Online supplemental materials

Contents

Supplemental Digital Appendix 1. Search strategy

Search last run on: February 21, 2020

PubMed/MEDLINE (Inception – Present)

Limits/expanders applied: None

((("teaching"[MeSH:noexp] OR "models, educational"[MeSH:noexp] OR "programmed instruction as topic"[MeSH:noexp] OR "computer-assisted instruction"[MeSH:noexp] OR "simulation training"[MeSH:noexp] OR "remedial teaching"[MeSH:noexp] OR "high fidelity simulation training"[MeSH:noexp] OR "computer user training"[MeSH:noexp] OR "teaching materials"[MeSH:noexp] OR "educational measurement"[MeSH:noexp] OR "curriculum"[MeSH:noexp] OR "learning"[MeSH:noexp] OR "Internet"[MeSH:noexp] OR "curriculum"[tiab] OR "curriculum"[ot] OR "curriculums"[tiab] OR "curriculums"[ot] OR "curricula"[tiab] OR "curricula"[ot] OR "curriculas"[tiab] OR "teaching"[tiab] OR "teaching"[ot] OR "instruction"[tiab] OR "instruction"[ot] OR "tutorial"[tiab] OR "tutorial"[ot] OR "tutorials"[tiab] OR "tutorials"[ot] OR "self-directed"[tiab] OR "self-directed "[ot] OR "learning"[tiab] OR "learning"[ot] OR "workshop"[tiab] OR "workshop"[ot] OR "workshops"[tiab] OR "workshops"[ot] OR "lecture"[tiab] OR "lecture"[ot] OR "lectures"[tiab] OR "lectures"[ot] OR "small-group"[tiab] OR "small-group"[ot] OR "web-based"[tiab] OR "web-based"[ot] OR "Internet"[tiab] OR "Internet"[ot] OR "e-learning"[tiab] OR "e-learning"[ot] OR "seminar"[tiab] OR "seminar"[ot] OR

"seminars"[tiab] OR "seminars"[ot] OR "online"[tiab] OR "online"[ot])) AND ("interpretation"[tiab] OR "interpretation"[ot] OR "interpreted"[tiab] OR "interpreted"[ot] OR "reading"[tiab] OR "reading"[ot] OR "interpret"[tiab] OR "interpret"[ot])) AND ("ECG"[tiab] OR "ECG"[ot] OR "ECGs"[tiab] OR "ECGs"[ot] OR "EKG"[tiab] OR "EKG"[ot] OR "EKGs"[tiab] OR "electrocardiogram"[tiab] OR "electrocardiogram"[ot] OR "electrocardiograms"[tiab] OR "electrocardiograms"[ot] OR "electrocardiograph"[tiab] OR "electrocardiograph"[ot] OR "electrocardiographs"[tiab] OR "electrocardiography"[tiab] OR "electrocardiography"[ot] OR "electro cardiogram"[tiab] OR "electro cardiogram"[ot] OR "electro cardiograms"[tiab] OR "electro cardiograph"[tiab] OR "electro cardiographs"[tiab] OR "electro cardiography"[tiab] OR "electrocardiography"[MeSH:noexp])

EMBASE via Ovid (1974 – Present)

Limits/expanders applied: None
1. electrocardiography/
2. ECG.ti,ab,kw.
3. ECGs.ti,ab,kw.
4. EKG.ti,ab,kw.
5. EKGs.ti,ab,kw.
6. electrocardiogram.ti,ab,kw.
7. electrocardiograms.ti,ab,kw.
8. electrocardiograph$.ti,ab,kw.
9. electro cardiogram.ti,ab,kw.
10. electro cardiograms.ti,ab,kw.
11. electro cardiograph$.ti,ab,kw.
12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. interpret$.ti,ab,kw.
14. reading.ti,ab,kw.
15. 13 or 14
16. teaching/
17. education/
18. curriculum/
19. educational model/
20. computerized adaptive testing/
21. educational technology/
22. learning/
23. self directed learning/
24. reinforcement/
25. curricul$.ti,ab,kw.
26. teaching.ti,ab,kw.
27. instruction.ti,ab,kw.
28. tutorial$.ti,ab,kw.
29. self-directed.ti,ab,kw.
30. learning.ti,ab,kw.
31. workshop$.ti,ab,kw.
32. lecture$.ti,ab,kw.
33. small-group.ti,ab,kw.
34. web-based.ti,ab,kw.
35. internet.ti,ab,kw.

36. e-learning.ti,ab,kw.
37. seminar$.ti,ab,kw.
38. online.ti,ab,kw.
39. 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 32 or 33 or 34 or 35 or 36 or 37 or 38
40. 12 and 15 and 39

PsycINFO via Ovid (1806 – Present)

Limits/expanders applied: None
1. electrocardiography/
2. ECG.ti,ab,kw.
3. ECGs.ti,ab,kw.
4. EKG.ti,ab,kw.
5. EKGs.ti,ab,kw.
6. electrocardiogram.ti,ab,kw.
7. electrocardiograms.ti,ab,kw.
8. electrocardiograph$.ti,ab,kw.
9. electro cardiogram.ti,ab,kw.
10. electro cardiograms.ti,ab,kw.
11. electro cardiograph$.ti,ab,kw.
12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. interpret$.ti,ab,kw.
14. reading.ti,ab,kw.
15. 13 or 14
16. teaching/
17. education/
18. curriculum/
19. learning/
20. self directed learning/
21. reinforcement/
22. curricul$.ti,ab,kw.
23. teaching.ti,ab,kw.
24. instruction.ti,ab,kw.
25. tutorial$.ti,ab,kw.
26. self-directed.ti,ab,kw.
27. learning.ti,ab,kw.
28. workshop$.ti,ab,kw.
29. lecture$.ti,ab,kw.
30. small-group.ti,ab,kw.
31. web-based.ti,ab,kw.
32. internet.ti,ab,kw.
33. e-learning.ti,ab,kw.
34. seminar$.ti,ab,kw.
35. online.ti,ab,kw.
36. medical education/
37. adaptive testing/
38. computer assisted instruction/ or intelligent tutoring systems/

39. 16 or 17 or 18 or 19 or 20 or 21 or 36 or 37 or 38
40. 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 39
41. 12 and 40
42. 15 and 41

Cochrane CENTRAL via Ovid

 Limits/expanders applied: None
 1. electrocardiography/
2. ECG.ti,ab,kw.
3. ECGs.ti,ab,kw.
4. EKG.ti,ab,kw.
5. EKGs.ti,ab,kw.
6. electrocardiogram.ti,ab,kw.
7. electrocardiograms.ti,ab,kw.
8. electrocardiograph$.ti,ab,kw.
9. electro cardiogram.ti,ab,kw.
10. electro cardiograms.ti,ab,kw.
11. electro cardiograph$.ti,ab,kw.
12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. interpret$.ti,ab,kw.
14. reading.ti,ab,kw.
15. 13 or 14
16. teaching/
17. education/
18. curriculum/
19. learning/
20. self directed learning/
21. reinforcement/
22. curricul$.ti,ab,kw.
23. teaching.ti,ab,kw.
24. instruction.ti,ab,kw.
25. tutorial$.ti,ab,kw.
26. self-directed.ti,ab,kw.
27. learning.ti,ab,kw.
28. workshop$.ti,ab,kw.
29. lecture$.ti,ab,kw.
30. small-group.ti,ab,kw.
31. web-based.ti,ab,kw.
32. internet.ti,ab,kw.
33. e-learning.ti,ab,kw.
34. seminar$.ti,ab,kw.
35. online.ti,ab,kw.
36. medical education/
37. adaptive testing/
38. computer assisted instruction/ or intelligent tutoring systems/
39. 16 or 17 or 18 or 19 or 20 or 21 or 36 or 37 or 38
40. 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 39

41. 12 and 40
42. 15 and 41

CINAHL via Ebsco (1979 – 2017)

Limits/expanders applied: Also search within full text expander

(electrocardiogram* OR electrocardiograph* OR electro cardiograph* OR electro cardiogram* OR ECG* OR EKG* OR MH electrocardiography) AND (interpret* OR "reading") AND (MH Teaching OR MH Models, Educational OR MH Programmed Instruction OR MH Computer Assisted Instruction OR MH Computer Simulation OR MH Remedial Teaching OR MH Teaching Materials OR MH Educational Measurement OR MH Curriculum OR MH Learning OR MH Computerized Adaptive Testing OR MH Educational Technology OR MH Self Directed Learning OR curricul* OR "teaching" OR "instruction" OR tutorial* OR "self directed" OR "learning" OR workshop* OR lecture* OR web-based OR "internet" OR e-learning OR seminar* OR "online")

ERIC via ProQuest

Limits/expanders applied: None

((electrocardiograph* OR electrocardiogram* OR EKG* OR ECG* OR "electro cardiogram*" OR "electro cardiograph*") AND (interpret* OR reading) AND (teaching OR learning OR curricul* OR instruction OR tutorial* OR self-directed OR lecture* OR web-based OR small-group OR e-learning OR online OR internet OR seminar* OR workshop*))

Web of Science (Core Collection)

Limits/expanders applied: None

((electrocardiograph* OR electrocardiogram* OR EKG* OR ECG* OR "electro cardiogram*" OR "electro cardiograph*") AND (interpret* OR reading) AND (teaching OR learning OR curricul* OR instruction OR tutorial* OR self-directed OR lecture* OR web-based OR small-group OR e-learning OR online OR internet OR seminar* OR workshop*))

Supplemental Digital Appendix 2. Inter-rater agreement for abstracted features

Feature	Kappa (N=85 studies)
Instrument features	
Number of cases	0.64
Modality	0.75
Pass/fail standard	0.88
Supervised	0.82
Timed	0.67
ECG diagnoses tested	0.75
Case complexity	0.60
Case difficulty estimation	0.73
Who selected cases	0.70
Vignette included	0.81
Response format	0.70
Scoring rubric gives credit	0.70
How scoring rubric was created	1.0
Scoring rubric creation by a group	0.93
Who scored responses	0.86
Number of human scorers	1.0
Scorer training	0.87
Feedback given to learners	0.75
Methodological quality	
Number enrolled	1.0
MERSQI SD	0.84
MERSQI-instit	0.76
MERSQI-objective	1.0
MERSQI-outcome	0.02 (84/85 raw agreement*)
MERSQI-soph	0.71
MERSQI-approp	0.71
MERSQI-FU	0.68
Blinded	0.79
Geographic location	0.98
QUADAS-2 features	
Selection	0.70
Flow	0.77
Conduct	0.86
Applicability	0.84
Validity evidence	
Content	0.95
Internal structure	0.95
Relations with other variables	1.0
Response process	0.66
Consequences	0.41 (80/85 raw agreement*)

* Codes for these variables were heavily skewed toward a single response ("knowledge" for MERSQI-outcome and "none" for consequences evidence), such that even a very small number of disagreements leads to a low kappa (i.e., no better than chance).

Supplemental Digital Appendix 3. Methodological features of studies of tests of physicians' ECG interpretation skill, from a systematic review of literature, February 2020[a]

| First author, year[ref] | Participants: Type[b]; no. enrolled | Study design (purpose)[c] | Bias[d] | | | | Blinded scoring |
			Selection	Flow	Conduct	Applicability	
Owen, 1965[19]	PG, MedStud; 85	NR2 (Train)		OK			
Stretton, 1967[20]	MedStud; 85	NR2 (Train)		OK	OK		OK
Kingston, 1979[21]	PractMD, MedStud; 38	PP1 (Train)		OK		OK	
Pinkerton, 1981[22]	PG; 81	CS1 (Survey)	OK	OK		OK	
Fincher, 1987[23]	MedStud; 107	RCT (Train)	OK				OK
Hancock, 1987[24]	PG; 1,825	CS1 (Valid)	OK	OK			
Fincher, 1988[25]	MedStud; 83	RCT (Train)					OK
Dunn, 1990[26]	PractMD; 3	PP1 (Survey)		OK		OK	
Grum, 1993[27]	MedStud; 95	RCT (Train)	OK	OK		OK	
White, 1995[28]	PG; 11	PP1 (Train)		OK		OK	
Gillespie, 1996[29]	PG; 57	CS1 (Survey)	OK	OK		OK	
Hatala, 1996[30]	PG; 10	NR2 (Train)				OK	
Gruppen, 1997[31]	MedStud; 264	NR2 (Train)		OK		OK	
Devitt, 1998[32]	PractMD, PG, MedStud, Nurse; 72	RCT (Train)		OK		OK	
Lazzari, 1998[33]	PractMD; 6	CS1 (Valid)		OK		OK	
Hatala, 1999[34]	PractMD, PG, MedStud; 30	RCT (Survey)				OK	
Massel, 2000[35]	PractMD; 3	CS1 (Survey)		OK			
Sur, 2000[36]	PG; 61	CS1 (Survey)		OK		OK	
Brady, 2001[37]	PractMD, PG; 458	CS1 (Survey)		OK		OK	
Goodacre, 2001[38]	PG; 10	RCT (Survey)		OK	OK		OK
Little, 2001[39]	MedStud; 46	CS1 (Survey)		OK		OK	
Boltri, 2003[40]	PG; 52	PP1 (Survey)		OK			
Hatala, 2003[41]	MedStud; 71	NR2 (Train)		OK			OK
Lucas, 2003[42]	MedStud; 112	NR2 (Train)		OK			
Solomon, 2004[43]	MedStud; 5	CS1 (Valid)		OK		OK	
Berger, 2005[44]	PG; 120	CS1 (Survey)		OK	OK	OK	OK
Snyder, 2005[45]	PG; 132	CS1 (Survey)				OK	
Hoyle, 2007[46]	PG; 122	CS1 (Survey)	OK	OK	OK	OK	OK
Burke, 2008[47]	PG; 46	CS1 (Train)		OK		OK	

First author, year[ref]	Participants: Type[b]; no. enrolled	Study design (purpose)[c]	Bias[d]				Blinded scoring
			Selection	Flow	Conduct	Applicability	
Nilsson, 2008[48]	MedStud; 50	NR2 (Train)		OK			OK
Eslava, 2009[49]	PG; 52	CS1 (Survey)		OK	OK	OK	OK
Jericho, 2009[50]	PG; 76	PP1 (Train)	OK	OK			OK
Lever, 2009[51]	PG, MedStud; 102	CS1 (Survey)	OK	OK		OK	
Rubinstein, 2009[52]	MedStud; 15	NR2 (Train)		OK		OK	OK
Southern, 2009[53]	PG; 110	NR2 (Survey)		OK	OK		
Crocetti, 2010[54]	PG; 46	CS1 (Survey)		OK		OK	
de Jager, 2010[55]	PG; 50	CS1 (Survey)	OK	OK	OK	OK	OK
Gregory, 2011[56]	MedStud; 18	RCT (Train)		OK		OK	
Mahler, 2011[57]	MedStud; 234	RCT (Train)		OK		OK	
Sibbald, 2012[58]	PG; 30	RCT (Train)		OK		OK	
Raupach, 2013[59]	MedStud; 564	RCT (Valid)	OK	OK	OK	OK	OK
Yadav, 2013[93]	PG; 41	PP1 (Train)	OK	OK		OK	
Boulouffe, 2014[60]	PractMD, PG, MedStud; 52	CS1 (Valid)					
Jablonover, 2014[61]	PG, MedStud; 253	CS1 (Survey)	OK	OK		OK	
McAloon, 2014[62]	PG, MedStud; 46	RCT (Train)		OK			
Sibbald, 2014[63]	PG; 29	CS1 (Survey)	OK			OK	
Blissett, 2015[64]	MedStud; 29	RCT (Train)		OK		OK	
DeBonis, 2015[65]	PG; 30	PP1 (Train)					
Dong, 2015[66]	MedStud; 126	RCT (Train)	OK	OK			
Jheeta, 2015[67]	PractMD, PG, NPPA; 764	PP1 (Survey)			OK	OK	
Kopec, 2015[68]	MedStud; 536	CS1 (Survey)		OK		OK	
Novotny, 2015[69]	PG; 29	CS1 (Survey)		OK		OK	
Pourmand, 2015[70]	PG, MedStud; 183	PP1 (Train)				OK	
Quinn, 2015[71]	PG; 125	PP1 (Valid)			OK	OK	
Rolskov, 2015[72]	MedStud; 220	RCT (Valid)			OK	OK	
Sibbald, 2015[73]	PG; 16	RCT (Train)					
Zeng, 2015[74]	MedStud; 200	RCT (Train)	OK				
Chudgar, 2016[75]	MedStud; 101	NR2 (Train)					
Davies, 2016[76]	MedStud, Pharm; 39	RCT (Train)		OK			OK
Fent, 2016[77]	PG, MedStud; 168	RCT (Train)		OK		OK	
Hartman, 2016[78]	PG; 113	CS1 (Valid)		OK	OK	OK	
Montassier, 2016[79]	MedStud; 98	RCT (Train)			OK		
Porras, 2016[80]	PG; 28	PP1 (Train)		OK		OK	

First author, year[ref]	Participants: Type[b]; no. enrolled	Study design (purpose)[c]	Bias[d] Selection	Flow	Conduct	Applicability	Blinded scoring
Barthelemy, 2017[81]	PG; 39	RCT (Train)			OK	OK	
Liu, 2017[82]	PG; 39	PP1 (Train)				OK	
Mirtajaddini, 2017[83]	PG; 163	RCT (Train)				OK	
Monteiro, 2017[84]	MedStud; 80	RCT (Train)		OK			
Rui, 2017[85]	MedStud; 181	RCT (Train)					
Compiet, 2018[86]	PractMD; 70	CS1 (Survey)					
Isfahani, 2018[87]	PG; 140	NR2 (Train)	OK	OK			
Kellman, 2018[88]	PG, MedStud	PP1 (Train)			OK	OK	
Kopec, 2018[89]	MedStud; 60	RCT (Train)		OK		OK	
Nag, 2018[90]	MedStud; 70	RCT (Train)					
Riding, 2018[91]	PractMD, Nurse, Other; 10,512	PP1 (Train)			OK	OK	
Suresh, 2018[92]	PG; 33	PP1 (Train)		OK			
Aziz, 2019[94]	PG; 35	PP1 (Train)		OK		OK	
Hatala, 2019[95]	PG, MedStud; 444	CS1 (Train)			OK	OK	
Knoery, 2019[96]	PractMD, EMT, NPPA; 91	PP1 (Train)		OK	OK	OK	
Sibbald, 2019[97]	PG; 61	RCT (Train)		OK		OK	
Smith, 2019[98]	MedStud; 42	PP1 (Train)		OK	OK	OK	
Soares, 2019[99]	PractMD, PG; 35	PP1 (Train)		OK		OK	
Baral, 2020[100]	MedStud; 145	PP1 (Train)			OK	OK	
Kewcharoen, 2020[101]	MedStud; 80	RCT (Train)		OK		OK	
Mohyuddin, 2020[102]	PG; 61	NR2 (Train)		OK			
Thach, 2020[103]	MedStud; 65	RCT (Train)		OK		OK	

[a] "OK" indicates low risk of bias or of problems with applicability, or use of blinded scoring (i.e., stronger study methods). In addition to the methodological features detailed in this table, all studies included the review employed objective assessment and used knowledge outcomes. Flow bias appraised essentially the same aspects of design as retention of participants (follow-up) and these results were fully congruent; thus follow-up is not reported separately.

[b]*Participant type:* EMT, emergency medical technicians; MedStud, medical students; NPPA, nurse practitioners or physician assistants or students; Nurse, nurses or nursing students; PG, postgraduate trainees (residents); Pharm, pharmacists or pharmacy students; PractMD, physicians in practice.

[c]*Study design:* CS1, 1-group cross-sectional; NR2, 2-group nonrandomized; PP1, 1-group pre/postintervention; RCT, randomized controlled trial. *Purpose* of study (in parentheses): Survey, survey study; Train, evaluation of training or theory-building intervention; Valid, creation and validation of assessment.

[d]*Selection, Flow, Conduct (bias)* and *Applicability* all refer to appraisals using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)[14]; for further explanation, see main text and Supplemental Digital Appendix 3 at [LWW INSERT LINK] for further explanation.

Supplemental Digital Appendix 4. Operational definitions and detailed coding for studies of tests of physicians' ECG interpretation skill, using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2), from a systematic review of literature, February 2020 (N=85 studies)

QUADAS-2 criterion	Operational definition	Low risk No. (%)	High risk No. (%)	Unclear No. (%)
Selection, risk of bias	Could the selection of trainees have introduced *bias*? Consider: • Was a case-control design (participants enrolled based on a known characteristic, including enrollment for training level comparisons [expert-novice comparison]) <u>avoided</u>? • Was a consecutive or random sample, or a large proportion (≥75%), of eligible participants enrolled? • Did the study avoid inappropriate exclusions?	16 (19%)	28 (33%)	41 (48%)

Flow, risk of bias	Could the trainee flow have introduced *bias*? Consider:
	• Were a high proportion (≥75%) of enrolled trainees included in the analysis?
	• If comparison with a reference test:
	   • Was there an appropriate interval between the index test and reference standard?
	   • Did all trainees receive the same reference standard?
		59 (69%)	9 (11%)	17 (20%)

Conduct, risk of bias (index test)	Could the conduct or interpretation of the index test have introduced *bias*? Consider: • Were the index test results interpreted without knowledge of the results of the reference standard or trainee status? (ie, blinded) • Was >1 rater involved (if only single human rater, then high risk); computer scoring and multiple-choice questions (i.e., low subjectivity) would usually be low. • If a pass/fail threshold was used, was it pre-specified?	20 (24%)	12 (14%)	53 (62%)
Applicability (index test)	Is there concern that the index test, its conduct, or its interpretation differ from the review question (i.e., *conceptual alignment* with the construct [ECG interpretation])? Consider: • Was there variation over the course of the study in test technology, execution, scoring, cut score, or interpretation? • Was the number of test items sufficient to cover the topic? (we operationally defined this as ≥5 ECGs) • Was the topic and scope of the test items appropriate to the construct? (we operationally required a list of the ECG diagnoses [ischemia, rhythm, etc] included in the test) • Did the test assess interpretation accuracy (vs knowledge, etc)?	56 (66%)	7 (8%)	22 (26%)

QUADAS-2 criterion	Operational definition	Low risk No. (%)	High risk No. (%)	Unclear No. (%)
Conduct, risk of bias (reference test); N=3	Could the reference standard, its conduct, or its interpretation have introduced *bias*? Consider: • Were the reference standard results interpreted without knowledge of the results of the index test? (independent, blinded) • Is the reference standard likely to correctly classify the target condition (i.e., is there evidence to support validity of scores and interpretations)?	1 (33%)	0	2 (67%)
Applicability (reference test); N=3	Is there concern that the target condition as defined by the reference standard does not match the review question (i.e., *conceptual alignment* with the index test)? Consider: • The same questions outlined above for the index test. • Did the reference test assess a construct that is conceptually related with the index test construct (i.e., ECG interpretation)?	0	3 (100%)	0

See here[14] for further details on the QUADAS-2.