OPEN

A Systematic Review and Meta-analysis on the Impact of Proficiency-based Progression Simulation Training on Performance Outcomes

Elio Mazzone,† Stefano Puliatti, MD,†‡§✉ Marco Amato,†‡§ Brendan Bunting,¶ Bernardo Rocco,§ Francesco Montorsi,* Alexandre Mottrie,†‡ and Anthony G. Gallagher, PhD, DSc†¶||*

Objective: To analyze all published prospective, randomized, and blinded clinical studies on the proficiency-based progression (PBP) training using objective performance metrics.

Background: The benefit of PBP methodology to learning clinical skills in comparison to conventional training is not settled.

Methods: Search of PubMed, Cochrane library's Central, EMBASE, MEDLINE, and Scopus databases, from inception to 1st March 2020. Two independent reviewers extracted the data. The Medical Education Research Study Quality Instrument (MERSQI) was used to assess the methodological quality of included studies. Results were pooled using biased corrected standardized mean difference and ratio-of-means. 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. Secondary outcomes were the number of procedural steps completed and the time to complete the task/procedure.

Results: From the initial pool of 468 studies, 12 randomized clinical studies with a total of 239 participants were included in the analysis. In comparison to the non-PBP training, ratio-of-means results showed that PBP training reduced the number of performance errors by 60% ($P < 0.001$) and procedural time by 15% ($P = 0.003$) and increased the number of steps performed by 47% ($P < 0.001$).

Conclusions and Relevance: Our systematic review and meta-analysis confirms that PBP training in comparison to conventional or quality assured training improved trainees' performances, by decreasing procedural errors and procedural time, while increasing the number of correct steps taken when compared to standard simulation-based training.

Keywords: objective performance metrics, procedural errors, procedural steps, proficiency-based metrics, proficiency-based progression training, simulation-based training, Surgical training, technology-enhanced training

(*Ann Surg* 2021;274:281–289)

Simulation-based training had a strong foothold in safety conscious industries such as aviation,[1] nuclear-power,[2] and had been used in anesthesia for more than a decade to give the individuals or teams the experience of emergency situations before they were actually encountered in a real-life clinical situation.[3] In the field of surgery, the role of simulation-based training was first introduced by Satava[4] and collaborators who set-up the first prospective, randomized, and blinded clinical study [(ie, randomized clinical trial (RCT)] demonstrating that trainees who underwent a virtual reality (VR)-based simulation training pathway performed significantly better than traditionally trained surgeons, thus achieving an optimal performance level before starting their clinical practice in the operating room.[5] Of note, this study was the first to introduce the ''proficiency-based progression'' (PBP) training methodology which differs significantly from traditional training pathways. Specifically, the operative procedure is characterized in detail to identify intraoperative objective performance metrics for optimal and suboptimal performance.[6] After defining these objective metrics, trainees are required to continue training until they demonstrate a quantitatively predefined benchmark or proficiency level. During this practice, trainees receive continuous formative feedback in accordance to the concept of deliberate practice.[7] The level of proficiency is based on the mean performance of the experienced practitioners performing the same training tasks.[8]

During the last 2 decades, the PBP methodology evolved in terms of the robustness of the metric development and clinical validation evidence.[9,10] Where a VR simulator was not available the metrics were deployed using simulation models; for example, knot tying models,[11,12] silicon models,[13] or cadavers.[14] The requirement to demonstrate to a quantitatively predefined proficiency benchmark in training, combined with simulation-based practice, meant that PBP training was particularly effective; demonstrated performance improvements >40% in objectively assessed intraoperative errors in comparison to traditional skills-based training in the areas of laparoscopic surgery,[5,8,15] arthroscopic surgery,[16] endovascular interventions,[17] anesthesia,[18] and communication skills for deteriorating patients.[19]

Several focused reviews have attempted to delineate the impact of simulation-based training specifically for laparoscopic surgery.[20,21] However, each had limitations including ambiguous classification of comparison interventions, incomplete assessment of study quality, or no quantitative pooling to derive best estimates of effect or effect size, focused their evaluation on process measures such as knowledge, skill time, skill process etc with only 1 study on patient outcomes.[22] Process measures are related to the performance of the procedure, that is, how long it took, but gives no indication of the quality of procedure performance. The review reported here focuses on prospective, RCT specifically on PBP simulation training and evaluates the impact of PBP on learning clinical skills in comparison to the traditional approach to training.

METHODS

Study Identification and Evaluation

A systematic review of the literature was conducted using the PubMed, Cochrane library's Central, EMBASE, MEDLINE, and

From the *Division of Oncology/Unit of Urology, URI, IRCCS Ospedale San Raffaele, Milan, Italy; Vita-Salute San Raffaele University, Milan, Italy; †ORSI Academy, Melle, Belgium; ‡Department of Urology, OLV, Aalst, Belgium; §Department of Urology, University of Modena and Reggio Emilia, Modena, Italy; ¶School of Medicine, Faculty of Life and Health Sciences, Ulster University, Northern Ireland, UK; and ||Faculty of Medicine, KU Leuven, Leuven, Belgium.

✉stefanopuliatti@gmail.com.

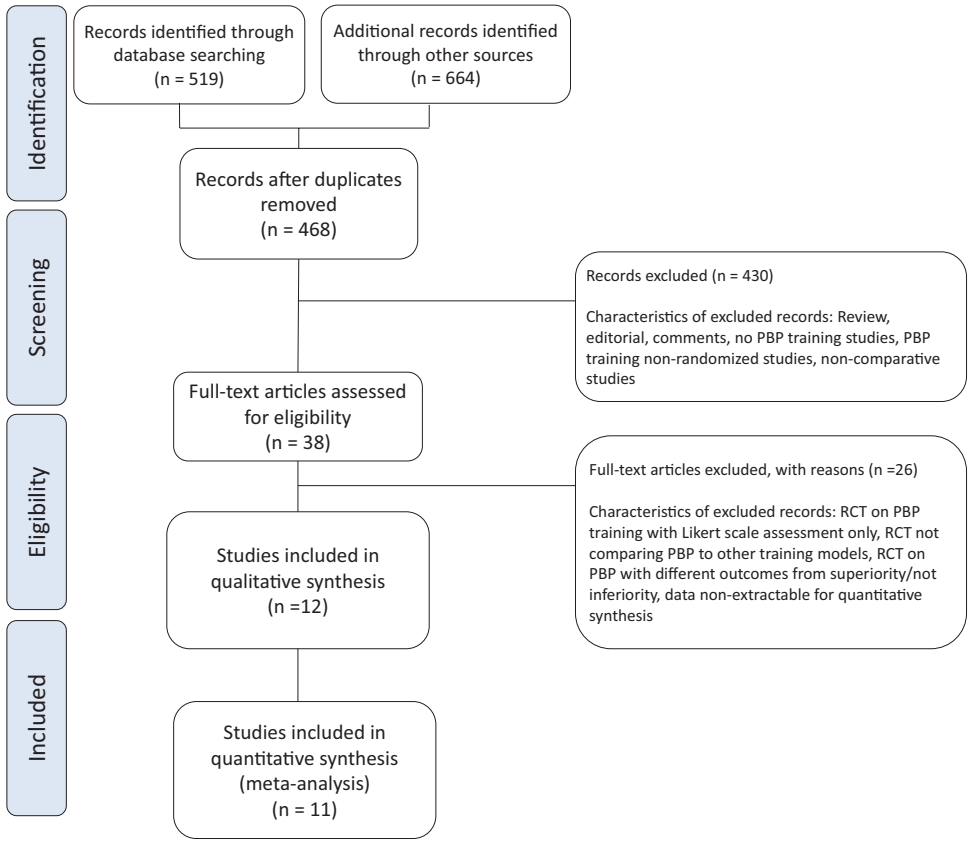
Elio Mazzone and Stefano Puliatti share first authorship.

The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

ISSN: 0003-4932/20/27402-0281

DOI: 10.1097/SLA.0000000000004650



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

FIGURE 1. Flow-chart of studies through the screening process according to the PRISMA methodology. PBP indicates proficiency-based progression; RCT, randomized clinical trial.

Scopus databases (Supplementary Material Appendix 1, http://links.lww.com/SLA/C803). We searched from inception of the databases up to 1st March 2020. All the references of key reviews on training were also screened. Keywords used for the research were: ''Proficiency-based AND progression AND training, Proficiency AND based AND progression, Proficiency-based AND training.'' This systematic review is reported in accordance with the preferred reporting items for systematic reviews and meta-analyses protocols (PRISMA-P) guidelines[23] and is registered with the international prospective registry of systematic reviews (PROSPERO, CRD42020182400).

Initial Screening, Eligibility Criteria, and Risk of Bias Assessment

After identifying all eligible studies, 2 independent reviewers (MA, SP) screened all titles and abstracts (or full text, for further clarification) for inclusion in the study. Literature reviews, editorial, comments, and non-PBP-based studies (other than as a control condition) were excluded at the initial screening (Fig. 1). Only studies that used objective binary scored performance-based metrics and a PBP methodology were included for the final quantitative synthesis.[8,11,15–17,19,24–27] Disagreements regarding eligibility were resolved by discussion between the 2 investigators until consensus was reached.

Methodological quality of the included studies was graded using the Medical Education Research Study Quality Instrument (MERSQI).[28] Two investigators (EM and SP) independently assessed the risk of bias for all studies and the inter-rater reliability (IRR) of the assessors was calculated (ie, IRR = Agreements/Agreements + Disagreements).[29]

Intervention and Comparison Arms

The training tasks/procedures considered for the meta-analytic comparison were categorized as, medical procedure, surgical procedure, basic skill, and clinical communication skill. The intervention outcome was considered to be the direct or post-training result related to the training pathway.

For meta-analytic evaluation, the PBP simulation-based training arm was considered as the experimental arm. The group which received a non-PBP simulation-based training represented the

TABLE 1. General Characteristics of 12 Randomized Clinical Trials Studies Included in the Final Qualitative Analysis of the Systematic Review

Study	Subjects N; Type	Comparison Arm	Task/Procedure Trained	Intraoperative Patient Performance	Outcomes Compared	Other Scale Used	MERSQI
Ahlberg et al	13; Residents	Swedish National Surgical Residency Training Program	Laparoscopic Cholecystectomies	Yes	Errors	—	16
Ahmed et al	18; Medicine Students	Self-Guided Ultrasound-Guided Peripheral Nerve Block Simulation Practice	Ultrasound-Guided Peripheral Nerve Block	No	Errors, Steps	—	15
Angelo et al	44; Residents	ACGME approved Orthopedic Residency & Arthroscopy Association of North America Shoulder Course	Arthroscopic Bankart Procedure	Yes	Errors, Steps, Time	—	16
Breen et al	90; Medicine and nursing students	National and certified ISBAR training Program	Clinical Communication	No	Errors, Steps	—	15
Cates et al	12; Attendings	Industry sponsored CASES education and training system	Carotid Artery Angiography	Yes	Errors, Time	—	15
Jensen et al	16; Residents	ESC Core Curriculum for the General Cardiologist	Coronary Angiography	No	Errors, Steps, Time	—	17
Palter, et al	25; Residents	ACGME approved General Surgery Residency Training Program	Laparoscopic Right Colectomy	Yes	Steps	OSATS	16
Pedowitz et al	44; Residents	ACGME approved Orthopedic Residency & Arthroscopy Association of North America Shoulder Course	Knot-Tying	No	Errors	—	14.5
Peeters, et al	10; Residents	National Obstetrics and Gynecology Residency Program	Fetoscopy Laser Surgery	No	Steps, Time	—	16.5
Seymour et al	16; Residents	ACGME approved General Surgery Residency Training Program	Laparoscopic Cholecystectomy	Yes	Errors, Time	—	15
Srinivasan et al	17; Residents	Irish National Anesthesia Training Program	Epidural Analgesia	Yes	Errors	GRS, TSCL	17
Van Sickle et al	22; Residents	ACGME approved General Surgery Residency Training Program	Nissen Fundoplication	Yes	Errors, Time	—	14.5

ACGME indicates Accreditation Council for Graduate Medical Education; ESC, European Society of Cardiology; GRS, Global rating scales; ISBAR, Identification, Situation, Background, Assessment, Recommendation; MERQI, Medical Education Research Study Quality Instrument CASES, Carotid Artery Stenting Education System; OSATS, objective structured assessment of technical skills; TSCL, Task-specific checklists.

comparison arm (Table 1). For both arms, studies including any simulation, VR simulators, other technology-enhanced training models, or human cadaveric specimens, were considered eligible.

Outcomes Definition

PBP training has been previously described in detail.[30–32] According to PBP-related definitions, metrics are explicitly defined units of measurement that characterize elements of procedure/task performance that are scored in a binary fashion (ie, occurred/did not occur). The metrics are quantitative assessments and are used for objective evaluations to make comparisons or to track performance. This included performance errors and steps as metrics and were objectively assessable. Error was defined as a ''deviation from the optimal performance.'' Steps were defined as component tasks and the series was aggregated because they constitute the completion of a specific procedure.[16] Only the studies that specified those parameters in their analysis were included within the qualitative analysis The quantitative analysis was limited to studies that specified step and error metrics in their analysis and used those metrics to define a proficiency benchmark that trainees were required to demonstrate before training was deemed completed. Time was also considered as additional outcome. Assessment by Likert scales was not included in the current analyses, because of the potential for inherent ambiguity.

All studies meeting these criteria are shown in Table 2. The primary outcome used for pooled meta-analysis was the number of procedural errors made since errors provide an objective measure of performance quality.[16,30,32,33] Secondary outcomes were the number of steps performed and time to completion of the task/procedure. Both are considered process measures of task/procedure performance.

Data Synthesis and Statistical Analysis

Data not suitable for meta-analytic evaluation was presented in narrative fashion (qualitative analysis). Reported results for continuous outcomes were pooled using biased corrected standardized

TABLE 2. Baseline Characteristics of Included Studies According to Participants Data and Outcomes Measures

Feature	Subgroup	All Outcomes	All outcomes -PBP group	All outcomes - Control group	Time	Errors	Steps	Proficiency
		No. Studies (No. Participants)	No. Studies (No. Participants)	No. Studies (No. Participants)	No. Studies (No. Participants)	No. Studies (No. Participants)	No. Studies (No. Participants)	No. Studies (No. Participants)
All		12 (239)	12 (127)	12 (112)	6 (100)	10 (211)	6 (134)	3 (102)
Design	RCT	12	12	12	6	10	6	3
Participants	Medical/nursing Students	2 (66)	2 (36)	2 (30)	0	2 (66)	2 (66)	1 (48)
	Residents	9 (161)	9 (85)	9 (76)	5 (88)	7 (133)	4 (68)	2 (54)
	Physicians in practice	1 (12)	1 (6)	1 (6)	1 (12)	1 (12)	0	0
Task or procedure	Skill	3 (70)	3 (37)	3 (33)	1 (22)	3 (70)	1 (18)	1 (30)
	Surgical procedure	4 (63)	4 (31)	4 (32)	3 (50)	3 (40)	2 (34)	0
	Medical procedure	4 (58)	4 (44)	4 (14)	2 (28)	3 (53)	2 (34)	1 (24)
	Not medical procedure	1 (48)	1 (25)	1 (23)	0	1 (48)	1 (48)	1 (48)
Clinical Relevance	Present	7 (117)	7 (73)	7 (44)	4 (84)	6 (99)	2 (42)	1 (24)
	Absent	5 (122)	5 (54)	5 (68)	2 (16)	4 (112)	4 (92)	2 (78)
Outcomes	Satisfaction, aptitude etc	0	0	0	0	0	0	0
	Knowledge, skills	3 (70)	3 (37)	3 (33)	1 (22)	3 (70)	1 (18)	0
	Behavior	8 (157)	8 (79)	8 (78)	5 (78)	6 (129)	5 (116)	3 (102)
	Patient/health system outcomes	1 (12)	1 (11)	1 (1)	0	1 (12)	0	0

PBP indicates proficiency-based progression, RCT, randomized clinical trials.

mean difference (SMD) (Hedges' g effect size) according to previous established methodology.[22,34] Thus, the bias-corrected SMD and odds ratio (OR) were used to compare continuous and dichotomous variables, respectively. Additionally, for continuous outcomes, ratio of means (ROM) was applied to provide an estimation of the pooled effect of PBP on the considered outcomes.[35,36] All results were reported with 95% confidence intervals. Preplanned subgroups analyses were performed in studies with or without intraoperative patient performance assessment.

Heterogeneity between studies was measured using the I^2 statistic [37] and the between-study variance (t^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[38]) were prioritized. For the assessment of small study effects and publication bias, values of the SMD or OR 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.[39] This error is caused because statistically significant study results generally have a higher likelihood of being published. Furthermore, Eggers asymmetry test[40] was used to explore statistically the presence of publication bias. Statistical significance for all analyses was defined as 2-sided $P < 0.05$. Statistical analysis was performed with the R software (version 3.6.3; http://www.r-project.org/).

RESULTS

Study Selection Flow-chart

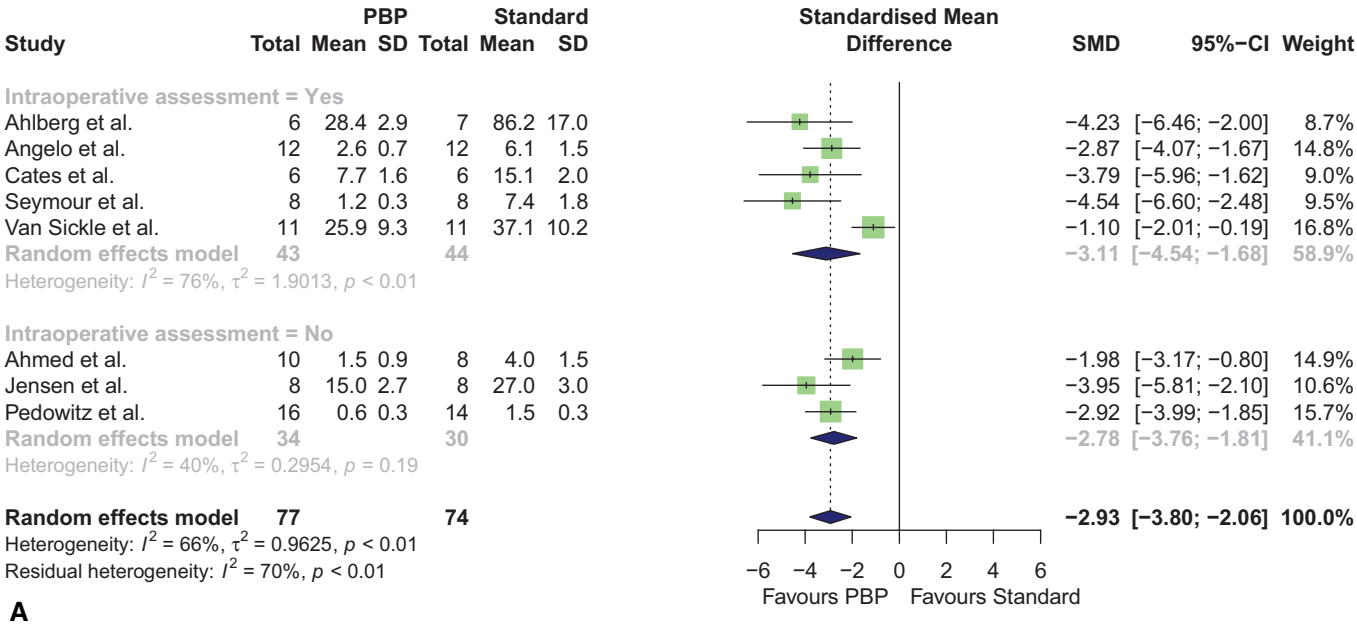
Figure 1 shows the flow of studies through the screening process. Five hundred nineteen papers were blindly screened by 2 reviewers (MA, SP) by reading all the titles and abstracts with 463 of these records included for further evaluation based on predefined eligibility criteria. Of these, 38 studies were considered eligible for final inclusion in qualitative analysis. At this point, final evaluation for the inclusion in the quantitative synthesis was carried out by 3 reviewers (AGG, EM, SP). At the end of the process, 12 and 11 manuscripts have been included for, respectively, the qualitative synthesis and the quantitative meta-analysis (Tables 1 and 2; Supplementary Material Appendix 3A, http://links.lww.com/SLA/C803). A summary of the 26 excluded manuscripts [41,42,51–60,43,61–65,44–50] is reported in Supplementary Material Appendix 3B, http://links.lww.com/SLA/C803

Study Quality and Risk of Bias

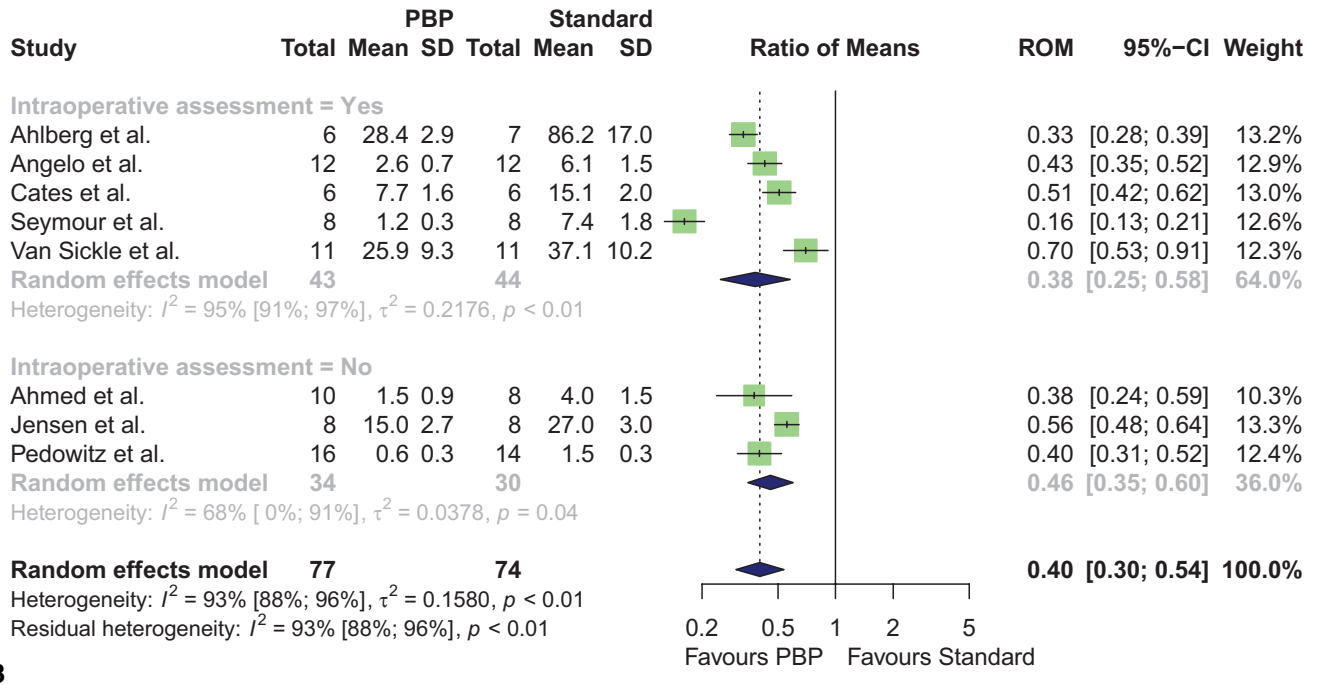
The Supplementary Material Appendix 2, http://links.lww.com/SLA/C803 summarises the quality criteria assessed for each RCT using the MERSQI tool. The overall methodological quality of the studies was high, with all the studies having low risk of bias. The overall mean score of the RCTs was 15.5 (range 14.5 and 17). The mean IRR of quality scores between assessors was 96% (range 90%–100%).

Evidence Synthesis

Tables 1 and 2 summarize general and design characteristics of the selected studies. Primary analysis included 12 papers for qualitative review and 11 studies for quantitative synthesis. The final screened manuscripts reported outcomes based on 5 full surgical procedures, 3 surgical skill tasks (ie, steps or part of a procedure, knotting and/or suturing), 3 nonsurgical medical procedures, and 1 clinical communication skill task. Overall, 12 attendings in practice (1 study), 161 residents (10 studies), and 66 medical students (2 studies) were evaluated in the included



A



B

FIGURE 2. Standardized mean difference (A) and ratio of means (B) between studies assessing the effect of proficiency-based progression versus standard training on procedural errors.

RCTs. Of these, 85 participants had been allocated to a PBP condition and n = 76 were in a non-PBP-based training pathway. According to the primary outcome (ie number of errors), 8 studies (151 participants) were included in the quantitative synthesis (ie, meta-analysis). For steps, time, and proficiency assessment on the procedure, 5 studies (86 participants), 6 studies (100 participants), and 2 studies (56 events) were included in the quantitative comparisons.

In quantitative synthesis testing for procedural errors, a pooled meta-analysis on 151 trainees was conducted (Fig. 2A,B), using random-effects models. Overall, PBP training reduced the number of errors when compared to standard training [SMD −2.93, 95% confidence intervals (CI): −3.80; −2.06; $P < 0.001$]. In a ROM analysis, PBP was estimated to reduce the mean rate of errors by approximately 60%, when compared to standard training (ROM 0.40, 95% CI: 0.30; 0.54; $P < 0.001$). Funnel plot and Egger linear

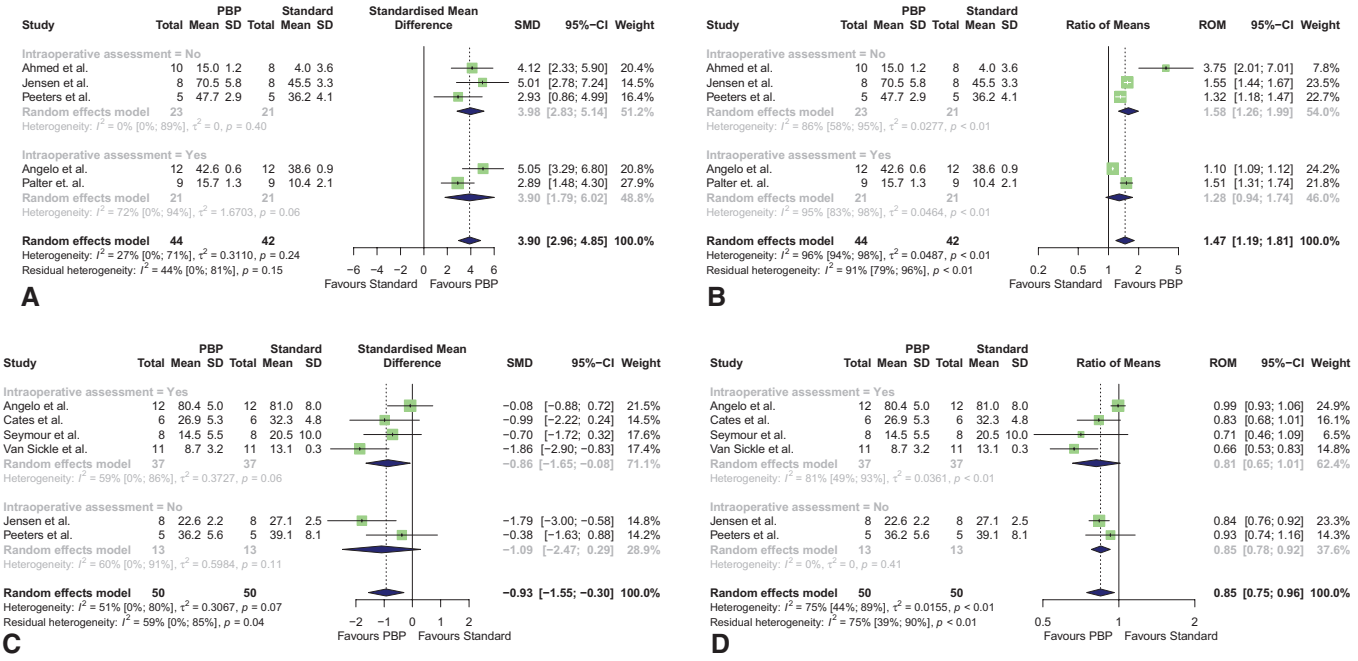


FIGURE 3. Standardized mean difference and ratio of means between studies assessing the effect of proficiency-based progression versus standard training on procedural steps (A,B) and procedural time (C,D).

regression estimates both showed evidence for potential publication bias (Supplementary Material Appendix 4A-B, http://links.lww.com/SLA/C803). In subgroup analyses, focusing on studies with intraoperative patient performance assessment (n = 87), PBP training outperformed standard training (SMD –3.11, 95% CI: –4.54; –1.68; $P < 0.001$), with an estimated reduction in mean rates of errors of 62% (ROM 0.38, 95% CI: 0.25; 0.58; $P < 0.001$).

For secondary outcomes, in quantitative synthesis testing for number of steps completed, a pooled meta-analysis on 86 trainees was conducted. Overall, trainees who completed PBP training performed more procedural steps than those who completed a standard training pathway (SMD 3.90, 95% CI: 2.96; 4.85; $P < 0.001$) (Fig. 3A). At ROM analysis, PBP increased the mean rate of steps performed by an average of 43%, when compared to standard training (ROM 1.47, 95% CI: 1.19; 1.81; $P < 0$ .001) (Fig. 3B). Funnel plot and Egger linear regression estimates recorded a marginal effect for potential publications bias (Supplementary Material Appendix 4C-D, http://links.lww.com/SLA/C803). In the two studies reporting the effect of PBP on steps performed in intraoperative patient procedure, PBP was shown to increase the number of steps performed (SMD 3.90, 95% CI: 1.79; 6.02; $P < 0.001$) but in ROM analysis such a difference failed to achieve statistical significance (ROM 1.28, 95% CI: 0.94; 1.74; $P = 0.1$).

In quantitative synthesis testing for procedural time, a pooled meta-analysis on 100 trainees was conducted. Overall, trainees who completed PBP training performed the task/procedure in less time than those who completed a standard training pathway (SMD –0.93, 95% CI: –1.55; –0.30; $P = 0.003$) (Fig. 3C). 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 0.85, 95% CI: 0.75–0.96, $P = 0.009$) (Fig. 3D). Funnel plot and Egger linear regression estimates demonstrate an absence of potential publication bias (Supplementary Material Appendix 4E-F, http://links.lww.com/SLA/C803). In subgroup analyses focusing on studies with intraoperative patient procedure assessment, PBP training slightly outperformed standard training (SMD –0.86, 95% CI: –1.65, –0.08; $P = 0.03$), with an estimated decrease in mean completion time of 19% (ROM 0.81, 95% CI 0.65; 1.01; $P = 0.06$).

Finally, in the quantitative synthesis testing for the rate of proficiency benchmark achievement on the procedure, a pooled meta-analysis on 56 trainees was conducted (Supplementary Material Appendix 5, http://links.lww.com/SLA/C803). Overall, trainees who completed PBP were more likely to reach the proficiency benchmark when compared to those who completed a standard training pathway (OR 6.92, 95% CI: 1.71; 28.02; $P < 0.001$ using a fixed-effect model). Funnel plot and Egger linear regression estimates demonstrated absence of potential publication bias (Supplementary Material Appendix 4G, http://links.lww.com/SLA/C803). Only one study reported results based on intraoperative patient procedure assessment, and it confirmed the protective effect of PBP training on achieving the final proficiency benchmark (OR 7.50, 95% CI 1.31; 43.03; $P = 0.02$ in a fixed effect model).

DISCUSSION

In this systematic review of peer-reviewed, published, prospective, randomized, and blinded clinical studies we report the meta-analysis and results from 12 studies. As measured with the MERSQI instrument the quality of the studies was high. PBP training consistently showed significant improvements in performance by trainees. Significant improvements in performance/procedure time

and procedure steps completed were observed. The largest and most consistent improvements, however, were found for error performance, particularly intra-operative errors on patients. In studies that evaluated intraoperative errors, we found a 60% reduction in comparison to the standard training group. For studies outside the operating room or clinical environment, we found a 50% reduction in errors. The aforementioned results are of particular importance if we consider the crucial impact that PBP exerted on procedural errors. Indeed, the number of steps completed by the clinician is fundamental to the completion of the procedure, and the completion of the procedure will inevitably take a certain amount of time. Both measures, however, provide little substantiation regarding the quality of performance. For example, all of the steps of a procedure may be completed, but done badly. Likewise, a procedure can be performed quickly but unsafely, or phases of the procedure can be omitted resulting in faster completion times.[16,30,32] Neither measures give a reliable indication of the quality of the operator's performance. In contrast, objectively assessed performance error in the PBP methodology gives direct, objective, transparent, and fair measures of quality.

All performance metrics in a PBP approach are developed with experienced surgeons/clinicians and cumulative validation evidence derived from them (e.g., a Delphi consensus meeting, objective assessment of performance),[66,67] following the recommendation of the American Psychological Association guidelines[68] (see Supplementary Material Appendix 6, http://links.lww.com/SLA/C803 for metric examples). The experienced clinicians reach a consensus on a safe and effective way for a trainee to learn to perform the procedure/task at the start of the learning curve (i.e., procedure steps). These are a sequence of actions that enable execution and completion of the procedure/task. Similarly, they reach consensus on performance errors (error or critical error). In laparoscopic surgery, for example, having the working end of an instrument out of view or the second instrument not assisting is not necessarily a serious slip-up but it is deemed to be (by very experienced clinicians) task execution ''that deviates from optimal performance.'' Likewise, the tip of the catheter scraping against the vessel wall, as it is being advanced in the majority of situations, leads to no serious consequences. It does, however, unnecessarily run the risk of dislodging plaque from the vessel wall which may travel up into the brain and cause a stroke. In contrast, poor catheter control and advancing into the lesion is a much more serious event. In the communication on a deteriorating patient (Supplementary Material Appendix 3A, http://links.lww.com/SLA/C803), information on an elevated white cell count should be conveyed but a much more serious error is to fail to communicate that the deteriorating patient has sepsis. These clinical situations are quite different but what the error metrics have in common is that they capture performance characteristics that are recognized by very experienced surgeons and clinicians' as suboptimal (ie, errors) or which compromise the integrity of the procedure or the safety of the patient and are thus more likely to impact on procedure quality and patient outcomes. Procedure steps and errors are units of performance that are specifically targeted in PBP training, but errors seem to be impacted more. This means that quantitative ''mathematization'' of the different steps and errors makes this methodology similarly applicable for different tasks. There was 1 study that directly assessed the impact of PBP training on a clinical outcome. Srinivasan et al[18] assessed the impact of PBP simulation training on the effectiveness and success of epidural analgesia administration during labor. They found that the PBP trained group had a 54% lower epidural failure rate than the simulation trained group.

The effectiveness of the PBP simulation training is probably accounted for by a number of factors. The first is that the performance characteristics on which training is based are derived from very experienced and practicing clinicians. They identify the characteristics and performances necessary for trainees at the start of their learning curve, and hence provide a reference approach to the successful performance of the procedure [10,33,69,70] and provide the basis for a performance metrics that can be systematically investigated with the collation of validation evidences.[24,31] Once supported with robust validation evidences, a proficiency benchmark is established based on the mean performance of experienced practitioners.[5,8,15–17,19,31,32] Another fundamental aspect of PBP training is that the detailed metrics are used to provide the trainees with objective, transparent and constructive feedback on their performance, thus affording trainees the opportunity to engage in deliberate practice training rather than repeated practice.[7] This said, PBP training is not complete until the trainee has demonstrated a level of proficiency-based on predefined benchmarks. They also have demonstrated that they can adequately undertake the task under conditions of a simulation or training model, and that they can achieve a quantitatively defined proficiency benchmark on simulations. The pretrained novice never completes the medical procedure on a live patient until they have shown that they can adequately perform the task within a training context. Evidence reviewed here suggests that PBP ensures that trainees are significantly better prepared than more traditionally trained clinicians.

There are other approaches to training skill that seem conceptually similar to PBP, for example, mastery learning (ML) and proficiency-based learning (PBL).[71] Like ML and PBL, PBP training starts with an online module and then the trainee progresses through practical tasks which may include high fidelity models such as porcine model, canine cadaver model, or human cadavers. Moreover, other fundamental concepts that are related with educational activities (ie, deliberate practice), formative testing, and advancement in performances are similarly applied to the ML and PBL methodologies. On the other hand, there seems to be residual heterogeneity in the different ML and PBL methodologies study designs reported in the literature.[71,72] For instance, ML is based on a different approach for establishing the performance goal: specifically, ML relies on the use of ''minimum passing mastery standard''[72] for each unit, whereas PBP uses a proficiency benchmark based on the objectively assessed performance of experienced surgeons as the quantitatively defined benchmark for trainees. Despite these differences, ML and PBL methodologies are based on concepts that are almost identical to those concerning PBP. Therefore, since the solidity of ML methodology was proven in previous publications,[72] the similarity between PBP and ML or PBL corroborates the efficacy of PBP and the importance of the results recorded in the current analysis.

Despite the strength of our findings, few limitations of the current systematic review need to be acknowledged. First, 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. Second, the limited number of studies included in the current review may reduce the generalizability of these findings and might increase the risk of residual biases. On the other hand, it is important to note that all the included studies were high-quality RCTs, a factor which corroborates the robustness of our findings.

CONCLUSIONS

Our systematic review and meta-analysis of RCTs confirms that PBP training improves trainees' performances when compared to high-quality simulation-based training programs. Notably, PBP decreases procedural errors by 60% compared to conventional/traditional training and such a positive impact on trainees'

performances is higher when focusing on intraoperative performance assessment. These results reinforce the need to fully implement PBP methodology in surgical and procedure-based medical treatment training pathways.

REFERENCES

1. Salas E, Bowers CA, Rhodenizer L. It is not how much you have but how you use it: toward a rational use of simulation to support aviation training. *Int J Aviat Psychol*. 1998;8:197–208.
2. Webster CS. The nuclear power industry as an alternative analogy for safety in anaesthesia and a novel approach for the conceptualisation of safety goals. *Anaesthesia*. 2005;60:1115–1122.
3. Gaba DM, DeAnda A. A comprehensive anesthesia simulation environment: re-creating the operating room for research and training. *Anesthesiology*. 1988;69:387–394.
4. Satava RM. Virtual reality surgical simulator. *Surg Endosc*. 1993;7:203–205.
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*. 2002;236:454–458.
6. Seymour NE, Gallagher AG, Roman SA, et al. Analysis of errors in laparoscopic surgical procedures. *Surg Endosc*. 2004;18:592–595.
7. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*. 2004;79(10 SUPPL):70–81.
8. 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*. 2007;193:797–804.
9. Mascheroni J, Mont L, Stockburger M, et al. International expert consensus on a scientific approach to training novice cardiac resynchronization therapy implanters using performance quality metrics. *Int J Cardiol*. 2019;289:63–69.
10. Angelo RL, Ryu RKN, Pedowitz RA, et al. Metric development for an arthroscopic bankart procedure: assessment of face and content validity. *Arthroscopy*. 2015;31:1430–1440.
11. Pedowitz RA, Nicandri GT, Angelo RL, et al. Objective assessment of knot-tying proficiency with the fundamentals of arthroscopic surgery training program workstation and knot tester. *Arthrosc - J Arthrosc Relat Surg*. 2015;31:1872–1879.
12. Ritter EM, McClusky DA 3rd, Gallagher AG, et al. Real-time objective assessment of knot quality with a portable tensiometer is superior to execution time for assessment of laparoscopic knot-tying performance. *Surg Innov*. 2005;12:233–237.
13. Angelo RL, Pedowitz RA, Ryu RKN, et al. The bankart performance metrics combined with a shoulder model simulator create a precise and accurate training tool for measuring surgeon skill. *Arthroscopy*. 2015;31:1639–1654.
14. Angelo RL, Ryu RKN, Pedowitz RA, et al. The bankart performance metrics combined with a cadaveric shoulder create a precise and accurate assessment tool for measuring surgeon skill. *Arthrosc*. 2015;31:1655–1670.
15. 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*. 2008;207:560–568.
16. Angelo RL, Ryu RKN, 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*. 2015;31:1854–1871.
17. Cates CU, Lönn L, 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 Simul Technol Enhanc Learn*. 2016;2:1–5.
18. Srinivasan KK, 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*. 2018;8:e020099.
19. Breen D, O'Brien S, McCarthy N, et al. Effect of a proficiency-based progression simulation programme on clinical communication for the deteriorating patient: a randomised controlled trial. *BMJ Open*. 2019;9:e025992.
20. Sutherland LM, Middleton PF, Anthony A, et al. Surgical simulation: a systematic review. *Ann Surg*. 2006;243:291–300.
21. Gurusamy K, Aggarwal R, Palanivelu L, et al. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *Br J Surg*. 2008;95:1088–1097.
22. Zendejas B, Brydges R, Hamstra SJ, et al. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. *Ann Surg*. 2013;257:586–593.
23. PRISMA statement. Available at: http://www.prisma-statement.org.
24. Ahmed OM, O'Donnell BD, Gallagher AG, et al. Development of performance and error metrics for ultrasound-guided axillary brachial plexus block. *Adv Med Educ Pract*. 2017;8:257–263.
25. Palter VN, Grantcharov TP. Development and validation of a comprehensive curriculum to teach an advanced minimally invasive procedure: a randomized controlled trial. *Ann Surg*. 2012;256:25–32.
26. Peeters SHP, Akkermans J, Slaghekke F, et al. Simulator training in fetoscopic laser surgery for twin-twin transfusion syndrome: a pilot randomized controlled trial. *Ultrasound Obstet Gynecol*. 2015;46:319–326.
27. Jensen UJ, Jensen J, Ahlberg G, et al. Virtual reality training in coronary angiography and its transfer effect to real-life catheterisation lab. *EuroIntervention*. 2016;11:1503–1510.
28. Reed DA, Cook DA, Beckman TJ, et al. Association between funding and quality of published medical education research. *JAMA*. 2007;298:1002–1009.
29. Kazdin AE. *Behavior Modification in Applied Settings*. 5th Ed. Belmont, CA: Thomson Brooks/Cole Publishing Co; 1994.
30. Gallagher AG. Metric-based simulation training to proficiency in medical education:- what it is and how to do it. *Ulster Med J*. 2012;81:107–113.
31. 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*. 2005;241:364–372.
32. Gallagher AG, O'Sullivan GC. *Fundamentals of Surgical Simulation: Principles and Practice*. Springer Publishing Company, Incorporated; 2011.
33. Crossley R, Liebig T, Holtmannspoetter M, et al. Validation studies of virtual reality simulation performance metrics for mechanical thrombectomy in ischemic stroke. *J Neurointerv Surg*. 2019;11:775–780.
34. Cook DA, Hatala R, Brydges R, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. 2011;306:978–988.
35. Friedrich JO, Adhikari NKJ, Beyene J. Ratio of means for analyzing continuous outcomes in meta-analysis performed as well as mean difference methods. *J Clin Epidemiol*. 2011;64:556–564.
36. Friedrich JO, Adhikari NKJ, Beyene J. The ratio of means method as an alternative to mean differences for analyzing continuous outcome variables in meta-analysis: a simulation study. *BMC Med Res Methodol*. 2008;8:32.
37. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557–560.
38. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177–188.
39. Shim SR, Kim SJ. Intervention meta-analysis: application and practice using R software. *Epidemiol Health*. 2019;41:e2019008.
40. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315:629–634.
41. Vargas MV, Moawad G, Denny K, et al. Transferability of virtual reality, simulation-based, robotic suturing skills to a live porcine model in novice surgeons: a single-blind randomized controlled trial. *J Minim Invasive Gynecol*. 2017;24:420–425.
42. Varley M, Choi R, Kuan K, et al. Prospective randomized assessment of acquisition and retention of SILS skills after simulation training. *Surg Endosc*. 2015;29:113–118.
43. Yang C, Kalinitschenko U, Helmert JR, et al. Transferability of laparoscopic skills using the virtual reality simulator. *Surg Endosc*. 2018;32:4132–4137.
44. Maertens H, Aggarwal R, Moreels N, et al. A Proficiency Based Stepwise Endovascular Curricular Training (PROSPECT) Program enhances operative performance in real life: a randomised controlled trial. *Eur J Vasc Endovasc Surg*. 2017;54:387–396.
45. Maertens H, Vermassen F, Aggarwal R, et al. Endovascular training using a simulation based curriculum is less expensive than training in the hybrid angiosuite. *Eur J Vasc Endovasc Surg*. 2018;56:583–590.
46. Martin JR, Anton N, Timsina L, et al. Performance variability during training on simulators is associated with skill transfer. *Surgery*. 2019;165:1065–1068.
47. Snyder CW, Vandromme MJ, Tyra SL, et al. Proficiency-based laparoscopic and endoscopic training with virtual reality simulators: a comparison of proctored and independent approaches. *J Surg Educ*. 2009;66:201–207.
48. Snyder CW, Vandromme MJ, Tyra SL, et al. Effects of virtual reality simulator training method and observational learning on surgical performance. *World J Surg*. 2011;35:245–252.
49. Stefanidis D, Scerbo MW, Montero PN, et al. Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial. *Ann Surg*. 2012;255:30–37.

50. Stoller J, Joseph J, Parodi N, et al. Are there detrimental effects from proficiency-based training in fundamentals of laparoscopic surgery among novices? An exploration of goal theory. *J Surg Educ*. 2016;73:215–221.
51. Ahlborg L, Hedman L, Nisell H, et al. Simulator training and non-technical factors improve laparoscopic performance among OBGYN trainees. *Acta Obstet Gynecol Scand*. 2013;92:1194–1201.
52. Bjerrum F, Sorensen JL, Konge L, et al. Randomized trial to examine procedure-to-procedure transfer in laparoscopic simulator training. *Br J Surg*. 2016;103:44–50.
53. Bjerrum F, Maagaard M, Led Sorensen J, et al. Effect of instructor feedback on skills retention after laparoscopic simulator training: follow-up of a randomized trial. *J Surg Educ*. 2015;72:53–60.
54. Brydges R, Carnahan H, Rose D, et al. Comparing self-guided learning and educator-guided learning formats for simulation-based clinical training. *J Adv Nurs*. 2010;66:1832–1844.
55. Dawidek MT, Roach VA, Ott MC, et al. Changing the learning curve in novice laparoscopists: incorporating direct visualization into the simulation training program. *J Surg Educ*. 2017;74:30–36.
56. De Win G, Van Bruwaene S, Kulkarni J, et al. An evidence-based laparoscopic simulation curriculum shortens the clinical learning curve and reduces surgical adverse events. *Adv Med Educ Pract*. 2016;7:357–370.
57. Franklin BR, Placek SB, Wagner MD, et al. Cost comparison of fundamentals of laparoscopic surgery training completed with standard fundamentals of laparoscopic surgery equipment versus low-cost equipment. *J Surg Educ*. 2017;74:459–465.
58. Gala R, Orejuela F, Gerten K, et al. Effect of validated skills simulation on operating room performance in obstetrics and gynecology residents: a randomized controlled trial. *Obstet Gynecol*. 2013;121:578–584.
59. Gauger PG, Hauge LS, Andreatta PB, et al. Laparoscopic simulation training with proficiency targets improves practice and performance of novice surgeons. *Am J Surg*. 2010;199:72–80.
60. Gershuni V, Woodhouse J, Brunt LM. Retention of suturing and knot-tying skills in senior medical students after proficiency-based training: Results of a prospective, randomized trial. *Surgery.* 2013;154:823–830.
61. Grierson LEM, Lyons JL, Dubrowski A. Gaze-down endoscopic practise leads to better novice performance on gaze-up displays. *Med Educ*. 2013;47:166–172.
62. Hashimoto DA, Petrusa E, Phitayakorn R, et al. A proficiency-based virtual reality endoscopy curriculum improves performance on the fundamentals of endoscopic surgery examination. *Surg Endosc*. 2018;32:1397–1404.
63. Hseino H, Nugent E, Lee MJ, et al. Skills transfer after proficiency-based simulation training in superficial femoral artery angioplasty. *Simul Healthc*. 2012;7:274–281.
64. Kiely DJ, Gotlieb WH, Lau S, et al. Virtual reality robotic surgery simulation curriculum to teach robotic suturing: a randomized controlled trial. *J Robot Surg*. 2015;9:179–186.
65. Lemke M, Lia H, Gabinet-Equihua A, et al. Optimizing resource utilization during proficiency-based training of suturing skills in medical students: a randomized controlled trial of faculty-led, peer tutor-led, and holography-augmented methods of teaching. *Surg Endosc*. 2020;34:1678–1687.
66. Puliatti S, Mazzone E, Dell'Oglio P. Training in robot-assisted surgery. *Curr Opin Urol*. 2020;30:65–72.
67. Puliatti S, Mazzone E, Amato M, et al. Development and validation of the objective assessment of robotic suturing and knot tying skills for chicken anastomotic model. *Surg Endosc*. 2020. doi:10.1007/s00464-020-07918-5.
68. Tippins N, Sackett P, Oswald F. Principles for the validation and use of personnel selection procedures. *Ind Organ Psychol*. 2018;11(S1):1–97.
69. Mascheroni J, Mont L, Stockburger M, et al. A validation study of intraoperative performance metrics for training novice cardiac resynchronization therapy implanters. *Int J Cardiol*. 2020;307:48–54.
70. Kojima K, Graves M, Taha W, et al. AO international consensus panel for metrics on a closed reduction and fixation of a 31A2 pertrochanteric fracture. *Injury.* 2018;49:2227–2233.
71. McGaghie WC, Issenberg SB, Barsuk JH, et al. A critical review of simulation-based mastery learning with translational outcomes. *Med Educ*. 2014;48:375–385.
72. Zendejas B, Cook DA, Bingener J, et al. Simulation-based mastery learning improves patient outcomes in laparoscopic inguinal hernia repair: a randomized controlled trial. *Ann Surg*. 2011;254:502–511.