclusion criteria, 155 records were excluded and a further 3 studies following removal of duplicates. Thus, the full texts of 15 articles were retrieved for screening. And 10 were subsequently included in the systematic review. A further three papers were identified by hand-searching of bibliographies, therefore a total of 13 studies were included in this review. This process is described in the flow diagram in Figure 6.

3.1 Quality and risk of bias assessment

Figure 7 summarises the quality criteria assessed for each study using the MERSQI tool. This tool includes 10 items, reflecting six domains of study quality: (a) study design, (b) sampling, (c) type of data, (d) validity, (e) data analysis, and (f) outcomes. The maximum domain score is 3, producing a maximum possible MERSQI score of 18 and potential range of 5-18. The overall methodological quality of the studies was high with a mean MERSQI score of 14.38 (range 11.5-17).

The quality of each trial was assessed using the Cochrane Collaboration's 'Risk of Bias' tool (http://handbook.cochrane.org) using Review Manager 5.3 software. Figure 8 shows the Risk of Bias Summary and Figure 9 shows the Risk of Bias Graph.

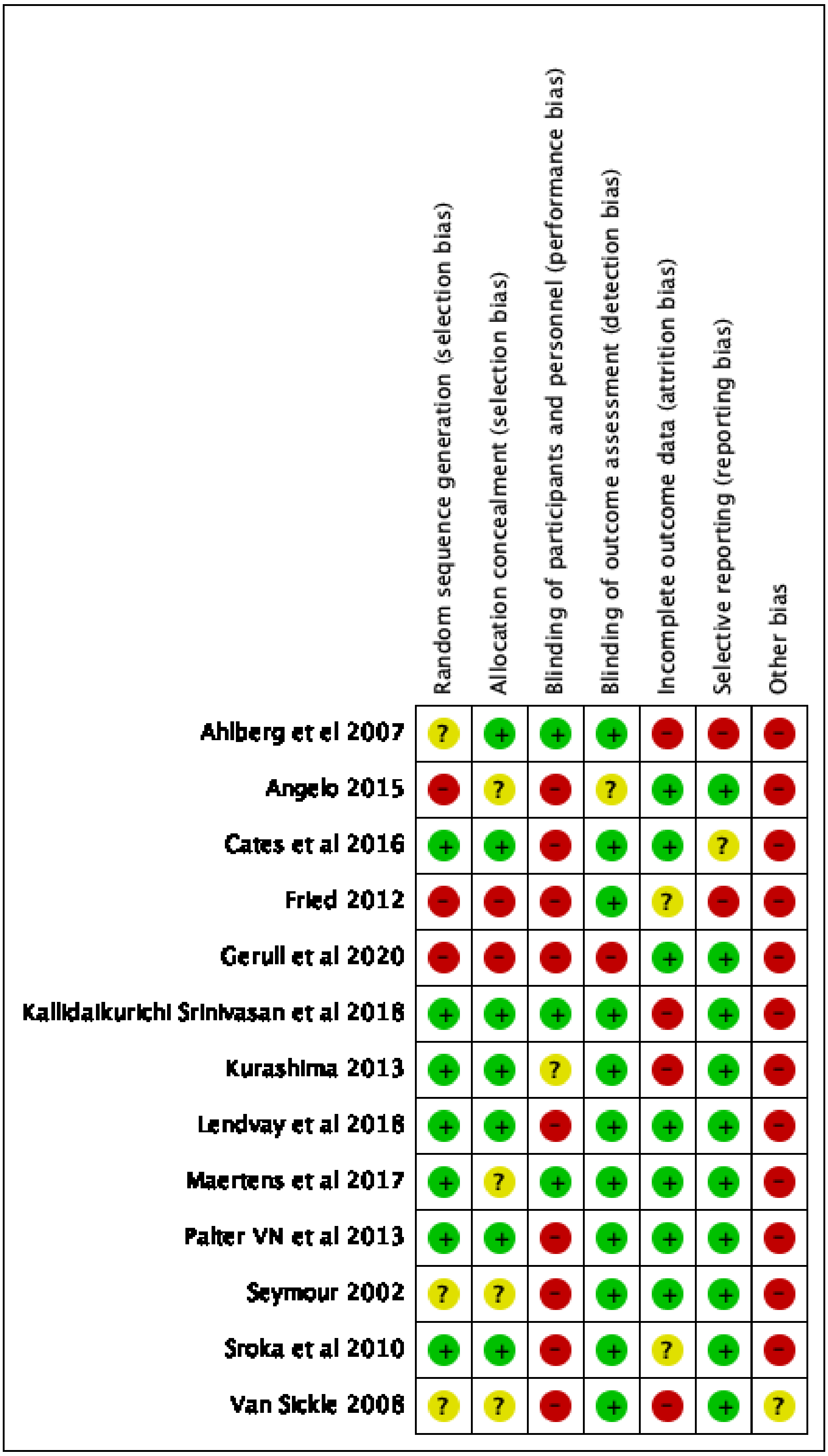


Figure 8. Risk of Bias Summary (Green indicates a low risk of bias; Yellow indicates an unclear risk of bias; Red indicates a high risk of bias)

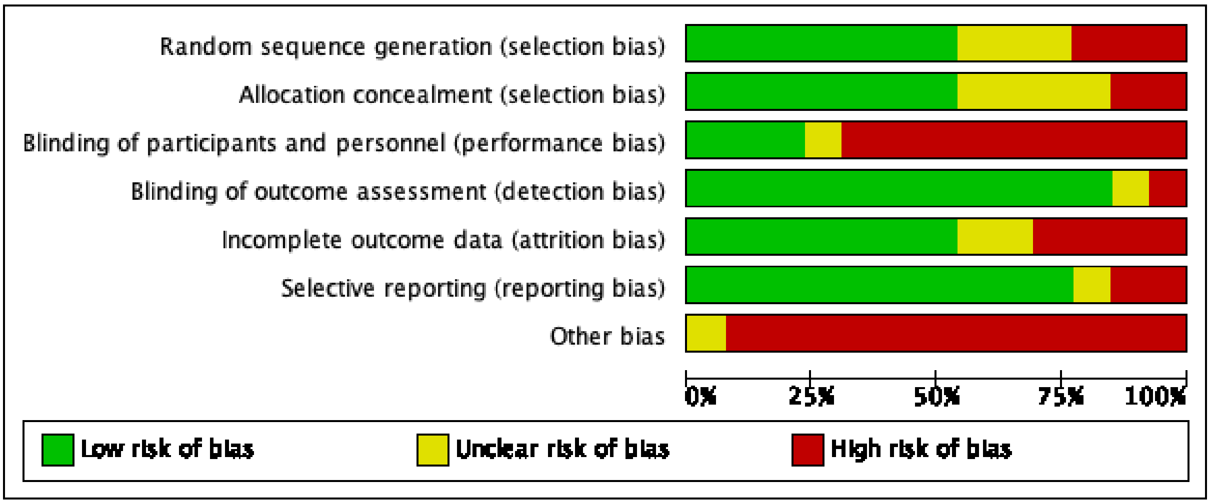


Figure 9. Risk of Bias Graph (green indicates a low risk of bias; yellow indicates an unclear risk of bias; red indicates a high risk of bias)

Most studies had a low risk of bias for random-sequence generation (53.8%), allocation concealment (53.8%) and incomplete outcome data (53.8%). Of the 13 studies, 12 (92.3%) had a high risk of bias for other types of bias. Performance bias was more difficult to assess in this review due to the difficulty associated in blinding participants in surgical clinical trials. The authors of this review decided where a genuine attempt was made to liken the training conditions in the control group to the intervention group a judgment of low risk would be assigned. Where there was no attempt to blind participants from their training status then a judgment of high risk would be acceptable. Most studies (10 [77%]) had a low risk of bias in the domain of selective outcome reporting.

The GRADE approach was also utilised to rate the certainty of evidence and to assess the quality of findings using the outcomes 1. No. of errors, 2. No. of steps, 3. Time and 4. Likert scale scoring (Table 3 shows the summary of findings table).

Outcomes	Standardised mean difference (95% CI) (between standard/traditional training group (control) and PBP training group)	Relative effect (95% CI) ROM	No. of participants (studies)	Quality of the evidence (GRADE)	Comments
No. of errors	3.11 (-4.54; -1.68)[1]	0.38 (0.25; 0.58)[2]	87 (5)	Moderate[3]	Lower number of errors indicates improved performance
No. of steps	3.90 (1.79; 6.02)[4]	1.28 (0.94; 1.74)[5]	42 (2)	Low[6]	Higher number of steps completed indicates improved performance
Time	0.81 (-1.40; -0.21)[7]	0.81 (0.66; 0.98)[8]	93 (5)	Moderate[9]	
Likert scale scoring	3.65 (1.40; 5.90)	1.52 (1.22; 1.90)[10]	49 (4)	Moderate[11]	

CI: Confidence Interval; ROM: Ratio of Means

1. Overall, PBP training reduced the number of errors when compared to standard training
2. PBP was estimated to reduce the mean rate of errors by 62%, when compared to standard training
3. All 5 included studies were RCTs but downgraded from high quality to moderate due to risk of bias assessment
4. Overall, trainees who completed PBP training performed more procedural steps than those who completed a standard training pathway
5. PBP did not statistically significantly increase the mean rate of steps performed when compared to standard training
6. Data is sparse (only two studies included in analysis)
7. The reduction of procedural time was less pronounced compared to other outcomes, such as the number of errors or steps completed
8. PBP reduced the mean procedural time by approximately 15%, when compared to standard training
9. All 5 included studies were RCTs but downgraded from high quality evidence to moderate due to incomplete outcome data in one study
10. Trainees who completed PBP achieved a mean Likert scale-based score 52% higher than those who completed a standard training pathway
11. Downgraded from high quality evidence to moderate due to high % of heterogeneity

Table 3. Summary of findings table (using the GRADE approach for quality of findings assessment)

3.2 Evidence synthesis

Table 2 summarizes the general and design characteristics for each study. (Appendix ii outlines the descriptive characteristics for each included study).

Participants in all studies were surgical residents (specialist trainees) varying from postgraduate years (PGYs) 1-6 except for the study by Cates et al [22] which assessed interventional cardiologists with no experience of the simulated procedure (carotid artery angiography), and the study by Srinivasan et al[23] where anaesthesia trainees were trained. Proficiency benchmark setting is described by all studies with proficiency benchmark levels based on mean expert operator performance (with the classification 'expert' defined by each study).

3.3 Outcome measures

A number of methods were used for data collection amongst included studies. Eight out of thirteen studies ([23-30]) used a validated assessment tool to measure the steps of procedure (metrics) completed.

Errors were specifically measured in nine studies [7,22,23,26,27,30-33], with Angelo et al also noting sentinel errors performed (an event or occurrence involving a serious deviation from optimal performance during a procedure that either (1) jeopardized the success or desired result of the procedure or (2) created significant iatrogenic insult to the patient's tissues). Assessment tools utilized in four studies [24,25,28,29] (RO-SCORE, OSATs, GOALS-GH) did not include a specific error score/domain.

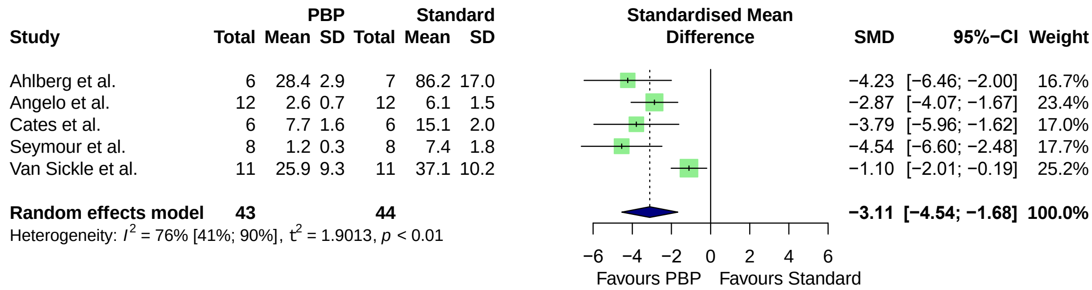
3.4 Inter-rater reliability

Eight out of thirteen studies included in this review [7,22,23,25-27,31,33] calculated Inter-Rater Reliability (IRR) to assess the degree of agreement among raters (of operator performance). In six out of these eight studies an IRR of 0.8 or 80% was deemed acceptable and achieved.

3.5 Results of quantitative synthesis

In quantitative synthesis testing for procedural errors, a pooled meta-analysis on 87 trainees was conducted (Fig. 10a-b), using random-effects models. Overall, PBP training reduced the number of errors when compared to standard training (SMD -3.11, 95% CI: -4.54; -1.68; $p < 0.001$). In a ROM analysis, PBP was estimated to reduce the mean rate of errors by 62%, when compared to standard training (ROM 0.38, 95% CI: 0.25; 0.58; $p < 0.001$). Funnel plots showed evidence for potential publications bias (Figure 14).

A.



B.

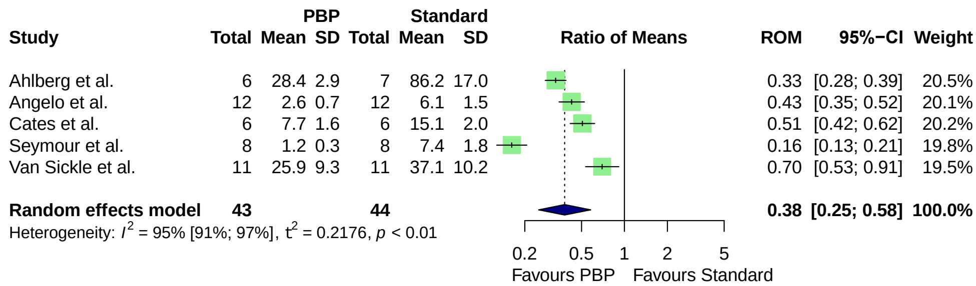
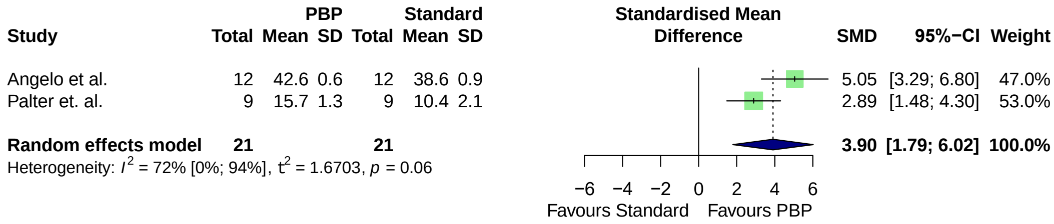


Figure 10. Standardized mean difference (6A) and ratio of means (6B) between studies assessing the effect of proficiency-based progression vs standard training on procedural errors

In quantitative synthesis testing for number of steps completed, a pooled meta-analysis on 42 trainees was conducted (Fig. 11a-b). Overall, trainees who completed PBP training performed more procedural steps than those who completed a standard training pathway (SMD 3.90, 95% CI: 1.79; 6.02; p < 0.001) (Fig. 11a). However, at ROM analysis, PBP did not statistically significantly increase the mean rate of steps performed when compared to standard training (ROM 1.28, 95% CI: 0.94; 1.74; p =0.1) (Fig. 11b). Funnel plots recorded a marginal effect for potential publications bias (Figure 14).

A.

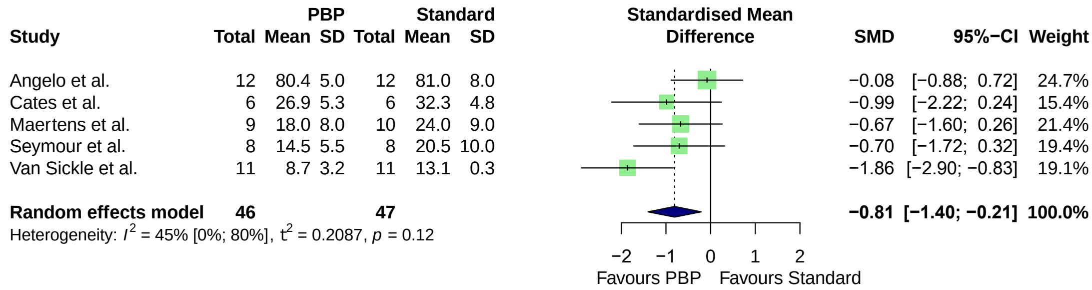


B.

Figure 11. Standardized mean difference (7A) and ratio of means (7B) between studies assessing the effect of proficiency-based progression vs standard training on the number of procedural steps

In quantitative synthesis testing for procedural time, a pooled meta-analysis on 98 trainees was conducted (Fig. 12a-b). Overall, trainees who completed PBP training performed the task/procedure in less time than those who completed a standard training pathway (SMD -0.81, 95% CI: -1.40; -0.21; p = 0.008) (Fig. 12a). The reduction of procedural time was less pronounced compared to other outcomes, such as the number of errors or steps completed. Indeed, at ROM analysis, PBP reduced the mean procedural time by approximately 15%, when compared to standard training (ROM 0.81, 95% CI: 0.66-0.98, p = 0.03) (Fig 12b). Funnel plot demonstrated presence of residual publications bias (Figure 14).

A.



B.

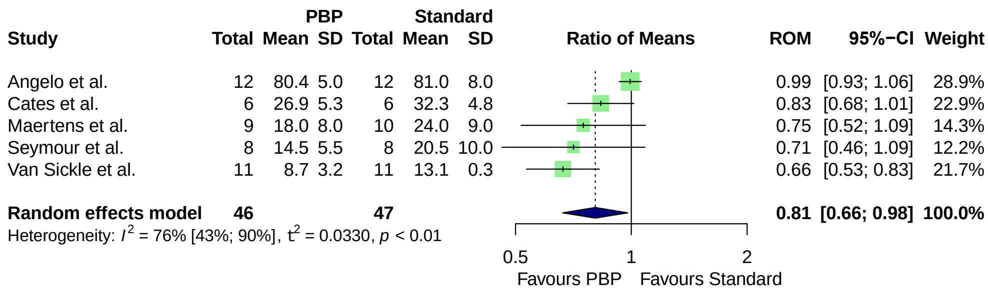
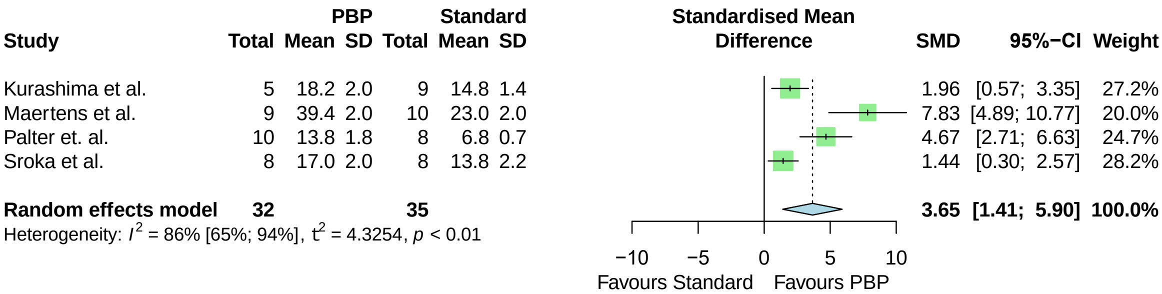


Figure 12. Standardized mean difference (8A) and ratio of means (8B) between studies assessing the effect of proficiency-based progression vs standard training on procedural time

Finally, in the quantitative synthesis testing for the average score at Likert scales evaluation, a pooled meta-analysis on 67 trainees was conducted (Figure 13a-b). Overall, at ROM analysis, trainees who completed PBP achieved a mean Likert scale-based score 52% higher than those who completed a standard training pathway (ROM 1.52, 95% CI: 1.22; 1.90; p = 0.01 using a random effect model), but such improvement was of lesser magnitude when compared to the reduction in objectively assessed performance errors. Funnel plots demonstrated presence of potential publications bias (Figure 14).

A.



B.

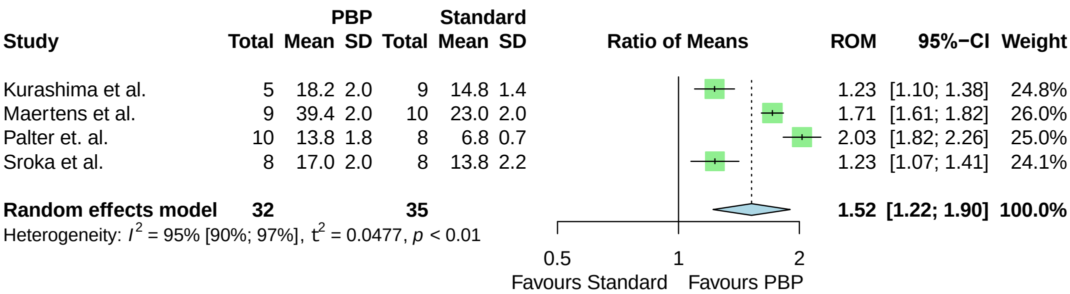


Figure 13. Standardized mean difference (9A) and ratio of means (9B) between studies assessing the effect of proficiency-based progression vs standard training on Likert Scale scoring

Figure 14. Funnel plots and Egger's asymmetry test evaluating publication biases according to different outcomes tested.

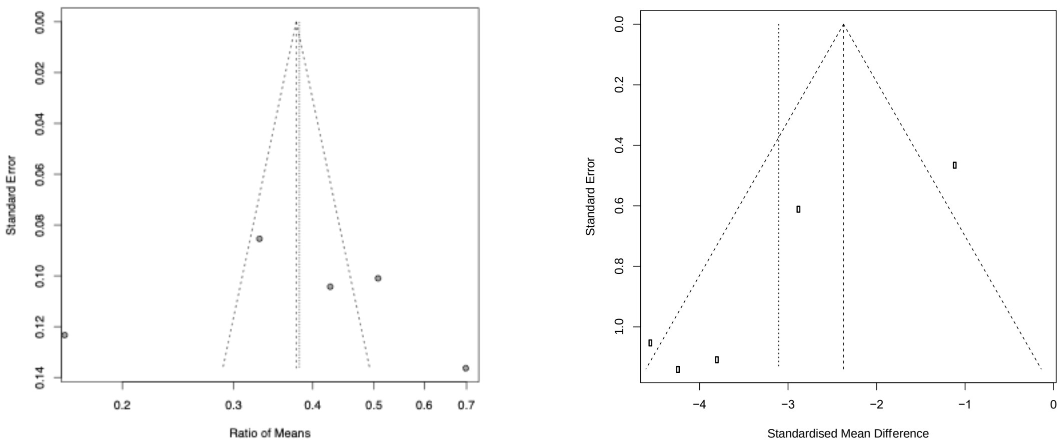


Figure 14 (A) Errors

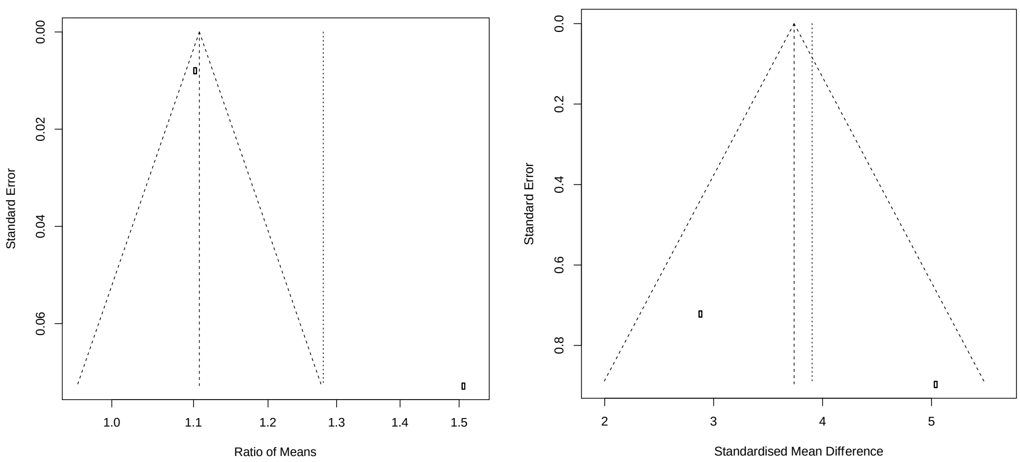


Figure 14 (B) Steps

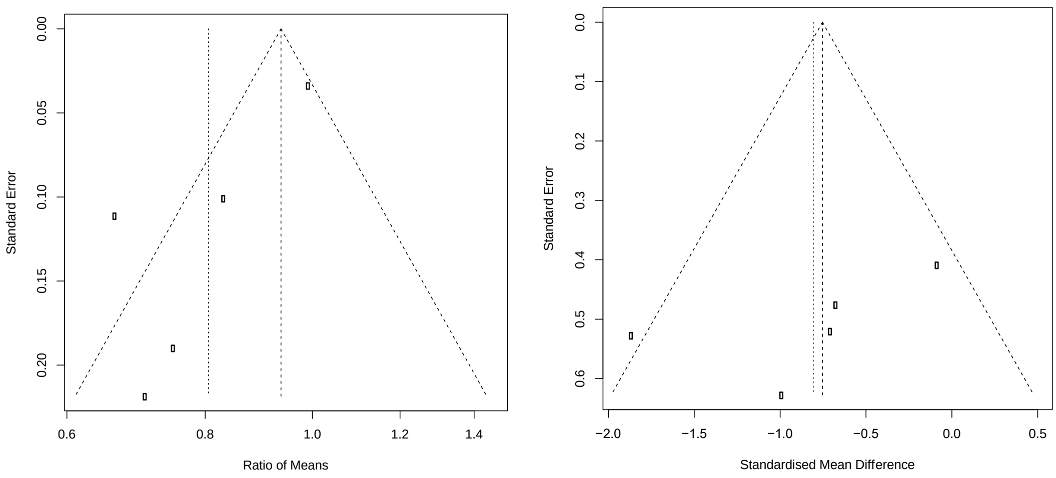


Figure 14 (C) Time

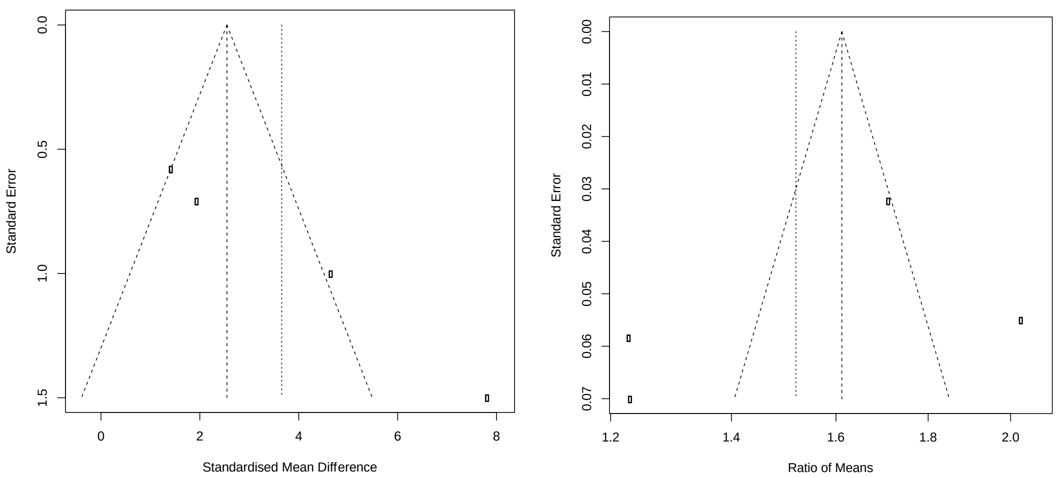


Figure 14 (D) Likert Scale Scoring

4. Discussion

Thirteen studies were included in the systematic review. In total 302 participants underwent surgical/operative training.

4.1 Quality and risk of bias assessment

As measured with the MERSQI instrument the quality of the studies was high. Although the differences between quality and risk of bias assessment are subtle, both quality assessment and risk of bias assessment were carried out for the following reasons: Quality assessment (MERSQI)- to assess the inclusion of methodological safeguards within individual studies [34]; Risk of Bias- to consider the implications of the inclusion/exclusion of such safeguards for study results [34]

4.2 Included study characteristics

Participants underwent assessment of fundamental abilities in all but one study [24]. Baseline assessments ranged from demographic questionnaires [25,27-29,32] to data on operative experience [23,25,28-32], baseline fundamental abilities [7,27,30,33] and cognitive knowledge [22,23,25,26,30].

The methods for defining proficiency benchmark levels varied between studies. However, for all studies this was based on mean 'expert' performance with the classification of 'expert' defined specifically for each study.

4.3 Outcomes

Proficiency-based progression (PBP) training consistently showed significant improvements in performance by trainees. Improvements in performance/procedure time, procedure steps and Likert Scale scores were observed.

Procedural errors

The largest and most substantial improvements, however, were found for error performance. In studies that evaluated procedural errors, we found a 62% reduction in comparison to the standard training group. This is clinically significant as objectively assessed error performance in PBP methodology gives direct, objective, transparent and fair measures of performance quality [16]. One study directly assessed the impact of PBP training on a clinical outcome. Srinivasan et al [23] assessed the impact of PBP simulation training on the effectiveness and success of epidural analgesia administration during labour. They found that the PBP trained group had a 54% lower epidural failure rate than the simulation trained group.

Number of steps, and time

The number of steps completed in a procedure is important in terms of procedure completion- however, the steps completed may be performed badly. Likewise, the procedural time taken is a less reliable measure of operator performance quality- a procedure can be performed more quickly if steps of the procedure are omitted, or can be done quickly but unsafely.

Likert-scale scoring

Likert-scale based scoring also saw an improvement following completion of PBP training with trainees who completed PBP training achieving scores 38% higher than that of standard training. This lesser improvement in Likert scale scoring between traditional training methods and metric-based proficiency training methods may be accounted for by the fact that Likert scale assessment tools exhibit less sensitivity and specificity than a metric-based training approach.

4.4 Proficiency-Based Progression Training

A number of factors may contribute to the effectiveness of Proficiency-Based-Progression training. Training is based upon strictly defined units of performance called metrics. Performance metrics are detailed descriptors of procedural steps that can also include errors or deviations from optimal performance [4,35]. Metrics are used to provide trainees with objective and transparent feedback during the training process. This allows trainees to engage with more deliberate practice rather than repeated practice [36]. Repetition of skills with deliberate practice is key to success and the defined metrics should be able to be replicated in various settings [6]. Another stipulation of PBP training is that that the performance characteristics on which training is based are derived from very experienced or expert clinicians. This provides a reference approach for optimum procedural performance [35,37]. Once these metrics have been subject to robust validation, they provide the basis for establishment of a quantitatively defined proficiency benchmark performance (based on mean experienced operator performance) [7,22,26,27,22,38.] Trainees subject to PBP training conditions are required to demonstrate a level of proficiency based on pre-defined simulation proficiency benchmarks before performing a procedure on a live patient. This meta-analysis shows that PBP-trained trainees can perform procedures with less errors and in less time than their traditionally trained peers, and that this has a positive impact on patient outcomes [23].

The surgical procedures included in this review were wide-ranging, from endoscopic sinus surgery to arthroscopic Bankart procedure, laparoscopic cholecystectomy to epidural analgesia. Proficiency-Based-Progression training has also been utilised in areas such as clinical communication by Breen et al [39] where it was shown to be a more effective way to teach clinical communication in the context of the deteriorating patient than e-learning either alone or in combination with standard simulation.

Given that Proficiency-Based progression training has been utilised across such a broad range of skill sets and has been shown in this review to be a superior training methodology to standard training programmes, particularly in relation to a reduction in the number of procedural errors by 62% when compared to standard training, it could be extrapolated from this that PBP training has the potential to be a valid, reliable and quality assured training methodology in dentistry.

5. Limitations

There are a number of limitations of this systematic review which should be recognised. These include the limited number of studies included which may reduce the generalizability of our findings and may increase the risk of residual biases. Also, as PBP training is a novel concept in dentistry it was not possible to directly demonstrate its effects on a dental procedure. However, the explicit and precise definition, validation and mathematisation of steps and errors (units of performance) in the PBP process makes the methodology applicable across various tasks and disciplines. Furthermore, a certain amount of heterogeneity (moderate-substantial) was expected. However, the pooling of study results for a joint interpretation seemed sensible in this instance. Therefore, meta-analysis with random effects was performed rather than fixed effect meta-analysis. The random effects model (DerSimonian and Laird method [20]) encompasses within-study as well as between-study variation[40]. Despite statistical adjustment using random-effect models, there is residual heterogeneity between studies due to differences in population, study protocols, and tasks/procedures which may have been remained unaccounted for. However, the studies included were high-quality RCTs which substantiates the robustness of our findings.

6. Conclusions

Proficiency-based simulation training using validated performance metrics (PBP) was found to improve the surgical/operative skills of trainees compared with standard/traditional training methods without a requirement to achieve proficiency standards. Proficiency Based Progression training decreased procedural errors by 62% compared with traditional training methods. There is still a need for further research in this area as there are insufficient good quality randomized clinical trials currently available (Table 3). Also, following quantitative analysis there is residual heterogeneity between studies despite statistical adjustment. However, there is sufficient evidence based on the results of this review to explore PBP training for use in dental skills training to determine whether similar clinical outcomes in terms of error reduction can be achieved with obvious benefits for patients. Further emphasis should be placed in future studies on the transfer of skills to the live patient setting and the clinical implications of improved operative performance for patients.

7. Directions for future research

This systematic review and meta-analysis is part of a wider study to develop a proficiency-based curriculum for clinical skills training in operative dentistry (including direct and indirect restorations) in Cork University Dental School and Hospital. In this programme we will develop an evidence-based, validated, quality assured training programme across dentistry. The aim is to produce dentists whose skill level is known at the time of graduation. Unlike surgery and other disciplines, dental practitioners are expected to be competent to operate independently on the date they graduate. Currently, dentistry uses simulation-based training without formalised metrics.

To date, a task analysis of the Class II composite preparation and restoration of an Upper Right 1st premolar has taken place. A list of metrics was produced; these were then operationally defined (face and content validation).

Subsequently, I co-ordinated and led a Delphi panel meeting of global dental experts (Ireland, UK, Singapore, Hong Kong, Australia) in July 2021. This meeting was a hybrid model with over 20 participants- some attended the meeting in person in CUDSH and some participants attended remotely via Zoom. This Delphi meeting led to the successful validation of the performance metrics previously established.

I am currently undertaking the 'construct' validation of these metrics which is the next stage of validation before the development of an online curriculum (via Canvas) using these validated metrics. Research of this kind is completely novel in the field of dentistry.

It has been shown that patients are exposed to increased risk of harm during the early part of a surgical trainee's learning curve[2]. We can extrapolate from this that the same would apply to a dental trainee. Leading on from validation of the performance metrics will be the establishment of a 'proficiency' benchmark both for the procedure itself and for the online curriculum which an undergraduate would be required to meet before deemed proficient in a dental procedure.

8. References

1. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med. Oct 2013;369(15):1434-42. doi:10.1056/NEJMsa1300625

2. Gallagher AG. Metric-based simulation training to proficiency in medical education:- what it is and how to do it. Ulster Med J. Sep 2012;81(3):107-13.

3. Satava RM. Virtual reality surgical simulator. The first steps. Surg Endosc. 1993 May-Jun 1993;7(3):203-5. doi:10.1007/BF00594110

4. Collins JW SA, Kelly J. Surgical Training Benefits from the Cumulative Outcomes of Proficiency-Based Training and Mentorship. JAMA Surgery2020. p. 616.

5. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg. Oct 2002;236(4):458-63; discussion 463-4. doi:10.1097/01.sla.0000028969.51489.b4

6. Dreyfus SE. The Five-Stage Model of Adult Skill Acquisition. Bulletin of Science Technology & Society. 2004; 24(177).

7. https://adee.org/projects/graduating-european-dentist/graduating-european-dentist-resource-pages

8. Zitzmann NU, Matthisson L, Ohla H, Joda T. Digital Undergraduate Education in Dentistry: A Systematic Review. Int J Environ Res Public Health. 05 07 2020;17(9)doi:10.3390/ijerph17093269

9. McGleenon EL, Morison S. Preparing dental students for independent practice: a scoping review of methods and trends in undergraduate clinical skills teaching in the UK and Ireland. Br Dent J. Jan 2021;230(1):39-45. doi:10.1038/s41415-020-2505-7

10. Field J, Dixon J, Towers A, et al. Defining dental operative skills curricula: An ADEE consensus paper. 2021. p. 405-414.

11. Mazzone E, Puliatti S, Amato M, et al. A Systematic Review and Meta-Analysis on the Impact of Proficiency-Based Progression Simulation Training on Performance Outcomes. Ann Surg. Dec 2020;doi:10.1097/SLA.0000000000004650

12. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. Jul 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097

13. Prictor M, Hill S. Cochrane Consumers and Communication Review Group: leading the field on health communication evidence. J Evid Based Med. 2013 Nov;6(4):216-20. doi: 10.1111/jebm.12066. PMID: 24325413.

14. Zendejas B, Brydges R, Hamstra SJ, Cook DA. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. Ann Surg. Apr 2013;257(4):586-93. doi:10.1097/SLA.0b013e318288c40b

15. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA. Sep 2011;306(9):978-88. doi:10.1001/jama.2011.1234

16. Mazzone E PS, Amato M, et al. A systematic review and Meta-Analysis on the Impact of Proficiency-Based Progression Simulation Training on Performance Outcomes. Ann Surg. 9000Publish Ahead of Print.

17. Friedrich JO AN, Beyene J. Ratio of means for analysing continuous outcomes in meta-analysis performed as well as means difference methods. J Clin Epidemiol.2011. p. 556-564.

18. Friedrich JO AN, Beyene J. The ratio of means method as an alternative to mean differences for analysing continuous outcome variables in meta-analysis: a simulation study. BMC Med Res Methodol.2008.

19. Higgins JPT TS, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ2003. p. 557-560.

20. DerSimonian R LN. Meta-analysis in clinical trials. 1986. p. 177-188.

21. Shim SR KS. Intervention meta-analysis: application and practice using R software. e2019008 ed. Epidemiol Health.2019.

22. Cates CU LL, Gallagher AG. Prospective, randomised and blinded comparison of proficiency-based progression full-physics virtual reality simulator training versus invasive vascular experience for learning carotid artery angiography by very experienced operators. BMJ Simulation and Technology Enhanced Learning. 2016;2:1-5.

23. Kallidaikurichi Srinivasan K, Gallagher A, O'Brien N, et al. Proficiency-based progression training: an 'end to end' model for decreasing error applied to achievement of effective epidural analgesia during labour: a randomised control study. BMJ Open. 10 2018;8(10):e020099. doi:10.1136/bmjopen-2017-020099

24. Gerull W, Zihni A, Awad M. Operative performance outcomes of a simulator-based robotic surgical skills curriculum. Surg Endosc. Oct 2020;34(10):4543-4548. doi:10.1007/s00464-019-07243-6

25. Maertens H, Aggarwal R, Moreels N, Vermassen F, Van Herzeele I. A Proficiency Based Stepwise Endovascular Curricular Training (PROSPECT) Program Enhances Operative Performance in Real Life: A Randomised Controlled Trial. Eur J Vasc Endovasc Surg. Sep 2017;54(3):387-396. doi:10.1016/j.ejvs.2017.06.011

26. Van Sickle KR, Ritter EM, Baghai M, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. J Am Coll Surg. Oct 2008;207(4):560-8. doi:10.1016/j.jamcollsurg.2008.05.007

27. Angelo RL, Ryu RK, Pedowitz RA, et al. A Proficiency-Based Progression Training Curriculum Coupled With a Model Simulator Results in the Acquisition of a Superior Arthroscopic Bankart Skill Set. Arthroscopy. Oct 2015;31(10):1854-71. doi:10.1016/j.arthro.2015.07.001

28. Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. Am J Surg. Jan 2010;199(1):115-20. doi:10.1016/j.amjsurg.2009.07.035

29. Kurashima Y, Feldman LS, Kaneva PA, et al. Simulation-based training improves the operative performance of totally extraperitoneal (TEP) laparoscopic inguinal hernia repair: a prospective randomized controlled trial. Surgical endoscopy. 2014;28(3):783-788.

30. Palter VN, Orzech N, Reznick RK, Grantcharov TP. Validation of a structured training and assessment curriculum for technical skill acquisition in minimally invasive surgery: a randomised controlled trial. Ann Surg. 2013;257(2):224-30.

31. Fried MP, Kaye RJ, Gibber MJ, et al. Criterion-based (proficiency) training to improve surgical performance. Arch Otolaryngol Head Neck Surg. Nov 1 2012;138(11):1024-9. doi:10.1001/2013.jamaoto.377

32. Lendvay TS, Brand TC, White L, et al. Virtual reality robotic surgery warm-up improves task performance in a dry laboratory environment: a prospective randomized controlled study. J Am Coll Surg. Jun 2013;216(6):1181-92. doi:10.1016/j.jamcollsurg.2013.02.012

33. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg. Jun 2007;193(6):797-804. doi:10.1016/j.amjsurg.2006.06.050

34. Furuya-Kanamori L, Xu C, Hasan SS, Doi SA. Quality versus Risk-of-Bias assessment in clinical research. J Clin Epidemiol. 01 2021;129:172-175. doi:10.1016/j.jclinepi.2020.09.044

35. Angelo RL, Ryu RK, Pedowitz RA, Gallagher AG. Metric Development for an Arthroscopic Bankart Procedure: Assessment of Face and Content Validity. Arthroscopy. Aug 2015;31(8):1430-40. doi:10.1016/j.arthro.2015.04.093

36. Ericsson KA, Krampe RT, C. T-R. The role of deliberate practice in the acquisition of expert performance. Psychol Rev1993. p. 363-406.

37. Kojima K, Graves M, Taha W, Cunningham M, Joeris A, Gallagher AG. AO international consensus panel for metrics on a closed reduction and fixation of a 31A2 pertrochanteric fracture. Injury. Dec 2018;49(12):2227-2233. doi:10.1016/j.injury.2018.09.019

38. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Ann Surg. Feb 2005;241(2):364-72.

39. Breen D, O'Brien S, McCarthy N, Gallagher A, Walshe N. Effect of a proficiency-based progression simulation programme on clinical communication for the deteriorating patient: a randomised controlled trial. BMJ Open. 07 09 2019;9(7):e025992. doi:10.1136/bmjopen-2018-025992

40. Bender R, Friede T, Koch A, et al. Methods for evidence synthesis in the case of very few studies. Res Synth Methods. Sep 2018;9(3):382-392. doi:10.1002/jrsm.1297

9. Appendices

Appendix I. Published paper in peer-reviewed Journal

Kehily E, Mazzone E, Coffey N, Allen F, Gallagher A, Roberts A. Proficiency Based Progression (PBP) training- the future model for dental operative skills training?: A systematic review and meta-analysis of existing literature. J Dent. 2022 Jan;116:103906.

CASRAI CRediT taxonomy

Kehily E.
Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, verification, visualization, writing- original draft, writing- review and editing
Mazzone E.
Formal analysis, methodology, writing- original draft
Coffey N.
Formal analysis, investigation, methodology, validation
Allen F.
Methodology, supervision, writing- original draft, writing- review and editing
Gallagher A.
Conceptualization, methodology, supervision
Roberts A.
Conceptualization, methodology, supervision, validation, writing- original draft, writing- review and editing

Journal of Dentistry 116 (2022) 103906

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent

Proficiency Based Progression (PBP) training- the future model for dental operative skills training?: A systematic review and meta-analysis of existing literature

Elaine Kehily [a,*], Elio Mazzone [b], Niamh Coffey [a], Finbarr Allen [c,d], Anthony Gallagher [e,f,g], Anthony Roberts [a]

[a] Cork University Dental School & Hospital, University College Cork, Ireland
[b] Division of Oncology, Unit of Urology URI, IRCCS Ospedale San Raffaele, Via Olgettina 60, Milan 20132, MI, Italy
[c] Faculty of Dentistry, National University of Singapore, Singapore
[d] National University Centre for Oral Health, Singapore, Singapore
[e] Faculty of Medicine, KU Leuven, Belgium
[f] Faculty of Health and Life Sciences, Ulster University, Northern Ireland
[g] ORSI Academy, Belgium

ARTICLE INFO

Keywords:
Preclinical skills
Simulation-based training
Dental education
Dental operative skills training

ABSTRACT

Objective: To evaluate the effectiveness of Proficiency-Based Progression (PBP) operative training using validated performance metrics, by comparing this to standard, conventional training methods.

Data: This systematic review was conducted in accordance with the guidelines of the Transparent Reporting of Systematic Reviews and Meta-Analyses (PRISMA). Study quality was assessed using the MERSQI tool and the Cochrane Risk of Bias tool. Results were pooled using biased corrected standardized mean difference and ratio-of-means (ROM). Summary effects were evaluated using a series of fixed and random effects models. The primary outcome was the number of procedural errors performed comparing PBP and non-PBP-based training pathways. In quantitative synthesis testing for procedural errors, a pooled meta-analysis on 87 trainees was conducted using random-effects models. In a ROM analysis, PBP was estimated to reduce the mean rate of errors by 62%, when compared to standard training (ROM 0.38, 95% CI: 0.25; 0.58; $p < 0.001$)

Sources: The electronic databases of PubMed, Embase, Web of Science, MEDLINE and Cochrane library's CENTRAL were searched from inception to 8/11/2021. Filters activated were Randomized Controlled trials, clinical trial.

Study selection: 13 studies were included for review with 11 included in the quantitative synthesis from 174 potentially relevant publications identified by the search strategy. Main inclusion criteria were studies comparing standard surgical/operative training with proficiency-based simulation training using validated metrics based on expert performance.

Conclusions: Our meta-analysis found that PBP training improved trainees' performances, by decreasing procedural errors. There is sufficient evidence to explore PBP training for use in dental skills training.

PBP training was estimated to reduce the mean rate of operative errors by 62%, when compared to standard training. Given that there is a direct correlation between operative skill and patient outcomes, these data suggest that there is sufficient evidence to explore PBP training for use in dental skills training.

1. Introduction

The acquisition of practical skills for clinicians providing treatment for their patients is clearly an important training consideration. Nowhere is this more evident than the training requirements for operative procedures by doctors, dentists and other health care workers with a direct correlation between operative skill and patient outcomes [1]. Simulation-based learning has a growing role to play in skill acquisition

* Corresponding author.
E-mail address: ekehily@ucc.ie (E. Kehily).

https://doi.org/10.1016/j.jdent.2021.103906
Received 27 September 2021; Received in revised form 18 November 2021; Accepted 22 November 2021
Available online 26 November 2021
0300-5712/

and re-validation [2] with Virtual Reality being first proposed by Satava almost 30-years ago for the acquisition of surgical skills [3]. More specifically, proficiency-based simulation has come to the forefront of surgical skill development and maintenance over recent years [4]. A study by Seymour et al [5] demonstrated that trainees who underwent a virtual reality (VR)-based simulation training pathway performed significantly better than traditionally trained surgeons. It was the first study to introduce the "Proficiency-Based Progression" (PBP) training methodology as an evidence-based alternative to more traditional training pathways. PBP is a robust methodology where the operative procedure in question is subject to a task analysis to identify performance metrics essential to the completion of the task (including steps, errors and critical errors). These performance metrics are then subject to a validation procedure (face, content, construct and predictive validity) and can be used to establish an objectively pre-defined proficiency benchmark performance to which trainees are trained. This proficiency benchmark performance is typically based on the mean performance of experienced practitioners.

The study by Seymour et al [5] was the first prospective, randomized, double-blind clinical study of simulation-based training for the operating room which demonstrated that surgical residents trained to a proficiency benchmark on a VR simulator made significantly fewer objectively assessed intra-operative errors when compared to the control group. These results have been replicated in other studies and there is increasing confidence that simulation-based training produces a superior skill set when compared to traditional training methods.

Unlike surgery and other medical disciplines, dental practitioners are expected to be competent to operate independently when they graduate. Traditionally in dentistry, undergraduates are trained to levels of competence with benchmarking based on a performance level reached by consensus amongst senior faculty staff members. Contrary to this, metric-based training to proficiency, or PBP, is benchmarked on the actual mean performance level of experienced or proficient practitioners. An important goal for dental educators is to establish evidence-based validated quality assured training programmes across dentistry. Our methods of assessment in clinical and pre-clinical training should exhibit high validity and reliability. Our assessment tools for dental undergraduate and postgraduate performance should exhibit construct validity (do more skilled individuals perform better on their assessments than less skilled or experienced individuals?) and predictive validity (do our assessments predict future skilled performance?) [2] and define a skill level that is transparent, objective and fair. Metric-based training to a pre-defined benchmark, such as proficiency, demonstrates all of these characteristics.

Conventionally skills in dentistry have been acquired through repeated practice in the clinical or simulation environment with summative feedback at the end of the procedure or when the trainee requests it to help facilitate learning. Using a feedback approach with metrics which are strictly defined and validated allows the trainee to know if they are performing a procedure incorrectly or in the wrong order. In PBP training, trainees should receive this feedback proximate to the performance error [2] with this approach being termed deliberate practice.

A recent systematic review of digital undergraduate education in dentistry [6] found that the use of Virtual Reality (VR) technologies for motor skills training is increasing. The review concluded that these technologies are valuable in dental undergraduate and postgraduate education but also highlighted the need for a better evidence base for the utility of VR in dental education. This coincides with reported shortages of suitable patients available for undergraduate students to treat [7] and a corresponding reduction in clinical experience. In a scoping review of methods and trends in undergraduate clinical skills teaching in Ireland and the UK, McGleenon and Morison [7] highlight that the use of clinical competencies without minimum experience requirements had the effect of "reduced confidence and perceived preparedness for practice". This evidence further underlines the importance of VR and simulation (including phantom head, computer-supported and haptic-enhanced VR simulators) in the future of dental skills training. However, without formalised performance metrics and effective feedback it's pedagogical effectiveness remains unclear. The systematic review by Zitzmann et al [6] found that operative performance on VR units, such as haptic simulators, with continuous feedback from clinical instructors led to better quality of tooth preparation than any real-time feedback from the VR simulator itself.

In a recent ADEE (Association for Dental Education in Europe) consensus paper [8] a special interest group focusing on the teaching of pre-operative clinical skills in dentistry made the following recommendations when planning operative skills courses:

- educators should employ and co-create resources with students that break down difficult tasks, and
- educators should consider whether to plan skills sessions in relation to deconstructed skills development rather than clinical presentation and complexity

Proficiency-Based-Progression training has the potential to fulfil both of these requirements when planning dental operative skills training.

As PBP training has yet to be utilised in dentistry, we can only look at how this training methodology is employed in surgical operative skills training and the effect that this has on both operator performance and patient outcomes. The simulators utilised (mannequins, box trainers, high fidelity simulators) in these surgical training programmes (Table 1) bear a close resemblance to the phantom head, computer-supported and haptic-enhanced VR simulators currently used in dental training globally and so the authors believe is comparable to dental undergraduate and postgraduate training.

A recent systematic review and meta-analysis carried out by Mazzone et al [9] showed that PBP training reduced the number of performance errors by 60% and procedural time by 15%. These data demonstrate significant advantages of PBP training, however their review did not focus specifically on the use of proficiency-based simulation training with validated performance metrics and did not include a Risk of Bias analysis or evaluate the impact of PBP training on Likert scale scoring.

Given the paucity of data on Proficiency-Based-Progression training in Dentistry, the aim of this systematic review was to evaluate the effectiveness of proficiency-based surgical/operative training (the closest comparator for dental clinical skills training) using validated performance metrics, by comparing this to standard conventional training methods. Uniquely, this study considers important aspects of the quality and risk of bias of the studies and, if the results are compelling, could justify the exploration of PBP training in dentistry.

2. Materials and methods

This systematic review was conducted in accordance with the guidelines of the Transparent Reporting of Systematic Reviews and Meta-Analyses (PRISMA) [10].

The primary aim of this systematic review was to answer the following PICO question: Does proficiency-based simulation training using validated performance metrics improve the surgical/operative skills of trainees compared with standard/conventional training methods without a requirement to achieve proficiency standards?

2.1. Search strategy and selection criteria

The electronic databases of PubMed, Embase, Web of Science, MEDLINE and the Cochrane library were searched (last search date 08/11/2021). The search terms entered were validat* AND ((performance) AND metrics) AND surgi* AND proficiency. The filters activated were Randomized Controlled trials, clinical trial. The databases were

Table 1
General characteristics of 13 randomized clinical trials studies included in the final qualitative analysis of the systematic review.

Study	Methods	Subjects N; Type	Comparison arm	Task/Procedure	Primary outcomes measured	Other scale used	Intraoperative patient performance	MERSQI score[h]
Ahlberg et al [31]	RCT	13; Residents (PGY 1–2)	Swedish National Surgical Residency Training Programme	Laparoscopic cholecystectomy	Errors, time	___	Yes	17
Angelo et al [25]	RCT	44; Residents (PGY4–5)	ACGME approved orthopedic Residency & Arthroscopy Association of North America Shoulder Course	Arthroscopic Bankart Procedure	Errors, steps, time	___	Yes	15.5
Cates et al [20]	RCT	12; interventional cardiologists	Industry sponsored CASES education and training system	Carotid artery angiography	Errors, time	___	Yes	15
Fried et al [29]	Prosp. cohort	14; residents (PGY1–3)	ACGME approved Otolaryngology Surgery Residency Programme	Endoscopic sinus surgery (ES3)	Errors, time	___	Yes	12.5
Gerull et al [22]	Prosp. cohort	31; residents	ACGME approved General Surgery, Urology & Obstetrics and gynecology Residency Programmes	RALS[a] case	___	RO-SCORE[b]; NTLX[c]	Yes	11.5
Kurashima et al [27]	RCT	16; residents (PGY 2–5)	General Surgery Residency Training Programme McGill University Canada	Inguinal hernia repair	___	GOALS-GH[d]	Yes	12.5
Lendvay et al [30]	RCT	51; residents (PGY 1–6)	ACGME accredited University of Washington Medical center and Madigan Army Medical center Residency Programmes	Laparoscopic procedure	Time, errors	___	Yes	13
Maertens et al [23]	RCT	32; residents (PGY 1–6)	Master of Medicine in Specialist Medicine (Surgery) training Programme, Ghent University	Endovascular procedure	Time, steps	OSATs[e]	Yes	16
Palter et al [28]	RCT	25; residents (PGY1–2)	ACGME approved General Surgery Residency Training Programme	Laparoscopic Right Colectomy	Time, errors, steps	OSATs	Yes	15
Seymour et al [5]	RCT	16; residents (PGY 1–4)	ACGME approved General Surgery Residency Training Program	Laparoscopic cholecystectomy	Time, errors, steps	___	Yes	14
Srinivasan et al [21]	RCT	17; residents	Irish National anesthesia Training Program	Epidural analgesia	Errors	GRS[f], TSCL[g]	Yes	15
Sroka et al [26]	RCT	17; residents (PGY 1–3)	General Surgery Residency Training Programme McGill University Canada	Laparoscopic cholecystectomy (FLS training programme)	___	GOALS	Yes	15
Van Sickle et al [24]	RCT	22; residents (PGY 3,5,6)	ACGME approved General Surgery Residency Training Program	Nissen fundoplication	Errors, time	___	Yes	15

a.RALS (Robotic Assisted Laparoscopic surgery); b. RO-SCORE (Robotic modification of Objective Structured Assessment of Technical skills); c. NTLX (NASA Task Load Index); d. GOALS-GH (Global Operative Assessment of Laparoscopic Skills-Groin Hernia); e. OSATs (The modified Ottawa Surgical Competency Operating Room Evaluation for surgical robotic cases); f. GRS (Global Rating Scales); g. TSCL, Task-Specific Checklists; h. MERSQI (Medical Education Research Study Quality Instrument);

searched with no publication date limits and no language restriction. Two reviewers conducted the search independently (EK and NC), and any conflict was resolved with discussion. The references quoted in the full text articles were hand searched for any further eligible studies. A PRISMA flow diagram describing the search process is shown in Fig. 1. At the end of the process, 13 and 11 studies have been included for, respectively, the qualitative synthesis and the quantitative meta-analysis.

2.2. Eligibility criteria for included studies

The main inclusion criteria were studies comparing standard surgical/operative training with proficiency-based simulation training using validated metrics based on expert performance were included. All Randomised Controlled Trials (RCTs), quasi-randomised controlled trials and non-Randomised Controlled Trials (CCTs) were considered in the search. Only studies using a validated surgical simulator were included.

Studies involving the training of novice or inexperienced operators in a specific procedure (e.g. surgical residents, specialist trainees) were included, where the proficiency benchmark for training was based on expert (experienced surgeons) performance.

The primary outcomes reported were metrics such as time, errors, the number of procedural steps achieved, and whether the proficiency benchmark was reached. Secondary outcome measures included baseline abilities (psychomotor skills), efficiency/economy of motion, the number of training repetitions required to achieve proficiency, economy of movement, tissue handling, instrument control, tool manipulation, tool path length, skill retention (at 6 months, 12 months, 18 months and 24 months), and the number of 'consultant' (senior clinician) takeovers. The characteristics of each study included in this study are presented in Table 1.

Operative metric validation studies were excluded from this review. Studies that used standard/manufacturer simulator metrics to measure outcomes were also excluded (metrics needed to be proficiency-based i.e. based on expert performance, and validated). Ongoing clinical trials registered on the Cochrane Central Register of Controlled Trials (CENTRAL) database were also excluded.

2.3. Data extraction and quality assessment

Study methods, participants, interventions and outcomes were extracted and recorded independently by two reviewer authors (EK and NC). These were compared, and a third reviewer (AR) was available for resolution of any differences.

The quality of each trial was assessed using the Medical Education Research Study Quality Instrument (MERSQI) tool and the Cochrane

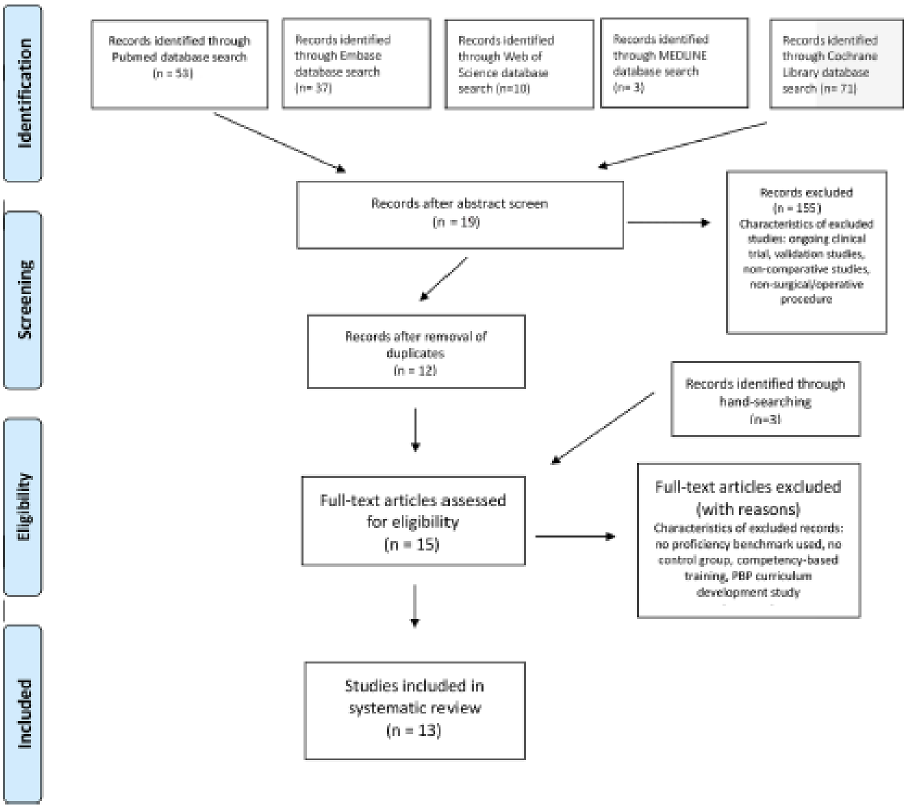


Fig. 1. Flow-chart of studies through the screening process according to the PRISMA methodology.

Collaboration's 'Risk of Bias' tool (http://handbook.cochrane.org). Items evaluated to generate this score included randomization, sequence generation, blinding, allocation concealment and incomplete outcome data. Fig. 2 shows the risk of bias table.

2.4. *Data synthesis and statistical analysis*

Data not suitable for meta-analytic evaluation was presented in narrative fashion (qualitative analysis). Meta-analysis was deemed appropriate as 11 included studies met the following criteria (as recommended by Cochrane Consumers and Communication Group [11]):

-outcomes are comparable (i.e. errors, time, steps, Likert scales) and could be pooled meaningfully
-Interventions and comparators are similar enough to be combined meaningfully
-correct data are available for the included studies.

In quantitative synthesis, the reported results for continuous outcomes were pooled using biased corrected standardized mean difference (SMD) (Hedges' g effect size) according to previous established methodology. [12,13] Moreover, as previously described, ratio of means (ROM) was applied to provide an estimation of the pooled effect of PBP on the considered outcomes. [14–16] All results were reported with 95% confidence intervals.

Heterogeneity between studies was measured using the I^2 statistic [17] and the between-study variance (τ^2) from the random-effect analyses. I^2 values >50% indicate large inconsistency. Unless otherwise indicated all models have allowed for different effect sizes (random effects). In case of large heterogeneity, random effect models (using the DerSimonian and Laird approach [18]) were prioritized.

For the assessment of small study effects and publication bias, values of the SMD or ROM were plotted against their standard error in a contour-enhanced funnel plot. The latter bias represents the error in connection with whether a study is published or not depending on the characteristics and result of individual studies. [19] This error is caused because statistically significant study results generally have a higher likelihood of being published. Statistical significance for all analysis was defined as two-sided $p < 0.05$. Statistical analysis was performed with the R software (version 3.6.3; http://www.r-project.org/).

3. Results

The search strategy identified 174 potentially relevant publications; 53 from PubMed, 37 from Embase, 10 from Web of Science, 3 from Medline and 71 from the Cochrane Library. Following screening of these abstracts and application of the inclusion and exclusion criteria, 155 records were excluded and a further 3 studies following removal of duplicates. Thus, the full texts of 15 articles were retrieved for screening. And 10 were subsequently included in the systematic review. A further three papers were identified by hand-searching of bibliographies, therefore a total of 13 studies were included in this review. This process is described in the flow diagram in Fig. 1.

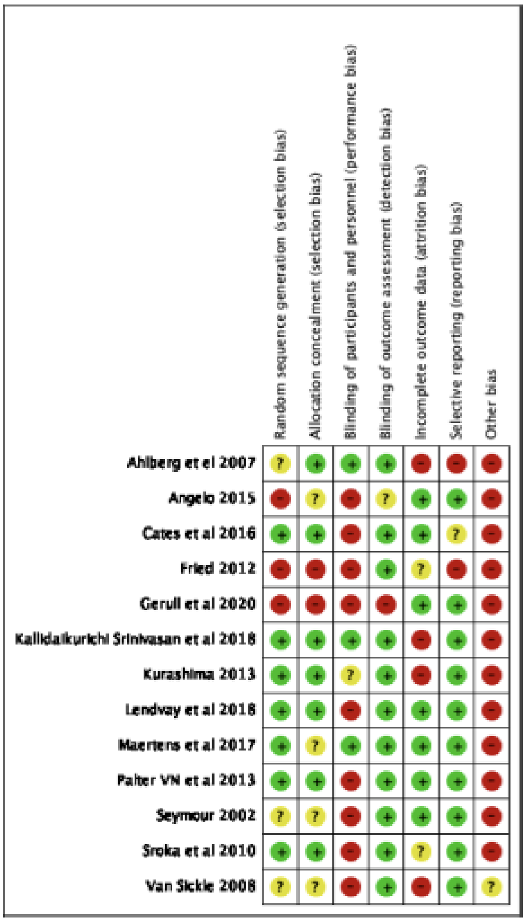


Fig. 2. Risk of bias summary (green indicates a low risk of bias; yellow indicates an unclear risk of bias; red indicates a high risk of bias) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

3.1. Quality and risk of bias assessment

eFig. 3 summarises the quality criteria assessed for each study using the MERSQI tool. This tool includes 10 items, reflecting six domains of study quality: (a) study design, (b) sampling, (c) type of data, (d) validity, (e) data analysis, and (f) outcomes. The maximum domain score is 3, producing a maximum possible MERSQI score of 18 and potential range of 5–18. The overall methodological quality of the studies was high with a mean MERSQI score of 14.38 (range 11.5–17).

The quality of each trial was assessed using the Cochrane Collaboration's 'Risk of Bias' tool (http://handbook.cochrane.org) using Review Manager 5.3 software. eFig. 2 shows the Risk of Bias Graph. Most studies had a low risk of bias for random-sequence generation (53.8%), allocation concealment (53.8%) and incomplete outcome data (53.8%). Of the 13 studies, 12 (92.3%) had a high risk of bias for other types of bias. Performance bias was more difficult to assess in this review due to the difficulty associated in blinding participants in surgical clinical trials. The authors of this review decided where a genuine attempt was made to liken the training conditions in the control group to the intervention group a judgment of low risk would be assigned. Where there was no attempt to blind participants from their training status then a judgment of high risk would be acceptable. Most studies had a low risk of bias in the domain of selective outcome reporting (10 [77%]).

3.2. Evidence synthesis

Table 1 summarizes the general and design characteristics for each study. eTable 1 outlines the descriptive characteristics for each included study.

Participants in all studies were surgical residents (specialist trainees) varying from postgraduate years (PGYs) 1–6 except for the study by Cates et al. [20] which assessed interventional cardiologists with no experience of the simulated procedure (carotid artery angiography), and the study by Srinivasan et al. [21] where anaesthesia trainees were trained. Proficiency benchmark setting is described by all studies with proficiency benchmark levels based on mean expert operator performance (with the classification 'expert' defined by each study).

3.3. Outcome measures

A number of methods were used for data collection amongst included studies. Eight out of 13 studies (21–28) used a validated assessment tool to measure the steps of procedure (metrics) completed.

Errors were specifically measured in nine studies [5,20,21,24,25,28–31], with Angelo et al also noting sentinel errors performed (an event or occurrence involving a serious deviation from optimal performance during a procedure that either (1) jeopardized the success or desired result of the procedure or (2) created significant iatrogenic insult to the patient's tissues). Assessment tools utilized in four studies [22,23,26,27] (RO-SCORE, OSATs, GOALS-GH) did not include a specific error score/domain.

3.4. Results of quantitative synthesis

In quantitative synthesis testing for procedural errors, a pooled meta-analysis on 87 trainees was conducted (Fig. 3a,b), using random-effects models. Overall, PBP training reduced the number of errors when compared to standard training (SMD -3.11, 95% CI: -4.54; -1.68; $p < 0.001$). In a ROM analysis, PBP was estimated to reduce the mean rate of errors by 62%, when compared to standard training (ROM 0.38, 95% CI: 0.25; 0.58; $p < 0.001$). Funnel plots showed evidence for potential publications bias (eFig. 4A).

In quantitative synthesis testing for number of steps completed, a pooled meta-analysis on 42 trainees was conducted (Fig. 4a,b). Overall, trainees who completed PBP training performed more procedural steps than those who completed a standard training pathway (SMD 3.90, 95% CI: 1.79; 6.02; $p < 0.001$) (Fig. 4a). However, at ROM analysis, PBP did not statistically significantly increase the mean rate of steps performed when compared to standard training (ROM 1.28, 95% CI: 0.94; 1.74; $p = 0.1$) (Fig. 4b). Funnel plots recorded a marginal effect for potential publications bias (eFig. 4B).

In quantitative synthesis testing for procedural time, a pooled meta-analysis on 98 trainees was conducted (Fig. 5a,b). Overall, trainees who completed PBP training performed the task/procedure in less time than those who completed a standard training pathway (SMD -0.81, 95% CI: -1.40; -0.21; $p = 0.008$) (Fig. 5a). The reduction of procedural time was less pronounced compared to other outcomes, such as the number of errors or steps completed. Indeed, at ROM analysis, PBP reduced the mean procedural time by approximately 15%, when compared to standard training (ROM 0.81, 95% CI: 0.66–0.98, $p = 0.03$) (Fig 5b). Funnel plot demonstrated presence of residual publications bias (eFig. 4C).

Finally, in the quantitative synthesis testing for the average score at Likert scales evaluation, a pooled meta-analysis on 67 trainees was conducted (Fig. 6a,b). Overall, at ROM analysis, trainees who completed

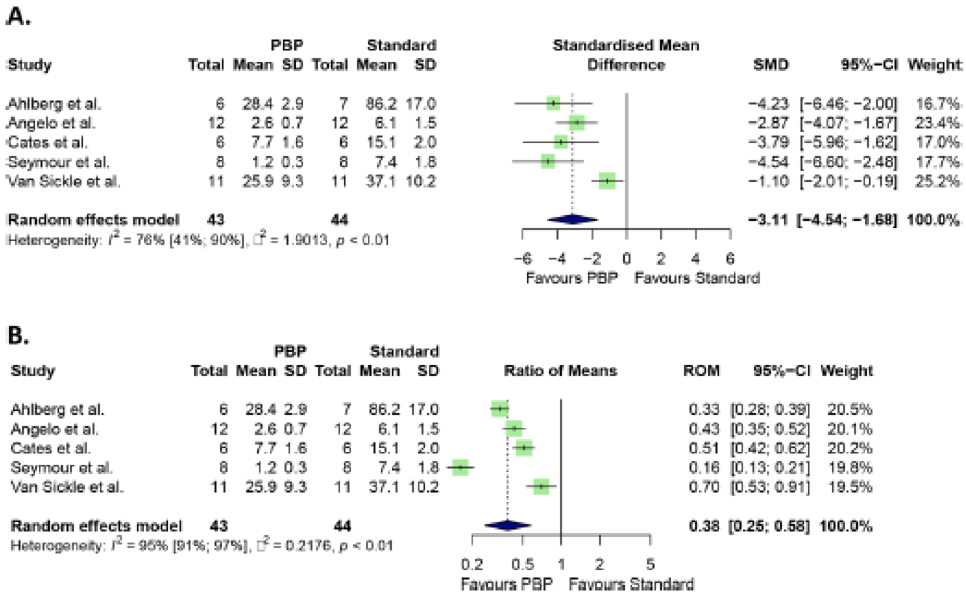


Fig. 3. Standardized mean difference (3A) and ratio of means (3B) between studies assessing the effect of proficiency-based progression vs standard training on procedural errors.

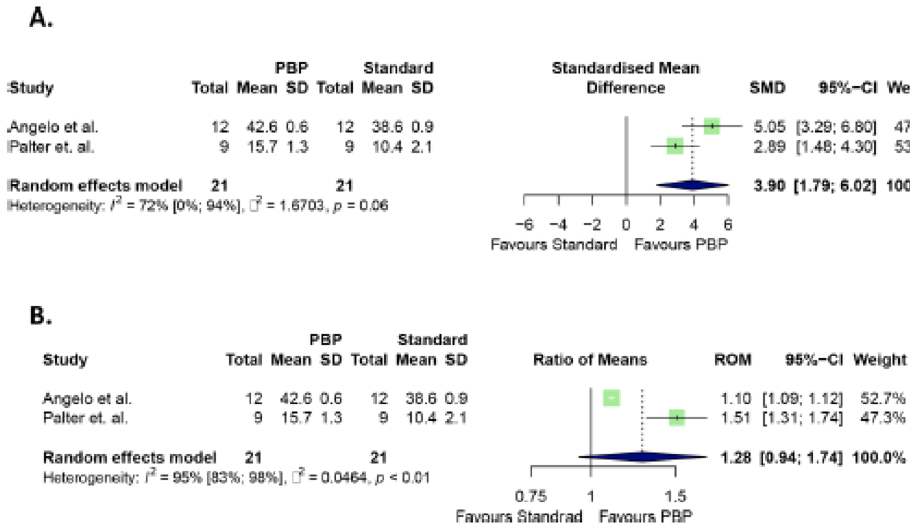


Fig. 4. Standardized mean difference (4A) and ratio of means (4B) between studies assessing the effect of proficiency-based progression vs standard training on the number of procedural steps.

PBP achieved a mean Likert scale-based score 52% higher than those who completed a standard training pathway (ROM 1.52, 95% CI: 1.22; 1.90; $p = 0.01$ using a random effect model), but such improvement was of lesser magnitude when compared to the reduction in objectively assessed performance errors. Funnel plots demonstrated presence of potential publications bias (eFig. 4D).

4. Discussion

Thirteen studies were included in the systematic review. In total 302 participants underwent surgical/operative training. As measured with the

MERSQI instrument the quality of the studies was high. Although the differences between quality and risk of bias assessment are subtle, both quality assessment and risk of bias assessment were carried out for the following reasons:

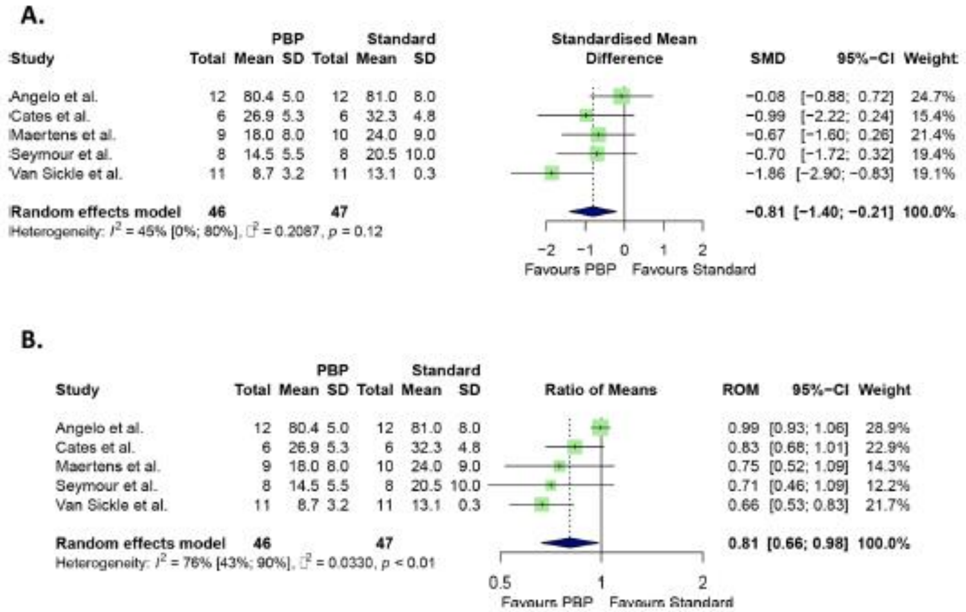


Fig. 5. Standardized mean difference (5A) and ratio of means (5B) between studies assessing the effect of proficiency-based progression vs standard training on procedural time.

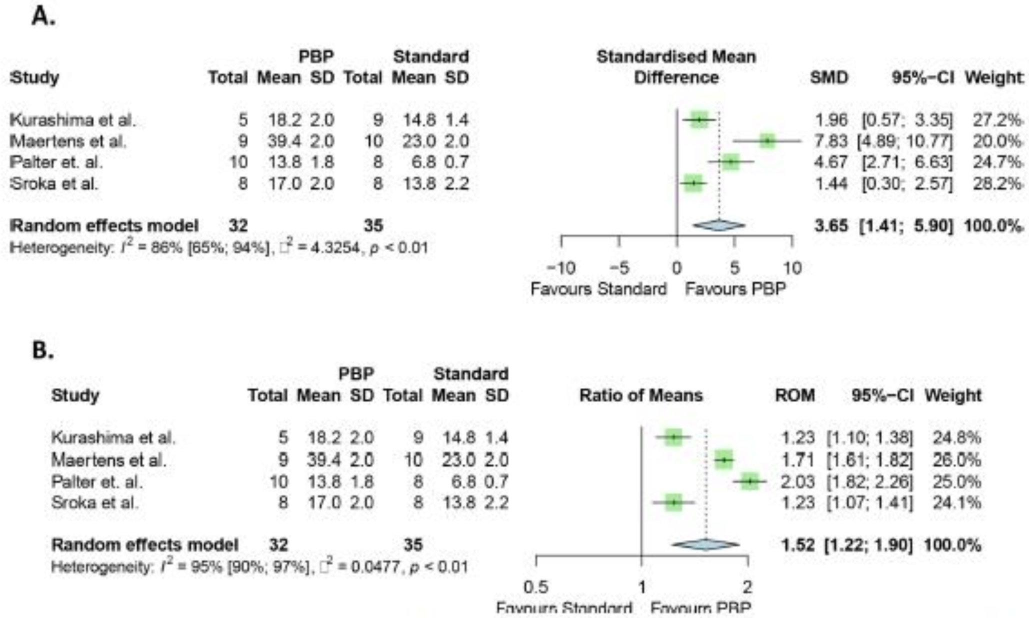


Fig. 6. Standardized mean difference (6A) and ratio of means (6B) between studies assessing the effect of proficiency-based progression vs standard training on Likert Scale scoring.

Quality assessment (MERSQI)- to assess the inclusion of methodological safeguards within individual studies [32]; Risk of Bias- to consider the implications of the inclusion/exclusion of such safeguards for study results [32]

Participants underwent assessment of fundamental abilities in all but one study [22]. Baseline assessments ranged from demographic questionnaires [23,25–27,30] to data on operative experience [21,23, 26–30], baseline fundamental abilities [5,25,28,31] and cognitive knowledge [20,21,23,24,28].

The methods for defining proficiency benchmark levels varied between studies. However, for all studies this was based on mean 'expert' performance with the classification of 'expert' defined specifically for each study.

Eight out of thirteen studies included in this review [5,20,21,23–25, 29,31] calculated Inter-Rater Reliability (IRR) to assess the degree of agreement among raters (of operator performance). In six out of these eight studies an IRR of 0.8 or 80% was deemed acceptable and achieved.

Proficiency-based progression training consistently showed significant improvements in performance by trainees. Improvements in performance/procedure time, procedure steps and Likert Scale scores were observed. The largest and most substantial improvements, however, were found for error performance. In studies that evaluated procedural errors, we found a 62% reduction in comparison to the standard training group. This is clinically significant as objectively assessed error performance in PBP methodology gives direct, objective, transparent and fair measures of performance quality [14]. One study directly assessed the impact of PBP training on a clinical outcome. Srinivasan et al[21] assessed the impact of PBP simulation training on the effectiveness and success of epidural analgesia administration during labour. They found that the PBP trained group had a 54% lower epidural failure rate than the simulation trained group.

The number of steps completed in a procedure is important in terms of procedure completion- however, the steps completed may be performed badly. Likewise, the procedural time taken is a less reliable measure of operator performance quality- a procedure can be performed more quickly if steps of the procedure are omitted, or can be done quickly but unsafely.

Likert-scale based scoring also saw an improvement following completion of PBP training with trainees who completed PBP training achieving scores 38% higher than that of standard training. This lesser improvement in Likert scale scoring between traditional training methods and metric-based proficiency training methods may be accounted for by the fact that Likert scale assessment tools exhibit less sensitivity and specificity than a metric-based training approach.

A number of factors may contribute to the effectiveness of Proficiency-Based-Progression training. Training is based upon strictly defined units of performance called metrics. Performance metrics are detailed descriptors of procedural steps that can also include errors or deviations from optimal performance [2,33]. Metrics are used to provide trainees with objective and transparent feedback during the training process. This allows trainees to engage with more deliberate practice rather than repeated practice [34]. Repetition of skills with deliberate practice is key to success and the defined metrics should be able to be replicated in various settings [4]. Another stipulation of PBP training is that the performance characteristics on which training is based are derived from very experienced or expert clinicians. This provides a reference approach for optimum procedural performance [33,35]. Once these metrics have been subject to robust validation, they provide the basis for establishment of a quantitatively defined proficiency benchmark performance (based on mean experienced operator performance) [5,20,24,25,20,36]. Trainees subject to PBP training conditions are required to demonstrate a level of proficiency based on pre-defined simulation proficiency benchmarks before performing a procedure on a live patient. This meta-analysis shows that PBP-trained trainees can perform procedures with less errors and in less time than their traditionally trained peers, and that this has a positive impact on patient outcomes [21].

The surgical procedures included in this review were wide-ranging, from endoscopic sinus surgery to arthroscopic Bankart procedure, laparoscopic cholecystectomy to epidural analgesia. Proficiency-Based-Progression training has also been utilised in areas such as clinical communication by Breen et al [37] where it was shown to be a more effective way to teach clinical communication in the context of the deteriorating patient than e-learning either alone or in combination with standard simulation.

Given that Proficiency-Based progression training has been utilised across such a broad range of skill sets and has been shown in this review to be a superior training methodology to standard training programmes, particularly in relation to a reduction in the number of procedural errors by 62% when compared to standard training, it could be extrapolated from this that PBP training has the potential to be a valid, reliable and quality assured training methodology in dentistry.

There are a number of limitations of this systematic review which should be recognised. These include the limited number of studies included which may reduce the generalizability of our findings and may increase the risk of residual biases. Also, as PBP training is a novel concept in dentistry it was not possible to directly demonstrate its effects on a dental procedure. However, the explicit and precise definition, validation and mathematisation of steps and errors (units of performance) in the PBP process makes the methodology applicable across various tasks and disciplines. Furthermore, a certain amount of heterogeneity (moderate-substantial) was expected. However, the pooling of study results for a joint interpretation seemed sensible in this instance. Therefore, meta-analysis with random effects was performed rather than fixed effect meta-analysis. The random effects model (DerSimonian and Laird method [18]) encompasses within-study as well as between-study variation [38]. Despite statistical adjustment using random-effect models, there is residual heterogeneity between studies due to differences in population, study protocols, and tasks/procedures which may have been remained unaccounted for. However, the studies included were high-quality RCTs which substantiates the robustness of our findings.

5. Conclusions

Proficiency-based simulation training using validated performance metrics (PBP) was found to improve the surgical/operative skills of trainees compared with standard/traditional training methods without a requirement to achieve proficiency standards. Proficiency Based Progression training decreased procedural errors by 62% compared with traditional training methods. There is still a need for further research in this area as there are insufficient good quality randomized clinical trials currently available (eTable 2). Also, following quantitative analysis there is residual heterogeneity between studies despite statistical adjustment. However, there is sufficient evidence based on the results of this review to explore PBP training for use in dental skills training to determine whether similar clinical outcomes in terms of error reduction can be achieved with obvious benefits for patients. Further emphasis should be placed in future studies on the transfer of skills to the live patient setting and the clinical implications of improved operative performance for patient outcomes.

Funding

This work was supported by the Health Research Board, Ireland (Grant number HPF2016–1675).

Declaration of Competing Interest

None to declare.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jdent.2021.103906.

References

[1] J.D. Birkmeyer, J.F. Finks, A. O'Reilly, et al., Surgical skill and complication rates after bariatric surgery, N. Engl. J. Med. 369 (15) (2013) 1434–1442, https://doi.org/10.1056/NEJMsa1300625.

[2] A.G. Gallagher, Metric-based simulation training to proficiency in medical education:- what it is and how to do it, Ulst. Med. J. 81 (3) (2012) 107–113.

[3] R.M. Satava, Virtual reality surgical simulator. The first steps, Surg. Endosc. 7 (3) (1993) 203–205, https://doi.org/10.1007/BF00594110.

[4] J.W Collins, A. Sridhar, J. Kelly, Surgical training benefits from the cumulative outcomes of proficiency-based training and mentorship, JAMA Surg. 155 (7) (2020) 616, https://doi.org/10.1001/jamasurg.2020.1041.

[5] N.E. Seymour, A.G. Gallagher, S.A. Roman, et al., Virtual reality training improves operating room performance: results of a randomized, double-blinded study, Ann. Surg. 236 (4) (2002) 458–463, https://doi.org/10.1097/01.sla.0000028969.51489.b4.

[6] N.U. Zitzmann, L. Matthisson, H. Ohla, T. Joda Digital undergraduate education in dentistry: a systematic review. *Int. J. Environ. Res. Public Health*. 05 07 2020;17(9) 10.3390/ijerph17093269.

[7] E.L. McGleenon, S. Morison, Preparing dental students for independent practice: a scoping review of methods and trends in undergraduate clinical skills teaching in the UK and Ireland, Br. Dent. J. 230 (1) (2021) 39–45, https://doi.org/10.1038/s41415-020-2505-7.

[8] J. Field, J. Dixon, A. Towers, R. Green, H. Albagami, G. Lambourn, J. Mallinson, W. Fokkinga, J. Tricio-Pesce, T. Crnić, S. Vital, Defining dental operative skills curricula: An ADEE consensus paper, Eur. J. Dent. Educ. 25 (2) (2021 May) 405–414, https://doi.org/10.1111/eje.12595.

[9] E. Mazzone, S. Puliatti, M. Amato, et al., A systematic review and meta-analysis on the impact of proficiency-based progression simulation training on performance outcomes, Ann. Surg. (2020), https://doi.org/10.1097/SLA.0000000000004650.

[10] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, P. Group, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009), e1000097, https://doi.org/10.1371/journal.pmed.1000097.

[11] Cochrane Consumers and Communication La Trobe University; Ryan, Rebecca; Hill, Sophie (2018): Meta-analysis. La Trobe. Journal contribution. 10.26181/5b57d2b96b326.

[12] B. Zendejas, R. Brydges, S.J. Hamstra, D.A. Cook, State of the evidence on simulation-based training for laparoscopic surgery: a systematic review, Ann. Surg. 257 (4) (2013) 586–593, https://doi.org/10.1097/SLA.0b013e318288c40b.

[13] D.A. Cook, R. Hatala, R. Brydges, et al., Technology-enhanced simulation for health professions education: a systematic review and meta-analysis, JAMA 306 (9) (2011) 978–988, https://doi.org/10.1001/jama.2011.1234.

[14] E. Mazzone, S. Puliatti, M. Amato, B. Bunting, B. Rocco, F. Montorsi, A. Mottrie, A. G.A Gallagher, Systematic review and meta-analysis on the impact of proficiency-based progression simulation training on performance outcomes, Ann. Surg. 1274 (2) (2021 Aug) 281–289.

[15] J.O. Friedrich, N.K. Adhikari, J Beyene, Ratio of means for analysing continuous outcomes in meta-analysis performed as well as means difference methods, J. Clin. Epidemiol. 64 (5) (2011) 556–564, https://doi.org/10.1016/j.jclinepi.2010.09.016.

[16] J.O. Friedrich, N.K. Adhikari, J Beyene, The ratio of means method as an alternative to mean differences for analysing continuous outcome variables in meta-analysis: a simulation study, BMC Med. Res. Methodol. 8 (32) (2008), https://doi.org/10.1186/1471-2288-8-32.

[17] J.P. Higgins, S.G. Thompson, J.J. Deeks, D.G. Altman, Measuring inconsistency in meta-analyses, BMJ 327 (7414) (2003) 557–560, https://doi.org/10.1136/bmj.327.7414.557.

[18] R. DerSimonian, N. Laird, Meta-analysis in clinical trials, Control Clin. Trials 7 (3) (1986 Sep) 177–188.

[19] S.R. Shim, S.J. Kim, Intervention meta-analysis: application and practice using R software, Epidemiol Health 41 (2019), e2019008.

[20] C.U. Cates, L. Lonn, A.G. Gallagher, Prospective, randomised and blinded comparison of proficiency-based progression full-physics virtual reality simulator training versus invasive vascular experience for learning carotid artery angiography by very experienced operators, BMJ Simul. Technol. Enhanc. Learn. 2 (2016) 1–5.

[21] K. Kallidaikurichi Srinivasan, A. Gallagher, N. O'Brien, et al., Proficiency-based progression training: an 'end to end' model for decreasing error applied to achievement of effective epidural analgesia during labour: a randomised control study, BMJ Open (10) (2018), e020099, https://doi.org/10.1136/bmjopen-2017-020099, 108.

[22] W. Gerull, A. Zihni, M. Awad, Operative performance outcomes of a simulator-based robotic surgical skills curriculum, Surg. Endosc. 34 (10) (2020) 4543–4548, https://doi.org/10.1007/s00464-019-07243-6.

[23] H. Maertens, R. Aggarwal, N. Moreels, F. Vermassen, I. Van Herzeele, A proficiency based stepwise endovascular curricular training (PROSPECT) program enhances operative performance in real life: a randomised controlled trial, Eur. J. Vasc. Endovasc. Surg. 54 (3) (2017) 387–396, https://doi.org/10.1016/j.ejvs.2017.06.011.

[24] K.R. Van Sickle, E.M. Ritter, M. Baghai, et al., Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying, J. Am. Coll. Surg. 207 (4) (2008) 560–568, https://doi.org/10.1016/j.jamcollsurg.2008.05.007.

[25] R.L. Angelo, R.K. Ryu, R.A. Pedowitz, et al. A proficiency-based progression training curriculum coupled with a model simulator results in the acquisition of a superior arthroscopic Bankart skill set. *Arthroscopy*. Oct 2015;31(10):1854–71. 10.1016/j.arthro.2015.07.001.

[26] G. Sroka, L.S. Feldman, M.C. Vassiliou, P.A. Kaneva, R. Fayez, G.M. Fried, Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial, Am J Surg 199 (1) (2010) 115–120, https://doi.org/10.1016/j.amjsurg.2009.07.035.

[27] Y. Kurashima, L.S. Feldman, P.A. Kaneva, et al., Simulation-based training improves the operative performance of totally extraperitoneal (TEP) laparoscopic inguinal hernia repair: a prospective randomized controlled trial, Surgical Endosc. 28 (3) (2014) 783–788.

[28] V.N. Palter, N. Orzech, R.K. Reznick, T.P. Grantcharov, Validation of a structured training and assessment curriculum for technical skill acquisition in minimally invasive surgery: a randomised controlled trial, Ann. Surg. 257 (2) (2013) 224–230.

[29] M.P. Fried, R.J. Kaye, M.J. Gibber, et al., Criterion-based (proficiency) training to improve surgical performance, Arch. Otolaryngol. Head Neck Surg. 138 (11) (2012) 1024–1029, https://doi.org/10.1001/2013.jamaoto.377.

[30] T.S. Lendvay, T.C. Brand, L. White, et al., Virtual reality robotic surgery warm-up improves task performance in a dry laboratory environment: a prospective randomized controlled study, J. Am. Coll. Surg. 216 (6) (2013) 1181–1192, https://doi.org/10.1016/j.jamcollsurg.2013.02.012.

[31] G. Ahlberg, L. Enochsson, A.G. Gallagher, et al., Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies, Am. J. Surg. 193 (6) (2007) 797–804, https://doi.org/10.1016/j.amjsurg.2006.06.050.

[32] L. Furuya-Kanamori, C. Xu, S.S. Hasan, S.A. Doi, Quality versus risk-of-bias assessment in clinical research, J. Clin. Epidemiol. 129 (2021) 172–175, https://doi.org/10.1016/j.jclinepi.2020.09.044.

[33] R.L. Angelo, R.K. Ryu, R.A. Pedowitz, A.G. Gallagher, Metric development for an arthroscopic bankart procedure: assessment of face and content validity, Arthroscopy 31 (8) (2015) 1430–1440, https://doi.org/10.1016/j.arthro.2015.04.093.

[34] K.A. Ericsson, R.T. Krampe, C. Tesch-Romer, The role of deliberate practice in the acquisition of expert performance, Psychol. Rev. (1993) 363–406.

[35] K. Kojima, M. Graves, W. Taha, M. Cunningham, A. Joeris, A.G. Gallagher, AO international consensus panel for metrics on a closed reduction and fixation of a 31A2 pertrochanteric fracture, Injury 49 (12) (2018) 2227–2233, https://doi.org/10.1016/j.injury.2018.09.019.

[36] A.G. Gallagher, E.M. Ritter, H. Champion, et al., Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training, Ann. Surg. 241 (2) (2005) 364–372.

[37] D. Breen, S. O'Brien, N. McCarthy, A. Gallagher, N. Walshe, Effect of a proficiency-based progression simulation programme on clinical communication for the deteriorating patient: a randomised controlled trial, BMJ Open. 9 (7) (2019), e025992, https://doi.org/10.1136/bmjopen-2018-025992.

[38] R. Bender, T. Friede, A. Koch, et al., Methods for evidence synthesis in the case of very few studies, Res. Synth. Methods 9 (3) (2018) 382–392, https://doi.org/10.1002/jrsm.1297.

Appendix II. General descriptive characteristics of 13 randomized clinical trials studies included in the final qualitative analysis of the systematic review.

Study	Methods	Participants	Intervention	Outcomes	Other
Ahlberg et al 2007	Prospective randomized study. The aim of this study was to assess the effect of proficiency-based VR training on the outcome of the first 10 entire cholecystectomies performed by novices	n=13 PGY 1-2 Laparoscopically inexperienced residents (no experience of LC)	Group 1: VR training group to proficiency Group 2: Control group Proficiency level set by assessing 5 experts in laparoscopic surgery over 6 tasks on the simulator (LapSim) All participants carried out basic skills training on simulator- Group1 then practised under supervision and received feedback from simulator and also the supervisor until they displayed proficiency on each 6 tasks twice All subjects then went on to in vivo practice (Group 1 once Prof had been achieved; Group 2 within two weeks after study commencement	Psychometric testing; baseline abilities tested on randomization Surgeries 1,5 and 10 were assessed for each subject: 1. Errors (for each stage of the procedure) 2. Time taken	Ethical approval obtained

Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Angelo et al 2015	Randomized controlled trial (non-random control group) Objective: To compare PBP training for ABR using simulation with the same curriculum without proficiency requirements and compared with the AANA resident course for learning ABR	PGY 4-5 residents n=44 from 21 ACGME approved orthopaedic residency training programs from across the US	Group 1: traditional weekend arthroscopy shoulder course- cadaveric repair Group 2: on-line material, suturing and knot tying course, Bankart shoulder model (simulation)- cadaveric repair Group 3: On-line material to proficiency, suturing and knot-tying course to proficiency, Bankart shoulder model to proficiency- cadaveric repair	1. Score on a validated assessment tool using a videotaped performance of an arthroscopic Bankart repair. Evaluation using previously validated performance metrics (Angelo et al. 2015) i.e. 45 key steps grouped into 1 of 13 phases of the procedure. 2. Errors 3. Sentinel errors 4. Time to complete an arthroscopic Bankart repair on a cadaver specimen	Registered with the National Institutes of Health (ClinicalTrials.gov No. NCT01921621)

Study	Methods	Participants	Intervention	Outcomes	Other
Cates et al 2016	Randomized controlled study Aim: To assess the transfer of training (ToT) of virtual reality simulation training compared to invasive vascular experience training for carotid artery angiography (CA) for highly experienced interventionists but new to carotid procedures	Experienced interventional cardiologists with no experience in carotid artery angiography (n=12)	Participants were randomised to train on virtual reality (VR) simulation to a quantitatively defined level of proficiency or to a traditional supervised in vivo patient case training Group 1 (VR trained n=6) Group 2 (Standard trained n=6)	1. Errors 2. Fluoroscopy time 3. Procedure time 4. Attending takeovers	Approved by IRB

Study	Methods	Participants	Intervention	Outcomes	Other
Gerull et al 2020	Prospective cohort (pilot) study	31 participants (surgical residents in general surgery, urology, obstetrics and gynaecology	Pre-training: all subjects participated in a live robot-assisted laporoscopic surgical (RALS) case prior to reaching proficiency on the novel da Vinci skills simulator curriculum-scored using the RO-SCORE (assessing operative performance) immediately afterwards and the subject completed the NTLX(mental workload) Robotic simulator training: following the RALS case, all subjects trained to pre-set proficiency goals on a da Vinci Skills Simulator with a novel skills curriculum (proficiency curriculum decided by an expert panel) Post-training: upon reaching proficiency, subjects participated in a live RALS case with the same attending surgeon that was present for the pre-training. RO-SCORE and NTLX completed immediately afterwards	Outcome measures (RO-SCORE domains): 1. Camera control 2. Energy control 3. Needle control 4. Tissue handling 5. Instrument control 6. Visuospatial 7. Efficiency 8. Communication 9. Overall Outcome measures (NTLX domains scored by subjects from 1-10): 1. Mental demand 2. Physical demand 3. Temporal demand 4. Performance 5. Effort 6. Frustration	IRB-approved protocol

Study	Methods	Participants	Intervention	Outcomes	Other
Fried et al 2012	Prospective cohort study/Non randomised study Objective: to compare performance levels of residents trained to proficiency using the Endoscopic Sinus Surgery Simulator (ES3) versus residents trained by performing a fixed number of surgical procedures	Otorhinolaryngology junior residents (n=14) (composed of 8 experimental (proficiency trained) subjects and 6 control (standard training) subjects PGY 1-3 otolaryngology residents) 6 attending surgeons established benchmark criteria Two academic medical centres in New York city Participant inclusion criteria: performed fewer than 5 ESS cases as the primary surgeon Strict surgical case inclusion criteria	Both groups completed preliminary questionnaire to assess fine motor skill disparities and had 3 test trials on the ES3 simulator in novice mode. All residents were videotaped performing the surgical procedure on patients. Experimental subjects then achieved benchmark proficiency criteria (3 consecutive trials of >93.9) on the Endoscopic Sinus Surgery Simulator (ES3-validated); control subjects repeated the surgical procedure twice All residents then video-taped again performing procedure on patient	Metrics assessed on videotape of surgical procedure at baseline and after training using specially designed software for this purpose (measured variables: time, case difficulty, tool manipulation, tissue respect, task completion rate, surgical confidence (10 point scale) and number of errors) The listed variables were applied to the three main tasks performed by subjects; Navigation, injection and dissection.	IRB clearance obtained

Study	Methods	Participants	Intervention	Outcomes	Other
Srinivasan et al 2018	Development and validation of metrics and subsequent application to training through a randomised controlled study (only the RCT is considered as part of this review) of labour epidural catheter placement	14 anaesthesia trainees (17 recruited originally- 3 were excluded from the study (one from Group S and two trainees from Group P- did not get an opportunity to perform labour epidural catheter placements within 2 weeks of completing training) Group S: n=8 Group P: n=6 Each trainee carried out 10 procedures (n=140)	Jan 2015-Sept 2016 Group S- standard simulation training group Group P- PBP group Part one of training: study material and assessment test (Group P only- 80% pre-defined pass percentage) Part two of training: a standardised workshop (didactic session and simulation training session) run for each participant within 4 weeks of receiving the study material (Group P to proficiency) Part three: clinical procedures on the labour ward in CUMH	Trainee participant characteristics were reported on 1. Proportion of epidural failures (Defined- primary outcome measure) 2. The proportion of patients who experienced pain during uterine contraction at 60 min from the time of epidural needle insertion 3. Mean error rate 4. Mean epidural failure rate (comparison of learning curves between groups)	Ethical approval granted

Study	Methods	Participants	Intervention	Outcomes	Other
Kurashima et al 2013	Randomized controlled trial Objective: to assess whether training to proficiency using a novel laparoscopic inguinal hernia repair (LIHR) simulation curriculum improves operating room performance for an inguinal hernia repair compared to standard residency training	General surgery residents (PGYs 2-5)at the Faculty of Medicine McGill University, Montreal, Canada participated in the study from September 2010 to May 2012	Seventeen residents underwent the baseline evaluation. One was excluded on the basis of GOALS-GH >17. The 16 remaining residents participated in a didactic LIHR course and then were randomized to a training group (n=7) or the control group (standard residency) (n=9) Training schedule was set by the participants themselves, repeating supervised practice (by a surgeon experienced in minimally invasive surgery) until proficiency was achieved Interval between the last VR training and final assessment was a maximum of 15 days	Primary outcome: the difference between baseline and final performance in the OR measured by the mean total GOALS-GH scores. Assessed at baseline in the OR and again within 3 months of baseline testing. The training group was additionally tested within 2 weeks after achievement of proficiency	IRB approved (A07-M71-10B)

Study	Methods	Participants	Intervention	Outcomes	Other
Maertens et al 2016	Prospective Randomised Controlled Trial Objective: to evaluate the impact of a PROSPECT (PROficiency based StePwise Endovascular Curricular Training) training programme on real life operative performance.	32 general surgery trainees from Ghent University (PGYs 1-6) Randomised into 3 groups- stratified by postgraduate level Conducted at an academic centre and nine general hospitals from Oct 2014- February 2016 Live patients involved (eligibility criteria- suffered from symptomatic arterial disease of the lower limbs; TASC type A or B of iliac and/or femoral arteries; more complex lesions excluded)	Group 1(n=11): standard practice training combined with PROSPECT and multimedia based training and simulation sessions (2 dropouts- n=9) Group 2(n=10): standard practice training combined with multimedia based training modules Group 3(n=11): standard practice only (1 dropout- n=10) No baseline differences in cognitive and technical skill level among the 3 groups Trainees required to perform two endovascular procedures 6 weeks after training programme- all cases were videotaped	Primary outcome measure was to measure the difference in technical performance during the real life procedures between the three groups using OSATs derived rating scales (metrics used: total procedure time, fluoroscopy time, amount of contrast used, radiation dose, number of consultant takeovers, peri-op and post-op complications (at 30 days)) Secondary outcomes: changes in knowledge (MCQ test) and technical skills (VR simulator) and skills retention after the training programme.	The trial was registered at clinical trials.gov (NCT01965860)

Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Lendvay et al 2018	2-center randomized trial Objective: to assess whether a brief virtual reality (VR) robotic warm-up (before a laporoscopic procedure) would enhance robotic task performance and reduce errors	PGY1-6, surgical fellow and faculty from the Dept of Urology, General Surgery and Gynecology at the University of Washington Medical Centre and Madigan Army Medical Centre. 51 participants Group 1 (n=25) Group 2 (n=26) Randomization stratified by site and surgical experience level Proficiency benchmark based on the performance of 2 experienced robotic surgeons (>150 procedures)	Both groups underwent a validated robotic surgery proficiency curriculum on a VR robotic simulator and on the daVinci surgical robot. Once successfully achieving performance benchmarks, each surgeons was randomized to one of two groups. Group 1 (n=25): no warm-up (read a leisure book for 10 minutes) Group 2: 3-5 minute VR warm-up on a simulator Required to complete 4 trial sessions on the da Vinci rocking pegboard and 1 session on FLS intracorporeal suturing (sessions were at least 24 hours apart to avoid one session 'warming-up' the surgeon for the next session)	1.Total task time (seconds) 2.Cognitive errors (total count): rings placed on incorrect pegs, incorrect sequence of pegs 3.Technical errors (total count): dropped rings, peg touches 4.Tool path length (total distance travelled by instruments (mm)) 5.Economy of motion: path length/task time (mm/s) 6.Errors (defined)	Institutional Review Board approval granted (#35096)

Study	Methods	Participants	Intervention	Outcomes	Other
Seymour et al 2002	Randomised controlled trial (stratified by PGY) Objective: Assessment of virtual reality training (MIST VR “manipulate and diathermy” task) for laparoscopic cholecystectomy using proficiency benchmarks plus standard training V standard training alone (VR v ST)-	PGY 1-4 surgical residents (n=16) 11 Male and 5 female	Virtual reality training (MIST VR “manipulate and diathermy” task) for laparoscopic cholecystectomy using proficiency benchmarks plus standard training V standard laparoscopic cholecystectomy training alone 3-8 one hour training sessions; no study duration stated	1. Achievement of benchmark proficiency 2. Completion of surgical task 3. Duration of the procedure 4. Errors occurring (as defined in Table 1) (Total number of each type of error and total number of errors per procedure)	Supported with a grant from the Fulbright Distinguished Scholar Program
Sroka et al 2010	Randomized controlled trial Objective: to assess whether training to proficiency with the Fundamentals of Laparoscopic Surgery (FLS) simulator would result in improved performance in the OR	General surgery residents PGY 1-3 at McGill University, Montreal n=17 one participant of Group 1 lost to follow up before the final evaluation	All participants underwent baseline FLS training and assessed using GOALS scores Group 1: n=9 (training group- used the FLS simulator in a supervised proficiency-based curriculum; continued with regular residency training) Group 2: n=8 (non-training group- continued with their regular residency training) Both groups documented their clinical lap. experience throughout the study	GOALS scores (depth perception, bimanual dexterity, efficiency, tissue handling, autonomy)	Research Ethics Board-Approved study (Project A03-E06-04A)

			At the end of the study period (145days) subjects were assessed again on the simulator and in the OR, using the same metrics.		
Palter et al 2013	Randomized, single-blinded prospective trial Objective: to develop and validate an ex vivo comprehensive curriculum for a basic laparoscopic procedure	PGY1 and PGY2 general surgery residents (novice laparoscopists- <10 laparoscopic procedures) n=20	Operative experience recorded as well as baseline abilities and cognitive knowledge (MCQ) Group 1: STAC group-case-based learning, proficiency-based VR training, lap. box training, and OR participation Group 2: conventional residency training	Time, path length, angular path, errors OSATS scores (camera navigation, instrument handling, coordination, grasping, cutting)	IRB approval granted

Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Van Sickle et al 2008	Randomized double-blind trial Objective: to demonstrate that a structured, stepwise curriculum for MIS suturing and knot tying based on the concept of training to expert performance levels resulted in improved operative performance	Subject enrolment: Jan2003 to July 2005 22 participants PGY level 3,5 or 6 2 subject groups: Group1: Curriculum training group (VR training, box trainer suturing and knot tying) Group2: Standard training group (standard operating room instruction and self-guided practice)	All subject shown a 15 min training video on lap. suturing and knot tying (followed by an examination where 100% was required to continue with the study) Group1: MIST-VR and box trainer simulator to proficiency levels (based on attending surgeons and clinical fellows (experts) performance of 5 repetitions each on MIST-VR Task 3, the foam Nissen suturing model and the intracorporeal slip-square knot) with supervised training Group2: Access to MIST-VR but without supervised training Intraoperative performance assessed on fundal suturing portion of a lap. Nissen fundoplication (standardized) and video-recorded	Demographics and baseline abilities recorded 1. Time 2. Errors 3. Needle manipulations	IRB approval granted

Appendix III. Supplementary materials (published online only)

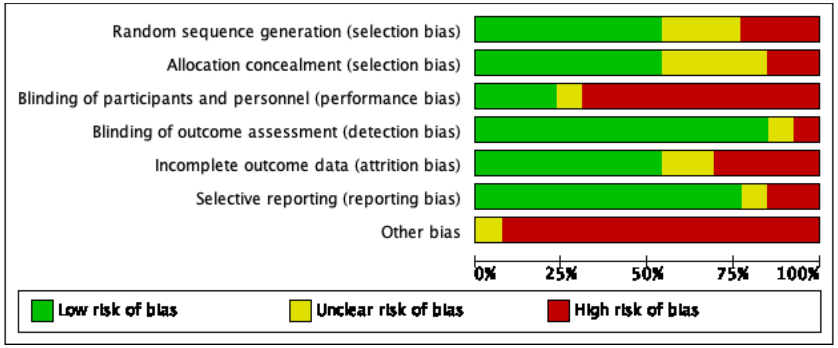
Online-only Supplemental Material

Table of contents

eFigure 1. Five-Stage Model of Adult Skill Acquisition- Dreyfus[7]

STAGE	PERFORMANCE CHARACTERISTICS
Expert	• Source of knowledge and information for others • Continually looks for better methods • Work primarily from intuition • Being forced to follow rules degrades performance
Proficient	• Seeks to understand larger context • Frustrated by over-simplification • Can self-correct performance • Can learn from experience of others
Competent	• Can troubleshoot problems on his/her own • Seeks out expert user advice • Develops conceptual models
Advanced Beginner	• Starts trying tasks on his/her own • Has difficulty troubleshooting • Begins to formulate principles, but without holistic understanding
Novice	• Has little or no previous experience • Doesn’t know how to respond to mistakes • Needs rules to function

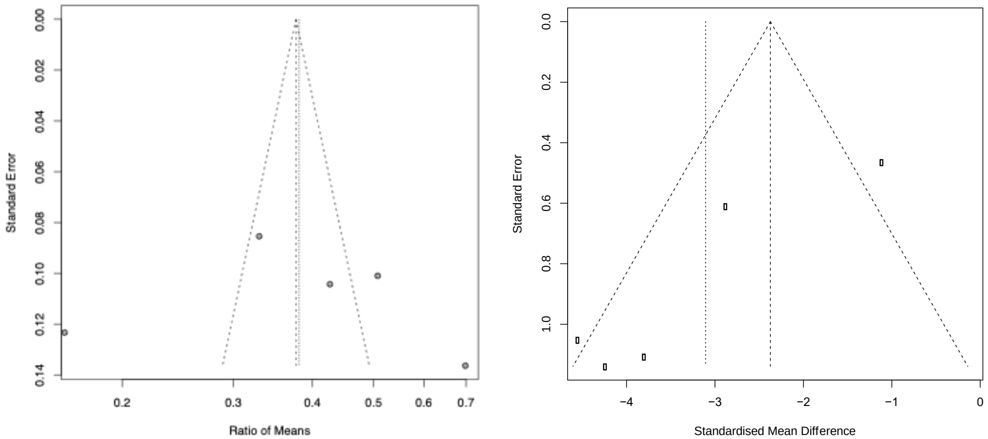
eFigure 2. Risk of Bias Graph (green indicates a low risk of bias; yellow indicates an unclear risk of bias; red indicates a high risk of bias)



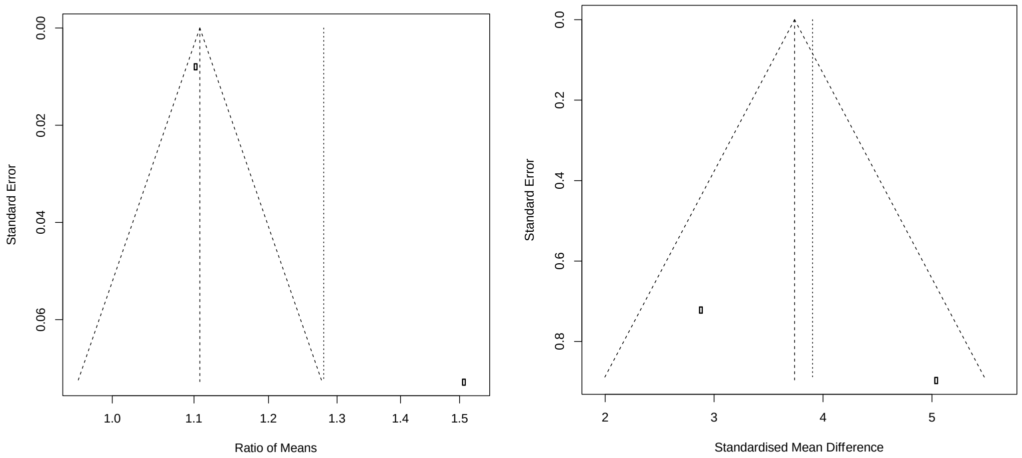
eFigure 3. Medical Education Research Study Quality Instrument (MERSQI) tool

Domain	MERSQI Item	Score	Max Score
Study design	Single group cross-sectional or single group posttest only	1	3
	Single group pretest & posttest	1.5	
	Nonrandomized, 2 groups	2	
	Randomized controlled trial	3	
Sampling	*Institutions studied:*		3
	1	0.5	
	2	1	
	3	1.5	
	Response rate, %:		
	Not applicable		
	<50 or not reported	0.5	
	50-74	1	
	≥75	1.5	
Type of data	Assessment by participants	1	3
	Objective measurement	3	
Validity of evaluation instrument	*Internal structure:*		3
	Not applicable		
	Not reported	0	
	Reported	1	
	Content:		
	Not applicable		
	Not reported	0	
	Reported	1	
	Relationships to other variables:		
	Not applicable		
	Not reported	0	
	Reported	1	
Data analysis	*Appropriateness of analysis:*		3
	Inappropriate for study design or type of data	0	
	Appropriate for study design, type of data	1	
	Complexity of analysis:		
	Descriptive analysis only	1	
	Beyond descriptive analysis	2	
Outcomes	Satisfaction, attitudes, perceptions, opinions, general facts	1	3
	Knowledge, skills	1.5	
	Behaviors	2	
	Patient/health care outcome	3	
Total possible score*			**18**

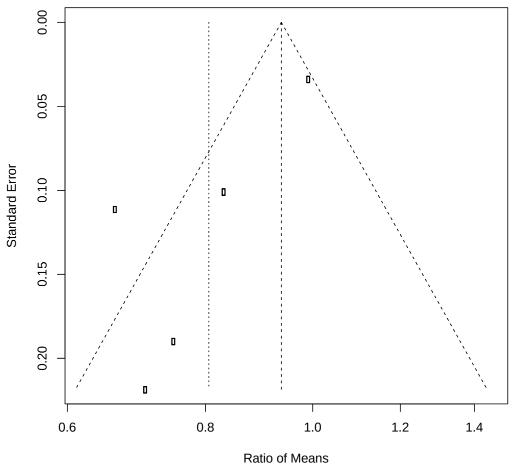
eFigure 4 – Funnel plots and Egger's asymmetry test evaluating publication biases according to different outcomes tested.

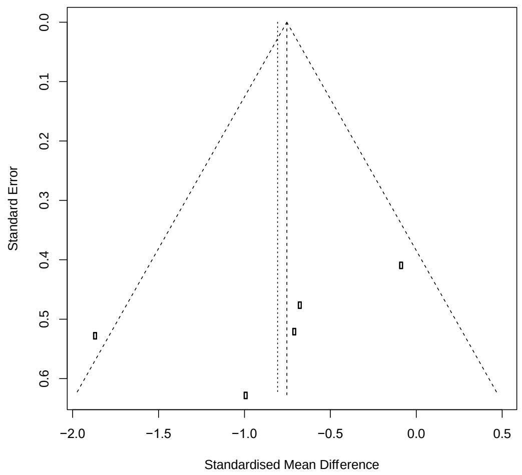


eFigure 4A. Errors

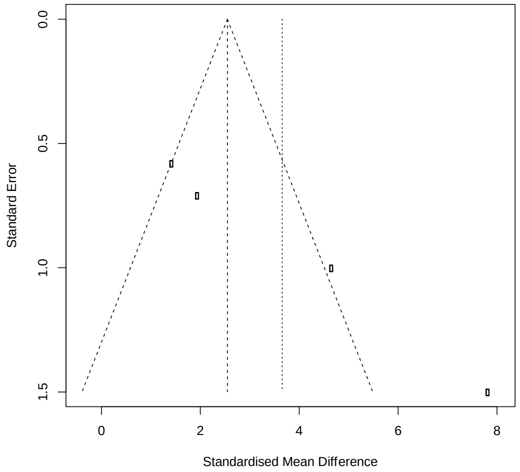


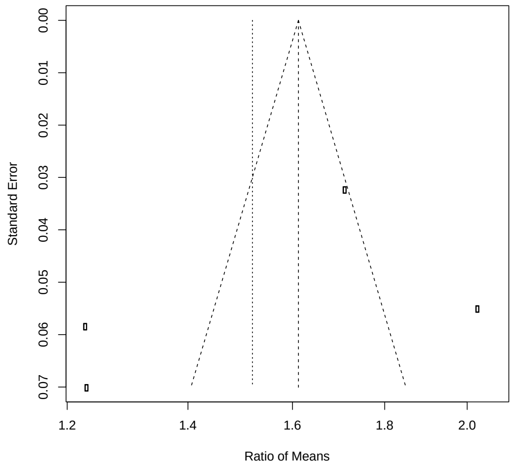
eFigure 4B. Steps





eFigure 4C. Time





eFigure 4D. Score

eTable 1. General descriptive characteristics of 13 randomized clinical trials studies included in the final qualitative analysis of the systematic review.

Study	Methods	Participants	Intervention	Outcomes	Other
Ahlberg et al 2007	Prospective randomized study. The aim of this study was to assess the effect of proficiency-based VR training on the outcome of the first 10 entire cholecystectomies performed by novices	n=13 PGY 1-2 Laparoscopically inexperienced residents (no experience of LC)	Group 1: VR training group to proficiency Group 2: Control group Proficiency level set by assessing 5 experts in laparoscopic surgery over 6 tasks on the simulator (LapSim) All participants carried out basic skills training on simulator- Group1 then practised under supervision and received feedback from simulator and also the supervisor until they displayed proficiency on each 6 tasks twice All subjects then went on to in vivo practice (Group 1 once Prof had been achieved; Group 2 within two weeks after study commencement	Psychometric testing; baseline abilities tested on randomization Surgeries 1,5 and 10 were assessed for each subject: 1. Errors (for each stage of the procedure) 2. Time taken	Ethical approval obtained
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**

Study	Methods	Participants	Intervention	Outcomes	Other
Angelo et al 2015	Randomized controlled trial (non-random control group) Objective: To compare PBP training for ABR using simulation with the same curriculum without proficiency requirements and compared with the AANA resident course for learning ABR	PGY 4-5 residents n=44 from 21 ACGME approved orthopaedic residency training programs from across the US	Group 1: traditional weekend arthroscopy shoulder course- cadaveric repair Group 2: on-line material, suturing and knot tying course, Bankart shoulder model (simulation)- cadaveric repair Group 3: On-line material to proficiency, suturing and knot-tying course to proficiency, Bankart shoulder model to proficiency- cadaveric repair	1. Score on a validated assessment tool using a videotaped performance of an arthroscopic Bankart repair. Evaluation using previously validated performance metrics (Angelo et al. 2015) i.e. 45 key steps grouped into 1 of 13 phases of the procedure. 2. Errors 3. Sentinel errors 4. Time to complete an arthroscopic Bankart repair on a cadaver specimen	Registered with the National Institutes of Health (ClinicalTrials.gov No. NCT01921621)

Cates et al 2016	Randomized controlled study Aim: To assess the transfer of training (ToT) of virtual reality simulation training compared to invasive vascular experience training for carotid artery angiography (CA) for highly experienced interventionists but new to carotid procedures	Experienced interventional cardiologists with no experience in carotid artery angiography (n=12)	Participants were randomised to train on virtual reality (VR) simulation to a quantitatively defined level of proficiency or to a traditional supervised in vivo patient case training Group 1 (VR trained n=6) Group 2 (Standard trained n=6)	1. Errors 2. Fluoroscopy time 3. Procedure time 4. Attending takeovers	Approved by IRB
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**

Gerull et al 2020	Prospective cohort (pilot) study	31 participants (surgical residents in general surgery, urology, obstetrics and gynaecology	Pre-training: all subjects participated in a live robot-assisted laporoscopic surgical (RALS) case prior to reaching proficiency on the novel da Vinci skills simulator curriculum-scored using the RO-SCORE (assessing operative performance) immediately afterwards and the subject completed the NTLX(mental workload) Robotic simulator training: following the RALS case, all subjects trained to pre-set proficiency goals on a da Vinci Skills Simulator with a novel skills curriculum (proficiency curriculum decided by an expert panel) Post-training: upon reaching proficiency, subjects participated in a live RALS case with the same attending surgeon that was present for the pre-training. RO-SCORE and NTLX completed immediately afterwards	Outcome measures (RO-SCORE domains): 10. Camera control 11. Energy control 12. Needle control 13. Tissue handling 14. Instrument control 15. Visuospatial 16. Efficiency 17. Communication 18. Overall Outcome measures (NTLX domains scored by subjects from 1-10): 7. Mental demand 8. Physical demand 9. Temporal demand 10. Performance 11. Effort 12. Frustration	IRB-approved protocol
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Fried et al 2012	Prospective cohort study/Non randomised study	Otorhinolaryngology junior residents (n=14) (composed of 8 experimental (proficiency trained) subjects and 6 control	Both groups completed preliminary questionnaire to assess fine motor skill disparities and had 3 test	Metrics assessed on videotape of surgical procedure at baseline and after training using	IRB clearance obtained

	Objective: to compare performance levels of residents trained to proficiency using the Endoscopic Sinus Surgery Simulator (ES3) versus residents trained by performing a fixed number of surgical procedures	(standard training) subjects PGY 1-3 otolaryngology residents) 6 attending surgeons established benchmark criteria Two academic medical centres in New York city Participant inclusion criteria: performed fewer than 5 ESS cases as the primary surgeon Strict surgical case inclusion criteria	trials on the ES3 simulator in novice mode. All residents were videotaped performing the surgical procedure on patients. Experimental subjects then achieved benchmark proficiency criteria (3 consecutive trials of >93.9) on the Endoscopic Sinus Surgery Simulator (ES3-validated); control subjects repeated the surgical procedure twice All residents then video-taped again performing procedure on patient	specially designed software for this purpose (measured variables: time, case difficulty, tool manipulation, tissue respect, task completion rate, surgical confidence (10 point scale) and number of errors) The listed variables were applied to the three main tasks performed by subjects; Navigation, injection and dissection.	
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Srinivasan et al 2018	Development and validation of metrics and subsequent application to training through a randomised controlled study (only the RCT is considered	14 anaesthesia trainees (17 recruited originally- 3 were excluded from the study (one from Group S and two trainees from Group P- did not get an opportunity to perform labour epidural	Jan 2015-Sept 2016 Group S- standard simulation training group Group P- PBP group	Trainee participant characteristics were reported on	Ethical approval granted

	as part of this review) of labour epidural catheter placement	catheter placements within 2 weeks of completing training) Group S: n=8 Group P: n=6 Each trainee carried out 10 procedures (n=140)	Part one of training: study material and assessment test (Group P only- 80% pre-defined pass percentage) Part two of training: a standardised workshop (didactic session and simulation training session) run for each participant within 4 weeks of receiving the study material (Group P to proficiency) Part three: clinical procedures on the labour ward in CUMH	1. Proportion of epidural failures (Defined- primary outcome measure) 2. The proportion of patients who experienced pain during uterine contraction at 60 min from the time of epidural needle insertion 3. Mean error rate 4. Mean epidural failure rate (comparison of learning curves between groups)	
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Kurashima et al 2013	Randomized controlled trial Objective: to assess whether training to proficiency using a novel laparoscopic inguinal hernia repair (LIHR) simulation curriculum improves	General surgery residents (PGYs 2-5)at the Faculty of Medicine McGill University, Montreal, Canada participated in the study from September 2010 to May 2012	Seventeen residents underwent the baseline evaluation. One was excluded on the basis of GOALS-GH >17. The 16 remaining residents participated in a didactic LIHR course and then were randomized to a training group (n=7) or the control	Primary outcome: the difference between baseline and final performance in the OR measured by the mean total GOALS-GH scores. Assessed at baseline in the OR and again within 3 months of baseline	IRB approved (A07-M71-10B)

	operating room performance for an inguinal hernia repair compared to standard residency training		group (standard residency) (n=9) Training schedule was set by the participants themselves, repeating supervised practice (by a surgeon experienced in minimally invasive surgery) until proficiency was achieved Interval between the last VR training and final assessment was a maximum of 15 days	testing. The training group was additionally tested within 2 weeks after achievement of proficiency	
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Maertens et al 2016	Prospective Randomised Controlled Trial Objective: to evaluate the impact of a PROSPECT (PROficiency based StePwise Endovascular Curricular Training) training programme on real life operative performance.	32 general surgery trainees from Ghent University (PGYs 1-6) Randomised into 3 groups- stratified by postgraduate level Conducted at an academic centre and nine general hospitals from Oct 2014- February 2016	Group 1(n=11): standard practice training combined with PROSPECT and multimedia based training and simulation sessions (2 dropouts- n=9) Group 2(n=10): standard practice training combined with multimedia based training modules	Primary outcome measure was to measure the difference in technical performance during the real life procedures between the three groups using OSATs derived rating scales (metrics used: total procedure time, fluoroscopy time, amount of contrast used, radiation dose, number of consultant takeovers,	The trial was registered at clinical trials.gov (NCT01965860)

		Live patients involved (eligibility criteria- suffered from symptomatic arterial disease of the lower limbs; TASC type A or B of iliac and/or femoral arteries; more complex lesions excluded)	Group 3(n=11): standard practice only (1 dropout- n=10) No baseline differences in cognitive and technical skill level among the 3 groups Trainees required to perform two endovascular procedures 6 weeks after training programme- all cases were videotaped	peri-op and post-op complications (at 30 days)) Secondary outcomes: changes in knowledge (MCQ test) and technical skills (VR simulator) and skills retention after the training programme.	
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Lendvay et al 2018	2-center randomized trial Objective: to assess whether a brief virtual reality (VR) robotic warm-up (before a laporoscopic procedure) would enhance robotic task performance and reduce errors	PGY1-6, surgical fellow and faculty from the Dept of Urology, General Surgery and Gynecology at the University of Washington Medical Centre and Madigan Army Medical Centre. 51 participants Group 1 (n=25) Group 2 (n=26)	Both groups underwent a validated robotic surgery proficiency curriculum on a VR robotic simulator and on the daVinci surgical robot. Once successfully achieving performance benchmarks, each surgeons was randomized to one of two groups. Group 1 (n=25): no warm-up (read a leisure book for 10 minutes)	1.Total task time (seconds) 2.Cognitive errors (total count): rings placed on incorrect pegs, incorrect sequence of pegs 3.Technical errors (total count): dropped rings, peg touches	Institutional Review Board approval granted (#35096)

		Randomization stratified by site and surgical experience level Proficiency benchmark based on the performance of 2 experienced robotic surgeons (>150 procedures)	Group 2: 3-5 minute VR warm-up on a simulator Required to complete 4 trial sessions on the da Vinci rocking pegboard and 1 session on FLS intracorporeal suturing (sessions were at least 24 hours apart to avoid one session 'warming-up' the surgeon for the next session)	4.Tool path length (total distance travelled by instruments (mm)) 5.Economy of motion: path length/task time (mm/s) 6.Errors (defined)	
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Seymour et al 2002	Randomised controlled trial (stratified by PGY) Objective: Assessment of virtual reality training (MIST VR "manipulate and diathermy" task) for laparoscopic cholecystectomy using proficiency benchmarks plus standard training V standard training alone (VR v ST)-	PGY 1-4 surgical residents (n=16) 11 Male and 5 female	Virtual reality training (MIST VR "manipulate and diathermy" task) for laparoscopic cholecystectomy using proficiency benchmarks plus standard training V standard laparoscopic cholecystectomy training alone 3-8 one hour training sessions; no study duration stated	1. Achievement of benchmark proficiency 2. Completion of surgical task 3. Duration of the procedure 4. Errors occurring (as defined in Table 1) (Total number of each type of error and total number of errors per procedure)	Supported with a grant from the Fulbright Distinguished Scholar Program

Sroka et al 2010	Randomized controlled trial Objective: to assess whether training to proficiency with the Fundamentals of Laparoscopic Surgery (FLS) simulator would result in improved performance in the OR	General surgery residents PGY 1-3 at McGill University, Montreal n=17 one participant of Group 1 lost to follow up before the final evaluation	All participants underwent baseline FLS training and assessed using GOALS scores Group 1: n=9 (training group- used the FLS simulator in a supervised proficiency-based curriculum; continued with regular residency training) Group 2: n=8 (non-training group- continued with their regular residency training) Both groups documented their clinical lap. experience throughout the study At the end of the study period (145days) subjects were assessed again on the simulator and in the OR, using the same metrics.	GOALS scores (depth perception, bimanual dexterity, efficiency, tissue handling, autonomy)	Research Ethics Board-Approved study (Project A03-E06-04A)
Palter et al 2013	Randomized, single-blinded prospective trial Objective: to develop and validate an ex vivo comprehensive curriculum for a basic laparoscopic procedure	PGY1 and PGY2 general surgery residents (novice laparoscopists- <10 laparoscopic procedures) n=20	Operative experience recorded as well as baseline abilities and cognitive knowledge (MCQ) Group 1: STAC group- case-based learning, proficiency-based VR	Time, path length, angular path, errors OSATS scores (camera navigation, instrument handling, coordination, grasping, cutting)	IRB approval granted

			training, lap. box training, and OR participation Group 2: conventional residency training		
Study	**Methods**	**Participants**	**Intervention**	**Outcomes**	**Other**
Van Sickle et al 2008	Randomized double-blind trial Objective: to demonstrate that a structured, stepwise curriculum for MIS suturing and knot tying based on the concept of training to expert performance levels resulted in improved operative performance	Subject enrolment: Jan2003 to July 2005 22 participants PGY level 3,5 or 6 2 subject groups: Group1: Curriculum training group (VR training, box trainer suturing and knot tying) Group2: Standard training group (standard operating room instruction and self-guided practice)	All subject shown a 15 min training video on lap. suturing and knot tying (followed by an examination where 100% was required to continue with the study) Group1: MIST-VR and box trainer simulator to proficiency levels (based on attending surgeons and clinical fellows (experts) performance of 5 repetitions each on MIST-VR Task 3, the foam Nissen suturing model and the intracorporeal slip-square knot) with supervised training	Demographics and baseline abilities recorded 4. Time 5. Errors 6. Needle manipulations	IRB approval granted

			Group2: Access to MIST-VR but without supervised training Intraoperative performance assessed on fundal suturing portion of a lap. Nissen fundoplication (standardized) and video-recorded		

eTable 2. Summary of findings table (using the GRADE approach for quality of findings assessment)

Outcomes	Standardised mean difference (95% CI) (between standard/traditional training group (control) and PBP training group)	Relative effect (95% CI) ROM	No. of participants (studies)	Quality of the evidence (GRADE)	Comments
No. of errors	3.11 (-4.54; -1.68)[1]	0.38 (0.25; 0.58)[2]	87 (5)	Moderate[3]	Lower number of errors indicates improved performance
No. of steps	3.90 (1.79; 6.02)[4]	1.28 (0.94; 1.74)[5]	42 (2)	Low[6]	Higher number of steps completed indicates improved performance
Time	0.81 (-1.40; -0.21)[7]	0.81 (0.66; 0.98)[8]	93 (5)	Moderate[9]	
Likert scale scoring	3.65 (1.40; 5.90)	1.52 (1.22; 1.90)[10]	49 (4)	Moderate[11]	

CI: Confidence Interval; ROM: Ratio of Means

12. Overall, PBP training reduced the number of errors when compared to standard training
13. PBP was estimated to reduce the mean rate of errors by 62%, when compared to standard training
14. All 5 included studies were RCTs but downgraded from high quality to moderate due to risk of bias assessment
15. Overall, trainees who completed PBP training performed more procedural steps than those who completed a standard training pathway
16. PBP did not statistically significantly increase the mean rate of steps performed when compared to standard training
17. Data is sparse (only two studies included in analysis)
18. The reduction of procedural time was less pronounced compared to other outcomes, such as the number of errors or steps completed
19. PBP reduced the mean procedural time by approximately 15%, when compared to standard training
20. All 5 included studies were RCTs but downgraded from high quality evidence to moderate due to incomplete outcome data in one study
21. Trainees who completed PBP achieved a mean Likert scale-based score 52% higher than those who completed a standard training pathway
22. Downgraded from high quality evidence to moderate due to high % of heterogeneity

Appendix IV. PRISMA abstract checklist

Section and Topic	Item #	Checklist item	Reported (Yes/No)
TITLE			
Title	1	Identify the report as a systematic review.	Yes
BACKGROUND			
Objectives	2	Provide an explicit statement of the main objective(s) or question(s) the review addresses.	Yes
METHODS			
Eligibility criteria	3	Specify the inclusion and exclusion criteria for the review.	Yes
Information sources	4	Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.	Yes
Risk of bias	5	Specify the methods used to assess risk of bias in the included studies.	Yes
Synthesis of results	6	Specify the methods used to present and synthesise results.	Yes
RESULTS			
Included studies	7	Give the total number of included studies and participants and summarise relevant characteristics of studies.	Yes
Synthesis of results	8	Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).	Yes
DISCUSSION			
Limitations of evidence	9	Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).	No
Interpretation	10	Provide a general interpretation of the results and important implications.	Yes
OTHER			
Funding	11	Specify the primary source of funding for the review.	No
Registration	12	Provide the register name and registration number.	Not registered

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

Appendix V. PRISMA checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Completed
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 4-5
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 6
METHODS			
Eligibility criteria	5	Specify (A) the inclusion and exclusion criteria for the review and (B)how studies were grouped for the syntheses.	(A) Pg 7 (B) Pg 9
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 6-7
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 6-7; Figure 1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 7
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 8 and 9
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 8,12
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 8, Page 12 Supplementary materials eTable 1
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 8-10 Figure 2; eFigure 2 and eFigure 3

Section and Topic	Item #	Checklist item	Location where item is reported
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 9-10
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Table 1; eTable 1
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Page 10
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Table 1; eTable 1; Figure 2; eFigure 2; Figure 3-6
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 9-10
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Page 9-10
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Page 10-11
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Page 9
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 10; Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	
Study characteristics	17	Cite each included study and present its characteristics.	Table 1; eTable 1
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Figure 2; eFigure 2
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Figure 3-6
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Table 1
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Page 12-13 Figures 3-6
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Figures 3-6

Section and Topic	Item #	Checklist item	Location where item is reported
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Figures 3-6
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	eFigure 4
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Figures 3-6
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 16
	23b	Discuss any limitations of the evidence included in the review.	Page 16
	23c	Discuss any limitations of the review processes used.	
	23d	Discuss implications of the results for practice, policy, and future research.	Page 16-17
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Not registered
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Protocol not prepared
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	NA
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	NA
Competing interests	26	Declare any competing interests of review authors.	NA
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	NA

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

Appendix VI. Data collection template for included studies

	PBP training			Standard training		
	Mean	Standard deviation	Num of participants	Mean	Standard deviation	Num of participants
Time						
Errors						
Subjective scale score (OSATS, GOALS etc.)						
Other objective measure available??						

Appendix VII. Postgraduate training completed to date

1. Epigeum research integrity training (external)
2. Completed Certificate and Diploma in Teaching & Learning in Higher Education, University College Cork (30 credits)
3. ST6013 Statistics and Data Analysis for Postgraduate Research students (10 credits), University College Cork
4. PG7016 Systematic Reviews for Health Sciences (10 credits), University College Cork
5. PG6009 Graduate Information Literacy Skills (5 credits), University College Cork
6. DH6014 Digital Skills for Research Postgraduates (5 credits), University College Cork
7. PG7048 Generic and Transferrable Skills Portfolio (5 credits), University College Cork
8. CSTAR SPSS Training, University College Dublin