Systematic Review and Meta-Analysis in Medical Education

心臟血管外科

吳協兆醫師

2024/02/24

緣起

- 教學之即時性需求高，研究耗費時日
- 醫學教育多樣性高，方式眾多
- 個人可能未能完全掌握所有重點

系統性回顧研究
Systematic Review

- 一種科學性的調查
- 聚焦於一個特定問題
- 使用一個清楚且事先特定的科學方法去確認、選擇、評估與總結一些相似研究的所見

系統性回顧研究的優勢

- 憑藉一個清楚定義的流程，可以對於某一議題進行確認、評值與篩選一些相關的研究論文，而可提出作為上的建議。
- 主要目的：對於一個被清楚定義的族群，在接受某種介入之後的結果做比較分析，而能回答一個焦點性的特定問題。
- 憑藉清楚的指引所完成的系統性回顧研究可以減少偏差、增加透明度與提升重現度。

系統性回顧研究的缺點

- 因系統性回顧研究有其焦點，故無法提供廣泛的綜合論述。
- 系統性回顧研究通常不是用來回答「如何」與「為何」某種介入的有效或無效。
- 醫教族群的多樣性高，不易嚴格定義，而介入處置亦會有少許甚至明顯的差異；所謂的控制組亦不易嚴格控制。
- 在某議題的論文發表數不夠多時，亦不易進行系統性回顧研究。
- 研究耗時，會影響被應用的及時性。

Systematic Review
Meta-Analysis

- 以系統性回顧的方式，對於同一議題之多個研究進行整合的方法
- 一個嘗試得到整體結論的統計方法
- 為醫療方式比較與選擇的常用研究方法之一
- Systematic review：系統性回顧
- Meta-analysis：統合分析

醫療論文 VS 醫學教育論文

- 醫療論文：較為強調醫療效果之比較與醫療方式之選擇
- 醫學教育論文：較為強調探討與分享醫學教育的方法與成效

實證醫學


Levels of Evidence
Meta Analyses
Randomised Controlled Trials
Cohort Studies
Case-Control Studies
Case-Series or Case-Reports
Editorials and Expert Opinion

EBM (evidence based medicine)

實證醫學

vs

EBE (evidence based education)

實證教育

醫學教育論文

- 醫學教育的方法多多益善
- 醫學教育的方法並不需選擇使用，常常可以併用。

Systematic Review in醫學教育論文

- Systematic review：可以蒐集到對於某一議題之眾人意見與作法，可強化對於某一議題之認識，從而增進教學方法之多樣性與教學內容之完整性

- 醫教論文之常見論文型態之一

AI於醫療照護與醫學教育之應用

Table 1. Characteristics of included studies

Authors	Year of publication	Title	Main research objectives	Applications	Level of Evidence*	Study results
Wang Yan [14]	2022	Reform status and exploration of higher medical education under the background of artificial intelligence.	Proposes important measures to promote medical education in the age of AI.	Complementary medicine; Medical robot; Intelligent health management.	V	It is recommended that the government introduce relevant policies to support the development of AI.
Zou Luxi [15]	2021	Research on the current situation and problems of applying artificial intelligence in medical education.	Advancing research on AI in the field of medical education.	Apple Watch; Smartphone monitoring system; Predicting disease risk.	V	Increase investment and focus in "AI+Healthcare Education".
Dai Shaochun [16]	2021	Development Prospect of Artificial Intelligence in Medical Assistant Education.	Introducing the future of AI in paramedical education.	Analysis of the learning situation; Personalized Learning.	V	Summarized the possible development direction of future artificial intelligence.
Ai Feiyan [31]	2021	The Application of Artificial Intelligence in Diagnostics Teaching.	Analyze the disadvantages of AI teaching and the advantages of teaching.	Teaching Diagnostics.	V	Academic hardware and software equipment still needs to be strengthened, and teachers need to fully understand artificial intelligence.
Liu Dalu [23]	2021	Cultivating practical literacy of machine learning for medical students.	Describe the current state of machine learning practice goals.	Auxiliary diagnosis; Help with treatment decisions.	IV	Work still needs to continue on adding machine learning to medical student practice.
Li Xinchun [24]	2021	Application Prospects of Artificial Intelligence Assisting Teaching Mode in Medical Imaging.	Discussing the value of AI to inform training in Integrative Medicine.	Promote the integration and optimization of impact expertise and educational resources; Help promote the construction of a new type of education; Promoting faculty development.	V	Artificial intelligence + medical education will be the direction of the times.

VR運用於醫學教育的經驗

Table 1: Listed here are the results of the search and a summary of each of the major studies included from the systematic review

Authors	Name of paper	Topic (s) tested	Study size	Groups	Results
Hu *et al.*	Impact of virtual reality anatomy training on ultrasound competency development: A randomized controlled trial	Three stations: - Urinary structures/great vessels - Cardiovascular structures - Hepatobiliary structures	n=101 (52 males, 49 females, down 14 students from original n=115 due to students receiving informal ultrasound training after recruitment. Original gender balance was 61 males, 54 females)	Two groups: Control: n=54 (down 4 from original n=58 due to students receiving informal ultrasound training after recruitment, changing eligibility status) Intervention: n=47 (down 10 from original n=57 due to students receiving informal ultrasound training after recruitment, changing eligibility status)	Results demonstrate the overall benefit from using virtual reality to teach ultrasonography to new learners. Time spent for tasks was slightly less in intervention group though not significantly ($p=0.12$), the experimental group had significantly higher scores for identification of structures via ultrasonography compared to control ($p<0.01$), the intervention group showed greater improvement in ultrasonography test scores than control group and in non-ultrasound anatomy image test scores despite the latter scores still being significantly lower than the control group. Additional results can be found in the original article.
Imai *et al.*	Incorporation of virtual reality in the clinical training of medical students studying esophageal and mediastinal anatomy and surgery.	Esophageal/ mediastinal anatomy and esophageal surgery	n=60 4th–5th year medical students, narrowed down from a larger unspecified number of medical students. Elimination was due to rescinding previously provided consent to participate in study	Two groups: 3D (Control): n=30 virtual reality (Intervention): n=30	The virtual reality group performed significantly better on understanding 3D CT images and interpreting surgical images. The virtual reality group was also significantly more interested in anatomy and surgery. No significant difference was found between groups regarding opinions of the learning process, ease of understanding lectures, development of knowledge bases, or ongoing desire to learn non-mediastinal anatomy. Additional results can be found in the original article.

教育對於醫學生的同理心與熱情之影響

TABLE 1 | A tabulated summary of the 24 studies in this systematic review.

Study	Study population	Study design	Curriculum design	Empathy topics addressed	Primary outcome (effect size and *P*-value where available)[a]	Quality asses (MERSQI)[b]
Physician-patient interaction						
Beard et al. (51)	***N*:** 10 **Level of training:** Third year	Controlled Trail 2 groups	**Modality:** Longitudinal integrated clerkships—VALUE **Frequency/Duration:** 10 months	A respect for a patient's values and preferences/a clear patient physician communication/A well-coordinated care	**Patient outcome**–A greater sense of satisfaction reported by VALUE patients with their health care providers in terms of explanations provided, knowledge of patients' history, and their best interests ($P <$ 0.05)	16.5
Collins et al. (52)	***N*:** 45 **Level of training**: Third and fourth year	Controlled trial (control-no intervention)	**Modality**: Student hot spotting/IPE/apprenticeship/supervision **Frequency/Duration**:6 months	Patient centered approach/Partnership for a personalized self-management plan	**Knowledge:** ATHI, JSE; A higher post-test score in terms of self-efficacy and empathy (participants Vs. controls) ($P =$ 0.05).	10.5
D'souza et al. (53)	***N*:** 82 **Level of training:** Second year	RCT	**Modality**: Didactic PowerPoint, video clips, and roleplay and simulation **Frequency/Duration**: Single session –2 h	Empathetic communication	**Self-report**: JSE: a difference in empathy score (control vs. intervention) ($p = 0.014$) with a decline at 3-week follow-up ($p =$ 0.020)	12
Kataoka (62)	***N*:** 69 **Level of training**: Year 1–6	Single group, pre and posttest	**Modality:** didactics case-based discussions; simulation with standardized patients, feedback provisions **Frequency/Duration:** Three 4-h workshops over a period of two years	Communication skills and medical interviewing	**Self-report:** JSE: an immediate significant increase (SD = 10.0) in post-test mean score ($p <$ 0.0001), however, the mean score bounced back to the pre-test level in year 5 (SD =1 2.9) and year 6 (SD = 13.8)	10
Modi et al. (65)	***N*:** 188 **Level of training**: First to third year	Controlled trial (control-no intervention)	**Modality:** Service learning experience—student run free clinic—socialization-mentorship **Frequency/Duration**: Weekly student run clinics over a period of 3 years	Early and consistent exposure to poor and underserved Patients—hidden curriculum—implicit to explicit	**Self-report**: JSE: A drop in mean empathy scores for both volunteers (2.2 points) ($P = 0.07$, effect size = 0.20), and non-volunteers (3.1 points) ($P =$ 0.009, effect size > 0.25)	10

TABLE 4 | Educational interventions with frequency and duration used in the selected studies ($n = 24$).

Teaching modalities	No. (percentages) of studies		Publications references	
	Total no. (%)	Effective no. (%)[a]	[a]Effective	[b]Ineffective
Single modality	4/24 (16.6%)	3/4 (75%)	(60, 69, 73)	(71)
Multimodality (Didactics, workshops, simulation, reflection)	20/24 (83.3%)	18/20 (90%)	(23, 51, 52, 54–59, 61, 63–68, 70, 72)	(53, 62)
Didactics (lectures, presentations, power-point, assessments, seminars, discussions)	9/24 (37.5%)	6/9 (66.6%)	(59, 61, 64, 66, 70, 73)	(53, 62, 71)
Small group/Case-Based discussion/workshops	8/24 (33.3%)	7/8 (87.5%)	(56–58, 61, 63, 66–68)	(62)
Simulation	7/24 (29.1%)	5/7 (71.4%)	(57, 59, 66, 68, 72)	(53, 62)
Role modeling/mentorship /interprofessional education	6/24 (25%)	6/6 (100%)	(23, 51, 52, 55, 65, 67)	

Meta-Analysis in醫學教育論文

- 有較強的統計與比較之意涵

- 分析 vs 比較

Table 1. Effect of Intervention on knowledge level.

Study number	Author, year & country	Outcome of interest	Summary statistics (Percentage of mark/ Score by using mobile application (Mean±SD, P Value))	Sample size (I: Intervention; C: Control)	Duration	Effectiveness of mobile application
1	Bonabi et al, 2019, Iran [14]	Change in knowledge level	**Mean knowledge score:** Cont: Pre test: 8.17 ± 2.03; Post test: 10.43 ± 1.8 (**P<0.001**); Int: Pre test: 7.51 ± 1.7; Post test: 10.7 ± 2.1 (**P<0.001**)	107 (I: 57; C:50); Final evaluation: 86 (I: 43; C:43)	4 months	Effective
2	Velasco et al, 2015, Brazil [15]	Change in knowledge level	**Mean score in the intervention:** Pre test: 4.8±3; Post test: 7.5 ± 2 (**P = 0.000**) **Mean score in the control:**Pre test: 5.9 ± 3; Post test: 7.5±3 (**P = 0.005**)	66 (I: 33; C: 33)	2 weeks	Effective
3	Clavier et al, 2019, France [16]	Change in knowledge level	**Post test scores:** **SCT:** Int: 60±9%; Cont: 68±11%; (**P = 0.006**) **MCQs:** Int: 18±4; Cont: 16±4; (P = 0.22)	62 (I: 32; C: 30); Final evaluation: 44 (I:22; C:22)	3 weeks	Not effective
4	Noll et al, 2017, Germany [17]	Change in knowledge level	**Average improvement in score (Immediately after learning):** Group A: 3.59±1.48; Group B: 3.86±1.51; (P = 0.1) **After 14 days follow-up:** Average decrease of the number of correct answers as follows; Group A: 0.33±1.62; Group B: 1.14±1.30	44 (Group A:22; Group B: 22)	45 min	Effective
5	Samra et al, 2016, USA [18]	Change in knowledge level	**Range of score in intervention (Out of 16):** Pre test: 0–1; Post test: 0–12 (**P = 0.01**) **Range of score in control (Out of 16):** Pre test: 0–2; Post test: 0–4 (P = 0.08); **Average improvement in score:** Int: 5.4 points (range, 0–12 points); Cont: 0.5 points (range, –1 to +1 points) (**P = .0286**)	29 (I: 15; C: 14); Final evaluation 21 (I: 7, C:14)	8 weeks	Effective

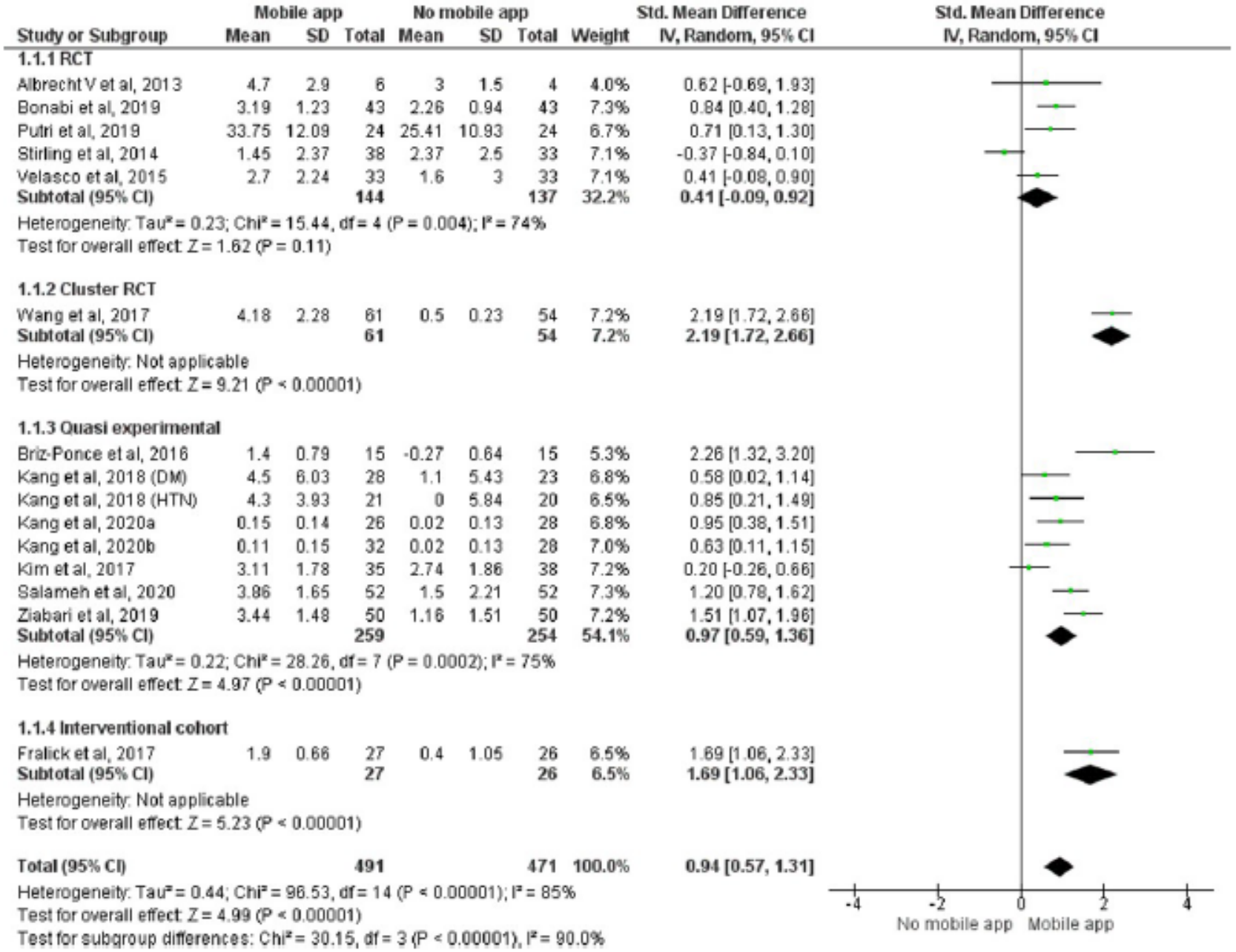

Fig 2. Forest plot: Effect of intervention on knowledge of HCPs.

CBL對於醫學生的效果(課業與感覺)

Citation	Country	Curriculum Area	Sample	Curriculum Year	Outcome Measures	Design
Bhardwaj et al., 2015 [87]	India	Community Medicine Microbiology Internal Medicine Obstetrics and Gynecology	CBL: 90 DL: NR	3rd year: 90 4th year: NR	Academic performance	Pre–Post
Carrasco et al., 2018 [88]	USA	Microbiology Immunology Oncology	CBL: 23 TBL: 23	1st year: 23	Other perceptions	Inter–Post (crossover)
Cendan et al., 2011 [18]	USA	General Surgery	CBL:130 DL:130	3rd year: 260	Academic performance	Post (crossover)
Chengyi et al., 2017 [89]	China	Otorhinolaryngology	PBL + CBL:50 DL: 50	NR: 100	Academic performance Intra-individual perceptions Student–faculty perceptions	Post
Ciraj et al., 2010 [13]	India	Microbiology Immunology Oncology	CBL:166 DL: 166	2nd year: 166	Academic performance	Inter–Post
Diwan et al., 2017 [90]	India	Endocrinology	CBL: 13 Tutorial method: 13	1st year: 26	Academic performance	Pre–Post (crossover)
Fortun et al., 2017 [91]	USA	Molecular Biology Biochemistry Medical Genetics	CBL: 154 Independent readings: 97	1st year: 130	Academic performance	Pre–Post

(a)

	D1	D2	D3	D4	D5	D6
Adiga et al. (2011)	+	✗	✗	-	+	✗
Ahmad et al. (2017)	✗	-	✗	-	✗	-
Alimoglu et al. (2014)*	✗	✗	✗	✗	-	✗
Bhardwaj et al. (2015)	✗	✗	✗	-	✗	-
Carrasco et al. (2018)	✗	✗	+	-	+	-
Cendan et al. (2011)	✗	✗	✗	✗	✗	-
Chengyi et al. (2017)	+	+	✗	-	+	✗
Ciraj et al. (2010)	✗	+	✗	-	✗	-
Diwan et al. (2017)	✗	✗	✗	✗	+	✗
Fortun et al. (2017)	✗	+	+	-	✗	-
Grover et al. (2020)	+	✗	✗	-	✗	-
Hansen et al. (2005)	✗	✗	✗	✗	+	-
Hashim et al. (2015)	+	✗	✗	✗	✗	-
Hempel et al. (2016)	✗	+	✗	✗	+	-
Jamkar et al. (2007)	+	✗	✗	✗	+	-
Joshi et al. (2014)	+	+	✗	✗	+	-
Kamat et al. (2012)	+	+	✗	✗	✗	-
Kaur et al. (2020)	+	✗	✗	✗	+	-
Latif et al. (2014)	✗	+	✗	✗	✗	-
Lee et al. (2013)	+	+	+	-	✗	-
Ma et al. (2016)	+	✗	✗	✗	+	✗
Maas et al. (2018)	✗	✗	✗	✗	✗	✗
Montaldo et al. (2013)	+	✗	✗	✗	✗	-
Nagaiaha et al. (2014)	+	✗	✗	✗	✗	-
Nair et al. (2013)	✗	✗	✗	✗	+	-
Nordquist et al. (2012)*	✗	✗	✗	✗	✗	✗
Palappallil et al. (2019)	+	✗	✗	✗	+	✗
Panja et al. (2013)	+	✗	✗	✗	+	-
Patil et al. (2016)	+	+	✗	-	+	-
Rajan et al. (2016)	✗	✗	✗	✗	✗	✗
Schwartz et al. (2007)	+	+	+	+	+	-
Sudhakar et al. (2017)	✗	✗	✗	✗	✗	✗
Surapaneni et al. (2010)	+	+	✗	✗	+	-
Tathe et al. (2014)	✗	✗	✗	✗	+	✗
Turk et al. (2019)	✗	✗	✗	+	+	-
Vedi et al. (2021)	✗	✗	✗	-	✗	-
Vora et al. (2015)	+	+	✗	-	+	-
Waliany et al. (2019)	✗	✗	✗	✗	✗	-
Willis et al. (2020)	✗	✗	✗	✗	✗	✗
Yang et al. (2021)	+	+	✗	-	+	-
Zhao et al. (2020)	+	+	✗	✗	+	-

D1: Selection of participants (selection bias)
D2: Confounding variables (selection bias)
D3: Exposure measurement (performance bias)
D4: Blinding outcome assessment (detection bias)
D5: Incomplete outcome data (attrition bias)
D6: Selective outcome reporting (reporting bias)

Table 2. Quantitative synthesis of the comparison of CBL and other teaching methodologies.

Outcome Measure	CURRICULUM AREA	k, n	I^2, *p* Value	SMD (95% CI)	GRADE
Exam scores	**POOLED EFFECT**	40, 4263	**94%, $p < 0.001$**	**2.37 (1.25 to 3.49)**	⊕○○○ *†††‡
	Biochemistry	11, 886	**94%, $p < 0.001$**	**3.32 (1.44 to 5.19)**	
	General Surgery	2, 297	0%, $p = 0.78$	0.29 (−0.13 to 0.71)	
	Otorhinolaryngology	4, 96	0%, $p = 0.70$	**11.78 (9.65 to 13.92)**	
	Microbiology	4, 545	**85%, $p < 0.001$**	0.79 (−016 to 1.75)	
	Endocrinology	3, 396	**80%, $p < 0.001$**	0.20 (−2.00 to 2.41)	
	Molecular Biology / Biochemistry / Medical Genetics	2, 251	**82%, $p = 0.02$**	1.03 (−4.55 to 6.61)	
	Pathology	1, 70	NA	**1.56 (1.02 to 2.10)**	
	Pharmacology	5, 367	**90%, $p < 0.001$**	−0.07 (−1.10; 0.96)	
	Physiology	3, 427	74%, $p = 0.02$	0.83 (−0.03 to 1.70)	
	Immunology	1, 84	NA	**0.72 (0.28 to1.16)**	
	Oncology	1, 583	NA	**0.91 (0.73 to 1.08)**	
	Semiology / Internal Medicine	1, 64	NA	**0.86 (0.34 to 1.38)**	
	Radiology	1, 135	NA	**1.58 (1.19 to 1.97)**	
	Nephrology	1, 62	NA	**4.41 (3.47 to 5.36)**	
OSCE	**POOLED EFFECT**	3, 1988	46%, $p = 0.16$	0.30 (−0.07 to 0.67)	⊕○○○ *§
	Emergency Medicine	3, 1988	46%, $p = 0.16$	0.30 (−0.07 to 0.67)	
Self-learning	**POOLED EFFECT**	4, 654	**97%, $p < 0.001$**	0.87 (−0.49 to 2.23)	⊕○○○ *†‡§
	Pathology	1, 70	NA	**0.98 (0.49 to 1.48)**	
	Nuclear Medicine	1, 70	NA	0.31 (−0.16 to 0.78)	
	Immunology	1, 170	NA	0.15 (−0.15 to 0.45)	
	Endocrinology	1, 374	NA	**2.01 (1.75 to 2.27)**	
Critical thinking	**POOLED EFFECT**	3, 276	**96%, $p < 0.001$**	2.76 (−2.42 to 7.94)	⊕○○○ *†§§
	Pathology	1, 70	NA	**1.01 (0.51 to 1.51)**	
	Endocrinology	1, 344	NA	**2.26 (1.99 to 2.53)**	
	Nephrology	1, 62	NA	**5.12 (4.07 to 6.18)**	
Satisfaction with the teaching method	**POOLED EFFECT**	4, 354	**95%, $p < 0.001$**	1.31 (−1.60 to 4.23)	⊕○○○ *†‡§§
	Otorhinolaryngology	1, 100	NA	**0.39 (0.00 to 0.79)**	
	Nuclear Medicine	1, 70	NA	0.31 (−0.16 to 0.78)	
	Nephrology	1, 62	NA	**4.12 (3.22 to 5.02)**	

高品質Systematic Review的特色

- PRISMA報告方式
  - 搜尋數個醫教相關的databases(Web of Science, ERIC, PsycINFO等)
  - 搜尋策略
  - 理論的闡明與討論
  - 教學方式的資訊
  - 評質所納入研究論文的品質
    - MERSQI：Medical Education Research Study Quality Instrument
    - NOS-E：Newcastle-Ottawa Scale – Education

PRISMA流程圖

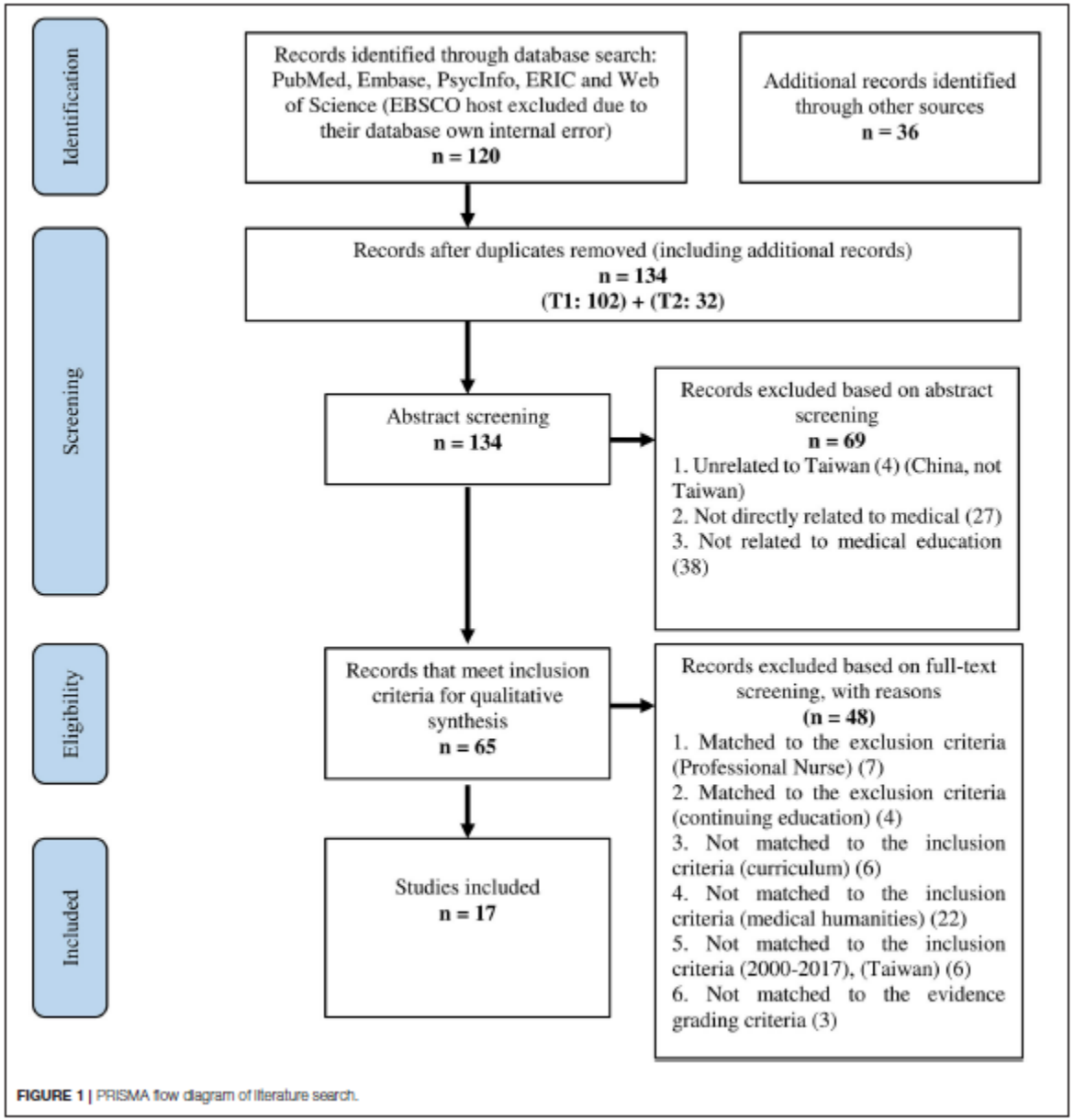

FIGURE 1 | PRISMA flow diagram of literature search.

MERSQI

Table 1. MERSQI Domain and Item Scores for 210 Medical Education Research Studies

Domain	MERSQI Item	Studies, No. (%)[a]	Score		Mean (SD)	
			Item	Maximum Domain	Item	Domain
Study design	1. Study design			3	1.28 (0.47)	1.28 (0.47)
	Single group cross-sectional or single group posttest only	140 (66.7)	1			
	Single group pretest and posttest	33 (15.7)	1.5			
	Nonrandomized, 2 group	31 (14.8)	2			
	Randomized controlled trial	6 (2.9)	3			
Sampling	2. No. of institutions studied			3	0.84 (0.46)	1.90 (0.65)
	1	135 (64.3)	0.5			
	2	8 (3.8)	1			
	>2	67 (31.9)	1.5			
	3. Response rate, %				1.06 (0.44)	
	Not applicable	30 (14.3)				
	<50 or not reported	60 (33.3)[b]	0.5			
	50-74	39 (21.7)[b]	1			
	≥75	81 (45.0)[b]	1.5			
Type of data	4. Type of data			3	1.91 (0.99)	1.91 (0.99)
	Assessment by study participant	114 (54.3)	1			
	Objective measurement	96 (45.7)	3			
Validity of evaluation instrument[c]	5. Internal structure			3	0.25 (0.44)	0.69 (0.93)
	Not applicable	25 (11.9)				
	Not reported	138 (74.6)[d]	0			
	Reported	47 (25.4)[d]	1			
	6. Content				0.29 (0.45)	
	Not applicable	25 (11.9)				
	Not reported	132 (71.4)[d]	0			
	Reported	53 (28.6)[d]	1			
	7. Relationships to other variables				0.15 (0.36)	
	Not applicable	25 (11.9)				
	Not reported	157 (84.9)[d]	0			
	Reported	28 (15.1)[d]	1			
Data analysis	8. Appropriateness of analysis	29 (13.8)		3	0.86 (0.35)	2.58 (0.65)
	Data analysis inappropriate for study design or type of data	181 (86.2)	0			
	Data analysis appropriate for study design and type of data		1			
	9. Complexity of analysis	58 (27.6)			1.72 (0.45)	
	Descriptive analysis only	152 (72.4)	1			
	Beyond descriptive analysis		2			
Outcomes	10. Outcomes	102 (48.6)		3	1.44 (0.50)	1.44 (0.50)
	Satisfaction, attitudes, perceptions, opinions, general facts	41 (19.5)	1			
	Knowledge, skills	62 (29.5)	1.5			
	Behaviors	5 (2.4)	2			
	Patient/health care outcome		3			
Total Score				**18**		9.95 (2.34)

NOS-E

Representativeness of intervention group	• Not representative: 0 • Very or somewhat representative of average learner in community: 1	• Representativeness is judged in relation to the community of eligible learners (e.g., the entire school year class, training program, or faculty). • "Very" representative indicates that all or a random sample of eligible learners enrolled. • "Somewhat" representative indicates that 75%–99% of eligible learners enrolled, or eligible but unenrolled learners are compared with those enrolled and found to be similar. • "Not representative" indicates either that sampling is not described or < 75% of eligible learners enrolled. • An intervention group defined by completion of the intervention (e.g., completers versus noncompleters) is "not representative" regardless of proportion.
Selection of comparison group	• No separate comparison group (e.g., single-group pretest–posttest): 0 • Drawn from a different community: 0 • Drawn from the same community: 1	• "Same" community indicates that there is no obvious reason to suspect systematic difference between intervention and comparison group. • A "different" community would include a different training program, a historical cohort with different exposures, or subjects substantially different in characteristics such as age, gender, performance, or desire to participate (e.g., volunteers versus nonvolunteers, experts versus novices).
Comparability of comparison group	• No separate comparison group: 0 • Randomized study: o Allocation not concealed: 1 o Allocation concealed: 2 • Nonrandomized study: o Controlled for 1 subject characteristic: 1 o Controlled for 2 or more subject characteristics: 2	• Randomized and nonrandomized studies are coded separately. • Allocation is considered concealed if enrollment, consent, or baseline assessment preceded randomization. • Controlling for subject characteristics requires statistical covariate analysis (e.g., including baseline scores or training level in a multivariate model); directly comparing characteristics between groups (e.g., t-test comparing baseline demographics) is insufficient. • Relevant subject characteristics include (but are not limited to) scores/grades on a pretest, standardized test, or earlier course, and grade point average/class rank.
Study retention[e]	• Poor retention could introduce bias: 0 • Retention unlikely to introduce bias: 1	• High if ≥ 75% of those enrolled provided outcome data, or if authors described those lost to follow-up. • Authors must report the number providing data (percentages or proportions completing the study are insufficient unless the denominator is specified).
Blinding of assessment[e]	• Outcome assessment not blinded: 0 • Outcome assessment blinded: 1	• Blinded if the assessor cannot be influenced by group assignment. • Assessments that do not require human judgment (e.g., multiple-choice tests or computer-scored performance) are considered to be blinded. • One-group studies are not blinded unless scoring does not require judgment or authors describe a plausible method for hiding the timing of assessment. • Participant-reported outcomes are never blinded.

執行步驟

- 提出欲研究的問題
- 組織研究團隊
- 對於Protocol進行註冊
- 設計與執行搜尋策略
- 納入與排除之篩選
- Data分析與綜合
- 撰寫文稿

醫教Systematic Review論文之12個重點提示

Med Teach 2019;41:1232-1238

(1)考慮是否選用Focused Review進行Systematic Review

- 焦點式回顧研究(Focused review)專注於一個特定問題
- 使用系統性的研究方法
- 必須透明，並能重現

(2)進行範疇界定搜尋(Scoping Search)

- 在撰寫protocol與決定研究問題之前，先進行範疇界定搜尋(Scoping Search)
- 這可協助評估研究想法的範圍與深度
- 若所搜尋的相關研究的命中比率太高，需考慮縮小範圍。
- 搜尋用語需與PICO一致(population, Intervention, Comparison, Outcome)
- Focused review需仔細確認研究之問題、納入標準與排除標準。

(3)形成焦點性的研究問題

- 在範疇界定搜尋(Scoping Search)之後才能確定最終的焦點性的研究問題
- 形成之問題需持續修正至可得有意義的答案為止
- 要能確認之前的systematic review的範疇、主要發現、限制與建議，以避免議題重複。

(4)Review量需適宜

- 避免事倍功半
- Review量過大會含有過多變數與無法控制的因素而導致無法產生清楚的結論； Review量過少則可能不足以下結論
- Review量是以Review的目的與所搜尋到的研究的品質所決定
- 合理量建議：小於30篇

(5)組織團隊與時間限定

- 研究計畫的範疇限定適當
- 適當的統籌
- 團隊合作
- 有可能在6個月內完成

(6)邀請圖書館員協助

- 協助選擇、撈取與過濾資訊
- 有助於訂定範疇限定的研究題目

(7)鎖定目標期刊

- 期刊之字數限定不同
- 文稿之寫作依照期刊之格式，節省時間。
- 研究題目切合期刊之所需，增加被刊登機會。

(8)使用PROSPERO資源

- systematic review註冊的國際database
  - 可避免重複發表相似的review，也可藉由protocol的註冊而減少研究的bias發生

(9)聯絡所搜尋到的研究的作者

- 可得到額外資訊
- 可能分享論文中未見的相關data
- 若回顧的鹹就數量較少，聯絡作者更有幫助

(10)實證綜合方式的選擇

- 焦點式回顧(focused review)仔細使用納入條件與排除條件會增加review的同質化
- 現實主義回顧(realist review)可用來發掘某特定介入如何與為何會成功或失敗
- 敘述式綜論(narrative synthesis)使用文字敘述來總結與說明所見，但有的期刊有字數限制
- 若搜尋到的研究論文篇數過少，不適合以界定範疇回顧(scoping review)作為實證綜合方式

(11)使用有趣的方式來表達所見

- 以圖片呈現所見
- 以視覺方式呈現data較容易快速得到理解

(12)焦點式系統回顧的限制

- Database的可近性(常搜尋英文論文)
  - 存有bias的研究結果
  - 其他來自焦點式回顧的過程
    - 用來限制review的因子需有合理的理由
    - 文獻搜尋的限制太多導致搜尋到的篇數太少
  - review的題目需是可應用的

結論

- 進行系統性回顧研究是一個耗費心力與時間的過程，但可以提供實證的綜合，以期回答特定的研究問題。
- 系統性回顧研究雖有其優勢與弱點，但也常被使用於醫教範疇，並廣泛被許多期刊所接受。
- 有興趣的話，可以在進行系統性回顧研究之前多參考一些相關指引。