Efficacy of serious games in healthcare professions education: a systematic review and meta-analysis.

Marc-André Maheu-Cadotte, RN, PhD(c), Sylvie Cossette, RN, PhD, Véronique Dubé, RN, PhD, Guillaume Fontaine, RN, MSN, PhD(c), Andréane Lavallée, RN, PhD(c), Patrick Lavoie, RN, PhD, Tanya Mailhot, RN, PhD, Marie-France Deschênes, RN, MSN, PhD(c)

This work should be attributed to the Faculty of Nursing of the University of Montreal.

Corresponding author:

Marc-André Maheu-Cadotte is a PhD candidate and FRQS Doctoral Fellow at the Faculty of Nursing of the Université de Montréal, (Montreal, Canada) and is also affiliated to the Montreal Heart Institute Research Center (Montreal, Canada) and the Research Center of the Université de Montréal Hospital Center (Montreal, Canada)

Montreal Heart Institute Research Center, S-2490

5000 Belanger Street, Montreal, Qc, Canada, H1T 1C8.

514-376-3330 ext. 2041

Fax: 514-593-7441

Email: marc-andre.maheu-cadotte@umontreal.ca

This is the accepted version of the manuscript that was published in the *Simulation in Healthcare* journal and accessible here: https://doi.org/10.1097/sih.0000000000000512
Page numbers in this file are not for citation purposes.

S. Cossette is a Full Professor and vice-dean at the Faculty of Nursing of the Université de Montréal (Montreal, Canada), and a researcher at the Montreal Heart Institute Research Center (Montreal, Canada).

V. Dubé is an Assistant Professor at the Faculty of Nursing of the Université de Montréal (Montreal, Canada), and a researcher at the Research Center of the Université de Montréal Hospital Center (Montreal, Canada).

G. Fontaine is a PhD candidate and Vanier Canada Graduate Scholar (CIHR) at the Faculty of Nursing of the Université de Montréal (Montreal, Canada), and is also affiliated to the Montreal Heart Institute Research Center (Montreal, Canada).

A. Lavallée is a PhD candidate at the Faculty of Nursing of the Université de Montréal (Montreal, Canada), and is also affiliated to the CHU Ste-Justine Research Center (Montreal, Canada).

P. Lavoie is an Assistant Professor at the Faculty of Nursing of the Université de Montréal (Montreal, Canada), and a researcher at the Montreal Heart Institute Research Center (Montreal, Canada).

T. Mailhot is a post-doctoral fellow at the Montreal Heart Institute Research Center (Montreal, Canada) and at the Department of Pharmacy and Health Systems Sciences of the Bouvé College of Health Sciences (Boston, Massachusetts).

M-F Deschênes is a PhD candidate at the Faculty of Nursing of the Université de Montréal (Montreal, Canada), and is also affiliated to the Center for Innovation in Nursing Education (Montreal, Canada).

SUMMARY STATEMENT

Serious games (SGs) are interactive and entertaining software designed primarily with an educational purpose. This systematic review synthesizes evidence from experimental studies regarding the efficacy of SGs for supporting engagement and improving learning outcomes in healthcare professions education. Randomized controlled trials (RCTs) published between January 2005 and April 2019 were included. Reference selection and data extraction were performed in duplicate, independently. Thirty-seven RCTs were found and 29 were included in random effect meta-analyses. Compared with other educational interventions, SGs did not lead to more time spent with the intervention (mean difference 23.21 minutes [95% confidence interval (CI) -1.25, 47.66]), higher knowledge acquisition (standardized mean difference (SMD) 0.16 [95%CI -0.20, 0.52]), cognitive (SMD 0.08 [95%CI -0.73, 0.89]) and procedural skills development (SMD 0.05 [95%CI -0.78, 0.87]), attitude change (SMD -0.09 [95%CI -0.38, 0.20]) nor behavior change (SMD 0.2 [95%CI -0.11, 0.51]). Only a small SMD of 0.27 [95%CI 0.01, 0.53)] was found in favor of SGs for improving confidence in skills.

ABBREVIATED TITLE

Serious games in health professions education

INTRODUCTION

Serious games (SGs) are interactive and entertaining software designed primarily with an educational purpose. Popularized in the beginning of the 2000s, SGs quickly became integrated into healthcare professions education as their entertaining factor showed the potential to engage learners and support their learning process.[1] SGs are thought to fulfill the needs of adult learners, such as autonomy, control, and a sense of achievement.[2,3] Moreover, authors report that learners are receptive to the visual and the interactive aspects of SGs, traditionally associated with video games.[4] Thus, their use by healthcare educators is expected to rise in both initial and continuing education.[5]

Learning in SGs typically occurs through a gameplay experience that combines challenges with various playful design elements, which can be seen as features or building blocks of SGs (e.g., scoring system, content unlocking, integration of a storyline).[6] Challenges allow learners to be actively involved in a decision-making process for which they can receive immediate feedback and see the results of their decisions.[7] For example, an SG can challenge learners to provide the correct management plan for a virtual patient. Points can be awarded, and learners can unlock a new game level if they provide the correct management plan. As such, SGs are often associated with a constructivist learning perspective where the learning progression is fueled by an interaction cycle between learners and the SG, and where the reception of feedback allows learners to reflect on new or better ways to take on a challenge.[6] Thus, one of the main objectives in designing SGs is to ensure that they support learners' engagement to take on the various challenges that are expected to lead to significant learning outcomes.[8]

Engagement can be defined as a bi-dimensional concept: a behavioral dimension (i.e., the extent of the learners' involvement while taking on the challenges; e.g., the total amount of time invested by learners' in the SG), and an experiential dimension (i.e., the subjective experience of the learners while taking on

the challenges; e.g., learners' affect while using the SG).[9] Proper integration of the challenges with the playful design elements while designing SGs can ensure that learners remain engaged towards the challenges they take on.[10]

Systematic reviews on the use of SGs in the healthcare professions report that their efficacy in supporting engagement and improving learning outcomes varies greatly.[1,11-15] However, reasons as to why SGs produce heterogeneous results have been left unexplored. Heterogeneity is often the product of diversity in the combination of studies with different research designs, populations, intervention designs, comparators, or outcomes evaluated. [16] Previous reviews of SGs combined the findings of quasi-experimental studies with experimental ones which could have induced biases in the results reported through the lack of a control group or randomization.[12,13] Other reviews combined studies evaluating SGs to ones evaluating gamification interventions (i.e., the application of gaming elements to non-gaming contexts) or commercial off-the-shelf games (i.e., games designed specially for entertainment but used for educational purpose) which could have induced heterogeneity in the results reported due to the different design of each of these interventions.[14,15]

Thus, in this systematic review, we focused on identifying, appraising, and synthesizing the results of experimental studies evaluating the efficacy of SGs on engagement and learning outcomes in healthcare education. Since the development of an SG can be expensive and time-consuming,[14] findings from this review will thus provide guidance to educators regarding the design and the adoption of SGs, and to researchers in the conduct of future works.

METHODS

Protocol and registration

This systematic review was based on the Cochrane Handbook for Systematic Reviews of Interventions.[16] We report this systematic review according to the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA) standards.[17] We prospectively registered (#CRD42017077424) and published the detailed review protocol.[18]

Eligibility criteria

We included randomized controlled trials (RCTs), cluster-RCTs, and crossover-RCTs published in English or in French from January 1, 2005, to April 24, 2019. An SG had to be assessed, as a stand-alone intervention or as part of a multi-component intervention, among healthcare professionals or students, from any level of education, either in an initial or a continuing education setting. For the purpose of this review, we defined SGs as interactive and entertaining software with a primary educational purpose that engage learners through challenges.[19-22] All types of comparator interventions were considered for inclusion. Studies had to report at least one measure of a learning outcome or one measure of engagement—behavioral (i.e., the duration of the educational intervention usage) or experiential (i.e., self-reported measures of learners' experience in using the educational intervention). Learning outcomes were defined after Kirkpatrick's model.[23] We considered all short-term and long-term measures of knowledge acquisition, skill development (subdivided as confidence in skills, cognitive skills, and procedural skills), attitude and behavior change, as well as clinical outcomes in healthcare system users.

Information sources and search

A librarian searched six bibliographical databases using keywords and MeSH terms related to: SGs (e.g., serious game(s), game-based learning/training, applied game(s)) healthcare professionals/healthcare students (e.g., physician(s), clinician(s), trainee(s)), and effect on engagement and

learning outcomes (e.g., efficacy, skills development, knowledge acquisition). These bibliographical databases were: Cumulative Index of Nursing and Allied Health (EBSCO), EMBASE (OVID), ERIC (ProQuest), PsycINFO (APA PsycNET), PubMed (NCBI), and Web of Science – SCI and SSCI (ISI – Thomson Scientific). We performed an initial search in these databases on December 13, 2017, and we updated our search on April 24, 2019 (see Text, Supplementary Digital Content 1, all search strategies are reported). To find additional articles, hand-searching was performed in scientific journals specialized in SGs (Games for Health Journal, Games, G|A|M|E The Italian Journal of Game Studies, International Journal of Computer Games Technology, International Journal of Serious Games, and JMIR Serious Games), in previous systematic reviews,[13,24] and in the reference lists of identified studies.

Identified references were imported and managed in EndNote (Version X8, Clarivate Analytics). We screened all references independently and in pairs, and all disagreements were resolved through discussion with a third author.

Data extraction process

We performed the data extraction process by using the Effective Practice and Organisation of Care template.[25] The extraction form was piloted by all review authors involved in this step using a single article. Authors then met to discuss issues they might have had while using the form. As no significant disparity was found between forms during the piloting phase, one author performed the initial data extraction and another one validated it.

Data items

For descriptive purposes, we extracted the following items: study aim; study design; population; attrition rate; name of SGs evaluated; theoretical framework used for the SGs development; cost and

duration of the SGs development; clinical topics addressed; methods of delivery of the comparator intervention (i.e., classroom learning, written material, e-learning, another serious game, simulation/virtual simulation); duration and frequency of use of the interventions; unit of measurement; time points measured; instruments; validity and reliability of the instruments.

For quantitative synthesis purposes, we extracted the following items: sample size; outcome data; risk of bias data.

Assessment of risk of bias

Two authors independently assessed the risk of bias of each included study using the Cochrane Collaboration's tool for assessing risk of bias[26] and all disagreements were resolved with the help of a third author. A high risk of bias diminishes the reliability of the study results. The following aspects are considered during assessment: random sequence generation, allocation concealment, measurement of study group characteristics and baseline outcomes, incomplete outcome data, blinding, contamination, and selective outcome reporting. For each criterion, we judged studies at "low risk", "high risk", or "unclear risk" of bias. We considered studies at high risk of bias if they were judged at high or unclear risk of bias on either of these three criteria: randomization sequence generation, allocation concealment, or blinding of assessors to participants' group assignment as these criteria are likely to significantly bias the results.[27]

Assessment of selective reporting of outcomes

We compared the outcomes reported in the articles with the outcomes reported in the research protocol or, if no protocol was available, with the trial prospective registration form. If the trial was not prospectively registered, we compared the outcomes presented in the methods section with the ones reported in the results section.

Assessment of reporting biases

We constructed a funnel plot in RevMan 5.3 (The Cochrane Collaboration, Copenhagen, Denmark) and visually inspected it to assess reporting biases (e.g., due to publication, language, or citation biases) at the body of literature level. We considered an asymmetrical funnel plot at visual inspection as an indicator of reporting biases.

Data synthesis

Efficacy of serious games in supporting behavioral and experiential engagement

To evaluate the efficacy of SGs in supporting behavioral engagement, we used meta-analytical methods to compare the duration of SG use versus the comparator intervention use. All meta-analyses in this systematic review were performed in RevMan 5.3 (The Cochrane Collaboration, Copenhagen, Denmark) using an inverse variance approach with random-effect models to combine continuous data. At least two studies had to contribute to a single meta-analysis for it to be conducted.[28] No minimal number of participants was required. All results are expressed with 95% confidence intervals (CIs) and statistically significant results are defined as a two-sided alpha of 0.05.

Regarding the efficacy of SGs to support behavioral engagement, the result is expressed as a mean difference (in minutes). Moreover, we compared narratively the expected frequency and duration of SG use, according to study authors, to the observed ones.

Regarding experiential engagement, this concept encompasses many aspects of the learners' experience with SGs, and we expected that authors would measure a diverse set of these aspects. As such, we employed an analysis approach where we let these aspects emerged from the data (and not from prespecified categories). This allowed us both to identify all aspects of experiential engagement that were

evaluated in included studies, and to compare between studies the results obtained regarding each identified aspect. First, we extracted the items composing each instrument used to assess and to compare the experiential engagement in learners between groups. Second, two authors independently analyzed and categorized all items into aspects that were refined iteratively through the data analysis process. Third, the proposition of each author was contrasted to reach a consensus on aspects of experiential engagement that were measured. The efficacy of SGs on experiential engagement was finally evaluated for each aspect.

Efficacy of serious games in improving learning outcomes

We conducted meta-analyses to evaluate the efficacy of SGs compared to any other educational intervention in improving learning outcomes. Meta-analyses included all studies with enough data to compute a standardized mean difference (SMD) regarding at least one outcome (i.e., post-test means, medians, or odds ratios; standard deviations, first and third quartiles, standard errors, p values, or t value; number of participants in each group).

We also performed meta-analyses of studies evaluating the efficacy of SGs versus passive comparators. However, the emphasis was kept on studies with active education comparisons as educators usually seek to find the best educational intervention, and not if an educational intervention is better than doing nothing. As such, results of these analyses are not reported here but online (see Figures, Supplementary Digital Content 2, meta-analyses are presented there).

Subgroup and sensitivity analyses

Statistical heterogeneity was assessed using the I^2 statistic. A value superior to 50% was considered as a high level of heterogeneity for all meta-analyses. We explored statistical heterogeneity by performing subgroup and sensitivity analyses. Subgroup analyses were conducted regarding the study

population (i.e., healthcare professionals v. healthcare students), the comparator intervention (i.e., classroom-learning, written material, e-learning, simulation or virtual simulation, or a non-active comparator), and the publication year. As for this last subgroup analysis, we prospectively retained 2014 as a cut-off year (i.e., before or in 2014 v. after 2014) as the New Media Consortium declared that year that SGs were to be greatly developed and evaluated by educational institutions in the next 2 to 3 years.[29] Sensitivity analyses were performed to restrict meta-analyses to studies with larger sample sizes. Smaller studies tend to be associated with larger standard errors and different intervention effects, which could introduce statistical heterogeneity.[16,30,31] We considered a study "small" if its sample size fell under the first quartile when looking at the distribution of all study sample sizes included in a single meta-analysis. The median threshold from which we considered a study sample size "small" was 46 participants and the range varied between 28 and 74.

Assessment of the overall quality of the evidence

The overall quality of the evidence regarding the efficacy of SGs on each outcome was assessed by using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.[32] GRADE formalizes the evaluation of the overall quality of evidence and the formulation of recommendations. Quality of evidence depends on risks of bias, inconsistencies, imprecisions, and indirectness in the results of the studies. For each outcome, there are four levels of quality of evidence (very low, low, moderate, high) which represent our confidence in the pooled SMDs (i.e., the findings of this review). Two authors independently assessed the quality of the evidence and all disagreements were resolved through consensus.

RESULTS

Descriptive results of included studies

From a pool of 3173 unique references, 37 studies were included in the systematic review, and 29 studies (78%; all percentages presented are out of the 37 studies included) provided enough data to be included in a meta-analysis (see Figure 1). Descriptive data regarding included studies are reported in Table 1 (see Text, Supplementary Digital Content 3 and 4, the lists of included studies and of excluded studies at the full-text assessment stage are reported).

The median publication year was 2017. Twenty-eight studies (76%) were conducted exclusively among healthcare students, eight studies (22%) exclusively among healthcare professionals, and one study (3%) among both healthcare professionals and students. Regarding the professions, many studies were conducted in the medical profession (n=24; 65%). The median sample size was 91 participants (interquartile range (IQR) 99). Median attrition rates were 6.61% (IQR 20.89) at post-test assessment and 20% (IQR 25.28) at a follow-up period (i.e., between 6-week and 6-month post-intervention). E-learning interventions (n=14; 38%) were the most frequent types of comparator intervention.

Three studies (8%) compared SGs between one another. It was shown in one of the included studies that a voluntarily poor decision-making in an SG compared to a "normal" one did not influence the improvement of cognitive skills (SMD 0.00 [95%CI -0.31, 0.31]).[33] Another study showed that more frequent, but lighter sessions of SG usage, led to higher knowledge acquisition than fewer but more intensive sessions of SG usage (SMD 0.43 [95%CI 0.30, 0.56]).[34] The last one focused on the evaluation of two similar SGs; the experimental group received a SG an educational content aligned with the learning objectives as the control group received a SG with a similar, but irrelevant educational content, to avoid compensatory equalization in this group.[35] The group that received the SG that was aligned with the learning objective had significantly higher procedural skills compared with the control group (SMD 1.30 [95%CI 0.85, 1.74]).

Otherwise, in ten studies (27%), the control group received no intervention or no intervention other than one shared with the experimental group (e.g., both experimental and control groups shared the same classroom-learning activity).

Risk of bias, selective reporting of outcomes in included studies, and reporting biases

Seven studies (19%) were judged at low risk of bias. The risk of bias graph is presented in Figure 2 (also see Figure, Supplementary Digital Content 5, the risk of bias summary for each study is presented). Other studies were judged at high risk of bias, mainly due to reporting or methodological issues at study level regarding the randomization sequence generation (n=17; 46%), and the allocation concealment (n=29; 78%). Only six studies (16%) were prospectively registered or had published a protocol before the publication of the results.

A funnel plot was constructed for the "Knowledge" outcome (see Figure, Supplementary Digital Content 6, the funnel plot is presented). Visual inspection of the funnel plot revealed no serious reporting biases at the body of literature level.

Description of the serious games

Most SGs were exclusively available on a computer (n=24; 65%), seven (19%) were offered exclusively on a portable or a handheld device, four (11%) were available on more than one platform, and two (5%) were available on an unspecified platform. Clinical topics were diverse; cardiac resuscitation (n=5; 14%), triaging (n=3; 8%), and anatomy (n=3; 8%) were the most frequent. About half of the included studies reported the expected frequency of using the SG (n=19; 51%) or its duration (n=23; 62%). In those, the median expected frequency of usage was one session (IQR 3) and the median expected duration of usage was 60 minutes (IQR 150). Data related to the cost and time of development of the SG

was not reported in any study (also see Table, Supplemental Digital Content 7, key information of the SGs assessed are presented).

Ten studies (27%) cited the theoretical framework that guided the design or the development of the SG; no theoretical framework was cited more than once across the 37 included studies. Two (5%) studies reported the use of a game-based learning theory to guide the design of the SG. The most frequent challenges in SGs included the assessment and/or management of a virtual patient presenting a health-related illness condition (n=17; 46%) and answering questions on a clinical topic (n=10; 27%).

Efficacy of serious games in supporting behavioral and experiential engagement

Our confidence in the results of all meta-analyses is presented in Table 2. Our confidence ranges from very low to low for almost all meta-analyses conducted due mostly to serious risks of bias in included studies, inconsistencies, and imprecisions in results.

Five studies (14%) were included in a meta-analysis (see Figure 3) to evaluate the efficacy of SGs on behavioral engagement (i.e., minutes spent with the interventions). A non-statistically significant result favoring SGs was found (mean difference (in minutes) 23.21 [95%CI -1.25, 47.66; I_2= 91%]). Heterogeneity remained high (> 50%) when conducting planned subgroup and sensitivity analyses (see Figures, Supplementary Digital Content 8, subgroup and sensitivity analysis graphs are presented).

Five studies (14%) reported enough data to allow a comparison between the expected behavioral engagement in the SG and the actual one in terms of time spent using the SG. In three studies, the actual time spent was shorter than expected,[36-38] in one it was longer,[39] and in one it was as expected.[40]

Experiential engagement was contrasted between groups in 11 studies (30%). Aspects of experiential engagement that were assessed are the following: perceived learning efficacy (n=9; 24%), enjoyment

(n=7; 19%), satisfaction (n=6; 16%), usability (n=5; 14%), appropriateness (n=4; 11%), focus (n=4; 11%), fidelity (n=4; 11%), difficulty (n=2; 5%), perceived learning efficiency (n=2; 5%), and stress (n=1; 3%). Results regarding the efficacy of SGs on each aspect of experiential engagement are reported in Table 3. Results were highly heterogeneous overall; SGs were rarely regarded as systematically superior to other educational interventions for any of the aspects identified.

Efficacy of serious games in improving learning outcomes

Knowledge

Fifteen studies (41%) assessed participants' acquisition of knowledge and eleven (30%) were included in meta-analysis (Figure 4). We observed a negligible and non-statistically significant SMD of 0.16 (95%CI -0.20, 0.52; I_2= 86%) in favor of SGs. The number of included studies in this meta-analysis allowed for subgroup and sensitivity analyses. Statistical heterogeneity remained high ($I^2 \geq 50\%$) in all subgroup analyses. However, a statistically significant difference (p=0.006) was found between the pooled SMD of studies conducted with healthcare students (SMD -0.19 [95%CI -0.66, 0.28]; I_2 85%) compared to the pooled SMD of studies conducted with healthcare professionals (SMD 0.80 [95%CI 0.27, 1.32], I_2=84%). When removing the studies falling under the first quartile in term of sample size (less than 48 participants) statistical heterogeneity also remained high. However, the result became statistically significant (SMD 0.48 [95%CI 0.19, 0.78], I_2=77%). Two studies (5%) assessed participants' retention of knowledge, one after a follow-up period of six weeks and the other one after six months. At six weeks, a negligible and non-statistically significant difference between a SG and written material was found (SMD 0.05 [95%CI -0.74, 0.83]).[41] At six months, a negligible and non-statistically significant difference between a SG and an e-learning intervention was found (SMD -0.14 [95%CI -0.60, 0.32]).[42]

Skills

Ten studies (27%) assessed participants' cognitive skills and five of them (11%) were included in a meta-analysis (see Figure 6). We observed a negligible and non-statistically significant SMD of 0.08 (95% CI -0.73, 0.89; I_2= 95%) in favor of SGs. Heterogeneity remained high (> 50%) when exploring the potential effect of study population, and publication year. Two studies included a follow-up period. At three months, a small and non-statistically significant difference (SMD 0.23 [95% CI -0.11, 0.57]) favoring the SG compared to a classroom learning intervention was found and, at a follow-up period of 6 months, a small and statistically significant difference (SMD 0.46 [95% CI 0.32, 0.68]) favoring the SG compared to an e-learning intervention was reported.[43,44]

Twelve studies (32%) assessed participants' procedural skills. Four studies (11%) were included in this meta-analysis (see Figure 7). We observed a negligible and non-statistically significant SMD of 0.05 (95%CI -0.78, 0.87; I_2= 88%). The heterogeneity remained high (>50%) when exploring the potential effect of the comparator intervention, study population, and publication year. One study (3%) included a four-month follow-up period and a negligible and non-statistically significant difference (SMD -0.16 [95%CI -0.62, 0.29]) favoring the e-learning intervention was found.[45]

Six studies (16%) assessed participants' confidence in their skills and four of them (11%) were included in a meta-analysis (see Figure 5). We observed a small and statistically significant SMD of 0.27 (95% CI 0.01, 0.53; I_2= 0%) in favor of SGs. Non-significant differences between groups were found while exploring the potential effect of study population, publication year, and the comparator intervention.

Attitude

Five studies (14%) assessed participants' attitude. Three studies (8%) were included in this meta-analysis (see Figure 8). We observed a negligible and non-statistically significant SMD of -0.09 (95%CI -0.38, 0.20; I_2= 47%) in favor of comparator educational interventions. The low number of included studies in the main meta-analysis precluded us from conducting other subgroup or sensitivity analyses.

Behavior

Three studies (8%) assessed the perception of behavior change in practice. Two studies (5%) were included in this meta-analysis (see Figure 9). We observed a negligible and non-statistically significant SMD of 0.2 (95%CI -0.11, 0.51; I_2= 0%) in favor of SGs.

Clinical outcomes

Only one study (3%) included the assessment of a clinical outcome in healthcare system users (i.e., number of days to blood pressure target in the patients who were taken care by the healthcare providers participating in the study) and the authors reported a statistically significant difference (p=0.018) favoring the SG compared (median = 142 days) to another e-learning intervention (median = 148 days).[46]

DISCUSSION

This systematic review examined the efficacy of SG in healthcare professions education. Most studies were published in the last three years, among students, in the medical profession, and compared an SG with another e-learning intervention. We found negligible and non statistically significant differences between SGs and other educational interventions regarding their effects on knowledge acquisition, cognitive and procedural skill development in a test setting, behavior change in clinical practice, or supporting engagement during the learning activities. Additionally, heterogeneous results were found

regarding the efficacy of SGs to support any of the identified aspects of experiential engagement. This systematic review adds to previous reviews on SGs in healthcare professions education by synthesizing the latest evidence of their efficacy, by evaluating the assumption that SGs are more engaging than other educational interventions, by quantifying their efficacy, and by exploring various sources of heterogeneity through meta-analytic methods.

Educators should be aware of the limited evidence supporting the engaging nature of SGs with healthcare professionals and students. Mixed findings regarding engagement are surprising considering that the decision to use an SG is often motivated by their potential to improve learners' engagement.[37,47,48] The concept of engagement has a large scope that makes it almost an umbrella term for multiple emotional or cognitive states and numerous behaviors.[9] As such, we remained inclusive of all measures used by authors that were linked either to the behavioral or experiential dimension of engagement while using the intervention. However, these findings are impeded by the small number of studies reporting engagement outcomes and the lack of information regarding the validity or the reliability of the assessment tools used in half of the studies. Authors should consider assessing learners' engagement across educational interventions using validated and reliable assessment tools, such as the evaluation questionnaire developed by Dankbaar et al.[49] in their study, or the usability questionnaire developed by Zaharias & Povlymenakou[50].

Non-significant differences between SGs and other educational interventions were found for most learning outcomes. Our findings are in line with the ones reported in previous reviews regarding the efficacy of SGs in improving learning outcomes.[12-14] Authors of these previous reviews underlined the mixed efficacy of SGs to improve learning outcomes compared to other educational interventions. Our meta-analyses showed that, for most learning outcomes and no matter what the comparator educational

intervention was, the overall body of the evidence did not support the claim that SGs were significantly more effective. The lack of theoretical framework to support SG design could serve as an explanation for these results as most authors did not explicate a theoretical framework for the design of their SGs and only two explicitly referred to a game-based learning theory.[6,20,46,51] Designing an SG through a theoretical lens holds the potential to greatly improve learners' engagement and learning outcomes.[52] Theoretical works should be undertaken and synthesized to explain the mechanisms through which SGs are expected to lead to learning outcomes.

Furthermore, authors of recent RCTs underlined the ongoing difficulty in identifying empirical data to support their design choices.[49,53] We had initially planned in the published protocol to evaluate the individual impact of SG design elements on engagement and learning outcomes.[18] Unfortunately, scarce data prevented us from doing so. Few included studies compared different versions of an SG between one another which is essential to isolate the impact of individual design choices.[33,35,46,54] Future studies should focus on evaluating the efficacy of different versions of a SG on engagement and learning outcomes.

Regarding the long-term retention of learning outcomes, only five studies included a follow-up period and three of them reported non-significant differences between groups. It should be noted that the median expected frequency and duration of usage is a single 60-minute session, and that learners have been shown in some studies to use the SGs less than expected [36,37]. It could be hypothesized that the duration of SG use is not enough to bring greater long-term changes in learning outcomes compared to other educational interventions. Future studies should consider assessing participants' long-term retention of learning outcomes, as there is insufficient evidence to support SG efficacy in the long-term compared to other educational interventions.

As the development of an SG can be a resource-intensive endeavor and as some SGs are commercialized following their evaluation, researchers should consider prospectively registering their trial or publishing their research protocol to improve the transparency in the reporting of their results and to avoid any suspicion of potential conflicts of interest.[55] This would facilitate the evaluation of the selective reporting of outcomes in result papers. Moreover, regarding the reporting, most studies were judged at high risk of bias as the reporting of the randomization sequence generation and the allocation concealment were unclear. Future studies should make sure to report all elements necessary to their assessment. The adoption of reporting grids, such as the CONSORT grid, by journals, and the use of them by researchers, could greatly improve the reporting of these trials.[56]

Strengths of this systematic review include the prospective publication of the protocol,[18] and the reporting of the results according to the PRISMA guidelines, enhancing the transparency of the research process.[57] Furthermore, the data extraction process was piloted, and all data extraction forms were validated by a second review author. Limits of this review include the selection of RCTs only and their relatively low number. Following the Cochrane guidance, we restricted this review to RCTs to minimize threats to internal validity, and as we were aware that the efficacy of SGs had already been evaluated in multiple RCTs.[28] Another limit include a potential language bias as only studies published in English and in French were considered. However, the visual inspection of the funnel plot did not allow for the identification of a significant language bias or other types of reporting biases. Furthermore, as the nature of what constitute an SG and how it differs from interventions such as virtual simulations is still a matter of debate,[58] we remained inclusive in our definition of SGs. To address potential ambiguities regarding the nature of study interventions, we screened all references independently and in pairs, and all disagreements

were resolved through discussion with a third author. Still, we recognize our inclusive definition of SGs as a potential limit to our work.

Compared with other educational interventions, SGs led to neither statistically better behavioral engagement, knowledge acquisition, cognitive and procedural skills development, attitude change, nor behavior change. Only a statistically significant but small SMD was found in favor of SGs to improve confidence in skills. Additionally, heterogeneous results were found regarding the efficacy of SGs to support any of the identified aspects of experiential engagement. Our findings are impeded by high or unclear risk of bias across studies, inconsistencies in the directions of effect, and imprecisions of study results. As such, our confidence ranges from very low to low regarding the results of almost all meta-analyses that were conducted. We recommend that authors base their SG design choices on a theoretical framework and that they report their results according to the CONSORT statement. Moreover, future research should focus on assessing if healthcare professionals' clinical practice changes occur in post SG training, clinical outcomes in patients under the care of healthcare professionals that used SGs, and long-term retention of learning outcomes.

FINANCIAL DISCLOSURE SUMMARY

We have no potential conflict of interest to disclose and this work received no specific funding.

REFERENCES

1. Akl EA, Pretorius RW, Sackett K, et al. The effect of educational games on medical students' learning outcomes: a systematic review: BEME Guide No 14. *Med Teach.* 2010;32(1):16-27.

2. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. *Am J Prev Med.* 2008;34(1):74-82.

3. Knowles MS. *L'apprenant adulte: vers un nouvel art de la formation.* Paris, France: Editions d'Organisation; 1995.

4. Yoder SL, Terhorst R, 2nd. "Beam me up, scotty": designing the future of nursing professional development. *J Contin Educ Nurs.* 2012;43(10):456-462.

5. Wynter L, Burgess A, Kalman E, Heron JE, Bleasel J. Medical students: what educational resources are they using? *BMC Med Educ.* 2019;19(1):36.

6. Kiili K. Digital game-based learning: Towards an experiential gaming model. *Internet High Educ.* 2005;8(1):13-24.

7. Tettegah S, McCreery M, Blumberg F. Toward a framework for learning and digital games research. *Educ Psychol.* 2016;50(4):253-257.

8. De Freitas S. *Learning in immersive worlds: a review of game-based learning.* Bristol, Royaume-Uni: Joint Information Systems Committee;2006.

9. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Transl Behav Med.* 2017;7(2):254-267.

10. Nevin CR, Westfall AO, Rodriguez JM, et al. Gamification as a tool for enhancing graduate medical education. *Postgraduate medical journal.* 2014:postgradmedj-2013-132486.

11. Gorbanev I, Agudelo-Londono S, Gonzalez RA, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. *Med Educ Online.* 2018;23(1):1438718.

12. Ricciardi F, De Paolis LT. A comprehensive review of serious games in health professions. *IJCGT.* 2014;2014:1-11.

13. Wang R, DeMaria S, Jr., Goldberg A, Katz D. A systematic review of serious games in training health care professionals. *Simul Healthc.* 2016;11(1):41-51.

14. Boyle EA, Hainey T, Connolly TM, et al. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Comput Educ.* 2016;94:178-192.

15. Gentry S, Gauthier A, L'Estrade Ehrstrom B, et al. Serious gaming and gamification education in health professions: a systematic review by the Digital Health Education collaboration. *J Med Internet Res.* 2019;21(3):e12994.

16. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.*: The Cochrane Collaboration; 2011.

17. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.

18. Maheu-Cadotte MA, Cossette S, Dube V, et al. Effectiveness of serious games and impact of design elements on engagement and educational outcomes in healthcare professionals and students: a systematic review and meta-analysis protocol. *BMJ Open.* 2018;8(3):e019871.

19. Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J, Edwards T. Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior.* 2016;54:170-179.

20. Salen K, Zimmerman E. *Rules of play: Game design fundamentals.* MIT press; 2004.

21. Bergeron B. *Developing serious games.* 1 ed. Newton Centre, Massachusetts: Charles River Media; 2006.

22. Stokes BG. Videogames have changed: Time to consider serious games? *Development Education Journal.* 2005;11(3):12.

23. Kirkpatrick DL. Luminary perspective: evaluating training programs. In: Biech E, ed. *ASTD Handbook for Workplace Learning Professionals* Alexndira, Virginia: ASTD Press; 2008:485-491.

24. Vlachopoulos D, Makri A. The effect of games and simulations on higher education: a systematic literature review. *International Journal of Educational Technology in Higher Education.* 2017;14(1):22.

25. Cochrane Effective Practice and Organisation of Care Review Group (EPOC). *Data collection checklist.* Ottawa, Canada: Institute of Population Health, University of Ottawa;2002.

26. Cochrane Effective Practice and Organisation of Care (EPOC). Suggested risk of bias criteria for EPOC reviews. EPOC Resources for review authors Web site. http://epoc.cochrane.org/resources/epoc-resources-review-authors. Published 2017. Accessed May 23, 2019.

27. Savovic J, Turner RM, Mawdsley D, et al. Association Between Risk-of-Bias Assessments and Results of Randomized Trials in Cochrane Reviews: The ROBES Meta-Epidemiologic Study. *Am J Epidemiol.* 2018;187(5):1113-1122.

28. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions.* Vol 4: John Wiley & Sons; 2011.

29. Johnson L, Adams Becker S, Estrada V, Freeman A. *NMC Horizon Report: 2014 Higher Education Edition.* Austin, Texas: The New Media Consortium;2014.

30. Turner RM, Bird SM, Higgins JPT. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PloS one.* 2013;8(3):e59202-e59202.

31. Schwarzer G, Carpenter JR, Rücker G. Small-Study Effects in Meta-Analysis. In: *Meta-Analysis with R.* Cham: Springer International Publishing; 2015:107-141.

32. Dijkers M. Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development. *KT Update.* 2013;1(5):1-9.

33. Buijs-Spanjers KR, Hegge HHM, Cnossen F, Hoogendoorn E, Jaarsma D, de Rooij SE. Dark play of serious games: effectiveness and features (G4HE2018). *Games Health J.* 2019;8(4):1-6.

34. Kerfoot BP, Baker H. An online spaced-education game for global continuing medical education: a randomized trial. *Ann Surg.* 2012;256(1):33-38.

35. Haubruck P, Nickel F, Ober J, et al. Evaluation of App-Based Serious Gaming as a Training Method in Teaching Chest Tube Insertion to Medical Students: Randomized Controlled Trial. *J Med Internet Res.* 2018;20(5):e195.

36. Courtier J, Webb EM, Phelps AS, Naeger DM. Assessing the learning potential of an interactive digital game versus an interactive-style didactic lecture: the continued importance of didactic teaching in medical student education. *Pediatr Radiol.* 2016;46(13):1787-1796.

37. Dankbaar ME, Alsma J, Jansen EE, van Merrienboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Adv Health Sci Educ Theory Pract.* 2016;21(3):505-521.

38. Harrington CM, Chaitanya V, Dicker P, Traynor O, Kavanagh DO. Playing to your skills: a randomised controlled trial evaluating a dedicated video game for minimally invasive surgery. *Surg Endosc.* 2018;32(9):3813-3821.

39. Mohan D, Farris C, Fischhoff B, et al. Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: randomized controlled trial. *BMJ.* 2017;359.

40. Mohan D, Fischhoff B, Angus DC, et al. Serious games may improve physician heuristics in trauma triage. *Proc Natl Acad Sci U S A.* 2018;115(37):9204-9209.

41. Rondon S, Sassi FC, de Andrade CRF. Computer game-based and traditional learning method: a comparison regarding students' knowledge retention. *BMC Medical Education.* 2013;13(1):30.

42. Sward KA, Richardson S, Kendrick J, Maloney C. Use of a Web-based game to teach pediatric content to medical students. *Academic Pediatrics.* 2008;8(6):354-359.

43. Diehl LA, Souza RM, Gordan PA, Esteves RZ, Coelho IC. InsuOnline, an electronic game for medical education on insulin therapy: a randomized controlled trial with primary care physicians. *Journal of Medical Internet Research.* 2017;19(3):e72.

44. Mohan D, Farris C, Fischhoff B, et al. Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: randomized controlled trial. *BMJ.* 2017;359:j5416.

45. Drummond D, Delval P, Abdenouri S, et al. Serious game versus online course for pretraining medical students before a simulation-based mastery learning course on cardiopulmonary resuscitation: a randomised controlled study. *European Journal of Anaesthesiology.* 2017;34(12):836-844.

46. Kerfoot BP, Turchin A, Breydo E, Gagnon D, Conlin PR. An online spaced-education game among clinicians improves their patients' time to blood pressure control: a randomized controlled trial. *Circ Cardiovasc Qual Outcomes.* 2014;7(3):468-474.

47. Graafland M, Bemelman WA, Schijven MP. Game-based training improves the surgeon's situational awareness in the operation room: a randomized controlled trial. *Surg Endosc.* 2017.

48. Diehl LA, Souza RM, Gordan PA, Esteves RZ, Coelho IC. InsuOnline, an Electronic Game for Medical Education on Insulin Therapy: A Randomized Controlled Trial With Primary Care Physicians. *J Med Internet Res.* 2017;19(3):e72.

49. Dankbaar ME, Richters O, Kalkman CJ, et al. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. *BMC Med Educ.* 2017;17(1):30.

50. Zaharias P, Poylymenakou A. Developing a usability evaluation method for e-learning applications: Beyond functional usability. *Intl Journal of Human–Computer Interaction.* 2009;25(1):75-98.

51. Tan AJQ, Lee CCS, Lin PY, et al. Designing and evaluating the effectiveness of a serious game for safe administration of blood transfusion: A randomized controlled trial. *Nurse Educ Today.* 2017;55:38-44.

52. Wu WH, Hsiao HC, Wu PL, Lin CH, Huang SH. Investigating the learning-theory foundations of game-based learning: a meta-analysis. *J Comput Assist Learn.* 2012;28(3):265-279.

53. Brull S, Finlayson S, Kostelec T, MacDonald R, Krenzischeck D. Using Gamification to Improve Productivity and Increase Knowledge Retention During Orientation. *J Nurs Adm.* 2017;47(9):448-453.

54. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. *Acad Med.* 2005;80(6):541-548.

55. Sim I, Chan A-W, Gülmezoglu AM, Evans T, Pang T. Clinical trial registration: transparency is the watchword. *The Lancet.* 2006;367(9523):1631-1633.

56. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine.* 2010;8(1):18.

57. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4(1):1.

58. Panzoli D, Lelardeux CP, Galaup M, Lagarrigue P, Minville V, Lubrano V. Interaction and communication in an immersive learning game: the challenges of modelling real-time collaboration in a virtual operating room. In: Ma M, Oikonomou A, Jain LC, eds. *Serious Games and Edutainment Applications.* Gewerbestrasse, Suisse: Springer; 2017:147-186.

Figure legends

Figure 1. PRISMA flow diagram

Figure 2. Risk of bias graph

Figure 3. Meta-analysis of the efficacy of serious games on supporting behavioral engagement

Note. CI: confidence interval; IV: inverse of the variance; SD: standard deviation

Figure 4. Meta-analysis of the efficacy of serious games on knowledge acquisition

Note. CI: confidence interval; IV: inverse of the variance; SD: standard deviation

Figure 5. Meta-analysis of the efficacy of serious games on improving confidence in skills

Note. CI: confidence interval; IV: inverse of the variance; SD: standard deviation

Figure 6. Meta-analysis of the efficacy of serious games on improving cognitive skills

Note. CI: confidence interval; IV: inverse of the variance; SD: standard deviation

Figure 7. Meta-analysis of the efficacy of serious games on improving procedural skills

Note. CI: confidence interval; IV: inverse of the variance; SD: standard deviation

Figure 8. Meta-analysis of the efficacy of serious games on attitude change

Note. CI: confidence interval; IV: inverse of the variance; SD: standard deviation

Figure 9. Meta-analysis of the efficacy of serious games on behavior change

Note. CI: confidence interval; IV: inverse of the variance; SD: standard deviation

Supplementary Digital Content

Supplementary Digital Content 1. pdf, All search strategies

Supplementary Digital Content 2. pdf, Meta-analyses of studies evaluating the efficacy of SGs versus passive comparators

Supplementary Digital Content 3. pdf, List of included studies

Supplementary Digital Content 4. pdf, List of excluded studies at full-text assessment stage

Supplementary Digital Content 5. pdf, Risk of bias summary

Supplementary Digital Content 6. pdf, Funnel Plot

Supplementary Digital Content 7. pdf, Key information of the serious games assessed

Supplementary Digital Content 8. pdf, Subgroup and sensitivity analyses

Table 1. Key information of included studies

First author-year, Country	Study design	Study participants	Outcomes †
Compared to classroom learning			
Courtier-2016 United States	Two-group cluster randomized controlled trial	48 fourth-year medical students	Experiential engagement Knowledge
Diehl-2017 Brazil	Two-group randomized controlled trial	170 primary care physicians	Attitudes§ Behaviors§ Cognitive skills§* Experiential engagement
Hannig-2013 Germany	Two-group randomized controlled trial	55 second-year dental students	Confidence
Knight-2010 United Kingdoms	Two-group randomized controlled trial	91 various healthcare professionals (e.g., medical doctors, nurses, paramedic)	Cognitive skills* Procedural skills
Compared to written material			
Boeker-2013 Germany	Two-group randomized controlled trial	145 third-year medical students	Experiential engagement Knowledge*
Polivka-2019 USA	Two-group randomized controlled trial	74 various healthcare professionals and students	Cognitive skills
Rondon-2013 Brazil	Two-group randomized controlled trial	29 second-year speech-language and hearing science students	Knowledge§
Compared to e-learning			
Adjedj-2017 France	Two-group randomized crossover trial	68 medical students	Experiential engagement*

First author-year, Country	Study design	Study participants	Outcomes [†]
Berger-2018 Switzerland	Two-group randomized controlled trial	117 second-year pharmacy students	Attitude Confidence Experiential engagement Knowledge
Buijs-Spanjers-2018 Netherlands	Three-group randomized controlled trial	176 third-year medical students	Attitude Cognitive skills* Experiential engagement*
Dankbaar-2017 Netherlands	Two-group randomized controlled trial	90 fourth-year medical students	Behavioral engagement* Behaviors Confidence Experiential engagement Knowledge
de Sena-2019 Brazil	Two-group randomized controlled trial	45 first-year medical students	Behavioral engagement* Knowledge Procedural skills
Drummond-2017 France	Two-group randomized controlled trial	82 second-year medical students	Procedural skills[§]
Gauthier-2015 Canada	Two-group randomized controlled trial	46 first-year medical anatomy students	Behavioral engagement Knowledge
Kerfoot-2014 USA	Two-group randomized controlled trial	111 physicians, nurse practitioners, and physician assistants	Behavioral engagement Clinical outcome in patients* Knowledge*
Mohan-2017 USA	Four-group randomized controlled trial	368 emergency medicine physicians	Behavioral engagement Cognitive skills[§]* Experiential engagement
Scales-2016 USA	Two-group randomized controlled trial	422 resident physicians from various training specialties	Knowledge

First author-year, Country	Study design	Study participants	Outcomes †
Sward-2008 USA	Two-group randomized controlled trial	100 third-year medical students	Experiential engagement Knowledge§
Compared to another serious game			
Buijs-Spanjers-2019 Netherlands	Two-group randomized controlled trial	159 third-year medical students	Attitudes Cognitive skills Experiential engagement
Haubruck-2018 Germany	Two-group randomized controlled trial	95 third-to-six-year medical students	Experiential engagement* Procedural skills*
Kerfoot-2012 USA	Two-group randomized controlled trial	1470 urologists from various countries	Knowledge*
Compared to simulation or virtual simulation			
Chee-2019 Singapore	Two-group randomized controlled trial	46 registered nurses	Confidence* Procedural skills*
Chien-2013 United States	Two-group randomized controlled trial	14 medical students	Procedural skills
Katz-2017 USA	Two-group randomized controlled trial	44 residents on liver transplant rotation	Procedural skills*
Compared to multiple interventions			
Brull-2017 United States	Three-group randomized controlled trial	115 newly graduated nurses at an urban community teaching hospital Compared to classroom learning and e-learning	Knowledge*
Dankbaar-2016 Netherlands	Three-group randomized controlled trial	79 fourth-year medical students	Behavioral engagement Cognitive skills Experiential engagement

First author-year, Country	Study design	Study participants	Outcomes †
		Compared to e-learning and to no intervention	
Mohan-2018 USA	Four-group randomized controlled trial	320 emergency medicine physicians Compared to e-learning, serious game, and no intervention	Behavioral engagement Cognitive skills* Experiential engagement
Compared to no intervention			
Boada-2015 Spain	Two-group randomized controlled trial	109 second-year nursing students	Procedural skills*
Cook-2012 United Kingdoms	Two-group randomized controlled trial	34 third-year nursing students	Procedural skills
Del Blanco-2017 Spain	Two-group randomized controlled trial	132 second- and third-year nursing and medicine students	Behaviors* Confidence
Foss-2014 Norway	Two-group randomized controlled trial	201 first- and second-year undergraduate nursing students	Cognitive skills
Graafland-2017 Netherlands	Two-group randomized controlled trial	31 first- or second-year residents in general surgical training	Cognitive skills
Harrington-2018 Ireland	Two-group randomized controlled trial	20 first-to-third-year medical students	Procedural skills*
Lagro-2014 Netherlands	Two-group randomized controlled trial	145 fifth-year medical students	Attitude Knowledge
Li-2015 China	Two-group randomized controlled trial	97 freshman medical students	Procedural skills*

First author-year, Country	Study design	Study participants	Outcomes †
Tan-2017 Singapore	Two-group cluster randomized controlled trial	111 second-year nursing students	Confidence* Knowledge* Procedural skills
Van Nuland-2014 Canada	Three-group crossover randomized controlled trial	67 kinesiology students	Knowledge*

Note. ‡ Comparator refers to an intervention solely received by the control group (i.e., an intervention that is not shared with the experimental group). § An outcome that was also measured at a follow-up period. * An outcome for which there was a statistically significant difference favoring the experimental group.

Table 2. Summary of our certainty in the quantitative evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

Outcomes	Standardized mean differences (SMDs)/ mean differences (MDs) with 95% confidence interval (CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Justification
Learning time with the intervention (in minutes)	MD 23.21 (95%CI -1.25, 47.66)	377 (4 RCTs)	⊕◯◯◯ Very low	Serious risk of bias and very serious imprecision. However, mostly consistent results
Knowledge	SMD 0.16 (95%CI -0.20, 0.52)	1047 (11 RCTs)	⊕◯◯◯ Very low	Serious risk of bias and inconsistency, and very serious imprecision of results
Confidence in skills	SMD 0.27 (95%CI 0.01, 0.53)	235 (4 RCTs)	⊕⊕⊕◯ Moderate	Serious risk of bias. However, consistent and precise results
Cognitive skills	SMD 0.08 (95%CI -0.73, 0.89)	634 (5 RCTs)	⊕◯◯◯ Very low	Serious risk of bias and inconsistency, and very serious imprecision of results
Procedural skills	SMD 0.05 (95%CI -0.78, 0.87)	210 (4 RCTs)	⊕◯◯◯ Very low	Serious risk of bias and inconsistency, and very serious imprecision of results
Attitudes	SMD -0.09 (95%CI -0.38, 0.20)	352 (3 RCTs)	⊕⊕◯◯ Low	Serious risk of bias and imprecision of results, and very serious inconsistency
Behaviors	SMD 0.2 (95%CI -0.11, 0.51)	164 (2 RCTs)	⊕⊕◯◯ Low	Serious risk of bias and inconsistency. However, precise results

Note. CI : Confidence interval; RCT : Randomized controlled trial; SMD : Standardized mean difference

Table 3. Participants' self-reported assessment regarding the following aspects of the experiential engagement.

	Comparator										
	Classroom learning			Written material	E-learning						
	Courtier-2016	Diehl-2017	Haubruck-2018	Boeker-2013	Adjed-2017	Berger-2018	Dankbaar-2016	Dankbaar-2017	Mohan-2018	Mohan-2017	Sward-2008
Appropriateness The intervention format and its content are judged appropriate and credible for learning.					+*	-					+*
Difficulty The level of knowledge and skills required to progress in the intervention is adequate to learners' expertise.					+		-*				
Enjoyment The intervention is enjoyable or pleasant for learners.	-*	+*		+*		-			○	-*	+*
Fidelity The intervention is representative of reality, as learners perceives it.		-	+*		+*	-*					
Focus The intervention allows learners to concentrate on the content presented.	-				+			+*	+*		
Learning efficacy The intervention allows learners to feel that their learning has progressed.	-*	+	+*	+*	+*	+	+*	+			+
Learning efficiency Learners perceive positively the ratio of time and effort invested in the intervention versus their learning progression.	-*										+

Satisfaction The intervention fulfills learners' overall expectations and needs.	- -	+*	+* - +
Stress The intervention is perceived as excessively demanding by learners.			-
Usability The intervention is perceived as easy to use by learners.			+ -* + -* -*

Note. +: The serious game was rated as superior to the comparator intervention; ○: No difference was reported; -: The serious game was rated as inferior to the comparator intervention; *: This difference reached statistical significance.

Figure 1 - PRISMA Flow diagram

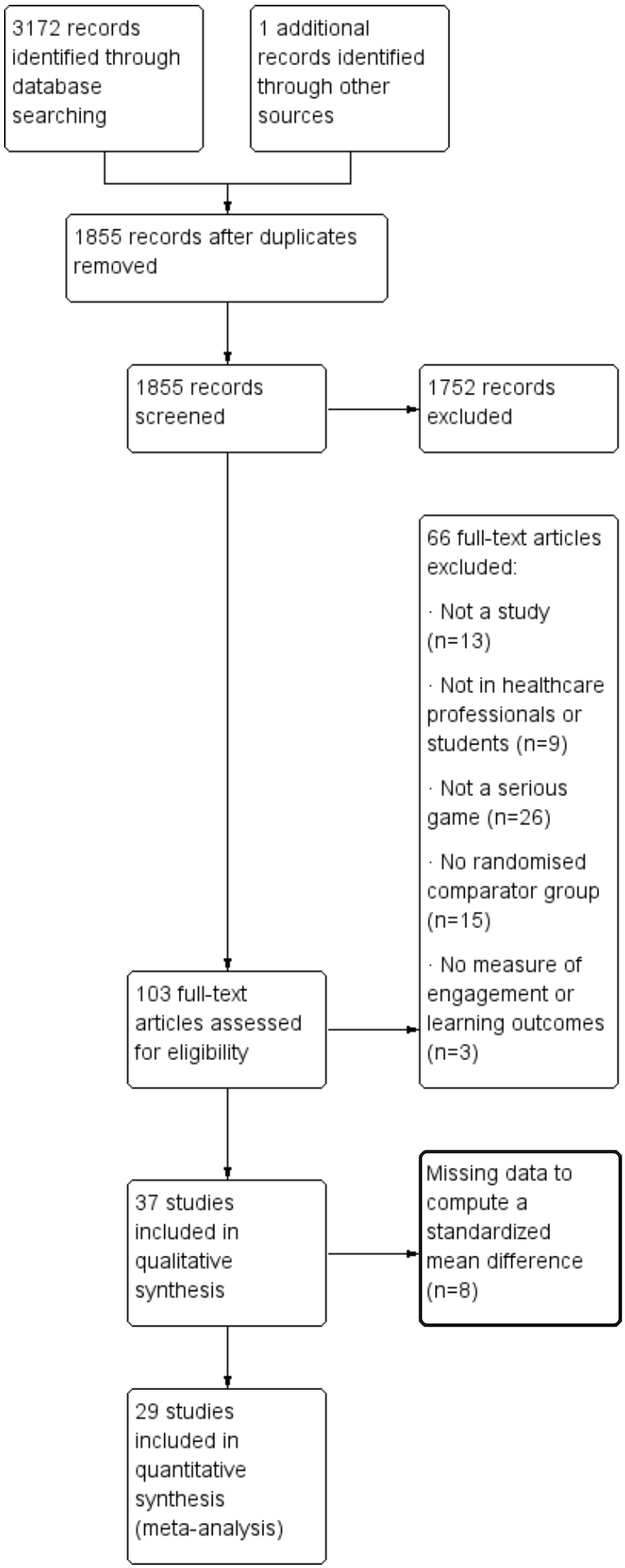

3172 records identified through database searching
1 additional records identified through other sources
1855 records after duplicates removed
1855 records screened
1752 records excluded
66 full-text articles excluded:
· Not a study (n=13)
· Not in healthcare professionals or students (n=9)
· Not a serious game (n=26)
· No randomised comparator group (n=15)
· No measure of engagement or learning outcomes (n=3)
103 full-text articles assessed for eligibility
37 studies included in qualitative synthesis
Missing data to compute a standardized mean difference (n=8)
29 studies included in quantitative synthesis (meta-analysis)

Figure 2 -
Risk of bias graph

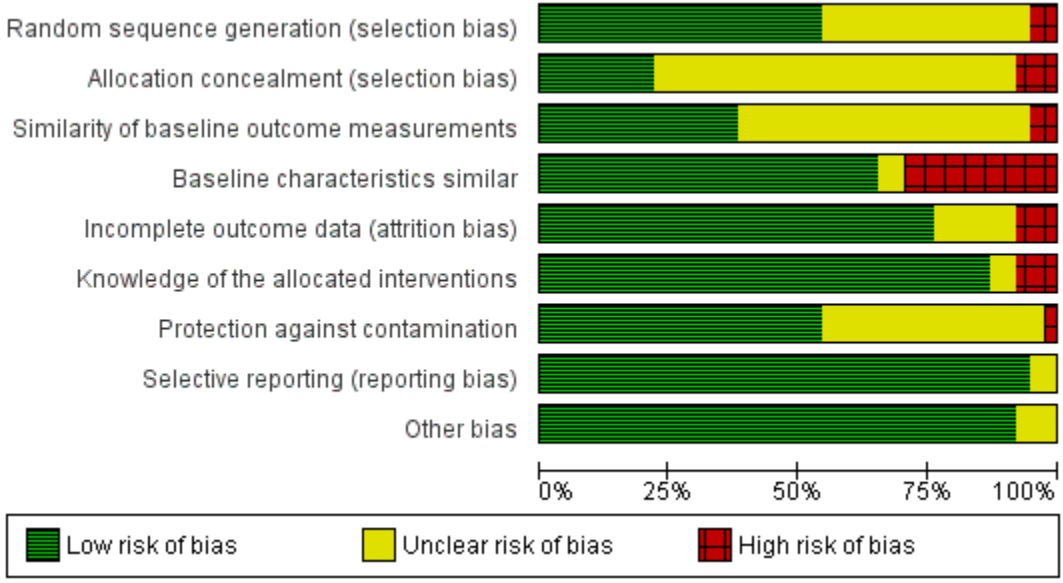

Random sequence generation (selection bias)
Allocation concealment (selection bias)
Similarity of baseline outcome measurements
Baseline characteristics similar
Incomplete outcome data (attrition bias)
Knowledge of the allocated interventions
Protection against contamination
Selective reporting (reporting bias)
Other bias
0%
25%
50%
75%
100%
Low risk of bias
Unclear risk of bias
High risk of bias

Figure 3 -
Meta-analysis on
time spent with
the intervention

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Mean Difference IV, Random, 95% CI
v. e-learning								
Dankbaar-2016	90	49	12	133	78	16	11.7%	-43.00 [-90.22, 4.22]
Mohan-2018	105	34.11	66	100	45.38	37	18.8%	5.00 [-11.78, 21.78]
De Sena-2019	18.57	0.66	23	7.41	0.43	22	20.7%	11.16 [10.84, 11.48]
Mohan-2018	117.67	50.78	66	100	45.38	36	18.3%	17.67 [-1.56, 36.90]
Dankbaar-2017	174	66	32	54	60	34	15.6%	120.00 [89.51, 150.49]
Subtotal (95% CI)			**199**			**145**	**85.1%**	**23.73 [-4.22, 51.69]**
Heterogeneity: Tau² = 852.99; Chi² = 54.96, df = 4 (P < 0.00001); I² = 93%								
Test for overall effect: Z = 1.66 (P = 0.10)								
v. simulation or virtual simulation								
Dankbaar-2016	90	49	13	70	46	20	14.9%	20.00 [-13.41, 53.41]
Subtotal (95% CI)			**13**			**20**	**14.9%**	**20.00 [-13.41, 53.41]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.17 (P = 0.24)								
Total (95% CI)			**212**			**165**	**100.0%**	**23.21 [-1.25, 47.66]**
Heterogeneity: Tau² = 753.93; Chi² = 55.23, df = 5 (P < 0.00001); I² = 91%								
Test for overall effect: Z = 1.86 (P = 0.06)								
Test for subgroup differences: Chi² = 0.03, df = 1 (P = 0.87), I² = 0%								

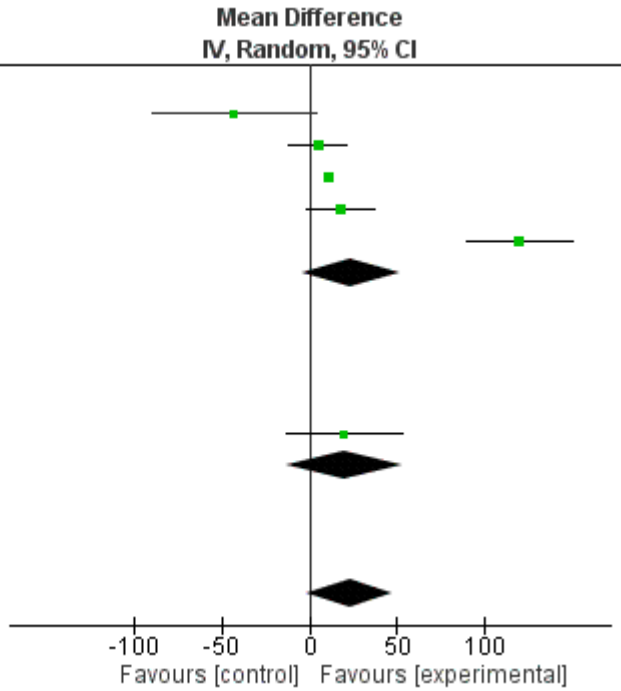


Figure 4 -
Meta-analysis on
knowledge
acquistion

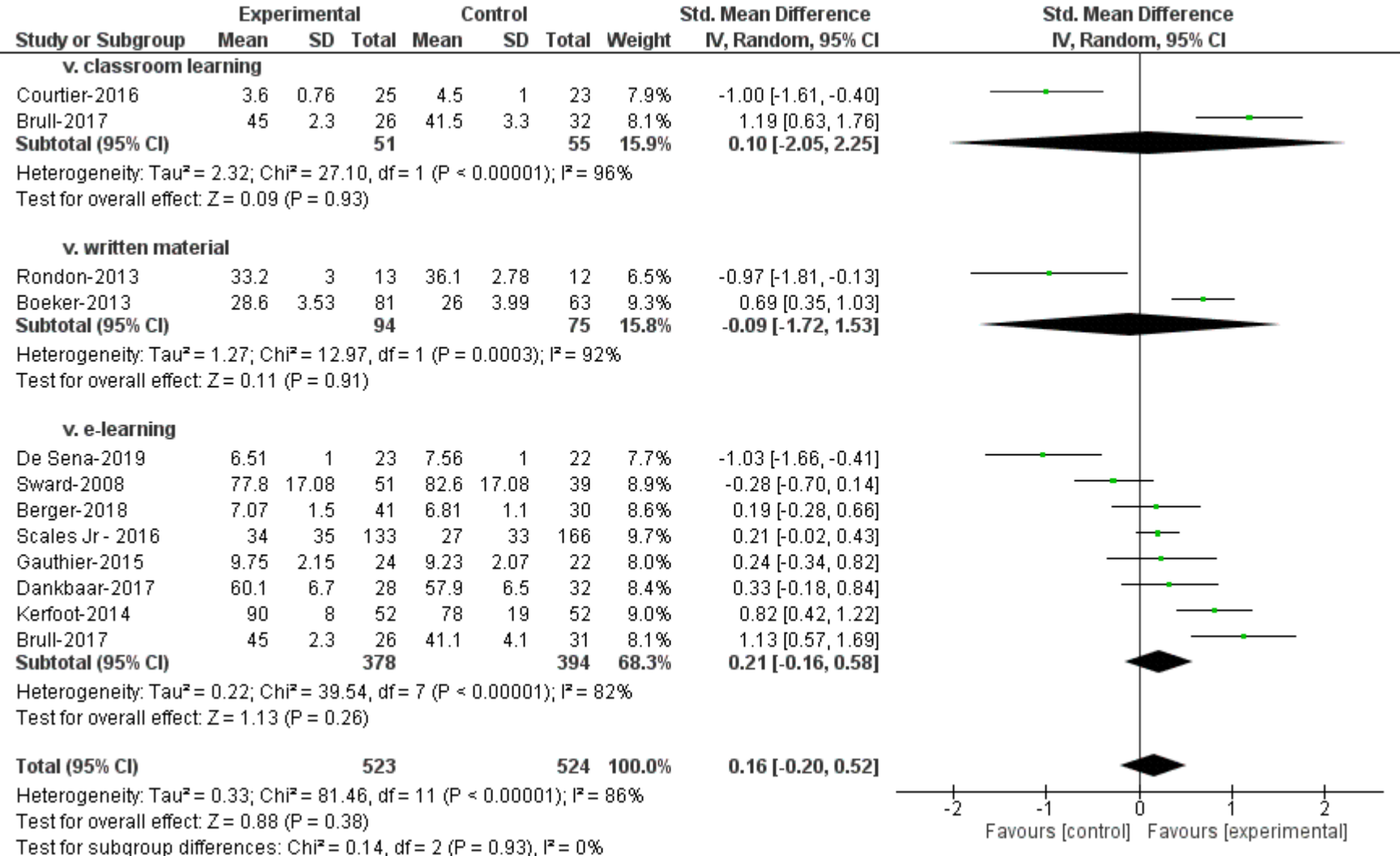

Experimental
Control
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
v. classroom learning
Courtier-2016 3.6 0.76 25 4.5 1 23 7.9% -1.00 [-1.61, -0.40]
Brull-2017 45 2.3 26 41.5 3.3 32 8.1% 1.19 [0.63, 1.76]
Subtotal (95% CI) 51 55 15.9% 0.10 [-2.05, 2.25]
Heterogeneity: Tau² = 2.32; Chi² = 27.10, df = 1 (P < 0.00001); I² = 96%
Test for overall effect: Z = 0.09 (P = 0.93)
v. written material
Rondon-2013 33.2 3 13 36.1 2.78 12 6.5% -0.97 [-1.81, -0.13]
Boeker-2013 28.6 3.53 81 26 3.99 63 9.3% 0.69 [0.35, 1.03]
Subtotal (95% CI) 94 75 15.8% -0.09 [-1.72, 1.53]
Heterogeneity: Tau² = 1.27; Chi² = 12.97, df = 1 (P = 0.0003); I² = 92%
Test for overall effect: Z = 0.11 (P = 0.91)
v. e-learning
De Sena-2019 6.51 1 23 7.56 1 22 7.7% -1.03 [-1.66, -0.41]
Sward-2008 77.8 17.08 51 82.6 17.08 39 8.9% -0.28 [-0.70, 0.14]
Berger-2018 7.07 1.5 41 6.81 1.1 30 8.6% 0.19 [-0.28, 0.66]
Scales Jr - 2016 34 35 133 27 33 166 9.7% 0.21 [-0.02, 0.43]
Gauthier-2015 9.75 2.15 24 9.23 2.07 22 8.0% 0.24 [-0.34, 0.82]
Dankbaar-2017 60.1 6.7 28 57.9 6.5 32 8.4% 0.33 [-0.18, 0.84]
Kerfoot-2014 90 8 52 78 19 52 9.0% 0.82 [0.42, 1.22]
Brull-2017 45 2.3 26 41.1 4.1 31 8.1% 1.13 [0.57, 1.69]
Subtotal (95% CI) 378 394 68.3% 0.21 [-0.16, 0.58]
Heterogeneity: Tau² = 0.22; Chi² = 39.54, df = 7 (P < 0.00001); I² = 82%
Test for overall effect: Z = 1.13 (P = 0.26)
Total (95% CI) 523 524 100.0% 0.16 [-0.20, 0.52]
Heterogeneity: Tau² = 0.33; Chi² = 81.46, df = 11 (P < 0.00001); I² = 86%
Test for overall effect: Z = 0.88 (P = 0.38)
Test for subgroup differences: Chi² = 0.14, df = 2 (P = 0.93), I² = 0%
-2
-1
0
1
2
Favours [control]
Favours [experimental]

Figure 4 -
Meta-analysis on
confidence
improvement

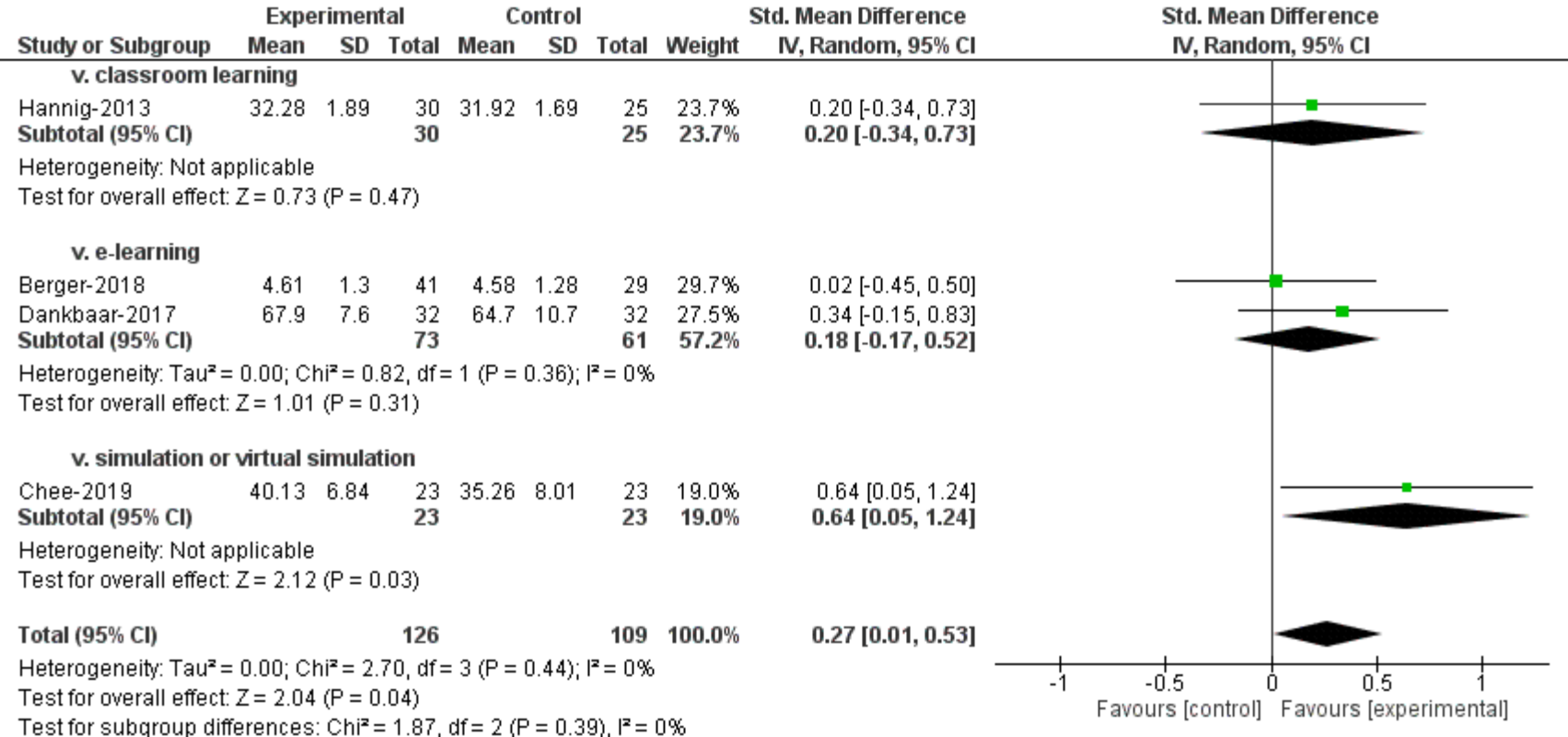

Experimental
Control
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
v. classroom learning
Hannig-2013 32.28 1.89 30 31.92 1.69 25 23.7% 0.20 [-0.34, 0.73]
Subtotal (95% CI) 30 25 23.7% 0.20 [-0.34, 0.73]
Heterogeneity: Not applicable
Test for overall effect: Z = 0.73 (P = 0.47)
v. e-learning
Berger-2018 4.61 1.3 41 4.58 1.28 29 29.7% 0.02 [-0.45, 0.50]
Dankbaar-2017 67.9 7.6 32 64.7 10.7 32 27.5% 0.34 [-0.15, 0.83]
Subtotal (95% CI) 73 61 57.2% 0.18 [-0.17, 0.52]
Heterogeneity: Tau² = 0.00; Chi² = 0.82, df = 1 (P = 0.36); I² = 0%
Test for overall effect: Z = 1.01 (P = 0.31)
v. simulation or virtual simulation
Chee-2019 40.13 6.84 23 35.26 8.01 23 19.0% 0.64 [0.05, 1.24]
Subtotal (95% CI) 23 23 19.0% 0.64 [0.05, 1.24]
Heterogeneity: Not applicable
Test for overall effect: Z = 2.12 (P = 0.03)
Total (95% CI) 126 109 100.0% 0.27 [0.01, 0.53]
Heterogeneity: Tau² = 0.00; Chi² = 2.70, df = 3 (P = 0.44); I² = 0%
Test for overall effect: Z = 2.04 (P = 0.04)
Test for subgroup differences: Chi² = 1.87, df = 2 (P = 0.39), I² = 0%
-1
-0.5
0
0.5
1
Favours [control]
Favours [experimental]

Figure 4 -
Meta-analysis on
cognitive skill
development

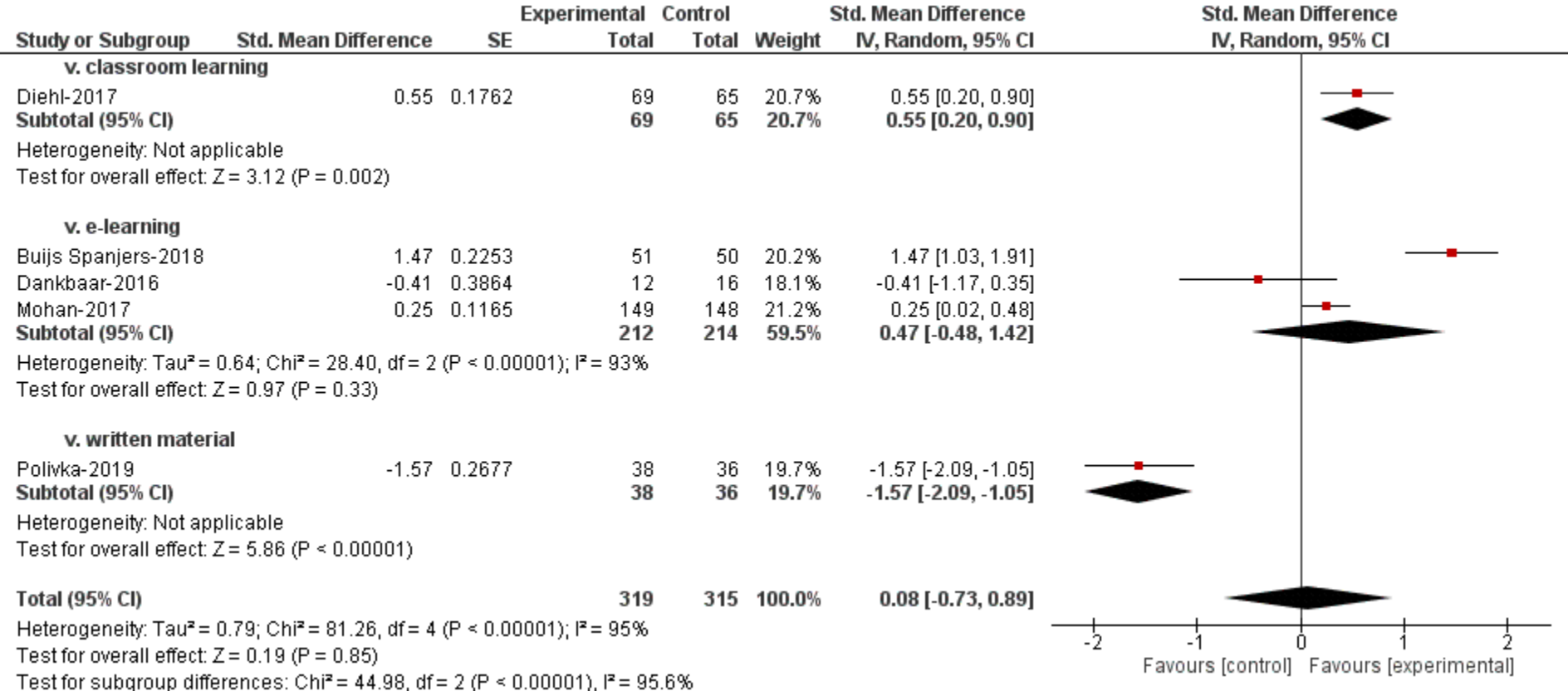

Experimental
Control
Std. Mean Difference
Std. Mean Difference
Study or Subgroup
Std. Mean Difference
SE
Total
Total
Weight
IV, Random, 95% CI
IV, Random, 95% CI
v. classroom learning
Diehl-2017 0.55 0.1762 69 65 20.7% 0.55 [0.20, 0.90]
Subtotal (95% CI) 69 65 20.7% 0.55 [0.20, 0.90]
Heterogeneity: Not applicable
Test for overall effect: Z = 3.12 (P = 0.002)
v. e-learning
Buijs Spanjers-2018 1.47 0.2253 51 50 20.2% 1.47 [1.03, 1.91]
Dankbaar-2016 -0.41 0.3864 12 16 18.1% -0.41 [-1.17, 0.35]
Mohan-2017 0.25 0.1165 149 148 21.2% 0.25 [0.02, 0.48]
Subtotal (95% CI) 212 214 59.5% 0.47 [-0.48, 1.42]
Heterogeneity: Tau² = 0.64; Chi² = 28.40, df = 2 (P < 0.00001); I² = 93%
Test for overall effect: Z = 0.97 (P = 0.33)
v. written material
Polivka-2019 -1.57 0.2677 38 36 19.7% -1.57 [-2.09, -1.05]
Subtotal (95% CI) 38 36 19.7% -1.57 [-2.09, -1.05]
Heterogeneity: Not applicable
Test for overall effect: Z = 5.86 (P < 0.00001)
Total (95% CI) 319 315 100.0% 0.08 [-0.73, 0.89]
Heterogeneity: Tau² = 0.79; Chi² = 81.26, df = 4 (P < 0.00001); I² = 95%
Test for overall effect: Z = 0.19 (P = 0.85)
Test for subgroup differences: Chi² = 44.98, df = 2 (P < 0.00001), I² = 95.6%
-2
-1
0
1
2
Favours [control]
Favours [experimental]

Figure 5 -
Meta-analysis on
procedural skill
development

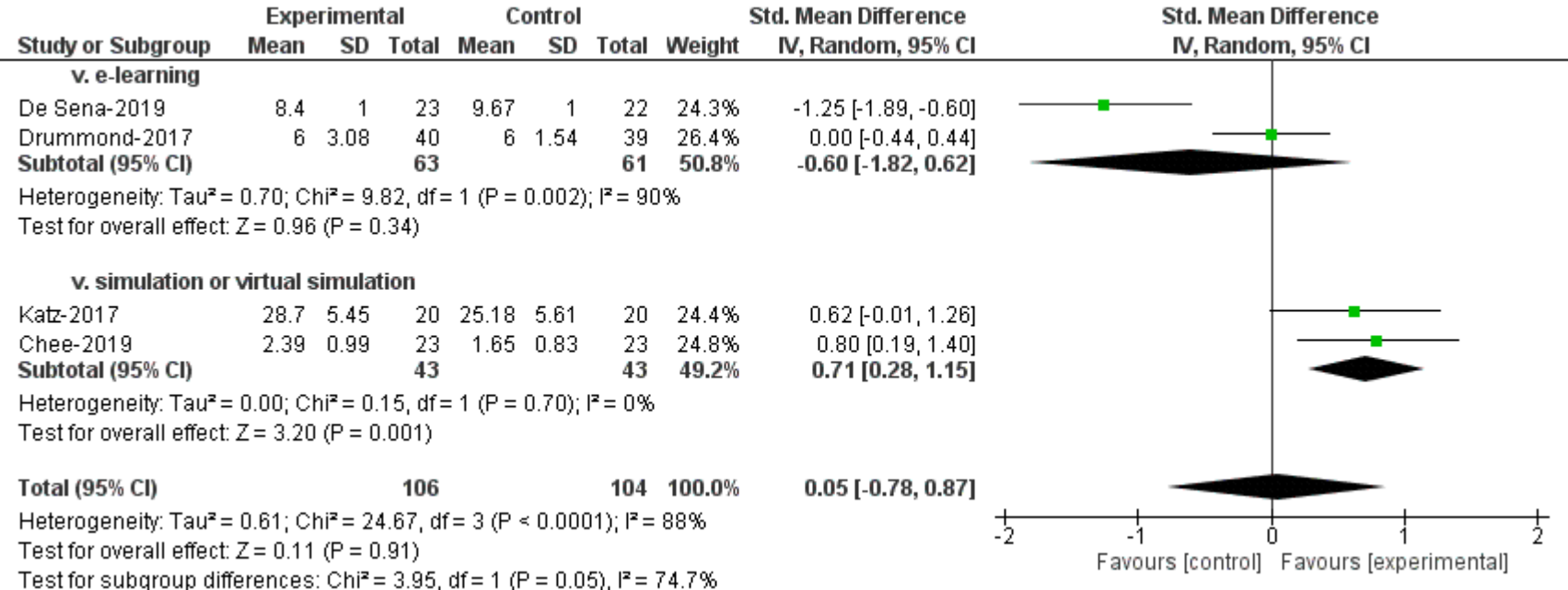

Experimental
Control
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
v. e-learning
De Sena-2019 8.4 1 23 9.67 1 22 24.3% -1.25 [-1.89, -0.60]
Drummond-2017 6 3.08 40 6 1.54 39 26.4% 0.00 [-0.44, 0.44]
Subtotal (95% CI) 63 61 50.8% -0.60 [-1.82, 0.62]
Heterogeneity: Tau² = 0.70; Chi² = 9.82, df = 1 (P = 0.002); I² = 90%
Test for overall effect: Z = 0.96 (P = 0.34)
v. simulation or virtual simulation
Katz-2017 28.7 5.45 20 25.18 5.61 20 24.4% 0.62 [-0.01, 1.26]
Chee-2019 2.39 0.99 23 1.65 0.83 23 24.8% 0.80 [0.19, 1.40]
Subtotal (95% CI) 43 43 49.2% 0.71 [0.28, 1.15]
Heterogeneity: Tau² = 0.00; Chi² = 0.15, df = 1 (P = 0.70); I² = 0%
Test for overall effect: Z = 3.20 (P = 0.001)
Total (95% CI) 106 104 100.0% 0.05 [-0.78, 0.87]
Heterogeneity: Tau² = 0.61; Chi² = 24.67, df = 3 (P < 0.0001); I² = 88%
Test for overall effect: Z = 0.11 (P = 0.91)
Test for subgroup differences: Chi² = 3.95, df = 1 (P = 0.05), I² = 74.7%
-2
-1
0
1
2
Favours [control]
Favours [experimental]

Figure 6 -
Meta-analysis on
attitude change

Study or Subgroup	Std. Mean Difference	SE	Experimental Total	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
v. classroom learning						
Diehl-2017	0.13	0.173	69	65	35.9%	0.13 [-0.21, 0.47]
Subtotal (95% CI)			**69**	**65**	**35.9%**	**0.13 [-0.21, 0.47]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.75 (P = 0.45)						
v. e-learning						
Buijs Spanjers-2018	-0.384	0.2009	51	50	30.6%	-0.38 [-0.78, 0.01]
Berger-2018	-0.0476	0.1849	59	58	33.5%	-0.05 [-0.41, 0.31]
Subtotal (95% CI)			**110**	**108**	**64.1%**	**-0.21 [-0.54, 0.12]**
Heterogeneity: $Tau^2 = 0.02$; $Chi^2 = 1.52$, df = 1 (P = 0.22); $I^2 = 34\%$						
Test for overall effect: Z = 1.23 (P = 0.22)						
Total (95% CI)			**179**	**173**	**100.0%**	**-0.09 [-0.38, 0.20]**
Heterogeneity: $Tau^2 = 0.03$; $Chi^2 = 3.79$, df = 2 (P = 0.15); $I^2 = 47\%$						
Test for overall effect: Z = 0.59 (P = 0.56)						
Test for subgroup differences: $Chi^2 = 1.95$, df = 1 (P = 0.16), $I^2 = 48.7\%$						

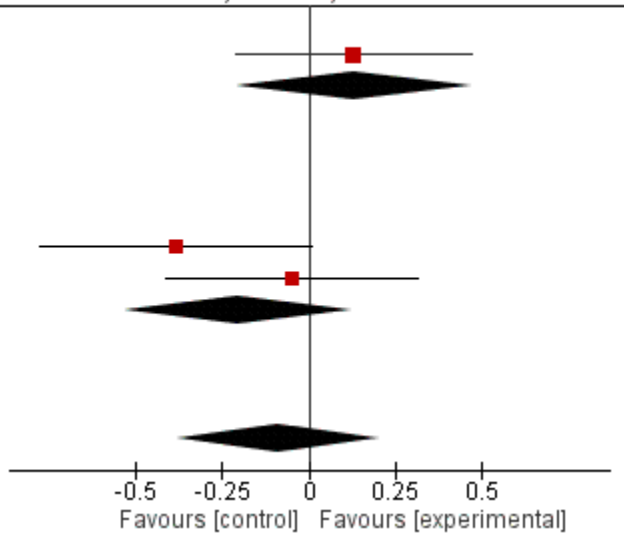
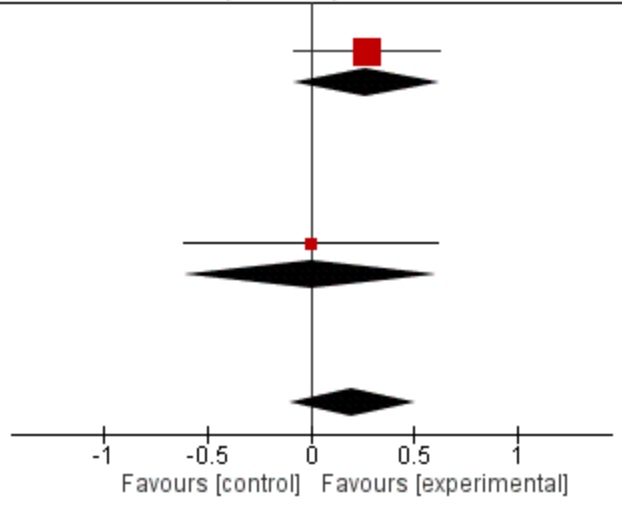


Figure 7 -
Meta-analysis on
behavior change

Study or Subgroup	Std. Mean Difference	SE	Experimental Total	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
v. classroom learning						
Diehl-2017	0.27	0.1815	65	58	74.8%	0.27 [-0.09, 0.63]
Subtotal (95% CI)			**65**	**58**	**74.8%**	**0.27 [-0.09, 0.63]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.49 (P = 0.14)						
v. e-learning						
Dankbaar-2017	0	0.3124	20	21	25.2%	0.00 [-0.61, 0.61]
Subtotal (95% CI)			**20**	**21**	**25.2%**	**0.00 [-0.61, 0.61]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 0.00 (P = 1.00)						
Total (95% CI)			**85**	**79**	**100.0%**	**0.20 [-0.11, 0.51]**
Heterogeneity: Tau² = 0.00; Chi² = 0.56, df = 1 (P = 0.45); I² = 0%						
Test for overall effect: Z = 1.29 (P = 0.20)						
Test for subgroup differences: Chi² = 0.56, df = 1 (P = 0.45), I² = 0%						



Supplementary content 1 - Search strategy

CINAHL

1. TI ("serious gam*" OR "applied gam*" OR ((simulation OR training OR teaching OR educational OR education OR learning OR interactive) AND (((online OR electronic OR digital OR "overthecounter" OR commercial OR computer OR virtual OR "mobile application*" OR "mobile app") AND (game OR games OR gamification OR gaming)) OR (videogame* OR "video game*")))) OR AB ("serious gam*" OR "applied gam*" OR ((simulation OR training OR teaching OR educational OR education OR learning OR interactive) AND (((online OR electronic OR digital OR "over-the-counter" OR commercial OR computer OR virtual OR "mobile application*" OR "mobile app") AND (game OR games OR gamification OR gaming)) OR (videogame* OR "video game*"))))

2. (MH "Video Games+")

3. 1 OR 2

4. TI ("Health Personnel" OR "Health professional*" OR "Health care profession*" OR "Healthcare profession*" OR "Medical student*" OR "Medical assistant*" OR "health worker*" OR Audiologist* OR Chiropractor* OR Dentist* OR Dietitian* OR Dermatolog* OR End ocrinologist* OR Gastroenterolog* OR Gynecolog* OR Radiolog* OR "Medical Staff" OR Midwife* OR neurologi* OR nutritionist* OR Nurse* OR nursing OR Optometrist* OR "Occupational Therapist*" OR Patholog* OR Paramedic* OR Paediatric* OR pediatrician* OR Paediatrician* OR pediatric* OR Pharmacist* OR Pharmaconomist* OR Pharmacologist* OR "Pharmacy technician*" OR Phlebotomist* OR Physician* OR Podiatrist* OR Psychologist* OR Psychotherapist* OR psychiatrist* OR "Physical therapist*" OR physiotherapist* OR "Respiratory therapist*" OR Surgeon* OR surgical OR Clinician* OR Cardiologist* OR "medical technician*" OR "emergency doctor*" OR emergentologist* OR "clinical officer*" OR "Community health worker*" OR Radiographer* OR technologist* OR Radiotherapist* OR Anesthetist* OR Resident* OR trainee* OR intern*) OR AB ("Health Personnel" OR "Health professional*" OR "Health care profession*" OR "Healthcare profession*" OR "Medical student*" OR "Medical assistant*" OR "health worker*" OR Audiologist* OR Chiropractor* OR Dentist* OR Dietitian* OR Dermatolog* OR endocrinologist* OR Gastroenterolog* OR Gynecolog* OR Radiolog* OR "Medical Staff" OR Midwife* OR neurologi* OR nutritionist* OR Nurse* OR nursing OR Optometrist* OR "Occupational Therapist*" OR Patholog* OR Paramedic* OR Paediatric* OR pediatrician* OR Paediatrician* OR pediatric* OR Pharmacist* OR Pharmaconomist* OR Pharmacologist* OR "Pharmacy technician*" OR Phlebotomist* OR Physician* OR Podiatrist* OR Psychologist* OR Psychotherapist* OR psychiatrist* OR "Physical therapist*" OR physiotherapist* OR "Respiratory therapist*" OR Surgeon* OR surgical OR Clinician* OR Cardiologist* OR "medical technician*" OR "emergency doctor*" OR emergentologist* OR "clinical officer*" OR "Community health worker*" OR Radiographer* OR technologist* OR Radiotherapist* OR Anesthetist* OR Resident* OR trainee* OR intern*)

5. (MH "Health Personnel+") OR (MH "Students, Medical") OR (MH "Students, Nursing+")

6. (MH "Education, Premedical") OR (MH "Education, Medical+") OR (MH "Education, Nursing+") OR (MH "Education, Pharmacy")

7. 4 OR 5 OR 6

Gabrielle Mathieu-Dupuis
Librarian - Consultant

8. TI (Knowledge* OR Aptitude* OR accuracy OR abilit* OR capacity* OR confidence OR competenc* OR impact* OR skill* OR performance* OR "Learning outcome*" OR "training outcome*" OR effectiveness OR efficacy OR improvement* OR innovat* OR retention OR "randomi?ed controlled trial") OR AB (Knowledge* OR Aptitude* OR accuracy OR abilit* OR capacity* OR confidence OR competenc* OR impact* OR skill* OR performance* OR "Learning outcome*" OR "training outcome*" OR effectiveness OR efficacy OR improvement* OR innovat* OR retention OR "randomi?ed controlled trial")

9. (MH "Knowledge+") OR (MH "Clinical Competence+") OR (MH "Quality Improvement+") OR(MH "Learning+") OR (MH "Educational Measurement+") OR (PT "randomized controlled trial")

10. 8 OR 9

11. 3 AND 7 AND 10

12. 11 AND LA ((english OR french)) AND DT 20000101-20171231 AND PT Journal Article

Gabrielle Mathieu-Dupuis
Librarian - Consultant

EMBASE (OVID)

1. ("serious gam*" OR "applied gam*" OR ((simulation OR training OR teaching OR educational

OR education OR learning OR interactive) AND (((online OR electronic OR digital OR "overthecounter" OR commercial OR computer OR virtual OR "mobile application*" OR "mobile app") AND (game OR games OR gamification OR gaming)) OR (videogame* OR "video game*")))).ti,ab.

2. exp video game/

3. 1 OR 2

4. ("Health Personnel" OR "Health professional$1" OR "Health care profession*" OR "Healthcare profession*" OR "Medical student$1" OR "Medical assistant$1" OR "health worker$1" OR Audiologist$1 OR Chiropractor$1 OR Dentist$1 OR Dietitian$1 OR Dermatolog* OR endocrinologist$1 OR Gastroenterolog* OR Gynecolog* OR Radiolog* OR "Medical Staff" OR Midwife$1 OR neurologi*OR nutritionist$1 OR Nurse$1 OR nursing OR Optometrist$1 OR "Occupational Therapist$1" OR Patholog* OR Paramedic$1 OR Paediatric$1 OR pediatrician$1 OR Paediatrician$1 OR podiatrist$1 OR pediatric$1 OR Pharmacist$1 OR Pharmaconomist$1 OR Pharmacologist$1 OR "Pharmacy technician$1" OR Phlebotomist$1 OR Physician$1 OR Podiatrist$1 OR Psychologist$1 OR Psychotherapist$1 OR psychiatrist$1 OR "Physical therapist$1" OR physiotherapist$1 OR "Respiratory therapist$1" OR Surgeon$1 OR Surgical OR Clinician$1 OR Cardiologist$1 OR "medical technician$1" OR "emergency doctor$1" OR emergentologist$1 OR "clinical officer$1" OR "Community health worker$1" OR Radiographer$1 OR technologist$1 OR Radiotherapist$1 OR Anesthetist$1 OR Resident$1 OR trainee$1 OR intern$1).ti,ab.

5. exp health care personnel/ OR exp premedical student/ OR exp medical student/ OR exp nursing student/

6. exp medical education/ OR exp nursing education/ OR exp clinical education/

7. 4 OR 5 OR 6

8. (Knowledge$1 OR Aptitude$1 OR accuracy OR abilit* OR capacity* OR confidence OR competenc*OR impact$1 OR skill$1 OR performance$1 OR "Learning outcome$1" OR "training outcome*" OR effectiveness OR efficacy OR improvement$1 OR innovat* OR retention OR

"randomi?ed controlled trial").ti,ab.

9. exp clinical competence/ OR *total quality management/ OR exp learning curve/ OR exp knowledge/ OR exp Randomized Controlled Trials as Topic/

10. 8 OR 9

11. 3 AND 7 AND 10

12. 2000:2017.dp. AND (english OR french).la. AND Journal: Article.pt.

13. 11 AND 12

Gabrielle Mathieu-Dupuis
Librarian - Consultant

ERIC (ProQuest)

1. TI,AB("serious gam*" OR "applied gam*" OR ((simulation OR training OR teaching OR educational OR education OR learning OR interactive) AND (((online OR electronic OR digital OR "over-the-counter" OR commercial OR computer OR virtual OR "mobile application*" OR "mobile app") AND (game OR games OR gamification OR gaming)) OR (videogame* OR "video game*"))))

2. SU.EXACT("Video Games")

3. 1 OR 2

4. TI,AB("Health Personnel" OR "Health professional$1" OR "Health care profession*" OR "Healthcare profession*" OR "Medical student$1" OR "Medical assistant$1" OR "health worker$1" OR Audiologist$1 OR Chiropractor$1 OR Dentist$1 OR Dietitian$1 OR Dermatolog* OR endocrinologist$1 OR Gastroenterolog* OR Gynecolog* OR Radiolog* OR "Medical Staff" OR Midwife$1 OR neurologi*OR nutritionist$1 OR Nurse$1 OR nursing OR Optometrist$1 OR "Occupational Therapist$1" OR Patholog* OR Paramedic$1 OR Paediatric$1 OR pediatrician$1 OR Paediatrician$1 OR podiatrist$1 OR pediatric$1 OR Pharmacist$1 OR Pharmaconomist$1 OR Pharmacologist$1 OR "Pharmacy technician$1" OR Phlebotomist$1 OR Physician$1 OR Podiatrist$1 OR Psychologist$1 OR Psychotherapist$1 OR psychiatrist$1 OR "Physical therapist$1" OR physiotherapist$1 OR "Respiratory therapist$1" OR Surgeon$1 OR Surgical OR Clinician$1 OR Cardiologist$1 OR "medical technician$1" OR "emergency doctor$1" OR emergentologist$1 OR "clinical officer$1" OR "Community health worker$1" OR Radiographer$1 OR technologist$1 OR Radiotherapist$1 OR Anesthetist$1 OR Resident$1 OR OR trainee$1 OR intern$1)

5. SU.EXACT.EXPLODE("Health Personnel") OR SU.EXACT("Premedical Students")OR SU.EXACT.EXPLODE("Medical Students") OR SU.EXACT.EXPLODE("Nursing Students")

6. SU.EXACT.EXPLODE("Pharmaceutical Education") OR SU.EXACT.EXPLODE("Medical Education") OR SU.EXACT.EXPLODE("Nursing Education") OR SU.EXACT.EXPLODE("Clinical Experience")

7. 4 OR 5 OR 6

8. TI,AB(Knowledge$1 OR Aptitude$1 OR accuracy OR abilit* OR capacity* OR confidence OR competenc*OR impact$1 OR skill$1 OR performance$1 OR "Learning outcome$1" OR "training outcome*" OR effectiveness OR efficacy OR improvement$1 OR innovat* OR retention OR "randomi?ed controlled trial")

9. SU.EXACT.EXPLODE("Learning Processes") OR SU.EXACT.EXPLODE("Knowledge Level") OR SU.EXACT.EXPLODE("Skill Development") OR SU.EXACT.EXPLODE("Outcomes of Education")

10. 8 OR 9

11. 3 AND 7 AND 10

12. PD(2000-2017) AND LA(english OR french) AND DTYPE(journal articles)

13. 11 AND 12

Gabrielle Mathieu-Dupuis
Librarian - Consultant

PsychINFO (APA PsychNet)

1. **Title** : ("serious gam*" OR "applied gam*" OR ((simulation OR training OR teaching OR educational OR education OR learning OR interactive) AND (((online OR electronic OR digital OR "over-the-counter" OR commercial OR computer OR virtual OR "mobile application*" OR "mobile app") AND (game OR games OR gamification OR gaming)) OR (videogame* OR "video game*")))) *OR* **Abstract** : ("serious gam*" OR "applied gam*" OR ((simulation OR training OR teaching OR educational OR education OR learning OR interactive) AND (((online OR electronic OR digital OR "over-the-counter" OR commercial OR computer OR virtual OR "mobile application*" OR "mobile app") AND (game OR games OR gamification OR gaming)) OR (videogame* OR "video game*")))) *OR* **Index terms**: {Computer Games}

2. **Title** : ("Health Personnel" OR "Health professional*" OR "Health care profession*" OR "Healthcare profession*" OR "Medical student*" OR "Medical assistant*" OR "health worker*" OR Audiologist* OR Chiropractor* OR Dentist* OR Dietitian* OR Dermatolog* OR endocrinologist* OR Gastroenterolog* OR Gynecolog* OR Radiolog* OR "Medical Staff" OR Midwife* OR neurologi* OR nutritionist* OR Nurse* OR nursing OR Optometrist* OR "Occupational Therapist*" OR Patholog* OR Paramedic* OR Paediatric* OR pediatrician* OR Paediatrician* OR pediatric* OR Pharmacist* OR Pharmaconomist* OR Pharmacologist* OR "Pharmacy technician*" OR Phlebotomist* OR Physician* OR Podiatrist* OR Psychologist* OR Psychotherapist* OR psychiatrist* OR "Physical therapist*" OR physiotherapist* OR "Respiratory therapist*" OR Surgeon* OR surgical OR Clinician* OR Cardiologist* OR "medical technician*" OR "emergency doctor*" OR emergentologist* OR "clinical officer*" OR "Community health worker*" OR Radiographer* OR technologist* OR Radiotherapist* OR Anesthetist* OR Resident* OR trainee* OR intern*) *OR* **Abstract:** ("Health Personnel" OR "Health professional*" OR "Health care profession*" OR "Healthcare profession*" OR "Medical student*" OR "Medical assistant*" OR "health worker*" OR Audiologist* OR Chiropractor* OR Dentist* OR Dietitian* OR Dermatolog* OR endocrinologist* OR Gastroenterolog* OR Gynecolog* OR Radiolog* OR "Medical Staff" OR Midwife* OR neurologi* OR nutritionist* OR Nurse* OR nursing OR Optometrist* OR "Occupational Therapist*" OR Patholog* OR Paramedic* OR Paediatric* OR pediatrician* OR Paediatrician* OR pediatric* OR Pharmacist* OR Pharmaconomist* OR Pharmacologist* OR "Pharmacy technician*" OR Phlebotomist* OR Physician* OR Podiatrist* OR Psychologist* OR Psychotherapist* OR psychiatrist* OR "Physical therapist*" OR physiotherapist* OR "Respiratory therapist*" OR Surgeon* OR surgical OR Clinician* OR Cardiologist* OR "medical technician*" OR "emergency doctor*" OR emergentologist* OR "clinical officer*" OR "Community health worker*" OR Radiographer* OR technologist* OR Radiotherapist* OR Anesthetist* OR Resident* OR trainee* OR intern*) *OR* **Index terms**: {Allied Health Personne} OR {Health Personne} OR {Medical Personnel} OR {Mental Health Personnel} OR {Medical Students{ OR {Nursing Students} *OR* **Index terms**: {Medical Education} OR {Nursing Education} OR {Medical Internship} OR {Medical Residency} OR {Psychiatric Training}

3. **Title** : (Knowledge* OR Aptitude* OR accuracy OR abilit* OR capacity* OR confidence OR competenc* OR impact* OR skill* OR performance* OR "Learning outcome*" OR "training outcome*" OR effectiveness OR efficacy OR improvement* OR innovat* OR retention OR "randomi?ed controlled trial") *OR* **Abstract** : (Knowledge* OR Aptitude* OR accuracy OR abilit* OR capacity* OR confidence OR competenc* OR impact* OR skill* OR performance* OR "Learning outcome*" OR "training outcome*" OR effectiveness OR efficacy OR

improvement* OR innovat* OR retention OR "randomi?ed controlled trial") *OR* **Index Terms:** {Declarative Knowledge} OR {Health Knowledge} OR {Job Knowledge} OR {Knowledge (General)} OR {Learning} OR {Procedural Knowledge} OR {Professional Competence} OR {Skill Learning} OR {Educational Measurement}

4. 1 AND 2 AND 3

5. **Language**:(english OR french) AND **Document type**: Journal Article AND **Year**: 2000 to 2017

6. 4 AND 5

Gabrielle Mathieu-Dupuis
Librarian - Consultant

PubMed

1. serious gam*[TIAB] OR applied gam*[TIAB] OR ((simulation[TIAB] OR training [TIAB] OR teaching[TIAB] OR educational[TIAB] OR education[TIAB] OR learning[TIAB] OR interactive[TIAB]) AND (((online[TIAB] OR electronic[TIAB] OR digital[TIAB] OR "over-the-counter"[TIAB] OR commercial[TIAB] OR computer[TIAB OR virtual[TIAB] OR mobile application*[TIAB] OR mobile app[TIAB]) AND (game[TIAB] OR games[TIAB] OR gamification[TIAB] OR gaming[TIAB] OR game-based[TIAB])) OR (videogame*[TIAB] OR video game*[TIAB)))
2. "Video Games"[MH]
3. #1 OR #2
4. Health Personnel*[TIAB] OR Health professional*[TIAB] OR Health care profession*[TIAB] OR Healthcare profession*[TIAB] OR Medical student*[TIAB] OR Medical assistant*[TIAB] OR health worker*[TIAB] OR Audiologist*[TIAB] OR Chiropractor*[TIAB] OR Dentist[TIAB] OR Dentists[TIAB] OR Dietitian*[TIAB] OR Dermatolog*[TIAB] OR endocrinologist*[TIAB] OR Gastroenterolog*[TIAB]OR Gynecolog*[TIAB]OR Radiolog*[TIAB] OR Medical Staff[TIAB] OR Midwife*[TIAB] OR neurologi*[TIAB] OR nutritionist*[TIAB] OR Nurse[TIAB] OR Nurses[TIAB] OR nursing[TIAB] OR Optometrist*[TIAB] OR Occupational Therapist*[TIAB] OR Patholog*[TIAB] OR Paramedic[TIAB] OR Paediatric[TIAB] OR pediatrician*[TIAB] OR Paediatrician*[TIAB] OR pediatrist*[TIAB] OR pediatric[TIAB] OR Pharmacist*[TIAB] OR Pharmaconomist*[TIAB] OR Pharmacologist*[TIAB] OR Pharmacy technician*[TIAB] OR Phlebotomist*[TIAB] OR Physician OR Podiatrist*[TIAB] OR Psychologist*[TIAB] OR Psychotherapist*[TIAB] OR psychiatrist*[TIAB] OR Physical therapist*[TIAB] OR physiotherapist*[TIAB] OR Respiratory therapist*[TIAB] OR Surgeon*[TIAB] OR surgical [TIAB] OR Clinician*[TIAB] OR Cardiologist*[TIAB] OR medical technician*[TIAB] OR emergency doctor*[TIAB] OR emergentologist*[TIAB] OR clinical officer*[TIAB] OR Community health worker*[TIAB] OR Radiographer*[TIAB] OR Radiotherapist*[TIAB] OR technologist[TIAB] Anesthetist*[TIAB] OR Resident[TIAB] OR residents[TIAB] OR trainee[TIAB] OR trainees[TIAB] OR intern[TIAB] OR interns[TIAB]
5. "Health Personnel"[MH] OR "Students, Premedical"[MH] OR "Students, Medical"[MH] OR "Students, Nursing"[Mesh]
6. "Education, Premedical"[MH] OR "Education, Medical"[MH] OR "Education, Nursing"[MH] OR "Education, Pharmacy"[MH] OR "Education, Public Health Professional"[MH] OR "Clinical Clerkship"[MH]
7. #4 OR #5 OR 6
8. knowledge*[TIAB] OR aptitude*[TIAB] OR accuracy[TIAB] OR ability[TIAB] OR abilities[TIAB] OR capacity [TIAB] OR capacities[TIAB] OR confidence[TIAB] OR compentency[TIAB] OR competencies[TIAB] OR impact*[TIAB] OR skill*[TIAB] OR performance*[TIAB] OR learning outcome*[TIAB] OR training outcome*[TIAB] OR effectiveness[TIAB] OR efficacy[TIAB] OR improvement*[TIAB] OR innovative*[TIAB] OR innovation*[TIAB] OR retention[TIAB] OR randomised controlled trial[TIAB] OR randomized controlled trial[TIAB]
9. "Clinical Competence"[MH] "Quality Improvement"[MH] OR "Learning Curve"[MH] OR Knowledge [MH] OR "Educational Measurement"[MH] OR "randomized controlled trial"[PT]

Gabrielle Mathieu-Dupuis
Librarian - Consultant

10. #8 OR 9
11. #3 AND #7 AND #10
12. (english[LA] OR french[LA]) AND 2000:2017[DP]
13. #11 AND #12

Gabrielle Mathieu-Dupuis
Librarian - Consultant

Web of Sciences

Science Citation Index Expanded (SCI-EXPANDED) --1945-present

Social Sciences Citation Index (SSCI) --1956-present

1. TS=("serious gam*" OR "applied gam*" OR ((simulation OR training OR teaching OR educational OR education OR learning OR interactive) AND (((online OR electronic OR digital OR "over-the-counter" OR commercial OR computer OR virtual OR "mobile application*" OR "mobile app") AND (game OR games OR gamification OR gaming)) OR (videogame* OR "video game*"))))

2. TS= ("Health Personnel" OR "Health professional$" OR "Health care profession*" OR "Healthcare profession*" OR "Medical student$" OR "Medical assistant$" OR "health worker$" OR Audiologist$ OR Chiropractor$ OR Dentist$ OR Dietitian$ OR Dermatolog* OR endocrinologist$ OR Gastroenterolog* OR Gynecolog* OR Radiolog* OR "Medical Staff" OR Midwife$ OR neurologi* OR nutritionist$ OR Nurse$ OR nursing OR Optometrist$ OR "Occupational Therapist$" OR Patholog* OR Paramedic$ OR Paediatric$ OR pediatrician$ OR Paediatrician$ OR pediatric$ OR Pharmacist$ OR Pharmaconomist$ OR Pharmacologist$ OR "Pharmacy technician$" OR Phlebotomist$ OR Physician$ OR Podiatrist$ OR Psychologist$ OR Psychotherapist$ OR psychiatrist$ OR "Physical therapist$" OR physiotherapist$ OR "Respiratory therapist$" OR Surgeon$ OR surgical OR Clinician$ OR Cardiologist$ OR "medical technician$" OR "emergency doctor$" OR emergentologist$ OR "clinical officer$" OR "Community health worker$" OR Radiographer$ OR technologist$ OR Radiotherapist$ OR Anesthetist$ OR Resident$ OR trainee$ OR intern$)

3. TS=("Clinical Clerkship" OR ((Clinical OR medical OR premedical OR pharma* OR nurse$) NEAR/3 (education OR training))

4. 2 OR 3

5. TS=(Knowledge$ OR Aptitude$ OR accuracy OR abilit* OR capacity* OR confidence OR competenc* OR impact$ OR skill$ OR performance$ OR "Learning outcome$" OR "training outcome*" OR effectiveness OR efficacy OR improvement$ OR innovat* OR retention OR "randomi?ed controlled trial")

6. 1 AND 2 AND 4

7. (PY=(2000-2017)) AND LANGUAGE: (English OR French) AND DOCUMENT TYPES: (Article OR Review)

8. 6 AND 7

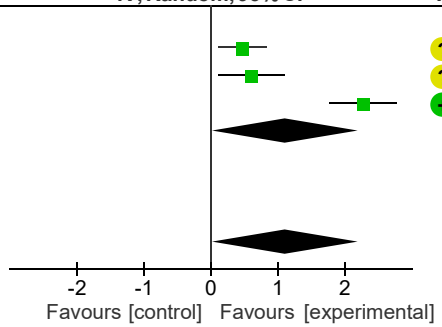
Gabrielle Mathieu-Dupuis
Librarian - Consultant

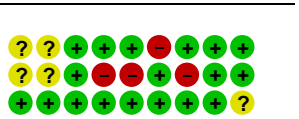
Supplementary content 2 - Meta-analyses on the efficacy of serious games versus passive comparators

1 Serious games v. passive comparators

1.1 Knowledge

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.1.1 v. none								
Lagro-2014	-13.4	4.6	67	-15.4	3.6	54	34.1%	0.47 [0.11, 0.84]
Van Nuland-2014	82.69	10.3	42	75.75	12.52	27	33.0%	0.61 [0.12, 1.11]
Tan-2017	16.46	1.86	57	11.76	2.26	46	32.9%	2.28 [1.78, 2.78]
Subtotal (95% CI)			**166**			**127**	**100.0%**	**1.11 [0.02, 2.21]**
Heterogeneity: Tau² = 0.88; Chi² = 35.26, df = 2 (P < 0.00001); I² = 94%								
Test for overall effect: Z = 2.00 (P = 0.05)								
Total (95% CI)			**166**			**127**	**100.0%**	**1.11 [0.02, 2.21]**
Heterogeneity: Tau² = 0.88; Chi² = 35.26, df = 2 (P < 0.00001); I² = 94%								
Test for overall effect: Z = 2.00 (P = 0.05)								
Test for subgroup differences: Not applicable								



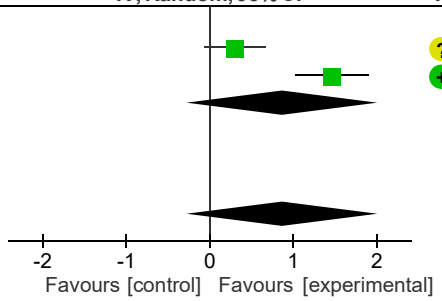


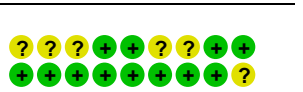
Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

1.2 Confidence in skills

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.2.1 v. none								
del Blanco-2017	19.14	5.39	58	17.43	5.96	56	50.5%	0.30 [-0.07, 0.67]
Tan-2017	31.56	8.77	57	19.39	7.57	46	49.5%	1.46 [1.02, 1.90]
Subtotal (95% CI)			**115**			**102**	**100.0%**	**0.87 [-0.27, 2.02]**
Heterogeneity: Tau² = 0.63; Chi² = 15.82, df = 1 (P < 0.0001); I² = 94%								
Test for overall effect: Z = 1.50 (P = 0.13)								
Total (95% CI)			**115**			**102**	**100.0%**	**0.87 [-0.27, 2.02]**
Heterogeneity: Tau² = 0.63; Chi² = 15.82, df = 1 (P < 0.0001); I² = 94%								
Test for overall effect: Z = 1.50 (P = 0.13)								
Test for subgroup differences: Not applicable								



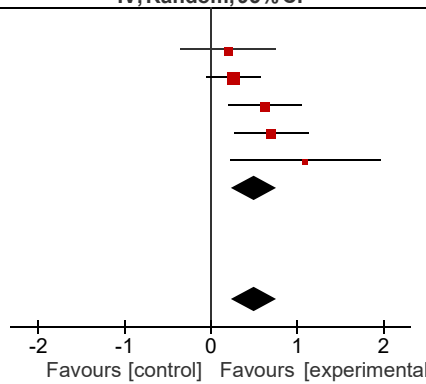


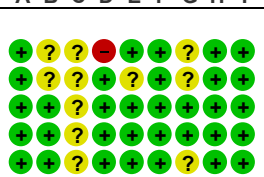
Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

1.3 Cognitive skills

Study or Subgroup	Std. Mean Difference	SE	Experimental Total	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.3.1 v. none						
Dankbaar-2016	0.2	0.2823	13	20	16.1%	0.20 [-0.35, 0.75]
Foss-2014	0.26	0.1616	82	73	31.2%	0.26 [-0.06, 0.58]
Mohan-2018	0.63	0.218	66	33	22.7%	0.63 [0.20, 1.06]
Mohan-2018	0.7	0.221	63	33	22.3%	0.70 [0.27, 1.13]
Graafland-2017	1.0933	0.4432	12	12	7.8%	1.09 [0.22, 1.96]
Subtotal (95% CI)			**236**	**171**	**100.0%**	**0.50 [0.24, 0.76]**
Heterogeneity: Tau² = 0.03; Chi² = 6.20, df = 4 (P = 0.18); I² = 36%						
Test for overall effect: Z = 3.73 (P = 0.0002)						
Total (95% CI)			**236**	**171**	**100.0%**	**0.50 [0.24, 0.76]**
Heterogeneity: Tau² = 0.03; Chi² = 6.20, df = 4 (P = 0.18); I² = 36%						
Test for overall effect: Z = 3.73 (P = 0.0002)						
Test for subgroup differences: Not applicable						



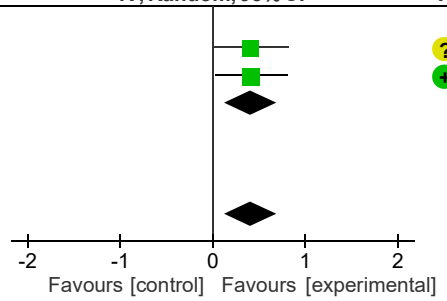


Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

1.4 Procedural skills

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.4.1 v. none								
Boada-2015	47.5	28.6	51	35.67	28.6	42	47.5%	0.41 [-0.00, 0.82]
Tan-2017	24.91	5.04	57	22.8	5.14	46	52.5%	0.41 [0.02, 0.80]
Subtotal (95% CI)			**108**			**88**	**100.0%**	**0.41 [0.13, 0.70]**
Heterogeneity: Tau² = 0.00; Chi² = 0.00, df = 1 (P = 1.00); I² = 0%								
Test for overall effect: Z = 2.83 (P = 0.005)								
Total (95% CI)			**108**			**88**	**100.0%**	**0.41 [0.13, 0.70]**
Heterogeneity: Tau² = 0.00; Chi² = 0.00, df = 1 (P = 1.00); I² = 0%								
Test for overall effect: Z = 2.83 (P = 0.005)								
Test for subgroup differences: Not applicable								



Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

Supplementary content 3 -
Lisf of included studies

List of included studies

1. Adjedj J, Ducrocq G, Bouleti C, et al. Medical student evaluation with a serious game compared to multiple choice questions assessment. *JMIR serious games.* 2017;5(2):e11.
2. Berger J, Bawab N, De Mooij J, et al. An open randomized controlled study comparing an online text-based scenario and a serious game by Belgian and Swiss pharmacy students. *Currents in pharmacy teaching & learning.* 2018;10(3):267-276.
3. Boada I, Rodriguez-Benitez A, Garcia-Gonzalez JM, Olivet J, Carreras V, Sbert M. Using a serious game to complement CPR instruction in a nurse faculty. *Computer methods and programs in biomedicine.* 2015;122(2):282-291.
4. Boeker M, Andel P, Vach W, Frankenschmidt A. Game-based e-learning is more effective than a conventional instructional method: a randomized controlled trial with third-year medical students. *PloS one.* 2013;8(12):e82328.
5. Brull S, Finlayson S, Kostelec T, MacDonald R, Krenzischeck D. Using gamification to improve productivity and increase knowledge retention during orientation. *J Nurs Adm.* 2017;47(9):448-453.
6. Buijs-Spanjers KR, Hegge HH, Jansen CJ, Hoogendoorn E, de Rooij SE. A Web-Based Serious Game on Delirium as an Educational Intervention for Medical Students: Randomized Controlled Trial. *JMIR serious games.* 2018;6(4):e17.
7. Buijs-Spanjers KR, Hegge HHM, Cnossen F, Hoogendoorn E, Jaarsma D, de Rooij SE. Dark play of serious games: effectiveness and features (G4HE2018). *Games Health J.* 2019;8(4):1-6.

8. Chee EJM, Prabhakaran L, Neo LP, et al. Play and Learn with Patients-Designing and Evaluating a Serious Game to Enhance Nurses' Inhaler Teaching Techniques: A Randomized Controlled Trial. *Games Health J.* 2019.

9. Chien JH, Suh IH, Park SH, Mukherjee M, Oleynikov D, Siu KC. Enhancing fundamental robot-assisted surgical proficiency by using a portable virtual simulator. *Surgical innovation.* 2013;20(2):198-203.

10. Cook NF, McAloon T, O'Neill P, Beggs R. Impact of a web based interactive simulation game (PULSE) on nursing students' experience and performance in life support training--a pilot study. *Nurse Educ Today.* 2012;32(6):714-720.

11. Courtier J, Webb EM, Phelps AS, Naeger DM. Assessing the learning potential of an interactive digital game versus an interactive-style didactic lecture: the continued importance of didactic teaching in medical student education. *Pediatric radiology.* 2016;46(13):1787-1796.

12. Dankbaar ME, Alsma J, Jansen EE, van Merrienboer JJ, van Saase JL, Schuit SC. An experimental study on the effects of a simulation game on students' clinical cognitive skills and motivation. *Advances in health sciences education : theory and practice.* 2016;21(3):505-521.

13. Dankbaar ME, Richters O, Kalkman CJ, et al. Comparative effectiveness of a serious game and an e-module to support patient safety knowledge and awareness. *BMC medical education.* 2017;17(1):30.

14. de Sena DP, Fabricio DD, da Silva VD, Bodanese LC, Franco AR. Comparative evaluation of video-based on-line course versus serious game for training medical students in cardiopulmonary resuscitation: A randomised trial. *PloS one.* 2019;14(4):e0214722.

15. Del Blanco A, Torrente J, Fernandez-Manjon B, Ruiz P, Giner M. Using a videogame to facilitate nursing and medical students' first visit to the operating theatre. A randomized controlled trial. *Nurse Educ Today.* 2017;55:45-53.

16. Diehl LA, Souza RM, Gordan PA, Esteves RZ, Coelho IC. InsuOnline, an electronic game for medical education on insulin therapy: a randomized controlled trial with primary care physicians. *J Med Internet Res.* 2017;19(3):e72.

17. Drummond D, Delval P, Abdenouri S, et al. Serious game versus online course for pretraining medical students before a simulation-based mastery learning course on cardiopulmonary resuscitation: A randomised controlled study. *Eur J Anaesthesiol.* 2017;34(12):836-844.

18. Foss B, Lokken A, Leland A, Stordalen J, Mordt P, Oftedal BF. Digital Game-Based Learning: A Supplement for Medication Calculation Drills in Nurse Education. *E-Learning and Digital Media.* 2014;11(4):342-349.

19. Gauthier A, Corrin M, Jenkinson J. Exploring the influence of game design on learning and voluntary use in an online vascular anatomy study aid. *Comput Educ.* 2015;87:24-34.

20. Graafland M, Bemelman WA, Schijven MP. Game-based training improves the surgeon's situational awareness in the operation room: a randomized controlled trial. *Surgical endoscopy.* 2017;31(10):4093-4101.

21. Hannig A, Lemos M, Spreckelsen C, Ohnesorge-Radtke U, Rafai N. Skills-O-Mat: Computer Supported Interactive Motion- and Game-Based Training in Mixing Alginate in Dental Education. *J Educ Comput Res.* 2013;48(3):315-343.

22. Harrington CM, Chaitanya V, Dicker P, Traynor O, Kavanagh DO. Playing to your skills: a randomised controlled trial evaluating a dedicated video game for minimally invasive surgery. *Surgical endoscopy.* 2018;32(9):3813-3821.

23. Haubruck P, Nickel F, Ober J, et al. Evaluation of app-based serious gaming as a training method in teaching chest tube insertion to medical students: randomized controlled trial. *J Med Internet Res.* 2018;20(5):e195.

24. Katz D, Zerillo J, Kim S, et al. Serious gaming for orthotopic liver transplant anesthesiology: A randomized control trial. *Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society.* 2017;23(4):430-439.

25. Kerfoot BP, Baker H. An online spaced-education game for global continuing medical education: a randomized trial. *Annals of surgery.* 2012;256(1):33-38.

26. Kerfoot BP, Turchin A, Breydo E, Gagnon D, Conlin PR. An online spaced-education game among clinicians improves their patients' time to blood pressure control: a

randomized controlled trial. *Circulation Cardiovascular quality and outcomes.* 2014;7(3):468-474.

27. Knight JF, Carley S, Tregunna B, et al. Serious gaming technology in major incident triage training: a pragmatic controlled trial. *Resuscitation.* 2010;81(9):1175-1179.
28. Lagro J, van de Pol MH, Laan A, Huijbregts-Verheyden FJ, Fluit LC, Olde Rikkert MG. A randomized controlled trial on teaching geriatric medical decision making and cost consciousness with the serious game GeriatriX. *Journal of the American Medical Directors Association.* 2014;15(12):957 e951-956.
29. Li J, Xu Y, Xu Y, et al. 3D CPR Game Can Improve CPR Skill Retention. *Studies in health technology and informatics.* 2015;216:974.
30. Mohan D, Farris C, Fischhoff B, et al. Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: randomized controlled trial. *BMJ.* 2017;359:j5416.
31. Mohan D, Fischhoff B, Angus DC, et al. Serious games may improve physician heuristics in trauma triage. *Proceedings of the National Academy of Sciences of the United States of America.* 2018;115(37):9204-9209.
32. Polivka BJ, Anderson S, Lavender SA, et al. Efficacy and Usability of a Virtual Simulation Training System for Health and Safety Hazards Encountered by Healthcare Workers. *Games Health J.* 2019;8(2):121-128.

33. Rondon S, Sassi FC, Furquim de Andrade CR. Computer game-based and traditional learning method: a comparison regarding students' knowledge retention. *BMC medical education.* 2013;13:30.

34. Scales CD, Jr., Moin T, Fink A, et al. A randomized, controlled trial of team-based competition to increase learner participation in quality-improvement education. *International journal for quality in health care : journal of the International Society for Quality in Health Care.* 2016;28(2):227-232.

35. Sward KA, Richardson S, Kendrick J, Maloney C. Use of a Web-based game to teach pediatric content to medical students. *Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association.* 2008;8(6):354-359.

36. Tan AJQ, Lee CCS, Lin PY, et al. Designing and evaluating the effectiveness of a serious game for safe administration of blood transfusion: A randomized controlled trial. *Nurse Educ Today.* 2017;55:38-44.

37. Van Nuland SE, Roach VA, Wilson TD, Belliveau DJ. Head to head: The role of academic competition in undergraduate anatomical education. *Anat Sci Educ.* 2015;8(5):404-412.

Supplementary content 4 -
List of excluded studies at the full text assessment stage

Efficacy of serious games in healthcare professions education

List of excluded studies at the full-text assessment stage

Not a study

1. Borro-Escribano B, Martinez-Alpuente I, Blanco AD, Torrente J, Fernandez-Manjon B, Matesanz R. Application of game-like simulations in the Spanish Transplant National Organization. *Transplantation proceedings.* 2013;45(10):3564-3565.
2. Cowan B, Sabri H, Kapralos B, Moussa F, Cristancho S, Dubrowski A. A serious game for off-pump coronary artery bypass surgery procedure training. *Studies in health technology and informatics.* 2011;163:147-149.
3. del Blanco A, Fernandez-Manjon B, Ruiz P, Giner M. Using videogames facilitates the first visit to the operating theatre. *Med Educ.* 2013;47(5):519-520.
4. Gallagher AG, Traynor O. Simulation in surgery: opportunity or threat? *Ir J Med Sci.* 2008;177(4):283-287.
5. Heldal I, Backlund P, Johannesson M, Lebram M, Lundberg L. Connecting the Links: Narratives, Simulations and Serious Games in Prehospital Training. *Studies in health technology and informatics.* 2017;235:343-347.
6. Johnston CL, Whatley D. Pulse!!--A virtual learning space project. *Studies in health technology and informatics.* 2006;119:240-242.
7. Kerfoot BP, Kissane N. The use of gamification to boost residents' engagement in simulation training. *JAMA surgery.* 2014;149(11):1208-1209.

8. Magro A, Swarz J, Ousley A. CancerSPACE: An Interactive E-learning Tool Aimed to Improve Cancer Screening Rates. *Journal of Computer-Mediated Communication.* 2010;15(3):482-499.
9. Semeraro F, Frisoli A, Ristagno G, et al. Relive: a serious game to learn how to save lives. *Resuscitation.* 2014;85(7):e109-110.
10. Shewaga R, Knox A, Ng G, Kapralos B, Dubrowski A. Z-DOC: a serious game for Z-plasty procedure training. *Studies in health technology and informatics.* 2013;184:404-406.
11. Smith-Stoner M, Willer A. Innovative use of the Internet and intranets to provide education by adding games. *Computers, informatics, nursing : CIN.* 2005;23(5):237-241.
12. Truchot-Cardot D. ["Games, to nurse? Is this legitimate?"]. *Krankenpfl Soins Infirm.* 2016;109(2):8-11, 58-61, 82-15.
13. Young J. Using a Role-Play Simulation Game to Promote Systems Thinking. *J Contin Educ Nurs.* 2018;49(1):10-11.

Not with healthcare professionals or students

1. Asadipour A, Debattista K, Chalmers A. Visuohaptic augmented feedback for enhancing motor skills acquisition. *Visual Computer.* 2017;33(4):401-411.
2. Bauer KN, Brusso RC, Orvis KA. Using Adaptive Difficulty to Optimize Videogame-Based Training Performance: The Moderating Role of Personality. *Military Psychology.* 2012;24(2):148-165.

3. Chan WY, Qin J, Chui YP, Heng PA. A serious game for learning ultrasound-guided needle placement skills. *IEEE Trans Inf Technol Biomed.* 2012;16(6):1032-1042.
4. Olson DK, Scheller A, Larson S, Lindeke L, Edwardson S. Using gaming simulation to evaluate bioterrorism and emergency readiness education. *Public health reports (Washington, DC : 1974).* 2010;125(3):468-477.
5. Simic G, Jevremovic A, Kostic Z, Dordevic D. Assessment based on Serious Gaming Interactive Questions (SGIQ). *Journal of Computer Assisted Learning.* 2015;31(6):623-637.
6. Sotomayor TM. Teaching tactical combat casualty care using the TC3 Sim Game-based simulation: a study to measure training effectiveness. *Studies in Health Technology & Informatics.* 2010;154:176-179.
7. Sterkenburg PS, Vacaru VS. The effectiveness of a serious game to enhance empathy for care workers for people with disabilities: A parallel randomized controlled trial. *Disabil Health J.* 2018;11(4):576-582.
8. van Dijk T, Spil T, van der Burg S, Wenzler I, Dalmolen S. Present or Play. *International Journal of Game-Based Learning.* 2015;5(2):55-69.
9. Yu FY, Han C, Chan TW. Experimental comparisons of face-to-face and anonymous real-time team competition in a networked gaming learning environment. *Cyberpsychol Behav.* 2008;11(4):511-514.

Not a serious game

1. Amer RS, Denehy GE, Cobb DS, Dawson DV, Cunningham-Ford MA, Bergeron C. Development and evaluation of an interactive dental video game to teach dentin bonding. *J Dent Educ.* 2011;75(6):823-831.
2. Ang ET, Chan JM, Gopal V, Li Shia N. Gamifying anatomy education. *Clin Anat.* 2018;31(7):997-1005.
3. Ankay Yilbas A, Canbay O, Akca B, et al. The effect of playing video games on fiberoptic intubation skills. *Anaesthesia, critical care & pain medicine.* 2018.
4. Ankay Yilbas A, Canbay O, Akca B, et al. The effect of playing video games on fiberoptic intubation skills. *Anaesthesia, critical care & pain medicine.* 2018.
5. Blakely G, Skirton H, Cooper S, Allum P, Nelmes P. Use of educational games in the health professions: a mixed-methods study of educators' perspectives in the UK. *Nursing & health sciences.* 2010;12(1):27-32.
6. Chalhoub M, Khazzaka A, Sarkis R, Sleiman Z. The role of smartphone game applications in improving laparoscopic skills. *Advances in medical education and practice.* 2018;9:541-547.
7. Chang TP, Raymond T, Dewan M, et al. The effect of an International competitive leaderboard on self-motivated simulation-based CPR practice among healthcare professionals: A randomized control trial. *Resuscitation.* 2019;138:273-281.
8. Cowan B, Rojas D, Kapralos B, Moussa F, Dubrowski A. Effects of sound on visual realism perception and task performance. *Visual Computer.* 2015;31(9):1207-1216.

9. Creutzfeldt J, Hedman L, Fellander-Tsai L. Effects of pre-training using serious game technology on CPR performance--an exploratory quasi-experimental transfer study. *Scandinavian journal of trauma, resuscitation and emergency medicine.* 2012;20:79.

10. Diehl LA, Gordan PA, Esteves RZ, Coelho IC. Effectiveness of a serious game for medical education on insulin therapy: a pilot study. *Archives of endocrinology and metabolism.* 2015;59(5):470-473.

11. El-Beheiry M, McCreery G, Schlachta CM. A serious game skills competition increases voluntary usage and proficiency of a virtual reality laparoscopic simulator during first-year surgical residents' simulation curriculum. *Surgical endoscopy.* 2017;31(4):1643-1650.

12. Evans DA, Curtis AR. Animosity, antagonism, and avatars: teaching conflict management in second life. *The Journal of nursing education.* 2011;50(11):653-655.

13. Fonseca LM, Aredes ND, Fernandes AM, et al. Computer and laboratory simulation in the teaching of neonatal nursing: innovation and impact on learning. *Revista latino-americana de enfermagem.* 2016;24:e2808.

14. Gunn T, Jones L, Bridge P, Rowntree P, Nissen L. The use of virtual reality simulation to improve technical skill in the undergraduate medical imaging student. *Interactive Learning Environments.* 2017;26(5):613-620.

15. Hashimoto DA, Gomez ED, Beyer-Berjot L, et al. A Randomized Controlled Trial to Assess the Effects of Competition on the Development of Laparoscopic Surgical Skills. *Journal of surgical education.* 2015;72(6):1077-1084.

16. Hedman L, Schlickum M, Fellander-Tsai L. Surgical novices randomized to train in two video games become more motivated during training in MIST-VR and GI Mentor II than students with no video game training. *Studies in health technology and informatics.* 2013;184:189-194.

17. Jalink MB, Heineman E, Pierie JP, ten Cate Hoedemaker HO. The effect of a preoperative warm-up with a custom-made Nintendo video game on the performance of laparoscopic surgeons. *Surgical endoscopy.* 2015;29(8):2284-2290.

18. Lin CC, Huang SC, Lin HH, Huang WJ, Chen WS, Yang SH. Naked-eye box trainer and training box games have similar training effect as conventional video-based box trainer for novices: A randomized controlled trial. *Am J Surg.* 2018;216(5):1022-1027.

19. Mak WW, Cheng SS, Law RW, Cheng WW, Chan F. Reducing HIV-related stigma among health-care professionals: a game-based experiential approach. *AIDS care.* 2015;27(7):855-859.

20. McGrath J, Kman N, Danforth D, et al. Virtual alternative to the oral examination for emergency medicine residents. *The western journal of emergency medicine.* 2015;16(2):336-343.

21. McMullan M, Jones R, Lea S. The effect of an interactive e-drug calculations package on nursing students' drug calculation ability and self-efficacy. *International journal of medical informatics.* 2011;80(6):421-430.

22. Moglia A, Perrone V, Ferrari V, et al. Influence of videogames and musical instruments on performances at a simulator for robotic surgery. *Minimally invasive therapy & allied*

technologies : MITAT : official journal of the Society for Minimally Invasive Therapy. 2017;26(3):129-134.

23. Orwoll B, Diane S, Henry D, et al. Gamification and Microlearning for Engagement With Quality Improvement (GAMEQI): A Bundled Digital Intervention for the Prevention of Central Line-Associated Bloodstream Infection. *American journal of medical quality : the official journal of the American College of Medical Quality.* 2018;33(1):21-29.
24. Rosser JC, Jr., Gentile DA, Hanigan K, Danner OK. The effect of video game "warm-up" on performance of laparoscopic surgery tasks. *JSLS : Journal of the Society of Laparoendoscopic Surgeons.* 2012;16(1):3-9.
25. Sapkaroski D, Baird M, McInerney J, Dimmock MR. The implementation of a haptic feedback virtual reality simulation clinic with dynamic patient interaction and communication for medical imaging students. *Journal of medical radiation sciences.* 2018;65(3):218-225.
26. Schlickum MK, Hedman L, Enochsson L, Kjellin A, Fellander-Tsai L. Systematic video game training in surgical novices improves performance in virtual reality endoscopic surgical simulators: a prospective randomized study. *World journal of surgery.* 2009;33(11):2360-2367.

No randomized comparator group

1. Albright G, Adam C, Goldman R, Serri D. A Game-Based Simulation Utilizing Virtual Humans to Train Physicians to Screen and Manage the Care of Patients with Mental Health Disorders. *Games Health J.* 2013;2(5):269-273.
2. Butt AL, Kardong-Edgren S, Ellertson A. Using Game-Based Virtual Reality with Haptics for Skill Acquisition. *Clin Simul Nurs.* 2018;16:25-32.
3. Buttussi F, Pellis T, Cabas Vidani A, Pausler D, Carchietti E, Chittaro L. Evaluation of a 3D serious game for advanced life support retraining. *International journal of medical informatics.* 2013;82(9):798-809.
4. Cowan B, Sabri H, Kapralos B, et al. A serious game for total knee arthroplasty procedure, education and training. *Journal of Cyber Therapy and Rehabilitation.* 2010;3(3):285-298.
5. Cutumisu M, Brown MRG, Fray C, Schmolzer GM. Growth Mindset Moderates the Effect of the Neonatal Resuscitation Program on Performance in a Computer-Based Game Training Simulation. *Frontiers in pediatrics.* 2018;6:195.
6. Dankbaar ME, Roozeboom MB, Oprins EA, et al. Preparing Residents Effectively in Emergency Skills Training With a Serious Game. *Simulation in healthcare : journal of the Society for Simulation in Healthcare.* 2017;12(1):9-16.
7. Hannig A, Kuth N, Ozman M, Jonas S, Spreckelsen C. eMedOffice: a web-based collaborative serious game for teaching optimal design of a medical practice. *BMC medical education.* 2012;12:104.

8. Kang J, Suh EE. Development and Evaluation of "Chronic Illness Care Smartphone Apps" on Nursing Students' Knowledge, Self-efficacy, and Learning Experience. *Computers, informatics, nursing : CIN.* 2018;36(11):550-559.

9. Kanthan R, Senger JL. The impact of specially designed digital games-based learning in undergraduate pathology and medical education. *Archives of pathology & laboratory medicine.* 2011;135(1):135-142.

10. McKenzie K. A comparison of the effectiveness of a game informed online learning activity and face to face teaching in increasing knowledge about managing aggression in health settings. *Advances in health sciences education : theory and practice.* 2013;18(5):917-927.

11. Middeke A, Anders S, Schuelper M, Raupach T, Schuelper N. Training of clinical reasoning with a Serious Game versus small-group problem-based learning: A prospective study. *PloS one.* 2018;13(9):e0203851.

12. O'Neill E, Reynolds PA, Hatzipanagos S, Gallagher JE. Graphic (games research applied to public health with innovative collaboration)--designing a serious game pilot for dental public health. *Bull Group Int Rech Sci Stomatol Odontol.* 2013;51(3):e30-31.

13. Qin J, Chui YP, Pang WM, Choi KS, Heng PA. Learning blood management in orthopedic surgery through gameplay. *IEEE computer graphics and applications.* 2010;30(2):45-57.

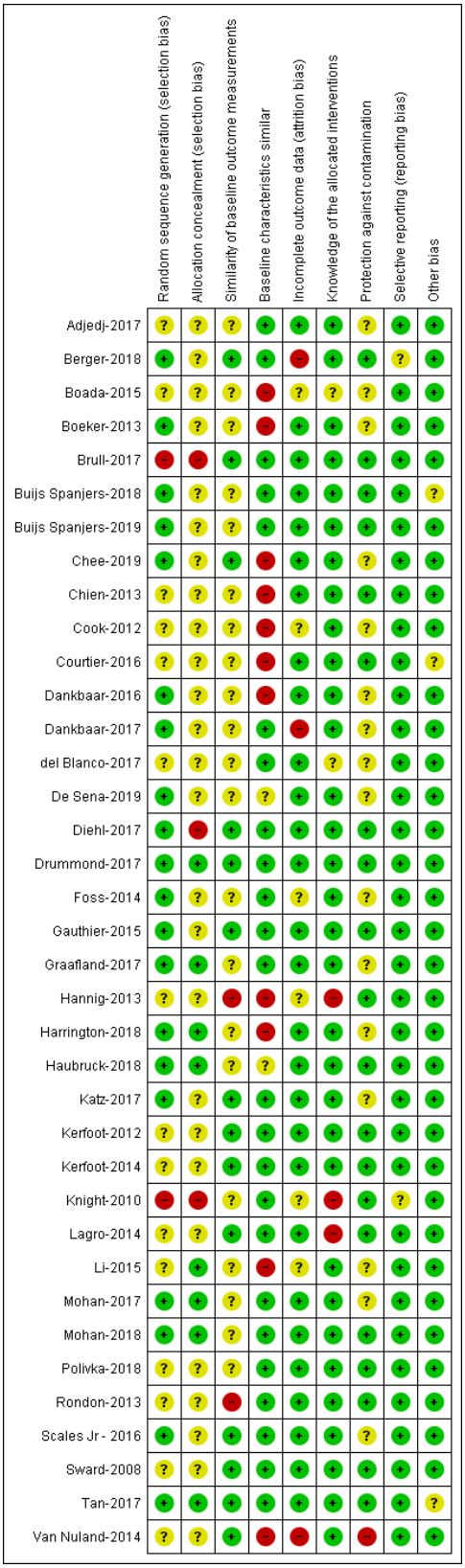
14. Savazzi F, Isernia S, Jonsdottir J, Di Tella S, Pazzi S, Baglio F. Engaged in learning neurorehabilitation: Development and validation of a serious game with user-centered design. *Computers & Education.* 2018;125:53-61.
15. Verkuyl M, Romaniuk D, Mastrilli P. Virtual gaming simulation of a mental health assessment: A usability study. *Nurse Educ Pract.* 2018;31:83-87.

No engagement or learning outcome

1. Hawkins GE, Rae B, Nesbitt KV, Brown SD. Gamelike features might not improve data. *Behav Res Methods.* 2013;45(2):301-318.
2. Kowalewski KF, Hendrie JD, Schmidt MW, et al. Validation of the mobile serious game application Touch Surgery for cognitive training and assessment of laparoscopic cholecystectomy. *Surgical endoscopy.* 2017;31(10):4058-4066.
3. Mohan D, Angus DC, Ricketts D, et al. Assessing the validity of using serious game technology to analyze physician decision making. *PloS one.* 2014;9(8):e105445.

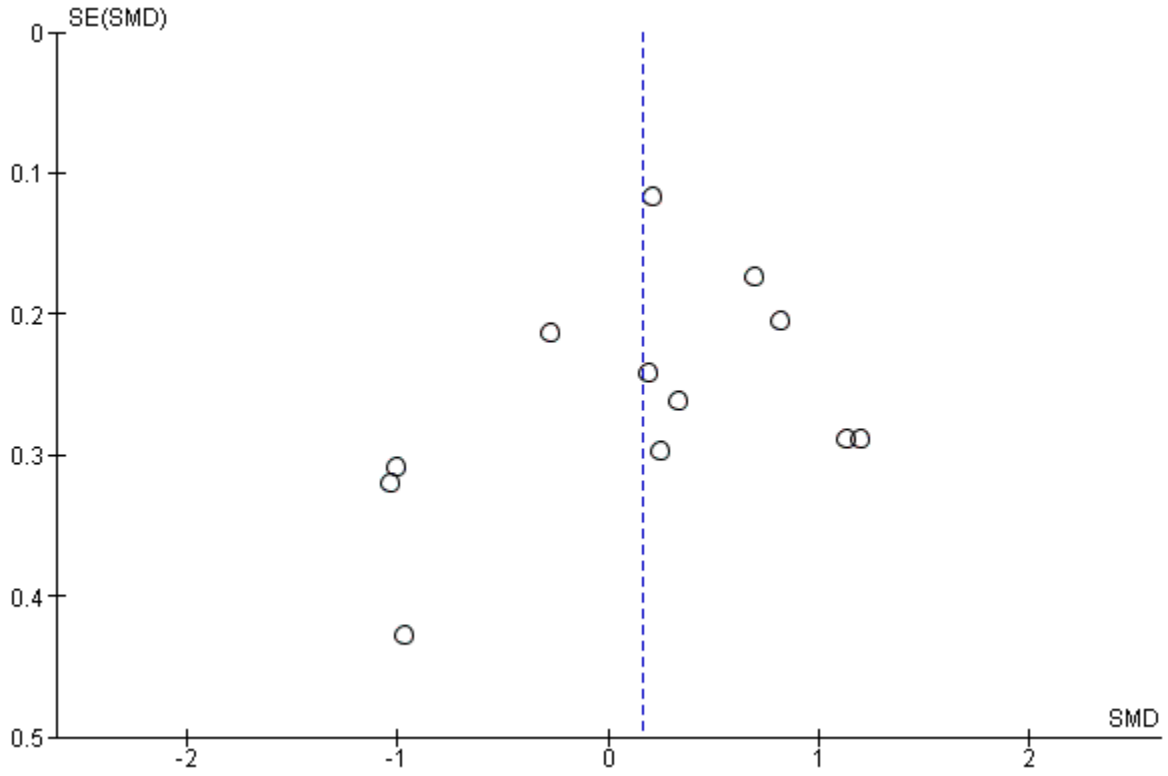
Supplementary content 5 -
Risk of bias summary

Risk of bias summary



Supplementary content 6 -
Funnel plot for knowledge
acquisition

Figure. Funnel plot of the knowledge acquisition outcome.



Note. SE: Standard error; SMD: Standardized mean difference.

Supplementary content 7 -
Key information of the
serious games assessed

Table. Key information of the serious games assessed

First author-year	Name of the serious game, platform	Clinical topic	Theoretical framework for the design or the development process of the serious game	Planned frequency (# of session) and duration (# of minutes) of usage	Challenge
Adjedj-2017	Name non-reported, computer and tablets	Atrial fibrillation management	Not reported	A single 30-minute session	To correctly examine various patients and provide them with the correct treatment in a medical office.
Berger-2018	Name non-reported, platform non-reported	Pharmacist triage	Not reported	A single session, duration not reported	To perform an adequate pharmacist triage and to provide a correct intervention in response.
Boada-2015	Life Support Simulation Activities (LISSA), computer	Cardiopulmonary resuscitation	Not reported	Not reported	To save a character from sudden cardiac arrest by performing cardiopulmonary resuscitation.
Boeker-2013	Uro-Island, computer	Phase contrast microscopic urinalysis	Not reported	Not reported	To free a character from an island by formulating clear diagnosis regarding various urine pathologies.
Brull-2017	World of Salus, computer	Pain management, wound management, and fall prevention and management.	Not reported	Frequency not reported, two to four hours of usage	Not reported

First author-year	Name of the serious game, platform	Clinical topic	Theoretical framework for the design or the development process of the serious game	Planned frequency (# of session) and duration (# of minutes) of usage	Challenge
Buijs-Spangers-2019	Delirium Experience, computer	Delirium management	Not reported	Not reported	To provide inadequate management of a patient diagnosed with delirium
Buijs-Spangers-2018	Delirium Experience, computer	Delirium management	Not reported	A single 20-minute session	To take proper care of a patient diagnosed with delirium
Chien-2013	Name non-reported, computer	Bimanual carrying and peg transfer	Not reported	A single 40-minute session	Not reported
Chee-2019	Play-learn inhalation game, computer and iOS	Administration of inhaled medication	Technology, pedagogy, and content knowledge framework	A single 10-minute session	To answer the most questions and to match the most pictures correctly.
Cook-2012	PULSE, computer, PlayStation Portable	Cardiopulmonary resuscitation	Not reported	Unlimited access for two weeks	To solve various clinical scenarios by using clinical equipment.
Courtier-2016	Tic-tac-toe, computer	Imaging study	Not reported	A single one-hour session	To form a tic-tac-toe before the other competing team by answering questions correctly.

First author-year	Name of the serious game, platform	Clinical topic	Theoretical framework for the design or the development process of the serious game	Planned frequency (# of session) and duration (# of minutes) of usage	Challenge
Dankbaar-2016	abcdeSIM, computer	Acutely-ill patients (e.g., bleeding)	Not reported	Frequency not reported, two to four hours of usage	To stabilize within 15 minutes patients presenting to an emergency department.
Dankbaar-2017	Air Medic Sky-1, computer	Patient safety awareness and personal stress management	Not reported	Frequency not reported, three to four hours of usage	To watch videos, to perform breathing exercises, and to diagnose and treat patients in a virtual flying hospital over the globe.
Del Blanco-2017	Operating Theater Game, computer	Functioning of an operating theater	Not reported	Not reported. However, participants had access to the game for a day.	Not reported
de Sena-2019	Name non-reported, iOS	Cardiopulmonary resuscitation	Not reported	A single 20-minute session	To identify a victim of cardiac arrest and to perform cardiopulmonary resuscitation.
Diehl-2017	InsuOnline, computer	Diabetes management	Adult- and problem-based learning	Frequency not reported, four hours of usage	To improve blood sugar management in patients diagnosed with diabetes

First author-year	Name of the serious game, platform	Clinical topic	Theoretical framework for the design or the development process of the serious game	Planned frequency (# of session) and duration (# of minutes) of usage	Challenge
Drummond-2017	*Staying Alive*, computer, electronic tablet	Sudden cardiac arrest	Not reported	Two sessions, 12 minutes of usage	To save a character from sudden cardiac arrest by performing cardiopulmonary resuscitation
Foss-2014	The Medication Game, computer	Basic mathematical concepts in medication calculation	Multiple intelligence theory	Not reported	To perform various mathematical calculations
Gauthier-2015	Vascular Invaders, computer	Human vascular anatomy	Evidence-centered design framework	Not reported. However, 35 days of free use.	To travel in a nanobot through various vascular vessels to destroy invaders.
Graafland-2017	Dr. Game, Surgeon Trouble, iOS and Android platforms	Equipment problems of the laparoscopic tower	Non-reported	Two 30-minute sessions	To align three similar titles (in a title-matching design) while solving laparoscopic equipment-related issues.
Hannig-2013	Skills-O-Mat, computer	Alginate mixing in dentistry	Peyton's method	Not reported	To mix alginate at the correct speed.
Harrington-2018	Underground, Nintendo WiiU	Laparoscopic technical skills	Not reported	Frequency not reported, 20 hours of usage	To build paths by moving objects in a maze.

First author-year	Name of the serious game, platform	Clinical topic	Theoretical framework for the design or the development process of the serious game	Planned frequency (# of session) and duration (# of minutes) of usage	Challenge
Haubruck-2019	Touch Surgery, iOS	Chest tube insertion	Not reported	A single 120-minute session	To answer the most questions correctly.
Katz-2017	Orthotopic liver transplant Trainer, iOS and Android platforms	Management of orthotopic liver transplant	Not reported	At least once per week for a month, duration of usage non-reported	To properly assess and manage a patient during the peri- and the per-operative periods to gain credits.
Kerfoot-2012	Name non-reported, computer (through e-mails)	Urology clinical practice guidelines	Spaced education	Two to four questions sent per e-mail every two to four days for 34 weeks	To answer the most questions correctly.
Kerfoot-2014	Name non-reported, computer (through e-mails)	Hypertension management	Salen and Zimmerman game design fundamentals Spaced education	One to two questions sent every three days for 52 weeks	To answer the most questions correctly.
Knight-2010	Triage trainer, computer	Basic incident triage sieve skills	Not reported	A single 60-minute session	As a first-respondent at a major incident scene, to assign the right priority to each casualty.

First author-year	Name of the serious game, platform	Clinical topic	Theoretical framework for the design or the development process of the serious game	Planned frequency (# of session) and duration (# of minutes) of usage	Challenge
Lagro-2014	GeriatriX, computer	Geriatric medical decision-making and cost consciousness	Not reported	Frequency not reported; 60 to 90 minutes of usage	To provide proper assessment and management to elderly patients.
Li-2015	3D CPR game, platform non-reported	Cardiopulmonary resuscitation	Not reported	Not reported	To save a character from sudden cardiac arrest by performing cardiopulmonary resuscitation.
Mohan-2018	Shift: The Next Generation, iOS	Trauma triage	Unspecified behavioral learning theories	Frequency not reported, two hours of usage	To triage a prespecified number of patients in the emergency department under time pressure.
Mohan-2017, 2018	Night Shift, electronic tablet	Trauma triage	Narrative engagement and unspecified behavioral learning theories	Mohan, 2017: A single one-hour session Mohan-2018: Frequency not reported, two hours of usage	To provide proper assessment and management to patients with severe injuries in the emergency department while solving an in-game mystery.

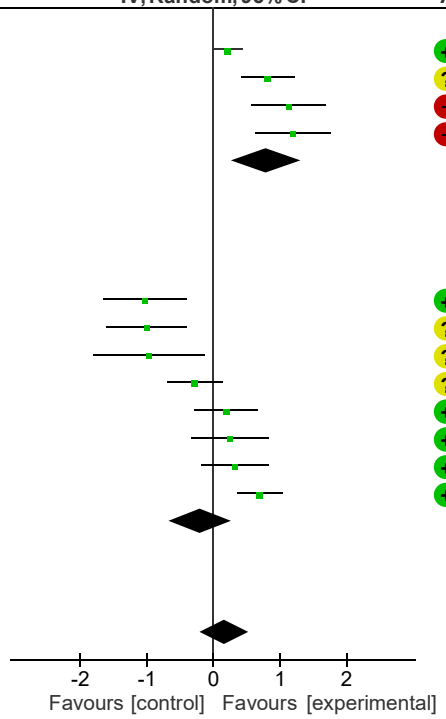
First author-year	Name of the serious game, platform	Clinical topic	Theoretical framework for the design or the development process of the serious game	Planned frequency (# of session) and duration (# of minutes) of usage	Challenge
Polivka-2018	HH-VSTS, computer	Health and safety hazards	Not reported	Not reported	To identify potential health and safety hazards in a home.
Rondon-2013	Anatesse 2.0, computer	Anatomy and physiology of the speech, language, hearing, and swallowing mechanisms	Not reported	Nine sessions (once per week), duration of usage not reported	To answer the most questions correctly.
Scales-2016	Name non-reported, computer (through e-mails)	Quality improvement and patient safety	Content retrieval	Two questions twice a week	To answer the most questions correctly.
Sward-2008	Name non-reported, computer	Pediatric	Not reported	Four one-hour sessions	To answer the most questions correctly to progress through an electronic game board.
Tan-2017	Name non-reported, computer	Blood transfusion	Experiential gaming model [6]	A single 30-minute session	Various challenges related to blood transfusion (e.g., choosing the correct material, checking if the blood product is right)

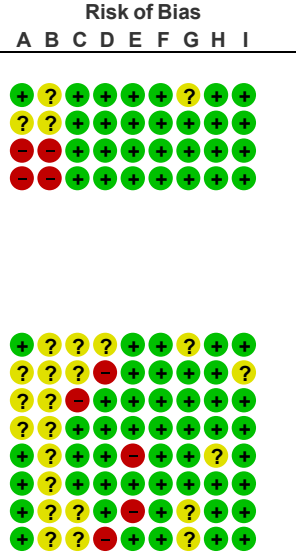
First author-year	Name of the serious game, platform	Clinical topic	Theoretical framework for the design or the development process of the serious game	Planned frequency (# of session) and duration (# of minutes) of usage	Challenge
Van Nuland-2014	Online Competitive Anatomy Tournament, computer	Functional anatomy	Not reported	A single 20-minute session	To answer the most questions correctly.

Supplementary content 8 -
Other sensitivity and
subgroup analyses

1.1 Knowledge

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.1.1 Healthcare professionals								
Scales Jr - 2016	34	35	133	27	33	166	9.7%	0.21 [-0.02, 0.43]
Kerfoot-2014	90	8	52	78	19	52	9.0%	0.82 [0.42, 1.22]
Brull-2017	45	2.3	26	41.1	4.1	31	8.1%	1.13 [0.57, 1.69]
Brull-2017	45	2.3	26	41.5	3.3	32	8.1%	1.19 [0.63, 1.76]
Subtotal (95% CI)			**237**			**281**	**34.8%**	**0.80 [0.27, 1.32]**
Heterogeneity: Tau² = 0.23; Chi² = 19.31, df = 3 (P = 0.0002); I² = 84%								
Test for overall effect: Z = 2.98 (P = 0.003)								
1.1.2 Healthcare students								
De Sena-2019	6.51	1	23	7.56	1	22	7.7%	-1.03 [-1.66, -0.41]
Courtier-2016	3.6	0.76	25	4.5	1	23	7.9%	-1.00 [-1.61, -0.40]
Rondon-2013	33.2	3	13	36.1	2.78	12	6.5%	-0.97 [-1.81, -0.13]
Sward-2008	77.8	17.08	51	82.6	17.08	39	8.9%	-0.28 [-0.70, 0.14]
Berger-2018	7.07	1.5	41	6.81	1.1	30	8.6%	0.19 [-0.28, 0.66]
Gauthier-2015	9.75	2.15	24	9.23	2.07	22	8.0%	0.24 [-0.34, 0.82]
Dankbaar-2017	60.1	6.7	28	57.9	6.5	32	8.4%	0.33 [-0.18, 0.84]
Boeker-2013	28.6	3.53	81	26	3.99	63	9.3%	0.69 [0.35, 1.03]
Subtotal (95% CI)			**286**			**243**	**65.2%**	**-0.19 [-0.66, 0.28]**
Heterogeneity: Tau² = 0.38; Chi² = 46.73, df = 7 (P < 0.00001); I² = 85%								
Test for overall effect: Z = 0.78 (P = 0.44)								
Total (95% CI)			**523**			**524**	**100.0%**	**0.16 [-0.20, 0.52]**
Heterogeneity: Tau² = 0.33; Chi² = 81.46, df = 11 (P < 0.00001); I² = 86%								
Test for overall effect: Z = 0.88 (P = 0.38)								
Test for subgroup differences: Chi² = 7.50, df = 1 (P = 0.006), I² = 86.7%								



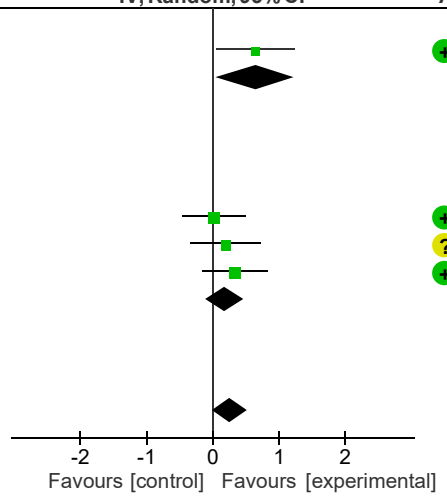


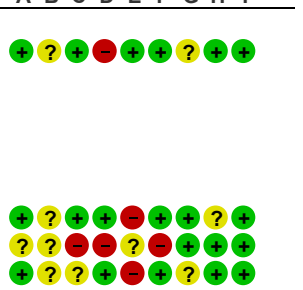
Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

1.2 Confidence in skills

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.2.1 Healthcare professionals								
Chee-2019	40.13	6.84	23	35.26	8.01	23	19.0%	0.64 [0.05, 1.24]
Subtotal (95% CI)			**23**			**23**	**19.0%**	**0.64 [0.05, 1.24]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.12 (P = 0.03)								
1.2.2 Healthcare students								
Berger-2018	4.61	1.3	41	4.58	1.28	29	29.7%	0.02 [-0.45, 0.50]
Hannig-2013	32.28	1.89	30	31.92	1.69	25	23.7%	0.20 [-0.34, 0.73]
Dankbaar-2017	67.9	7.6	32	64.7	10.7	32	27.5%	0.34 [-0.15, 0.83]
Subtotal (95% CI)			**103**			**86**	**81.0%**	**0.18 [-0.11, 0.47]**
Heterogeneity: Tau² = 0.00; Chi² = 0.83, df = 2 (P = 0.66); I² = 0%								
Test for overall effect: Z = 1.24 (P = 0.22)								
Total (95% CI)			**126**			**109**	**100.0%**	**0.27 [0.01, 0.53]**
Heterogeneity: Tau² = 0.00; Chi² = 2.70, df = 3 (P = 0.44); I² = 0%								
Test for overall effect: Z = 2.04 (P = 0.04)								
Test for subgroup differences: Chi² = 1.87, df = 1 (P = 0.17), I² = 46.5%								

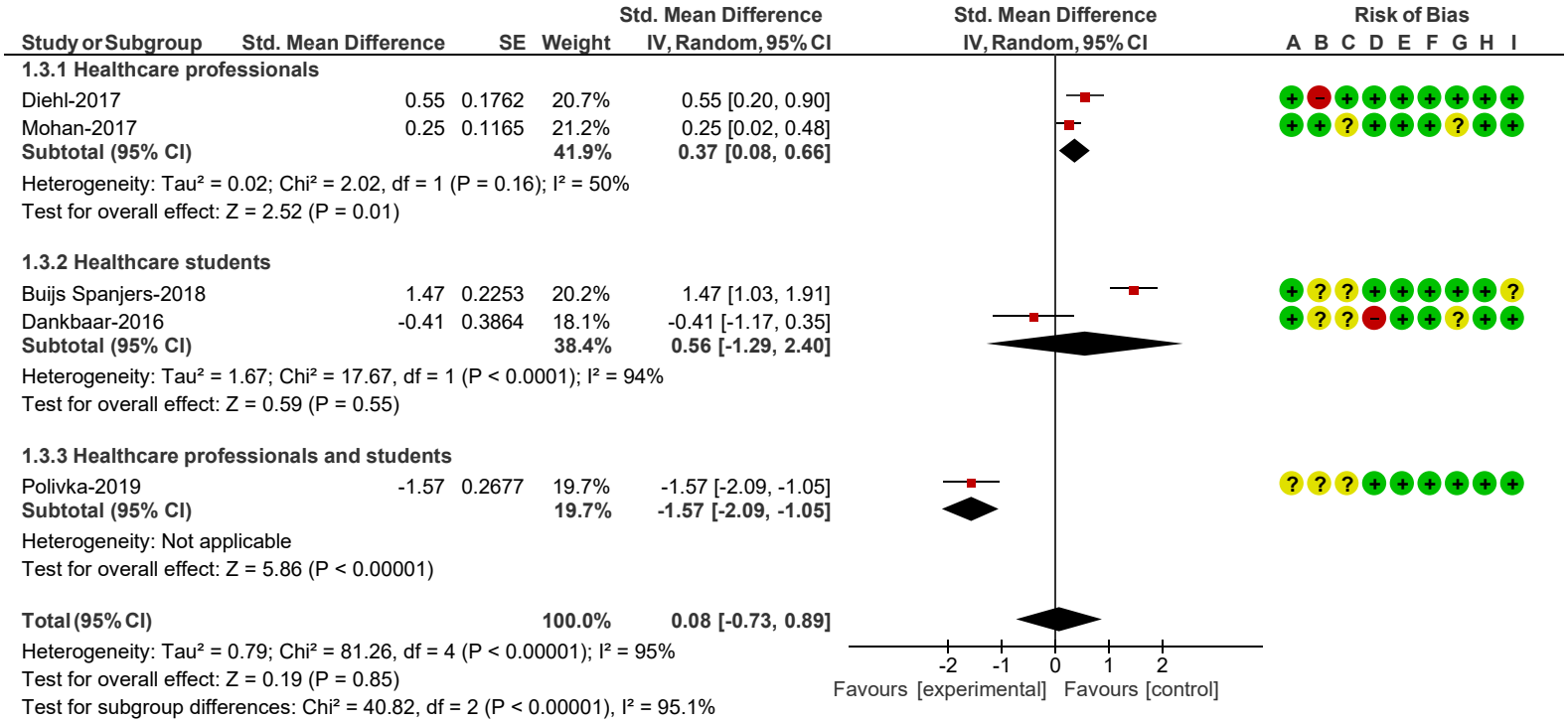




Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

1.3 Cognitive skills

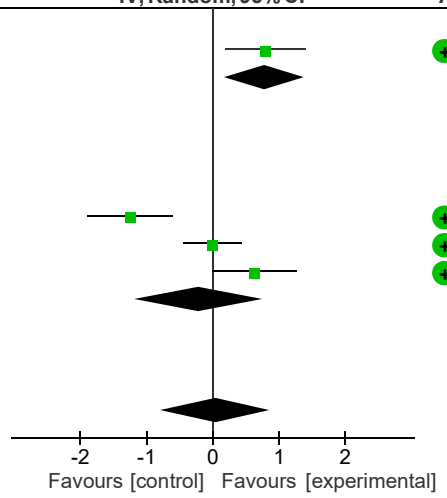


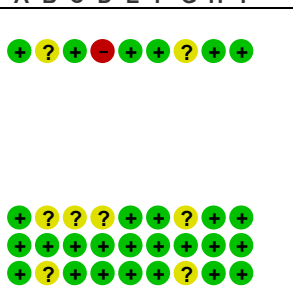
Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

1.4 Procedural skills

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.4.1 Healthcare professionals								
Chee-2019	2.39	0.99	23	1.65	0.83	23	24.8%	0.80 [0.19, 1.40]
Subtotal (95% CI)			**23**			**23**	**24.8%**	**0.80 [0.19, 1.40]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.59 (P = 0.010)								
1.4.2 Healthcare students								
De Sena-2019	8.4	1	23	9.67	1	22	24.3%	-1.25 [-1.89, -0.60]
Drummond-2017	6	3.08	40	6	1.54	39	26.4%	0.00 [-0.44, 0.44]
Katz-2017	28.7	5.45	20	25.18	5.61	20	24.4%	0.62 [-0.01, 1.26]
Subtotal (95% CI)			**83**			**81**	**75.2%**	**-0.20 [-1.17, 0.76]**
Heterogeneity: Tau² = 0.64; Chi² = 17.29, df = 2 (P = 0.0002); I² = 88%								
Test for overall effect: Z = 0.41 (P = 0.68)								
Total (95% CI)			**106**			**104**	**100.0%**	**0.05 [-0.78, 0.87]**
Heterogeneity: Tau² = 0.61; Chi² = 24.67, df = 3 (P < 0.0001); I² = 88%								
Test for overall effect: Z = 0.11 (P = 0.91)								
Test for subgroup differences: Chi² = 2.95, df = 1 (P = 0.09), I² = 66.1%								



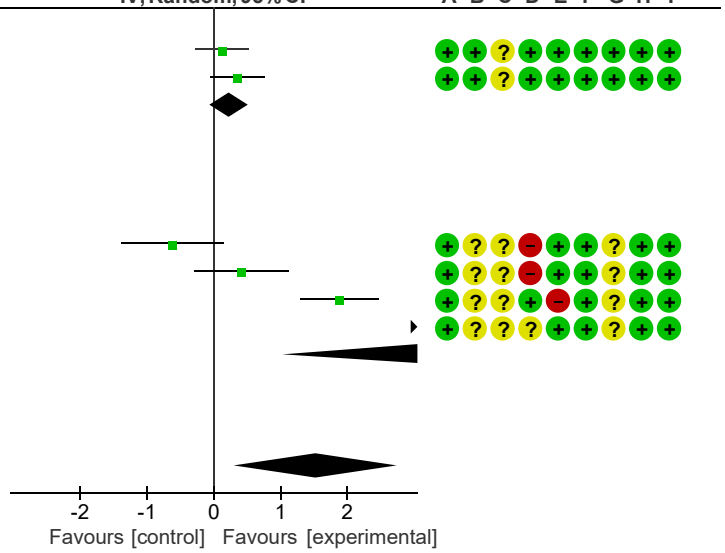


Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

1.5 Learning time with the intervention

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
1.5.1 Healthcare professionals								
Mohan-2018	105	34.11	66	100	45.38	37	19.3%	0.13 [-0.27, 0.53]
Mohan-2018	117.67	50.78	66	100	45.38	36	19.3%	0.36 [-0.05, 0.77]
Subtotal (95% CI)			**132**			**73**	**38.6%**	**0.24 [-0.05, 0.53]**
Heterogeneity: Tau² = 0.00; Chi² = 0.61, df = 1 (P = 0.43); I² = 0%								
Test for overall effect: Z = 1.65 (P = 0.10)								
1.5.2 Healthcare students								
Dankbaar-2016	90	49	12	133	78	16	18.3%	-0.62 [-1.39, 0.15]
Dankbaar-2016	90	49	13	70	46	20	18.5%	0.41 [-0.29, 1.12]
Dankbaar-2017	174	66	32	54	60	34	18.9%	1.88 [1.30, 2.47]
De Sena-2019	18.57	0.66	23	7.41	0.43	22	5.7%	19.59 [15.31, 23.87]
Subtotal (95% CI)			**80**			**92**	**61.4%**	**3.57 [1.03, 6.11]**
Heterogeneity: Tau² = 5.85; Chi² = 100.97, df = 3 (P < 0.00001); I² = 97%								
Test for overall effect: Z = 2.76 (P = 0.006)								
Total (95% CI)			**212**			**165**	**100.0%**	**1.53 [0.32, 2.75]**
Heterogeneity: Tau² = 1.94; Chi² = 109.90, df = 5 (P < 0.00001); I² = 95%								
Test for overall effect: Z = 2.48 (P = 0.01)								
Test for subgroup differences: Chi² = 6.52, df = 1 (P = 0.01), I² = 84.7%								



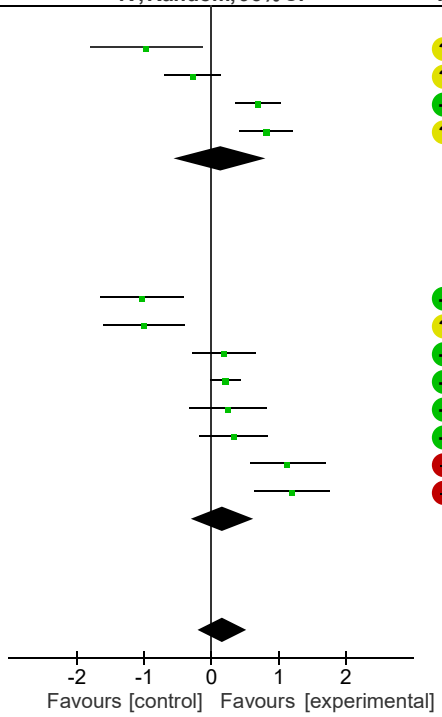
Risk of bias legend

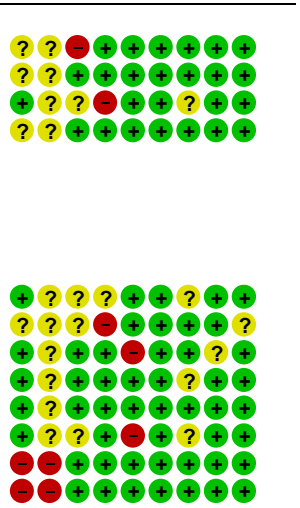
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

2 Publisher before or in 2014 v. published after 2014

2.1 Knowledge

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 Published before or in 2014								
Rondon-2013	33.2	3	13	36.1	2.78	12	6.5%	-0.97 [-1.81, -0.13]
Sward-2008	77.8	17.08	51	82.6	17.08	39	8.9%	-0.28 [-0.70, 0.14]
Boeker-2013	28.6	3.53	81	26	3.99	63	9.3%	0.69 [0.35, 1.03]
Kerfoot-2014	90	8	52	78	19	52	9.0%	0.82 [0.42, 1.22]
Subtotal (95% CI)			**197**			**166**	**33.6%**	**0.14 [-0.54, 0.82]**
Heterogeneity: Tau² = 0.41; Chi² = 27.26, df = 3 (P < 0.00001); I² = 89%								
Test for overall effect: Z = 0.39 (P = 0.70)								
2.1.2 Published after 2014								
De Sena-2019	6.51	1	23	7.56	1	22	7.7%	-1.03 [-1.66, -0.41]
Courtier-2016	3.6	0.76	25	4.5	1	23	7.9%	-1.00 [-1.61, -0.40]
Berger-2018	7.07	1.5	41	6.81	1.1	30	8.6%	0.19 [-0.28, 0.66]
Scales Jr - 2016	34	35	133	27	33	166	9.7%	0.21 [-0.02, 0.43]
Gauthier-2015	9.75	2.15	24	9.23	2.07	22	8.0%	0.24 [-0.34, 0.82]
Dankbaar-2017	60.1	6.7	28	57.9	6.5	32	8.4%	0.33 [-0.18, 0.84]
Brull-2017	45	2.3	26	41.1	4.1	31	8.1%	1.13 [0.57, 1.69]
Brull-2017	45	2.3	26	41.5	3.3	32	8.1%	1.19 [0.63, 1.76]
Subtotal (95% CI)			**326**			**358**	**66.4%**	**0.17 [-0.30, 0.63]**
Heterogeneity: Tau² = 0.38; Chi² = 52.73, df = 7 (P < 0.00001); I² = 87%								
Test for overall effect: Z = 0.71 (P = 0.48)								
Total (95% CI)			**523**			**524**	**100.0%**	**0.16 [-0.20, 0.52]**
Heterogeneity: Tau² = 0.33; Chi² = 81.46, df = 11 (P < 0.00001); I² = 86%								
Test for overall effect: Z = 0.88 (P = 0.38)								
Test for subgroup differences: Chi² = 0.01, df = 1 (P = 0.94), I² = 0%								



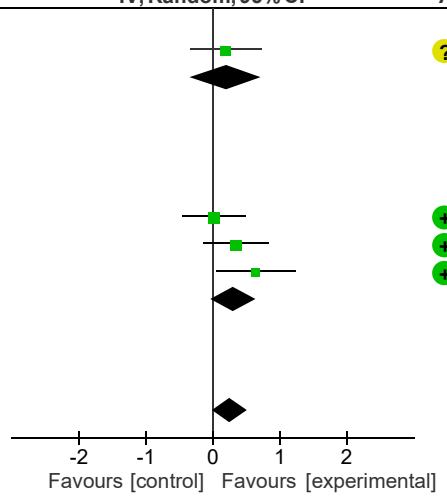


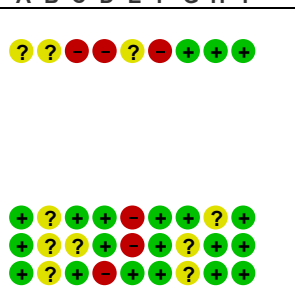
Risk of bias legend

(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

2.2 Confidence in skills

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 Published before or in 2014								
Hannig-2013	32.28	1.89	30	31.92	1.69	25	23.7%	0.20 [-0.34, 0.73]
Subtotal (95% CI)			**30**			**25**	**23.7%**	**0.20 [-0.34, 0.73]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.73 (P = 0.47)								
2.2.2 Published after 2014								
Berger-2018	4.61	1.3	41	4.58	1.28	29	29.7%	0.02 [-0.45, 0.50]
Dankbaar-2017	67.9	7.6	32	64.7	10.7	32	27.5%	0.34 [-0.15, 0.83]
Chee-2019	40.13	6.84	23	35.26	8.01	23	19.0%	0.64 [0.05, 1.24]
Subtotal (95% CI)			**96**			**84**	**76.3%**	**0.30 [-0.04, 0.64]**
Heterogeneity: Tau² = 0.02; Chi² = 2.61, df = 2 (P = 0.27); I² = 23%								
Test for overall effect: Z = 1.74 (P = 0.08)								
Total (95% CI)			**126**			**109**	**100.0%**	**0.27 [0.01, 0.53]**
Heterogeneity: Tau² = 0.00; Chi² = 2.70, df = 3 (P = 0.44); I² = 0%								
Test for overall effect: Z = 2.04 (P = 0.04)								
Test for subgroup differences: Chi² = 0.11, df = 1 (P = 0.75), I² = 0%								





Risk of bias legend

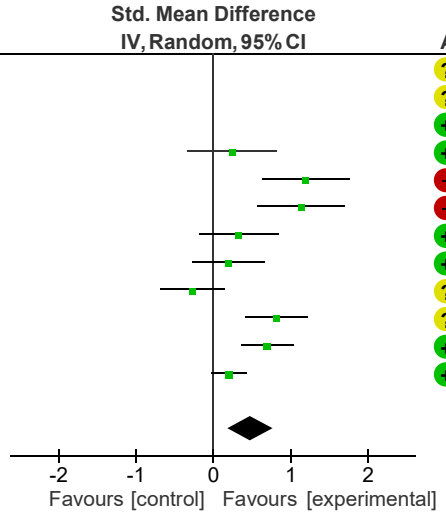
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

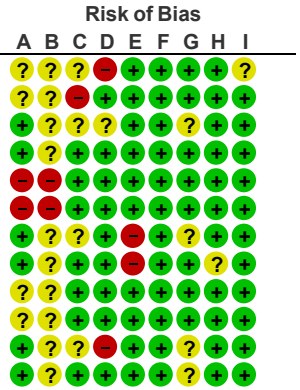
3 Sample size

3.1 Knowledge

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Courtier-2016	3.6	0.76	25	4.5	1	23		Not estimable
Rondon-2013	33.2	3	13	36.1	2.78	12		Not estimable
De Sena-2019	6.51	1	23	7.56	1	22		Not estimable
Gauthier-2015	9.75	2.15	24	9.23	2.07	22	9.5%	0.24 [-0.34, 0.82]
Brull-2017	45	2.3	26	41.5	3.3	32	9.7%	1.19 [0.63, 1.76]
Brull-2017	45	2.3	26	41.1	4.1	31	9.7%	1.13 [0.57, 1.69]
Dankbaar-2017	60.1	6.7	28	57.9	6.5	32	10.4%	0.33 [-0.18, 0.84]
Berger-2018	7.07	1.5	41	6.81	1.1	30	10.9%	0.19 [-0.28, 0.66]
Sward-2008	77.8	17.08	51	82.6	17.08	39	11.6%	-0.28 [-0.70, 0.14]
Kerfoot-2014	90	8	52	78	19	52	11.8%	0.82 [0.42, 1.22]
Boeker-2013	28.6	3.53	81	26	3.99	63	12.6%	0.69 [0.35, 1.03]
Scales Jr - 2016	34	35	133	27	33	166	13.9%	0.21 [-0.02, 0.43]
Total (95% CI)			**462**			**467**	**100.0%**	**0.48 [0.19, 0.78]**

Heterogeneity: Tau² = 0.15; Chi² = 34.96, df = 8 (P < 0.0001); I² = 77%
Test for overall effect: Z = 3.23 (P = 0.001)



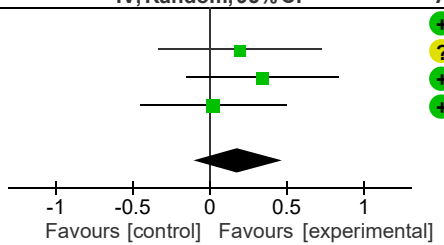


Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

3.2 Confidence in skills

Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Chee-2019	40.13	6.84	23	35.26	8.01	23		Not estimable
Hannig-2013	32.28	1.89	30	31.92	1.69	25	29.3%	0.20 [-0.34, 0.73]
Dankbaar-2017	67.9	7.6	32	64.7	10.7	32	34.0%	0.34 [-0.15, 0.83]
Berger-2018	4.61	1.3	41	4.58	1.28	29	36.7%	0.02 [-0.45, 0.50]
Total (95% CI)			**103**			**86**	**100.0%**	**0.18 [-0.11, 0.47]**

Heterogeneity: Tau² = 0.00; Chi² = 0.83, df = 2 (P = 0.66); I² = 0%
Test for overall effect: Z = 1.24 (P = 0.22)

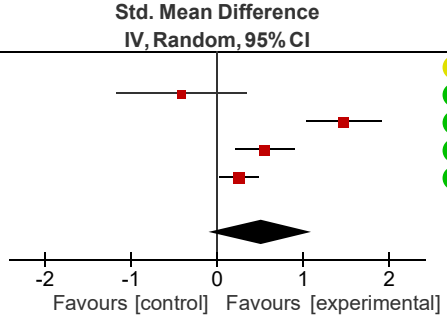


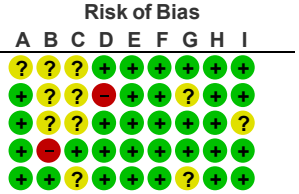
Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

3.3 Cognitive skills

Study or Subgroup	Std. Mean Difference	SE	Weight	Std. Mean Difference IV, Random, 95% CI
Polivka-2019	-1.57	0.2677		Not estimable
Dankbaar-2016	-0.41	0.3864	20.0%	-0.41 [-1.17, 0.35]
Buijs Spanjers-2018	1.47	0.2253	25.3%	1.47 [1.03, 1.91]
Diehl-2017	0.55	0.1762	26.7%	0.55 [0.20, 0.90]
Mohan-2017	0.25	0.1165	28.1%	0.25 [0.02, 0.48]
Total (95% CI)			**100.0%**	**0.51 [-0.09, 1.11]**

Heterogeneity: Tau² = 0.32; Chi² = 28.66, df = 3 (P < 0.00001); I² = 90%
Test for overall effect: Z = 1.65 (P = 0.10)



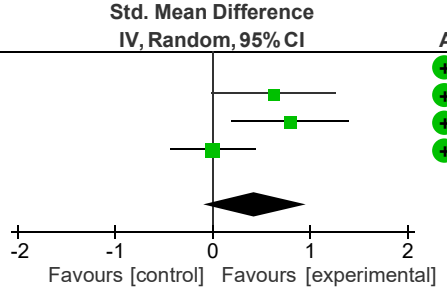


Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

3.4 Procedural skills

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
De Sena-2019	8.4	1	23	9.67	1	22		Not estimable
Katz-2017	28.7	5.45	20	25.18	5.61	20	29.8%	0.62 [-0.01, 1.26]
Chee-2019	2.39	0.99	23	1.65	0.83	23	31.3%	0.80 [0.19, 1.40]
Drummond-2017	6	3.08	40	6	1.54	39	38.9%	0.00 [-0.44, 0.44]
Total (95% CI)			**83**			**82**	**100.0%**	**0.44 [-0.08, 0.95]**

Heterogeneity: Tau² = 0.13; Chi² = 5.23, df = 2 (P = 0.07); I² = 62%
Test for overall effect: Z = 1.64 (P = 0.10)

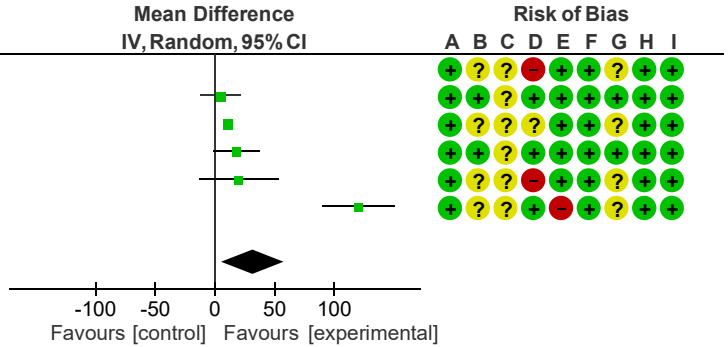


Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias

3.5 Learning time with the intervention

Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI
Dankbaar-2016	90	49	12	133	78	16		Not estimable
Mohan-2018	105	34.11	66	100	45.38	37	21.3%	5.00 [-11.78, 21.78]
De Sena-2019	18.57	0.66	23	7.41	0.43	22	23.5%	11.16 [10.84, 11.48]
Mohan-2018	117.67	50.78	66	100	45.38	36	20.7%	17.67 [-1.56, 36.90]
Dankbaar-2016	90	49	13	70	46	20	16.8%	20.00 [-13.41, 53.41]
Dankbaar-2017	174	66	32	54	60	34	17.6%	120.00 [89.51, 150.49]
Total (95% CI)			**200**			**149**	**100.0%**	**31.88 [6.20, 57.55]**

Heterogeneity: Tau² = 730.81; Chi² = 50.17, df = 4 (P < 0.00001); I² = 92%
Test for overall effect: Z = 2.43 (P = 0.01)



Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Similarity of baseline outcome measurements
(**D**) Baseline characteristics similar
(**E**) Incomplete outcome data (attrition bias)
(**F**) Knowledge of the allocated interventions
(**G**) Protection against contamination
(**H**) Selective reporting (reporting bias)
(**I**) Other bias