Accepted Manuscript

Virtual reality simulator versus box-trainer to teach minimally invasive procedures: a meta-analysis

Hugo Gonçalo Guedes, Zêmia Maria Câmara Costa Ferreira, Layra Ribeiro de Sousa Leão, Edna Frasson Souza Montero, José Pinhata Otoch, Everson Luiz de Almeida Artifon

PII: S1743-9191(18)31722-9

DOI: https://doi.org/10.1016/j.ijsu.2018.12.001

Reference: IJSU 4811

To appear in: *International Journal of Surgery*

Received Date: 1 September 2018

Revised Date: 23 November 2018

Accepted Date: 4 December 2018

Please cite this article as: Guedes HG, Câmara Costa Ferreira ZM, Ribeiro de Sousa Leão L, Souza Montero EF, Otoch JP, Luiz de Almeida Artifon E, Virtual reality simulator versus box-trainer to teach minimally invasive procedures: a meta-analysis, *International Journal of Surgery*, https://doi.org/10.1016/j.ijsu.2018.12.001.

Journal: International Journal of Surgery

Article Type: Meta-analysis.

Virtual reality simulator versus box-trainer to teach minimally invasive procedures: a meta-analysis.

Which better surgery model to teach laparoscopy?

Hugo Gonçalo Guedes[1], Zêmia Maria Câmara Costa Ferreira[2], Layra Ribeiro de Sousa Leão[3], Edna Frasson Souza Montero[1], José Pinhata Otoch[1], Everson Luiz de Almeida Artifon[1].

1. General Surgery Department, University of Sao Paulo School of Medicine, Carvalho Aguiar Street, number 255, São Paulo - SP 05422-090, Brazil.
2. Potiguar University, Senador Salgado Filho Avenue, number 1610, Lagoa Nova, Natal - RN, 59056-000, Brazil.
3. Hospital Israelita Albert Einstein, Albert Einstein Avenue, number 627/701, Morumbi, São Paulo - SP, 05652-900, Brazil.

Author contributions: Guedes HG, and Artifon ELA performed the systematic review and meta-analysis. Guedes HG, Ferreira ZMCC and Leão LRS wrote the article. Frasson E, Otoch JP and Artifon ELA revised the manuscript. The final version of the manuscript was approved by all authors.

Corresponding author and reprint address: Hugo Gonçalo Guedes, PhD, General Surgery Department, University of Sao Paulo School of Medicine, Carvalho Aguiar street, number 255, São Paulo - SP 05422-090, Brazil. hugogguedes@yahoo.com.br. **Telephone:** +5511998657143.

PRISMA 2009 Checklist statement: The authors have read the PRISMA 2009 Checklist, and the manuscript was prepared and revised according to the PRISMA 2009 Checklist.

Declarations of interest: none.

Financial support: The authors declare that they have no financial support.

Journal: International Journal of Surgery

Article Type: Meta-analysis.

Virtual reality simulator versus box-trainer to teach minimally invasive procedures: a meta-analysis.

ABSTRACT:

Background: To evaluate the effectiveness of virtual reality simulator (VRS) training compared to box-trainer training (BT) for learning outcomes in minimally invasive surgery (MIS) techniques. **Materials and Methods:** A systematic review of the literature was performed using CENTRAL, MEDLINE, EMBASE, Scopus, CINAHL, LILACS. The primary outcomes were time to perform MIS and performance score in MIS. After being selected, the articles were evaluated for methodological quality and risk of bias. The results were evaluated for quality of evidence and meta-analysis was performed. **Results:** 20 randomized clinical trials were included in the qualitative analysis and 14 were used in the meta-analysis. VRS training was more efficient than BT training ($P < 0.00001$, 95% CI: -35.08 to -25.01) when evaluating participant time needed to complete the peg task. In descriptive analysis, VRS training was better than BT training in participant performance score to perform MIS. There was no statistical difference in the meta-analysis in the time needed to perform surgery, time to complete basic or advanced tasks and performance score for basic or advanced tasks. **Conclusions:** VRS training was better than BT training in participant performance scores when performing MIS and in the time needed to complete the basic task of peg transfer. In all other outcomes, regardless of the student's level of experience or type of activity, the two forms of training were equivalent.

Key words: Simulation Training; Laparoscopy; Minimally Invasive Surgical Procedures; Training; Box trainer; Virtual Reality.

PROSPERO Registry: CRD42016046840.

1. INTRODUCTION:

Surgery is classically based on practical training of a surgical trainee under supervision and guidance of an experienced surgeon, thus minimizing the risk of iatrogenic outcomes. Before any intervention on the patient, a resident or a medical student can be trained using a reality simulator, giving them increased surgical expertise without submitted a patient to novices trainment risks(1).

The use of a virtual reality simulator (VRS) is interesting, since it allows training of interventionist skills. Nagendran M *et al*, in a Cochrane meta-analysis, showed that virtual reality training decreased operating time and improved operative performance when compared to box-trainer (BT) training or no supplementary training at all. However, the impact of these results on patients' health and healthcare funders in better outcomes or reduced costs is still unknown(2).

The introduction of virtual platforms and videos seems to be related to equivalent learning rates to traditional methods, diminished the need of supervision by a trained professional, diversified learning, lowered costs and increased demand for trainees to perform complex surgical techniques(3, 4).

The objective of this study is to evaluate the effectiveness of VRS compared to BT training for education of trainees in minimally invasive surgery (MIS).

2. MATERIAL AND METHODS:

2.1.Protocol and registry:

This systematic review of the literature was performed according to the methodology established by the *Cochrane Handbook for Systematic Reviews of Interventions*, by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)(5) and by Assessing the Methodological Quality of Systematic Reviews (AMSTAR). The study was registered in the International Prospective Register of Systematic Reviews (PROSPERO) under the code CDR42016046840.

2.2.Eligibility criteria:

Only randomized clinical trials (RCT) that compared VRS and BT on the learning of minimally invasive surgical techniques (MIST) were included in this study. All other learning variables, such as conventional teaching classes and curriculum, were controlled and similar in the 2 groups. Participants included medical students, doctors, resident doctors of surgical specialties and surgeons, independent of their experience with MIST. RCTs comparing any type of VRS versus any type of BT, regardless the form of training, were included.

2.3.Search methods for identification of studies:

The databases used for study identification were: Cochrane, MEDLINE, EMBASE, CINAHL, LILACS and SCOPUS. The search strategy included key terms: “education”, “virtual reality training”, “simulation”, “instruction”, “laparoscopy”, “minimally invasive surgical procedures”, “endoscopy” and “random”. For example, the MEDLINE search strategy was attached in the paper supplementary material.

Others complementary studies were researched in the grey literature research process. The lists of abstracts published at national and international surgery congresses in the last 10 years have been verified, by hand searching on paper and internet, including: Brazilian Digestive Disease Week, Digestive Disease Week - DDW, Sao Paulo University thesis library, to find another randomized controlled trials by the electronic searches, which included the terms: surgical training; simulators; minimally invasive surgery; trainment; laparoscopy; medical education.

2.4.Study selection and data collection:

Studies that complied with the eligibility criteria were subjected to full-text analysis. The opinion of a third, independent researcher was requested in case of disagreement about the inclusion of any study. Data gathering was performed independently by the two authors and computed in standard forms specific for collection. Data about the identification of the study, eligibility criteria, methodological aspects, participants, interventions, comparisons, outcomes and relevant results were included in this study.

2.5.Assessment of methodological quality and risk of bias in included studies:

The risk of bias in each selected study for this systematic review was assessed based on the recommendations of the Cochrane Collaboration using a tool called *Risk of Bias* (RoB tool) as described in chapter 8.5 of the *Cochrane Handbook for Systematic Reviews of Interventions(6).*

2.6.Data analysis:

The results were collected and summarized in forest plot graphics, generated by Review Manager 5.3 (Copenhagen: The Nordic Cochrane Center, The Cochrane Collaboration, 2011), if the studies were homogeneous and the data was available. For those outcomes with no possible quantitative analysis, the results were presented as a narrative.

2.7.Measures of treatment effect:

For continuous outcomes, the Mean Difference (MD) was calculated with a 95% Confidence Interval (CI). The Standardized Mean Difference (SMD) was used with a 95% CI in some selected outcomes where different scales may have been used, such as operative performance.

2.8.Assessment of heterogeneity:

The statistical heterogeneity was explored by using a Chi square test (Chi^2). Quantity of heterogeneity was measured by the I^2 as recommended by Higgins *et al(6)*, in which I^2 values over 50% are considered as evidence with significant heterogeneity between the studies. The analysis was performed using models with random effects. Clinical, statistical or methodological bias would be used to explain the possible reasons for heterogeneity.

2.9.Assessment of the quality of evidence:

The quality of the evidence in this systematic review was analyzed according to Grading of Recommendations Assessment, Development and Evaluation Working Group (GRADE)(6, 7).

3. RESULTS:

The results of the search are summarized in the study flow diagram below (**Figure 1**).

3.1.Included studies:

20 RCTs were included, involving 695 participants, of which 350 were for VRS and 345 for BT. The sample size ranged from 16(8-10) to 84(11) participants per study, while the number of participants per intervention group ranged from 7(8) to 42[9 (11)]. The details of the trials are shown in **table 1**.

Table 1. Descriptive data from the selected studies.

Author, year	Country	VRS	BT	Experience	Kind of Tasks	Performance score	VRS model	BT model	Training
Akdemir A, 2014	Turkey	20	20	Participants with no experience	Basic	OSA-LS	LapSim	Not specified	Theoretical class and 1h training per week for 4 weeks.
Borahay MA, 2013	Germany	8	8	Participants with no experience	Basic	-	dV-Trainer	TRLCD05 Laparoscopic MIT	4 tasks (1h each).
Brinkmann C, 2017	Germany	18	18	Participants with no experience	Advanced	GOALS	Lap Mentor II	Not specified	5 days of training, No further details.
Debes AJ, 2010	Norway	20	18	Participants with no experience	Basic	-	MIST-VR	D-Box	5 tasks drawn out of a total of 8 (basic tasks). 8 training sessions.
Hamilton EC, 2002	USA	24	25	Participants with no experience	Basic	GOALS	MIST-VR	SCMIS GEM	2 weeks. 10 sessions with 30 minutes each.
Hassan SO, 2015	USA	14	16	Participants with no experience	Basic	-	dVSSS	Not specified	Training with 2 tasks. Total of 5 sessions.
Hiemstra E, 2011	Netherlands	20	20	Participants with no experience	Both	-	SIMENDO VR Trainer	Box trainer	Only one 20-minute session.
Jensen K, 2014	Denmark	14	14	Participants with no experience	Advanced	-	SEP	D-Box Basic Simulator	No details about time and number of repetitions.
Kanumuri P, 2008	USA	8	8	Participants with no experience	Basic	-	MIST-VR	ProMIS Simulator	8 sessions in 4 weeks.
Kothari SN, 2002	USA	13	11	Novices	Basic	-	MIST-VR	Not specified	5 sessions in 5 days.
Lehmann KS, 2005	Germany	16	16	Novices and participants	Basic	-	VEST	Not specified	2-task training for 4 days.

				with no experience					
Loukas C, 2012	Greece	22	22	Participants with no experience	Basic	-	LapVR	Not specified	3 tasks repeated in 12 sessions.
Madan AK, 2007	USA	17	14	Participants with no experience	Advanced	-	MIST-VR	LTS 2000	10 sessions, 20 minutes each.
Nickel F, 2015	Germany	42	42	Participants with no experience	Advanced	OSATS	Lap Mentor II	Not specified	9 tasks in 12 hours of training.
Orlando MS, 2017	USA	20	20	Participants with no experience	Basic	Other	dV-Trainer	Box trainer	10 sessions of peg transfer tasks.
Pearson AM, 2002	USA	10	8	Participants with no experience	Advanced	-	MIST-VR	Not specified	10 sessions for each task.
Tanoue K, 2008	Japan	20	20	Participants with no experience	Basic	-	Procedicus MIST	Not specified	2-hour sessions for 2 days.
Thomaier L, 2016	USA	20	20	Participants with no experience	Basic	Other	dV-Trainer	Not specified	10 sessions of peg transfer tasks.
Yiasemidou M, 2017	United Kingdom	7	9	Participants with no experience	Advanced	-	Lap Mentor	Box trainer	At least 25 sessions in 6 weeks.
Youngblood PL, 2005	USA	17	16	Participants with no experience	Both	GOALS	LapSim	Tower Trainer	3 tasks, 10 sessions for each, in 12 days.

3.2.Risk of bias in included studies

Details about the risk of bias in the included trials are summarized in **Figure 2**. We considered studies as having a high risk of bias if more than 20% of patients were lost to follow up leading to incomplete outcomes, and if the study had financial support by the simulator's group.

3.3.Effects of intervention: VRS training versus BT training

It was not possible to pool the data of six studies(10, 12-16) in meta-analysis because the data needed (mean and standard deviation) was not available.

Time to complete (TTC) a MIS were referenced to one paper that compared a laparoscopic cholecystectomy and another study compared a laparoscopic oophorectomy. There was no significant difference between VRS and BT in terms of

TTC a MIS after training was performed (**Figure 3**). Additionally, the operative performance was found to be significantly better in the VRS group than in the BT group when performing a MIS. However, this is a descriptive analysis, since only one study reported the magnitude of the difference in a laparoscopic cholecystectomy(17).

The participants in the VRS group were able to perform the basic task peg transfer in a shorter time than those in the BT group. This difference was statistically significant (**Figure 3**). There was no statistically significant difference between VRS and BT training however in the TTC a ligation loop (basic Task) (**Figure 4**). Similarly, there was no statistically significant difference between VRS training and BT in the TTC a series of basic tasks (**Figure 4**). Last, there was no statistically significant difference between VRS training and BT training in regards the time needed to perform a set of advanced tasks (dieresis and synthesis) after the training session was performed (**Figure 4**).

There was found to be no statistically significant difference between VRS training and BT training regarding the participant performance score for the execution of basic tasks after the training session was performed (**Figure 5**). Similarly, there was no statistically significant difference between VRS training and BT training regarding the performance score for the execution of advanced tasks (**Figure 5**).

The assessment of the quality of evidence of the main results is summarized in **table 2** and is generated based on the GRADE classification. Three results had moderate quality (TTC a basic task peg transfer and ligation loop and TTC all basics tasks), while the others results had low quality. The BT risk column expresses an absolute value (in seconds, for example). In the other hand, the VRS risk column expresses a relative risk result with CI. The final quality of evidence was explained by reasons expressed in the legends (Heterogeneity > 50%, Small sample or randomization / allocation are not clear and / or wide CI).

Table 2. Classification for quality of the evidence of the results (GRADE classification).

	Potential absolute effects* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	BT risk	**VRS risk**				

	Potential absolute effects* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	BT risk	**VRS risk**				
TTC a MIS	The mean time to perform a MIS was **1596** seconds	The mean time to complete a MIS in the intervention group was 148.83 seconds higher (-203.54 to 501.2)	-	68 (2 RCTs)	⊕⊕◯◯ LOW [a,b,c]	-
TTC a basic task: peg transfer	The mean time to perform a peg transfer (basic task) was **104.07** seconds	The mean time to complete a peg transfer (basic task) in the intervention group was 35.08 seconds lower (45.15 to 25.01 lower)	-	203 (6 RCTs)	⊕⊕⊕◯ MODERATE [b]	-
TTC a ligation loop task (basic task)	The mean time to perform a ligation loop task was **311** seconds	The mean time to complete a ligation loop task in the intervention group was 2.58 seconds higher (-7.27 to 12.43)	-	60 (2 RCTs)	⊕⊕⊕◯ MODERATE [c]	-
TTC laparoscopic camera navigation (basic task)	The mean time to perform laparoscopic camera navigation (basic task) was **23.25** seconds	The mean time to complete laparoscopic camera navigation in the intervention group was 81.25 lower (-271.45 to 109.94)	-	44 (2 RCTs)	⊕⊕◯◯ LOW [a,b,c]	-
TTC all the basic tasks		SMD with **2.67** SD higher (-0.36 to 5.69)	-	105 (3 RCTs)	⊕⊕⊕◯ MODERATE [b,c]	-
Performance score when performing advanced tasks	-	SMD with **0.31 SD** lower (0.96 to 0.34 lower)	-	120 (2 RCTs)	⊕⊕◯◯ LOW [a,b]	-

	Potential absolute effects* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	BT risk	**VRS risk**				
Performance score when performing basic tasks	-	SMD **1.97** higher (-0.39 to 3.56)	-	202 (5 RCTs)	⊕⊕◯◯ LOW [a,b]	-

***The risk in the intervention group** (95% CI) is based on the assumed risk in the control group and the **relative effect** of the intervention (95% CI).

GRADE Working Group grades of evidence
High quality evidence: further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality of evidence: further research is unlikely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality of evidence: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality of evidence: any estimate of effect is very uncertain.

CI = confidence interval; SMD = standardized mean difference; RCT = randomized clinical trial; VRS = virtual reality simulator; BT = box-trainer; SD = standard deviation; TTC = Time to complete

[a] Heterogeneity > 50%; [b] Small sample or randomization / allocation are not clear; [c] wide CI.

4. DISCUSSION:

There is no systematic review and meta-analysis comparing VRS and BT in terms of learning outcomes. In 2013, Nagendran *et al(2)* published a systematic review and meta-analysis that assessed the effectiveness of VRS on the learning results of surgical trainees with limited laparoscopic experience and, in a subgroup, they compared VRS *versus* BT in terms of performance score. The data was not used in the meta-analysis because the heterogeneity could not be assessed, since only one trial reported the magnitude of the difference. However, the results follow the same conclusion of this meta-analysis, where the VRS group reached a better performance score than BT group. In another study, Reznick *et al(18)* submitted groups previously trained in VRS and BT to performance evaluations in anesthetized animal models and demonstrated that the VRS group was more effective than the BT group.

Contrary to the previously mentioned studies, Zendejas *et al(19)* published a systematic review and meta-analysis in which they stated that BT training is superior to the VRS training.

However, the operating time, and proportionally, the costs involved, increase between novice surgeons and experienced surgeons(20). Thus, using surgical training with simulators when surgeons reach a good level of performance and develop familiarity with MIST is no longer just a learning concern, but it is also relevant to the public health and a benefit to patients and healthcare systems.

The peg transfer is, among the tasks proposed by the *Fundamentals of Laparoscopic Surgery* (FLS), the simplest and purposely the first task to be trained. Probably for this reason, the favorable result of the virtual reality simulation did not affect the final outcomes when considering all the basic skills trained.

The results regarding the laparoscopic camera navigation are in accordance with the current literature which states that this ability can be acquired with any type of simulator(21-23).

Based on the methods of the selected studies and literature research, we can suggest that the ideal educational delivery of the training interventions using the simulated setting includes a minimum of 5 sessions for each task with a minimal time of 20 minutes, independent of simulation model. In each session, the novices perform training with basics tasks and can include advanced tasks.

This study has some limitations. The different types of simulation models, and the low number of participants per study can lead to inaccurate results, the bias of having only the published data (publication bias) and the fact that the reviewers are not blinded are important examples of limitations and biases. Despite the many different simulators available, the research was focused on the ability to learn basic or advanced tasks, based on a standard teaching system (FLS), which can be perfectly reproducible in any of the devices. For example, the basics tasks are parameters to dexterity learning. In fact, they are learned in a few simulators sessions, BT or VRS, and then performance plateaus off once the activity is mastered. The objective with regard to all simulation is its application to reality. Another limitation is that we could not find studies that proposed the gold standard parameters, like a uniform and complex proficiency score system. using performance speed as a measure of quality, perpetuates the myth that fast is good.

Regarding the quality of evidence, this systematic review and meta-analysis is the best evidence currently available, offering a substantial improvement in terms of conclusions, since the last Cochrane's systematic review about the theme had only a very low quality of evidence(2, 13, 14, 18-25, 32-34). The limitations that prevented a high quality of evidence, besides heterogeneity inherent to some results, were the small sample of the studies and the wide confidence intervals of the results. These numbers should generate a reflection about professional qualifications and a research interest in the basic areas of the surgical education, either with medical students or with surgical resident doctors from the first years.

In the attempt to reduce the publication bias in all the RCTs selected, this study used the protocol from clinicaltrials.gov with the aim of identifying the objectives and methods previously proposed by the RCTs. Since both simulators are capable to provide adequate training and because of no possibility of blinding, the low risk of bias was adopted.

5. CONCLUSION:

This systematic review and meta-analysis seeks to clarify an important gap in the medical education field. VRS training was better than BT training in terms of participant performance score when performing a MIS and time needed to complete the basic task of peg transfer. In all other outcomes, regardless of the student's level of experience or type of activity, the two forms of training are equivalent.

6. FUNDING:

The authors declare that they have no financial support.

Provenance and peer review

Not commissioned, externally peer-reviewed

7. REFERENCES:

1. Moorthy K, Munz Y, Jiwanji M, Bann S, Chang A, Darzi A. Validity and reliability of a virtual reality upper gastrointestinal simulator and cross validation using structured assessment of individual performance with video playback. Surg Endosc. 2004;18(2):328-33.
2. Nagendran M, Gurusamy KS, Aggarwal R, Loizidou M, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev. 2013;8:CD006575.
3. Eversbusch A, Grantcharov TP. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. Surg Endosc. 2004;18(10):1514-8.
4. Nousiainen M, Brydges R, Backstein D, Dubrowski A. Comparison of expert instruction and computer-based video training in teaching fundamental surgical skills to medical students. Surgery. 2008;143(4):539-44.
5. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009;62(10):1006-12.
6. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions.*: The Cochrane Collaboration.; 2011. Available from: www.handbook.cochrane.org.
7. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Going from evidence to recommendations. BMJ. 2008;336(7652):1049-51.
8. Yiasemidou M, Siqueira J, Tomlinson J, Glassman D, Stock S, Gough M. "Take-home" box trainers are an effective alternative to virtual reality simulators. Journal of surgical research [Internet]. 2017; 213:[69-74 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/667/CN-01335667/frame.html.
9. Kanumuri P, Ganai S, Wohaibi EM, Bush RW, Grow DR, Seymour NE. Virtual reality and computer-enhanced training devices equally improve laparoscopic surgical skill in novices. JSLS. 2008;12(3):219-26.
10. Borahay M, Haver M, Eastham B, Patel P, Kilic G. Modular comparison of laparoscopic and robotic simulation platforms in residency training: a randomized trial. Journal of minimally invasive gynecology [Internet]. 2013; 20(6):[871-9 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/094/CN-00909094/frame.html.
11. Nickel F, Brzoska JA, Gondan M, Rangnick HM, Chu J, Kenngott HG, et al. Virtual reality training versus blended learning of laparoscopic cholecystectomy: a randomized controlled trial with laparoscopic novices. Medicine (Baltimore). 2015;94(20):e764.
12. Debes AJ, Aggarwal R, Balasundaram I, Jacobsen MB. A tale of two trainers: virtual reality versus a video trainer for acquisition of basic laparoscopic skills. Am J Surg. 2010;199(6):840-5.
13. Hiemstra E, Terveer EM, Chmarra MK, Dankelman J, Jansen FW. Virtual reality in laparoscopic skills training: is haptic feedback replaceable? Minim Invasive Ther Allied Technol. 2011;20(3):179-84.
14. Loukas C, Nikiteas N, Schizas D, Lahanas V, Georgiou E. A head-to-head comparison between virtual reality and physical reality simulation training for basic skills acquisition. Surg Endosc. 2012;26(9):2550-8.
15. Pearson AM, Gallagher AG, Rosser JC, Satava RM. Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying. Surg Endosc. 2002;16(1):130-7.
16. Yiasemidou M, de Siqueira J, Tomlinson J, Glassman D, Stock S, Gough M. "Take-home" box trainers are an effective alternative to virtual reality simulators. J Surg Res. 2017;213:69-74.

17. Hamilton EC, Scott DJ, Fleming JB, Rege RV, Laycock R, Bergen PC, et al. Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. Surg Endosc. 2002;16(3):406-11.
18. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. Am J Surg. 1997;173(3):226-30.
19. Zendejas B, Brydges R, Hamstra SJ, Cook DA. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. Ann Surg. 2013;257(4):586-93.
20. Wilkiemeyer M, Pappas TN, Giobbie-Hurder A, Itani KM, Jonasson O, Neumayer LA. Does resident post graduate year influence the outcomes of inguinal hernia repair? Ann Surg. 2005;241(6):879-82; discussion 82-4.
21. Brinkmann C, Fritz M, Pankratius U, Bahde R, Neumann P, Schlueter S, et al. Box- or Virtual-Reality Trainer: Which Tool Results in Better Transfer of Laparoscopic Basic Skills?-A Prospective Randomized Trial. J Surg Educ. 2017.
22. Stefanidis D, Haluck R, Pham T, Dunne JB, Reinke T, Markley S, et al. Construct and face validity and task workload for laparoscopic camera navigation: virtual reality versus videotrainer systems at the SAGES Learning Center. Surg Endosc. 2007;21(7):1158-64.
23. Diesen D, Erhunmwunsee L, Bennett K, Ben-David K, Yurcisin B, Ceppa E, et al. Effectiveness of laparoscopic computer simulator versus usage of box trainer for endoscopic surgery training of novices. Journal of surgical education [Internet]. 2011; 68(4):[282-9 pp.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/061/CN-00811061/frame.html.
24. Hassan SO, Dudhia J, Syed LH, Patel K, Farshidpour M, Cunningham SC, et al. Conventional Laparoscopic vs Robotic Training: Which is Better for Naive Users? A Randomized Prospective Crossover Study. J Surg Educ. 2015;72(4):592-9.
25. Maniar HS, Council ML, Prasad SM, Chu C, Damiano RJ. Comparison of skill training with robotic systems and traditional endoscopy: implications on training and adoption. J Surg Res. 2005;125(1):23-9.
26. Madan AK, Frantzides CT. Prospective randomized controlled trial of laparoscopic trainers for basic laparoscopic skills acquisition. Surg Endosc. 2007;21(2):209-13.
27. Youngblood PL, Srivastava S, Curet M, Heinrichs WL, Dev P, Wren SM. Comparison of training on two laparoscopic simulators and assessment of skills transfer to surgical performance. J Am Coll Surg. 2005;200(4):546-51.
28. Orlando MS, Thomaier L, Abernethy MG, Chen CC. Retention of laparoscopic and robotic skills among medical students: a randomized controlled trial. Surg Endosc. 2017.
29. Thomaier L, Orlando M, Abernethy M, Paka C, Chen CC. Laparoscopic and robotic skills are transferable in a simulation setting: a randomized controlled trial. Surg Endosc. 2016.
30. Tanoue K, Ieiri S, Konishi K, Yasunaga T, Okazaki K, Yamaguchi S, et al. Effectiveness of endoscopic surgery training for medical students using a virtual reality simulator versus a box trainer: a randomized controlled trial. Surg Endosc. 2008;22(4):985-90.
31. Akdemir A, Zeybek B, Ergenoglu AM, Yeniel AO, Sendag F. Effect of spaced training with a box trainer on the acquisition and retention of basic laparoscopic skills. Int J Gynaecol Obstet. 2014;127(3):309-13.
32. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev. 2009(1):CD006575.
33. Gurusamy K, Nagendran M, Toon C, Davidson B. Laparoscopic surgical box model training for surgical trainees with limited prior laparoscopic experience. Hpb [Internet]. 2014; 16:[386 p.]. Available from: http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/113/CN-01011113/frame.html.
34. Matzke J, Ziegler C, Martin K, Crawford S, Sutton E. Usefulness of virtual reality in assessment of medical student laparoscopic skill. J Surg Res. 2017;211:191-5.

8. FIGURE LEGENDS:

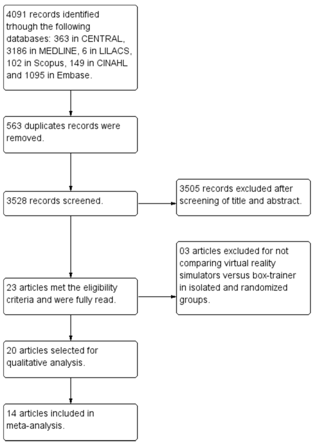
Figure 1. Literature research flow diagram.

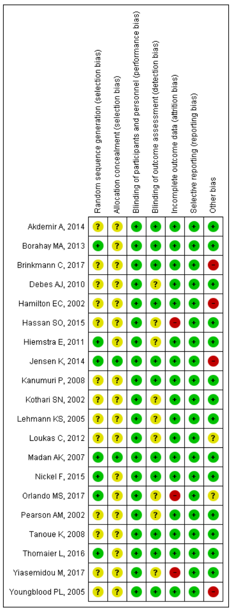
Figure 2. Representation of risk of bias: review authors' judgments about each risk of bias item presented as percentages across all included studies.

Figure 3. *Forests plots* comparing VRS training versus BT training in regards to the time required to perform a MIS after training and the time to execute a basic task (peg transfer). MD = mean difference; SD = standard deviation; 95% CI = 95% confidence interval; IV = inverse variation; Chi^2 = chi-square test; df = degrees of freedom; I^2 = statistical test for heterogeneity.

Figure 4. *Forests plots* comparing VRS training *versus* BT training in regards to the time needed to execute a ligation loop, time to perform all the basic tasks after the training session and time to perform an advanced task (run the tissue) after the training sessions.

Figure 5. *Forests plots* comparing the VRS training *versus* BT training in terms of the performance score for basic tasks and advanced tasks.

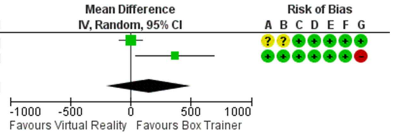

4091 records identified trhough the following databases: 363 in CENTRAL, 3186 in MEDLINE, 6 in LILACS, 102 in Scopus, 149 in CINAHL and 1095 in Embase.
563 duplicates records were removed.
3528 records screened.
3505 records excluded after screening of title and abstract.
03 articles excluded for not comparing virtual reality simulators versus box-trainer in isolated and randomized groups.
23 articles met the eligibility criteria and were fully read.
20 articles selected for qualitative analysis.
14 articles included in meta-analysis.


Random sequence generation (selection bias)
Allocation concealment (selection bias)
Blinding of participants and personnel (performance bias)
Blinding of outcome assessment (detection bias)
Incomplete outcome data (attrition bias)
Selective reporting (reporting bias)
Other bias
Akdemir A, 2014
Borahay MA, 2013
Brinkmann C, 2017
Debes AJ, 2010
Hamilton EC, 2002
Hassan SO, 2015
Hiemstra E, 2011
Jensen K, 2014
Kanumuri P, 2008
Kothari SN, 2002
Lehmann KS, 2005
Loukas C, 2012
Madan AK, 2007
Nickel F, 2015
Orlando MS, 2017
Pearson AM, 2002
Tanoue K, 2008
Thomaier L, 2016
Yiasemidou M, 2017
Youngblood PL, 2005

VRS *vs.* BT: time to perform a MIS

Study or Subgroup	Virtual Reality Mean	SD	Total	Box Trainer Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Akdemir A, 2014	340	170.93	20	340	149.56	20	59.3%	0.00 [-99.54, 99.54]
Jensen K, 2014	1,962	474	14	1,596	402	14	40.7%	366.00 [40.44, 691.56]
Total (95% CI)			**34**			**34**	**100.0%**	**148.83 [-203.54, 501.20]**

Heterogeneity: Tau² = 51892.65; Chi² = 4.44, df = 1 (P = 0.04); I² = 77%
Test for overall effect: Z = 0.83 (P = 0.41)

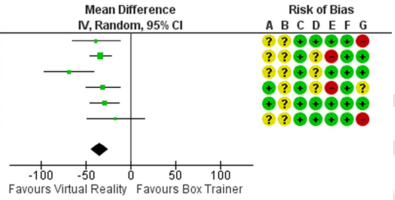


Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Selective reporting (reporting bias)
(**G**) Other bias

VRS *vs.* BT: TTC task peg transfer

Study or Subgroup	Virtual Reality Mean	SD	Total	Box Trainer Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Brinkmann C, 2017	88	17	18	127	55	18	11.2%	-39.00 [-65.59, -12.41]
Hassan SO, 2015	35	6	14	69	23	16	30.1%	-34.00 [-45.70, -22.30]
Kothari SN, 2002	37	17	13	106	45	11	10.2%	-69.00 [-97.15, -40.85]
Orlando MS, 2017	73	24	20	104	35	20	18.6%	-31.00 [-49.60, -12.40]
Thomaier L, 2016	71	23	20	100	29	20	21.9%	-29.00 [-45.22, -12.78]
Youngblood PL, 2005	87.24	28.37	17	104.07	59.84	16	8.2%	-16.83 [-49.10, 15.44]
Total (95% CI)			**102**			**101**	**100.0%**	**-35.08 [-45.15, -25.01]**

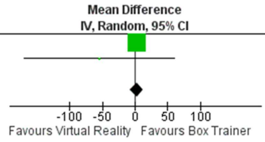
Heterogeneity: Tau² = 52.13; Chi² = 7.61, df = 5 (P = 0.18); I² = 34%
Test for overall effect: Z = 6.83 (P < 0.00001)



Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Selective reporting (reporting bias)
(**G**) Other bias

VRS *vs.* BT: TTC ligating loop

Study or Subgroup	Virtual Reality Mean	SD	Total	Box Trainer Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	A	B	C	D	E	F	G
Brinkmann C, 2017	49	17	18	46	13	18	99.3%	3.00 [-6.89, 12.89]	?	?	+	+	+	+	-
Kothari SN, 2002	256	140	13	311	147	11	0.7%	-55.00 [-170.49, 60.49]	?	?	+	?	+	+	+
Total (95% CI)			**31**			**29**	**100.0%**	**2.58 [-7.27, 12.43]**							

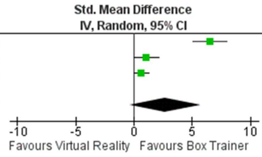


Heterogeneity: Tau² = 0.00; Chi² = 0.96, df = 1 (P = 0.33); I² = 0%
Test for overall effect: Z = 0.51 (P = 0.61)

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

VRS *vs.* BT: TTC overall basics tasks

Study or Subgroup	Virtual Reality Mean	SD	Total	Box Trainer Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	A	B	C	D	E	F	G
Hamilton EC, 2002	64.3	2.9	24	43.4	3.4	25	32.3%	6.50 [5.04, 7.95]	?	?	+	+	+	+	-
Kanumuri P, 2008	206	49.04	8	156	38.27	8	33.4%	1.07 [0.00, 2.14]	?	?	+	+	+	+	+
Tanoue K, 2008	208.4	107.2	20	153.9	59.2	20	34.3%	0.62 [-0.02, 1.25]	?	?	+	+	+	+	+
Total (95% CI)			**52**			**53**	**100.0%**	**2.67 [-0.36, 5.69]**							

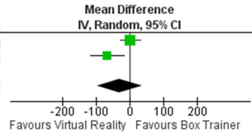


Heterogeneity: Tau² = 6.82; Chi² = 53.38, df = 2 (P < 0.00001); I² = 96%
Test for overall effect: Z = 1.73 (P = 0.08)

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

VRS *vs.* BT: TTC running the tissue

Study or Subgroup	Virtual Reality Mean	SD	Total	Box Trainer Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	A	B	C	D	E	F	G
Madan AK, 2007	106.3	37.7	17	105.9	46.4	14	54.3%	0.40 [-29.80, 30.60]	+	+	+	+	+	+	+
Youngblood PL, 2005	139.12	53.62	17	206.44	86.41	16	45.7%	-67.32 [-116.74, -17.90]	?	?	+	+	+	+	-
Total (95% CI)			**34**			**30**	**100.0%**	**-30.52 [-96.63, 35.60]**							



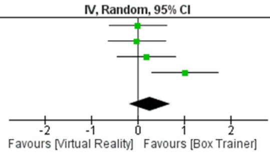
Heterogeneity: Tau² = 1856.41; Chi² = 5.25, df = 1 (P = 0.02); I² = 81%
Test for overall effect: Z = 0.90 (P = 0.37)

Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

VRS *vs.* BT: performance score overall basics tasks

	Virtual Reality			Box Trainer				Std. Mean Difference	Risk of Bias						
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	A	B	C	D	E	F	G
Akdemir A, 2014	17	4.27	20	17	2.13	20	26.0%	0.00 [-0.62, 0.62]	?	?	+	+	+	+	+
Orlando MS, 2017	21.92	7.5	20	22.09	4.4	20	26.0%	-0.03 [-0.65, 0.59]	+	?	+	?	−	+	?
Thomaier L, 2016	23.5	6.7	20	22.5	3.8	20	25.9%	0.18 [-0.44, 0.80]	+	?	+	+	+	+	+
Youngblood PL, 2005	3.31	0.8	17	2.27	1.16	17	22.1%	1.02 [0.30, 1.74]	?	?	+	+	+	+	−
Total (95% CI)			**77**			**77**	**100.0%**	**0.27 [-0.18, 0.71]**							

Heterogeneity: $Tau^2 = 0.10$; $Chi^2 = 5.82$, df = 3 (P = 0.12); $I^2 = 48\%$
Test for overall effect: Z = 1.16 (P = 0.25)

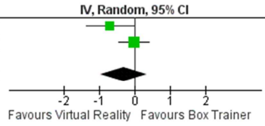


Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Selective reporting (reporting bias)
(**G**) Other bias

VRS *vs.* BT: performance score overall advanced tasks

	Virtual Reality			Box Trainer				Std. Mean Difference	Risk of Bias						
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	A	B	C	D	E	F	G
Brinkmann C, 2017	12.92	3.06	18	15.31	3.61	18	42.1%	-0.70 [-1.37, -0.02]	?	?	+	+	+	+	−
Nickel F, 2015	49.4	10.5	42	49.7	12	42	57.9%	-0.03 [-0.45, 0.40]	+	?	+	+	+	+	+
Total (95% CI)			**60**			**60**	**100.0%**	**-0.31 [-0.96, 0.34]**							

Heterogeneity: $Tau^2 = 0.14$; $Chi^2 = 2.71$, df = 1 (P = 0.10); $I^2 = 63\%$
Test for overall effect: Z = 0.93 (P = 0.35)



Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Selective reporting (reporting bias)
(**G**) Other bias

Journal: International Journal of Surgery

Article Type: Meta-analysis.

Virtual reality simulator versus box-trainer to teach minimally invasive procedures: a meta-analysis.

HIGHLIGHTS:

- Virtual Reality Simulator was better than Box Trainer in terms of participant performance scores in minimally invasive surgery.
- Virtual Reality Simulator training was better than Box Trainer training in time needed to complete peg transfer.
- In all other outcomes, regardless of the student's level of experience or type of activity, the two forms of training are equivalent.

International Journal of Surgery Author Disclosure Form

The following additional information is required for submission. Please note that failure to respond to these questions/statements will mean your submission will be returned. If you have nothing to declare in any of these categories then this should be stated.

Please state any conflicts of interest

The authors declare that they have no conflicts of interest or financial support.

Please state any sources of funding for your research

The authors declare that they have no conflicts of interest or financial support.

Please state whether Ethical Approval was given, by whom and the relevant Judgement's reference number

Ethical approval was not required for this article.

Research Registration Unique Identifying Number (UIN)

Please enter the name of the registry and the unique identifying number of the study. You can register your research at http://www.researchregistry.com to obtain your UIN if you have not already registered your study. This is mandatory for human studies only.

This article have not an unique identifying number, but it have a PROSPERO Registry number: CRD42016046840. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=46840

Author contribution

Please specify the contribution of each author to the paper, e.g. study design, data collections, data analysis, writing. Others, who have contributed in other ways should be listed as contributors.

Guedes HG, and Artifon ELA performed the systematic review and meta-analysis.

Guedes HG, Ferreira ZMCC and Leão LRS wrote the article. Frasson E, Otoch JP and Artifon ELA revised the manuscript. The final version of the manuscript was approved by all authors.

Guarantor

The Guarantor is the one or more people who accept full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Hugo Gonçalo Guedes.

Data Statement.