

**Assessment of competences in rheumatology training: results of a systematic literature review
to inform EULAR points to consider**

Online Supplementary Material

Table of Contents

	Page
Online Supplementary Text S1: Search strategy for assessment of competences in rheumatology	2
Online Supplementary Text S2: Search strategy for assessment of competences in other medical specialties	8
Online Supplementary Text S3: List of relevant medical specialties other than rheumatology	16
Online Supplementary Text S4: Exclusion criteria	17
Online Supplementary Table S1: Medical education research study quality instrument (MERSQI)	18
Online Supplementary Table S2: Summary of rheumatology articles on objective structured clinical examination	19
Online Supplementary Table S3: Summary of articles from other specialties evaluating mini-clinical examination and multisource feedback	22
Online Supplementary Table S4: Summary of articles from other specialties evaluating written exams and the script concordance test and corresponding references	24

Online Supplementary Text S1: Search strategy for the assessment of competences in rheumatology

MEDLINE (Ovid)

Dates searched: 1946 to February Week 2 2018 February 20, 2018 and updated 2018 to July 22, 2019

1. exp Musculoskeletal Diseases/
2. exp "skin and connective tissue diseases"/
3. Rheumatology/
4. Rheumat\$.tw.
5. Arthritis.tw.
6. Musculoskeletal.tw.
7. (Connective Tissue adj (diseases or disorders or conditions)).tw.
8. or/1-7
9. education/ or ed.fs.
10. (educat\$ or training or program\$).tw.
11. exp Training Support/
12. exp Education, Medical/
13. ((medical or doctor\$ or physician\$ or clinician\$ or specialty or specialist\$) adj (educat\$ or training or trainee\$ or fellow\$)).tw.
14. or/9-13
15. exp Educational Measurement/
16. "Surveys and Questionnaires"/
17. (assess\$ or self-assess\$ or debrief\$).tw.
18. ((written or oral or in-training) adj exam\$).tw.
19. (quizz\$ or essay\$ or question\$).tw.
20. viva\$.tw.
21. ((clinical or practical or examination) adj2 (competenc\$ or skill\$ or supervis\$ or exercise\$ or evaluat\$)).tw.
22. or/15-21
23. validation studies/
24. (validat\$ or verif\$).tw.
25. "Predictive Value of Tests"/
26. test characteristic\$.tw.
27. "reproducibility of results"/
28. reproducibil\$.tw.
29. reliability.tw.
30. agreement.tw.

- 31. feasibility.tw.
- 32. or/23-31
- 33. and/8,14,22,32

The Cochrane Library (www.thecochranelibrary.com)

Dates searched: February 21, 2018 and updated July 23, 2019

- #1 MeSH descriptor: [Musculoskeletal Diseases] explode all trees
- #2 MeSH descriptor: [Skin and Connective Tissue Diseases] explode all trees
- #3 MeSH descriptor: [Rheumatology] this term only
- #4 Rheumat*:ti,ab
- #5 Arthritis:ti,ab
- #6 Musculoskeletal:ti,ab
- #7 ("Connective Tissue" next (diseases or disorders or conditions)):ti,ab
- #8 #1 or #2 or #3 or #4 or #5 or #6 or #7
- #9 MeSH descriptor: [Education] this term only
- #10 Any MeSH descriptor with qualifier(s): [Education - ED]
- #11 (educat* or training or program*):ti,ab
- #12 MeSH descriptor: [Training Support] explode all trees
- #13 MeSH descriptor: [Education, Medical] explode all trees
- #14 ((medical or doctor* or physician* or clinician* or specialty or specailist*) next (educat* or training or trainee* or fellow*)):ti,ab
- #15 #9 or #10 or #11 or #12 or #13 or #14
- #16 MeSH descriptor: [Educational Measurement] explode all trees
- #17 MeSH descriptor: [Surveys and Questionnaires] this term only
- #18 (assess* or self-assess* or debrief*):ti,ab
- #19 ((written or oral or in-training) next exam*):ti,ab
- #20 (quiz* or essay* or question*):ti,ab
- #21 viva*:ti,ab
- #22 ((clinical or practical or examination) near/2 (competenc* or skill* or supervis* or exercise* or evaluat*)):ti,ab
- #23 #16 or #17 or #18 or #19 or #20 or #21 or #22
- #24 MeSH descriptor: [Validation Studies] this term only
- #25 (validat* or verif*):ti,ab
- #26 MeSH descriptor: [Predictive Value of Tests] this term only
- #27 "test characteristic":ti,ab
- #28 MeSH descriptor: [Reproducibility of Results] explode all trees
- #29 reproducibil*:ti,ab
- #30 reliability:ti,ab
- #31 agreement:ti,ab
- #32 feasibility:ti,ab
- #33 #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32
- #34 #8 and #15 and #23 and #33

EMBASE (Embase.com)

Dates searched: 1980 to 21 Feb 2018 and updated 2018 to 23 Jul 2019

#8 AND #13 AND #20 AND #30 AND [embase]/lim AND ([article]/lim OR [article in
#32 press]/lim OR [review]/lim)
#31 #8 AND #13 AND #20 AND #30
#30 #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29
#29 feasibility:ab,ti
#28 agreement.:ab,ti
#27 reliability:ab,ti
#26 reproducibil*:ab,ti
#25 'reproducibility'/de
#24 'test characteristic*':ti,ab
#23 'predictive value'/de
#22 validat*:ti,ab OR verif*:ti,ab
#21 'validation study'/de
#20 #14 OR #15 OR #16 OR #17 OR #18 OR #19
((clinical OR practical OR examination) NEAR/2 (competenc* OR skill* OR supervis*
#19 OR exercise* OR evaluat*)):ti,ab
#18 viva*:ti,ab
#17 quiz*:ti,ab OR essay*:ti,ab OR question*:ti,ab
#16 ((written OR oral OR 'in training') NEAR/2 exam*):ti,ab
#15 assess*:ti,ab OR 'self assess*':ti,ab OR debrief*:ti,ab
#14 'questionnaire'/exp
#13 #9 OR #10 OR #11 OR #12
((medical OR doctor* OR physician* OR clinician* OR specialty OR specialist*)
#12 NEAR/2 (educat* OR training OR trainee* OR fellow*)):ti,ab
#11 'medical education'/exp
#10 educat*:ti,ab OR training:ti,ab OR program*:ti,ab
#9 'education'/de
#8 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7
#7 ('connective tissue' NEAR/2 (diseases OR disorders OR conditions)):ti,ab
#6 musculoskeletal:ti,ab
#5 arthritis:ti,ab
#4 rheumat*:ti,ab
#3 'rheumatology'/de
#2 'connective tissue disease'/exp
#1 'musculoskeletal disease'/exp

CINAHL (EBSCOhost)

Dates searched: 1982 to February 2018 and updated 2018 to July 2019

S1 (MH "Musculoskeletal Diseases+")
 S2 (MH "Skin and Connective Tissue Diseases+")
 S3 (MH "Rheumatology")
 S4 TI Rheumat* OR AB Rheumat*
 S5 TI Arthritis OR AB Arthritis
 S6 TI Musculoskeletal OR AB Musculoskeletal
 S7 TI (("Connective Tissue" N2 (diseases OR disorders OR conditions)) OR AB (("Connective Tissue" N2 (diseases OR disorders OR conditions)))
 S8 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7
 S9 (MH "Education")
 S10 TI (educat* OR training OR program*) OR AB (educat* OR training OR program*)
 S11 S9 OR S10
 S12 S8 AND S11
 S13 (MH "Education, Medical+")
 S14 TI (((medical OR doctor* OR physician* OR clinician* OR specialty OR specailist*) N2 (educat* OR training OR trainee* OR fellow*))) OR AB (((medical OR doctor* OR physician* OR clinician* OR specialty OR specailist*) N2 (educat* OR training OR trainee* OR fellow*)))
 S15 S12 OR S13 OR S14
 S16 (MH "Educational Measurement+")
 S17 (MH "Instrument by Type+")
 S18 (MH "Data Collection Methods+")
 S19 TI (assess* OR self-assess* OR debrief*) OR AB (assess* OR self-assess* OR debrief*)
 S20 TI (((written OR oral OR in-training) N2 exam*):) OR AB (((written OR oral OR in-training) N2 exam*))
 S21 TI (quizz* OR essay* OR question*) OR AB (quizz* OR essay* OR question*)
 S22 TI viva* OR AB viva*
 S23 TI (((clinical OR practical OR examination) N2 (competenc* OR skill* OR supervis* OR exercise* OR evaluat*))) OR AB (((clinical OR practical OR examination) N2 (competenc* OR skill* OR supervis* OR exercise* OR evaluat*)))
 S24 S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23
 S25 (MH "Validation Studies")
 S26 TI (validat* OR verif*) OR AB (validat* OR verif*)
 S27 (MH "Predictive Value of Tests")
 S28 TI "test characteristic*" OR AB "test characteristic*"
 S29 MH "Reproducibility of Results")
 S30 TI reproducibil* OR AB reproducibil*
 S31 TI reliability OR AB reliability
 S32 TI agreement OR AB agreement
 S33 TI feasibility OR AB feasibility
 S34 S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33
 S35 S15 AND S24 AND S34 Limiters - Exclude MEDLINE records

ERIC (EBSCOhost)

Dates searched: 1966 to February 2018 and updated 2018 to July 2019

S1 DE "Medicine"
S2 DE "Physicians"
S3 TI Rheumat* OR AB Rheumat*
S4 TI Arthritis OR AB Arthritis
S5 TI Musculoskeletal OR AB Musculoskeletal
S6 TI (("Connective Tissue" N2 (diseases OR disorders OR conditions)) OR AB (("Connective Tissue" N2 (diseases OR disorders OR conditions))
S7 S1 OR S2 OR S3 OR S4 OR S5 OR S6
S8 DE "Education"
S9 TI (educat* OR training OR program*) OR AB (educat* OR training OR program*)
S10 S8 OR S9
S11 S7 AND S10
S12 DE "Medical Education"
S13 TI (((medical OR doctor* OR physician* OR clinician* OR specialty OR specialist*) N2 (educat* OR training OR trainee* OR fellow*))) OR AB (((medical OR doctor* OR physician* OR clinician* OR specialty OR specialist*) N2 (educat* OR training OR trainee* OR fellow*)))
S14 S11 OR S12 OR S13
S15 DE "Measurement" OR DE "Measurement Instruments (1966 1980)"
S16 DE "Surveys"
S17 DE "Questionnaires"
S18 TI (assess* OR self-assess* OR debrief*) OR AB (assess* OR self-assess* OR debrief*)
S19 TI (((written OR oral OR in-training) N2 exam*):) OR AB (((written OR oral OR in-training) N2 exam*)) OR AB (assess* OR self-assess* OR debrief*)
S20 TI (quizz* OR essay* OR question*) OR AB (quizz* OR essay* OR question*)
S21 TI viva* OR AB viva*
S22 TI (((clinical OR practical OR examination) N2 (competenc* OR skill* OR supervis* OR exercise* OR evaluat*))) OR AB (((clinical OR practical OR examination) N2 (competenc* OR skill* OR supervis* OR exercise* OR evaluat*)))
S23 S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22
S24 DE "Validity"
S25 TI (validat* OR verif*) OR AB (validat* OR verif*)
S26 DE "Predictor Variables"
S27 TI "test characteristic*" OR AB "test characteristic*"
S28 TI reproducibil* OR AB reproducibil*
S29 TI reliability OR AB reliability
S30 TI agreement OR AB agreement
S31 TI feasibility OR AB feasibility
S32 S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31
S33 S14 AND S23 AND S32

Web of Science

Dates searched: 1964 to February 21, 2018 and updated 2018 to July 23, 2019

#1 TOPIC: (Rheumat*) OR TOPIC: (Arthritis) OR TOPIC: (Musculoskeletal) OR TOPIC: ("Connective Tissue" NEXT (diseases or disorders or conditions))

#2 TOPIC: (educat* or training or trainee* or fellow*)

#3 TOPIC: (assess* or self-assess* or debrief*) OR TOPIC: ((written or oral or in-training) NEAR/2 exam*) OR TOPIC: (quiz* or essay* or question*) OR TOPIC: (viva*) OR TOPIC: (((clinical or practical or examination) NEAR/2 (competenc* or skill* or supervis* or exercise* or evaluat*)))

#4 TOPIC: (validat* or verif*) OR TOPIC: ("test characteristic*") OR TOPIC: (reproducibil*) OR TOPIC: (reliability) OR TOPIC: (agreement) OR TOPIC: (feasibility)

#5 #1 AND #2 AND #3 AND #4

ProQuest Dissertations & Theses Global

Dates searched: 1980 to February 2018 and updated to July 2019

Set#: S1

Searched for: all(Rheumat* OR Arthritis OR Musculoskeletal) OR all("Connective Tissue" NEXT (diseases or disorders or conditions))

Set#: S2

Searched for: all(educat* or training or trainee* or fellow*)

Set#: S3

Searched for: all(assess* or self-assess* or debrief*) OR all((written or oral or in-training) NEAR/2 exam*) OR all(quiz* or essay* or question*) OR all(viva*) OR all(((clinical or practical or examination) NEAR/2 (competenc* or skill* or supervis* or exercise* or evaluat*)))

Set#: S4

Searched for: all(validat* or verif*) OR all("test characteristic*") OR all(reproducibil*) OR all(reliability) OR all(agreement) OR all(feasibility)

Set#: S5

1 AND 2 AND 3 AND 4

Online Supplementary Text S2: Search strategy for the assessment of competences in other medical specialties

MEDLINE (Ovid)

Dates searched: 1946 to February Week 3 2018 February 25 and updated 2018 to July 22, 2019

1. exp Medical Staff/
2. exp medicine/
3. Allergolog\$.tw.
4. An?esthesiolog\$.tw.
5. Cardiolog\$.tw.
6. Psychiatr\$.tw.
7. ((medical or Clinical) adj genetic\$.tw.
8. (Dermatolog\$ or Venereolog\$.tw.
9. Emergency Medicine.tw.
10. Endocrinolog\$.tw.
11. Gastroenterolog\$.tw.
12. ((general adj (practic\$ or practitioner\$)) or gp\$.tw.
13. Geriatric\$.tw.
14. (Gyn?ecolog\$ or Obstetri\$.tw.
15. Infectious Disease\$.tw.
16. (Internal Medicine or internist\$.tw.
17. Oncolog\$.tw.
18. Nephrolog\$.tw.
19. Neurolog\$.tw.
20. Occupational medicine.tw.
21. Ophtalmolog\$.tw.
22. Orthop?edic\$.tw.
23. Otorhinolaryngolog\$.tw.
24. P?ediatric\$.tw.
25. ((Physical or rehabilitation) adj medicine).tw.
26. Pneumolog\$.tw.
27. Public Health.tw.
28. (radiolog\$ or Radiotherap\$.tw.
29. ((\$port or exercise) adj medicine).tw.
30. or/1-29
31. education/ or ed.fs.
32. (educat\$ or training or program\$.tw.

33. exp Training Support/
34. or/31-33
35. 30 and 34
36. exp Education, Medical/
37. ((medical or doctor\$ or physician\$ or clinician\$ or specialty or specialist\$) adj (educat\$ or training or trainee\$ or fellow\$)).tw.
38. 36 or 37
39. 35 or 38
40. exp Educational Measurement/
41. "Surveys and Questionnaires"/
42. (assess\$ or self-assess\$ or debrief\$).tw.
43. ((written or oral or in-training) adj exam\$).tw.
44. (quizz\$ or essay\$ or question\$).tw.
45. viva\$.tw.
46. ((clinical or practical or examination) adj2 (competenc\$ or skill\$ or supervis\$ or exercise\$ or evaluat\$)).tw.
47. or/40-46
48. validation studies/
49. (validat\$ or verif\$).tw.
50. "Predictive Value of Tests"/
51. test characteristic\$.tw.
52. "reproducibility of results"/
53. reproducibil\$.tw.
54. reliability.tw.
55. agreement.tw.
56. feasibility.tw.
57. or/48-56
58. and/39,47,57
59. ("review" or "review academic" or "review tutorial").pt.
60. (medline or medlars or embase or pubmed).tw,sh.
61. (scisearch or psychinfo or psycinfo).tw,sh.
62. (psychlit or psyclit).tw,sh.
63. cinahl.tw,sh.
64. ((hand adj2 search\$) or (manual\$ adj2 search\$)).tw,sh.
65. (electronic database\$ or bibliographic database\$ or computeri?ed database\$ or online database\$).tw,sh.
66. (pooling or pooled or mantel haenszel).tw,sh.
67. (retraction of publication or retracted publication).pt.

- 68. (peto or dersimonian or der simonian or fixed effect).tw,sh.
- 69. or/60-68
- 70. 59 and 69
- 71. meta-analysis.pt.
- 72. meta-analysis.sh.
- 73. (meta-analys\$ or meta analys\$ or metaanalys\$).tw,sh.
- 74. (systematic\$ adj5 review\$).tw,sh.
- 75. (systematic\$ adj5 overview\$).tw,sh.
- 76. (quantitativ\$ adj5 review\$).tw,sh.
- 77. (quantitativ\$ adj5 overview\$).tw,sh.
- 78. (quantitativ\$ adj5 synthesis\$).tw,sh.
- 79. (methodologic\$ adj5 review\$).tw,sh.
- 80. (methodologic\$ adj5 overview\$).tw,sh.
- 81. (integrative research review\$ or research integration).tw.
- 82. or/70-81
- 83. 58 and 82

The Cochrane Library (www.thecochranelibrary.com)

Dates searched: February 27, 2018 and updated 2018 to July 23, 2019

Note: The Cochrane Database of Systematic Reviews only was searched in the update as DARE and NHS EED ceased to be updated March 2015

- #1 MeSH descriptor: [Medical Staff] explode all trees
- #2 MeSH descriptor: [Medicine] explode all trees
- #3 Allergolog*:ti,ab
- #4 An?esthesiolog*:ti,ab
- #5 Cardiolog*:ti,ab
- #6 Psychiatr*:ti,ab
- #7 ((medical or Clinical) next genetic*):ti,ab
- #8 (Dermatolog* or Venereolog*):ti,ab
- #9 "Emergency Medicine":ti,ab
- #10 Endocrinolog*:ti,ab
- #11 Gastroenterolog*:ti,ab
- #12 ((general next (practic* or practitioner*)) or gp*):ti,ab
- #13 Geriatric*:ti,ab
- #14 (Gyn?ecolog* or Obstetri*):ti,ab
- #15 "Infectious Disease*":ti,ab
- #16 (Internal Medicine or internist*):ti,ab
- #17 Oncolog*:ti,ab

- #18 Nephrolog*:ti,ab
- #19 Neurolog*:ti,ab
- #20 "Occupational medicine":ti,ab
- #21 Ophtalmolog*:ti,ab
- #22 Orthop?edic*:ti,ab
- #23 Otorhinolaryngolog*:ti,ab
- #24 P?ediatric*:ti,ab
- #25 ((Physical or rehabilitation) next medicine):ti,ab
- #26 Pneumolog*:ti,ab
- #27 "Public Health":ti,ab
- #28 (radiolog* or Radiotherap*):ti,ab
- #29 ((Sport or exercise) next medicine):ti,ab
- #30 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #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
- #31 MeSH descriptor: [Education] this term only
- #32 Any MeSH descriptor with qualifier(s): [Education - ED]
- #33 (educat* or training or program*):ti,ab
- #34 MeSH descriptor: [Training Support] explode all trees
- #35 #31 or #32 or #33 or #34
- #36 MeSH descriptor: [Education, Medical] explode all trees
- #37 ((medical or doctor* or physician* or clinician* or specialty or specailist*) next (educat* or training or trainee* or fellow*)):ti,ab
- #38 #30 and #35
- #39 #36 or #37 or #38
- #40 MeSH descriptor: [Educational Measurement] explode all trees
- #41 MeSH descriptor: [Surveys and Questionnaires] this term only
- #42 (assess* or self-assess* or debrief*):ti,ab
- #43 ((written or oral or in-training) next exam*):ti,ab
- #44 (quiz* or essay* or question*):ti,ab
- #45 viva*:ti,ab
- #46 ((clinical or practical or examination) near/2 (competenc* or skill* or supervis* or exercise* or evaluat*)):ti,ab
- #47 #40 or #41 or #42 or #43 or #44 or #45 or #46
- #48 MeSH descriptor: [Validation Studies] this term only
- #49 (validat* or verif*):ti,ab
- #50 MeSH descriptor: [Predictive Value of Tests] this term only
- #51 "test characteristic":ti,ab
- #52 MeSH descriptor: [Reproducibility of Results] this term only
- #53 reproducibil*:ti,ab
- #54 reliability:ti,ab
- #55 agreement:ti,ab
- #56 feasibility:ti,ab
- #57 #48 or #49 or #50 or #51 or #52 or #52 or #53 or #54 or #55 or #56

#58 #39 and #47 and #57

EMBASE (Embase.com)

Dates searched: 1980 to 28 Feb 2018 and updated 2018 to 23 Jul 2019

#56. #29 AND #36 AND #43 AND #53 AND [embase]/lim AND ([systematic review]/lim OR [meta analysis]/lim)
#55. #29 AND #36 AND #43 AND #53 AND [embase]/lim
#54. #29 AND #36 AND #43 AND #53
#53. #44 OR #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52
#52. feasibility:ti,ab
#51. agreement:ti,ab
#50. reliability:ti,ab
#49. reproducibil*:ti,ab
#48. 'reproducibility'/de
#47. 'test characteristic*':ti,ab
#46. 'predictive value'/exp
#45. validat*:ti,ab OR verif*:ti,ab
#44. 'validation study'/exp
#43. #37 OR #38 OR #39 OR #40 OR #41 OR #42
#42. ((clinical OR practical OR examination) NEAR/2 (competenc* OR skill* OR supervis* OR exercise* OR evaluat*)):ti,ab
#41. viva*:ti,ab
#40. quizz*:ti,ab OR essay*:ti,ab OR question*:ti,ab
#39. ((written OR oral OR 'in training') NEAR/2 exam*):ti,ab
#38. assess*:ti,ab OR 'self assess*':ti,ab OR debrief*:ti,ab
#37. 'questionnaire'/exp
#36. #33 OR #34 OR #35
#35. (('medical or doctor*' OR physician* OR clinician* OR specialty OR specailist*) NEAR/2 (educat* OR training OR trainee* OR fellow*)):ti,ab
#34. 'medical education'/exp
#33. #29 AND #32
#32. #30 OR #31
#31. educat*:ti,ab OR training:ti,ab OR program*:ti,ab
#30. 'education'/de
#29. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #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
#28. ((sport OR exercise) NEAR/2 medicine):ti,ab
#27. radiolog*:ti,ab OR radiotherap*:ti,ab
#26. 'public health':ti,ab
#25. pneumolog*:ti,ab
#24. ((physical OR rehabilitation) NEAR/2 medicine):ti,ab

- #23. p?ediatric*:ti,ab
- #22. otorhinolaryngolog*:ti,ab
- #21. orthop?edic*:ti,ab
- #20. ophtalmolog*:ti,ab
- #19. 'occupational medicine':ti,ab
- #18. neurolog*:ti,ab
- #17. nephrolog*:ti,ab
- #16. oncolog*:ti,ab
- #15. 'internal medicine':ti,ab OR internist*:ti,ab
- #14. 'infectious disease*':ti,ab
- #13. gyn?ecolog*:ti,ab OR obstetri*:ti,ab
- #12. ((general NEAR/2 (practic* OR practitioner*)):ti,ab) OR gp*:ti,ab
- #11. gastroenterolog*:ti,ab
- #10. endocrinolog*:ti,ab
- #9. 'emergency medicine':ti,ab
- #8. dermatolog*:ti,ab OR venereolog*:ti,ab
- #7. ((medical OR clinical) NEAR/2 genetic*):ti,ab
- #6. psychiatr*:ti,ab
- #5. cardiolog*:ti,ab
- #4. an?esthesiolog*:ti,ab
- #3. allergolog*:ti,ab
- #2. 'medicine'/de
- #1. 'medical personnel'/exp

PsycINFO (Ovid)

Dates searched: 1806 to February Week 3 2018 and updated 2018 to July Week 3 2019

1. exp medical personnel/
2. family medicine/ or exp medical sciences/
3. Allergolog\$.tw.
4. An?esthesiolog\$.tw.
5. Cardiolog\$.tw.
6. Psychiatr\$.tw.
7. ((medical or Clinical) adj genetic\$.tw.
8. (Dermatolog\$ or Venereolog\$).tw.
9. Emergency Medicine.tw.
10. Endocrinolog\$.tw.
11. Gastroenterolog\$.tw.
12. ((general adj (practic\$ or practitioner\$)) or gp\$).tw.
13. Geriatric\$.tw.

14. (Gynecology or Obstetrics).tw.
15. Infectious Disease\$.tw.
16. (Internal Medicine or internist\$.tw.
17. Oncology\$.tw.
18. Nephrology\$.tw.
19. Neurology\$.tw.
20. Occupational medicine.tw.
21. Ophthalmology\$.tw.
22. Orthopedics\$.tw.
23. Otorhinolaryngology\$.tw.
24. Pediatric\$.tw.
25. ((Physical or rehabilitation) adj medicine).tw.
26. Pneumology\$.tw.
27. Public Health.tw.
28. (radiology or Radiotherapy).tw.
29. ((Sport or exercise) adj medicine).tw.
30. or/1-29
31. EDUCATION/
32. (education or training or program\$.tw.
33. 31 or 32
34. 30 and 33
35. exp medical education/
36. ((medical or doctor or physician or clinician or specialty or specialist) adj (education or training or trainee or fellow)).tw.
37. or/34-36
38. exp educational measurement/
39. exp measurement/
40. (assessment or self-assessment or debrief\$.tw.
41. ((written or oral or in-training) adj exam\$.tw.
42. (quiz or essay or question\$.tw.
43. viva\$.tw.
44. ((clinical or practical or examination) adj2 (competence or skill or supervision or exercise or evaluation)).tw.
45. exp test construction/
46. or/38-45
47. exp test validity/
48. test reliability/

49. (validat\$ or verif\$).tw.
50. test characteristic\$.tw.
51. psychometrics/
52. reproducibil\$.tw.
53. reliability.tw.
54. agreement.tw.
55. feasibility.tw.
56. or/47-55
57. 37 and 46 and 56
58. limit 57 to ("0830 systematic review" or 1200 meta analysis)

PubMed Health

Data searched: February 28, 2018

"clinical competencies"

"clinical skills"

Epistemonikos

Date searched: July 23, 2019.

Note: PubMed Health was discontinued on October 31, 2018 and Epistemonikos was searched as an alternative source of similar content

“clinical competencias” in title or abstract (all publication types)

“clinical skills” in title or abstract (all publication types)

Online Supplementary Text S3: List of relevant medical specialties other than rheumatology

Allergology

Anaesthesiology

Cardiology

Child and Adolescent Psychiatry

Clinical Genetics

Dermatology and Venereology

Emergency Medicine

Endocrinology

Gastroenterology

General practice

Geriatrics

Gynaecology and Obstetrics

Infectious Diseases

Internal Medicine

Medical Oncology

Nephrology

Neurology

Occupational Medicine

Ophthalmology

Orthopaedics

Otorhinolaryngology

Paediatrics

Physical Medicine and Rehabilitation

Pneumology

Psychiatry

Public Health Medicine

Radiation Oncology and Radiotherapy

Rheumatology

Sports and exercise medicine

Online Supplementary Text S4: Exclusion criteria

- Target population: undergraduates, postgraduates not yet in residency/training, consultants, other healthcare professionals, standardised trainees (to train and/or assess potential assessors)
- Specialties: cardiothoracic surgery, clinical neurophysiology, laboratory medicine/ medical biopathology, medical microbiology, neurosurgery, nuclear medicine, oro-maxillo-facial surgery, paediatric surgery, pathology, pharmacology, plastic, reconstructive and aesthetic surgery, radiology, surgery, thoracic surgery, urology
- Non-medical disciplines
- Assessment performed to compare teaching strategies, curricula, courses, other types of training
- Surveys to explore assessors or residents/trainees opinions
- Description of implementation of an assessment tool without exploring validity/reliability/feasibility
- Evaluation of specialty-specific tools not applicable/transferrable to rheumatology
- For SLRs: full list of papers included in the SLR not available/not retrievable

Online Supplementary Table S1 : Medical education research study quality instrument (MERSQI)

Domain	MERSQI item	Item	Score
			Max domain
Study design	1. Study design		3
	Single group cross-sectional or single group posttest only	1	
	Single group pretest and posttest	1.5	
	Nonrandomized, 2 group	2	
	Randomized controlled trial	3	
Sampling	2. No. of institutions studied		3
	1	0.5	
	2	1	
	>2	1.5	
	3. Response rate		
	Not applicable		
	<50 or not reported	0.5	
50-74	1		
>75	1.5		
Type of data	4. Type of data		3
	Assessment by study participant	1	
	Objective measurement	3	
Validity of evaluation instrument	5. Internal structure		3
	Not applicable		
	Not reported	0	
	Reported	1	
	6. Content		
	Not applicable		
	Not reported	0	
	Reported	1	
	7. Relationships to other variables		
	Not applicable		
Not reported	0		
Reported	1		
Data analysis	8. Appropriateness of analysis		3
	Data analysis inappropriate for study design or type of data	0	
	Data analysis appropriate for study design and type of data	1	
	9. Complexity of analysis		
	Descriptive analysis only	1	
Beyond descriptive analysis	2		
Outcomes	10. Outcomes		3
	Satisfaction, attitudes, perceptions, opinions, general facts	1	
	Knowledge, skills	1.5	
	Behaviors	2	
	Patient/health care outcome	3	
TOTAL SCORE			18

Online Supplementary Table S2: Summary of rheumatology articles on objective structured clinical examination

Author, country, year [ref]	Petrusa ER, USA, 1990 [19]	Joorabchi B, USA, 1991 [21]	Skinner BD, USA, 1997 [18]	Hilliard RI, Canada, 1998 [20]	Hergenroeder AC, USA, 2001 [24]	Kramer AWM, Netherlands 2002 [28]	Kramer AWM, Netherlands, 2003 [29]
Competence assessed	Clinical skills	Clinical skills	Clinical skills	Clinical skills	Clinical skills and knowledge	Knowledge and clinical skills	Clinical skills
Tool	OSCE	OSCE	OSCE	OSCE	OSCE	Written exam	OSCE
Comparator tool	Monthly ITE, ABIM exam score	RPRs and Academy of pediatrics ITE	ABFP- ITE and faculty global evaluation	ITE	None	KTS and an OSCE	Written exam
Sample size	74	29	17	43	58	147	73
Workplace based or simulation based	Simulation based	Simulation based	Simulation based	Simulation based	Simulation based		Simulation based
Specialty	Internal medicine	Pediatrics	General Practice	Pediatrics	Pediatrics	General practice	General practice
Current practice or research purpose	Current practice	Research purpose	Current practice	Research purpose	Research purpose	Current practice	Current practice
Control population	none	6 medical students	none	none	none	none	84 GP
Study duration	3 years	-	3 years	1 year	1 month	6 weeks	3 years
Internal consistency (Cronbach alpha unless otherwise stated)	-	0.80	0.12-0.32	0.69	0.80-0.99	0.70	-
Inter-rater reliability	-	-	r=0.60-0.95	-	-	nd	-
Test-retest reliability/intra-rater reliability	-	-	-	-	r=0.62-0.72	nd	-
Inter-method (or parallel forms) reliability /concurrent validity	No correlation of OSCE with ABIM or ITE	OSCE vs RPR r=0.41 OSCE vs ITE r=0.71	OSCE vs ITE r=0.30; OSCE vs faculty r=0.03-0.51	OSCE vs ITE subdomains r values up to 0.57	-	GKT vs OSCE r=0.22	Total OSCE score vs written exam r=0.54
Feasibility	Approximately 100 case presentations to reach a 0.80 g coefficient	-	-	-	-	-	-
Predictive validity	-	-	-	-	-	-	-
Risk of bias	Low	Unclear	Low	Low	High	Low	Low
OSCE, Objective structured clinical examination; ITE, In-training examination; ABIM, American Board of Internal Medicine; ABFP, American Board of Family Practice; RPRs, resident performance ratings; KTS, the written knowledge test of skills; GKT, general knowledge test; GP, general practitioner							

Online Supplementary Table S2 (continued): Summary of rheumatology articles on objective structured clinical examination

Author, country, year	Weller JL, New Zealand, 2005 [30]	Berkenstadt H, Israel, 2006 [17]	Williams JB, USA, 2009 [22]	Boudreau D, Canada, 1994 [27]
Competence assessed	Clinical skills	Clinical skills	Clinical skills	Communication skills
Tool	OSCE	OSCE	OSCE	OSCE
Comparator tool	None	Oral exam	None	None
Sample size	21	145	26	9
Workplace based or simulation based	Simulation based	Simulation based	Simulation based	Simulation based
Specialty	Anesthesiology	Anesthesiology	Emergency medicine	Respiratory medicine
Current practice or research purpose	Research purpose	Current practice	Research purpose	Research purpose
Control population	none	none	none	7 IM residents and 6 pulmonology attending physicians
Study duration	-	2 years	3 weeks	-
Internal consistency (Cronbach alpha unless otherwise stated)	ICC 0.40	r=0.27-0.32	0.8-0.9	0.69 0-0.65
Inter-rater reliability	r= 0.3-0.6	r=0.76-0.89	No difference in scores live or on videotapes	r=0.88
Test-retest reliability/intra-rater reliability	-	-	-	-
Inter-method (or parallel forms) reliability /concurrent validity	-	OSCE vs oral exam subitems r=0.14-0.54	-	-
Feasibility	10–12 cases with three to four judges each rating all the cases to achieve a g coefficient of 0.8	-	-	A g coefficient of 0.8 can be feasibly achieved with 14 cases
Predictive validity	-	-	-	-
Risk of bias	Unclear	Low	Unclear	Low

OSCE, Objective structured clinical examination; IM, internal medicine; ICC, intraclass correlation coefficient

Online Supplementary Table S2 (continued): Summary of rheumatology articles on objective structured clinical examination

Author, country, year [ref]	Savoldelli GL, Canada, 2006 [23]	Van Nuland M, Belgium, 2007 [26]	Leung KK, Taiwan, 2012 [25]
Competence assessed	Clinical and communication skills	Communication skills	ICS. Scales for nurses, patients, and the CAF for staff/peer/self (CAF-S/P/Se)
Tool	OSCE	OSCE	OSCE
Comparator tool	Oral examination	None	OSCE clinical scores
Sample size	20	80	38
Workplace based or simulation based	Simulation based	Simulation based	Simulation based
Specialty	Anesthesiology	General practice	General Practice
Current practice or research purpose	Research purpose	Research purpose	Current practice
Control population	none	none	none
Study duration	-	-	1 month
Internal consistency (Cronbach alpha unless otherwise stated)	0.93-0.98	0.78 and 0.89	ICS-C 0.57 and 0.58. ICS-P 0.96 ICS-N 0.95 CAF-S/P/Se f 0.97, 0.96, and 0.93
Inter-rater reliability	r=0.77-0.87	r=0.26 and 0.49	r = 0.72 for staff vs peer, r = 0.73 for staff vs nurse, r = 0.63 for peer vs nurse. No correlation between patient and self-scoring and any other ratings
Test-retest reliability/intra-rater reliability	-	-	-
Inter-method (or parallel forms) reliability /concurrent validity	OSCE stations scores vs oral examination r=0.52-0.53	-	OSCE evaluation was not correlated to any of the ratings
Feasibility	-	-	-
Predictive validity	-	-	-
Risk of bias	Unclear	Low	Low

OSCE, Objective structured clinical examination; DOPS, direct observation of practical skills; ICS, Interpersonal and communication skills; CAF, competency assessment form; CbD, Case-based Discussion, ACE, Assessment of Clinical Expertise, DOPS, Direct Observation of Procedural Skills; PAT, Peer Assessment Tool; PSQ, Patient Satisfaction Questionnaire, Case Conference (CP), JCP, Journal Club Presentation; ABIM, American Board of Internal Medicine; *Calculation method developed in James LR et al. J Appl Psychol. 1984;69:85-98 ¹Rated with the University of California, San Francisco form (9 point Likert scale); ²Rated with a 6-point score

Online Supplementary Table S3: Summary of articles from other specialties evaluating mini-clinical examination and multisource feedback

Author, country, year [ref]	Norcini JJ, USA, 1995 [36]	Durning SJ, USA, 2002 [37]	Hatala R, Canada, 2006 [38]	Wooliscroft JO, USA, 1994 [41]
Competence assessed	Clinical skills	Clinical skills	Clinical skills	Humanistic qualities
Tool	MiniCEX	MiniCEX	MiniCEX	MSF
Comparator tool	none	Modified ABIM-MEF and ITE	RCPSC IM examination	none
Sample size	88	23	22	70
Workplace based or simulation based	Workplace based	Workplace based	Workplace based	Workplace based
Specialty	Internal Medicine	Internal Medicine	Internal Medicine	Internal medicine and pulmonary medicine
Current practice or research purpose	Current practice	Current practice	Research purpose	Research purpose
Control population	none	none	none	none
Study duration	1 year	3 years	1 year	18 months
Internal consistency (Cronbach alpha unless otherwise stated)	0.65-0.81	0.90	0.79	0.96
Inter-rater reliability	-	-	-	Patients vs nurses, program supervisors and attending physicians r=0.2-0.3; program supervisors vs nurses r=0.5; nurses vs attending physicians r=0.22-0.62 Program supervisors vs attending physicians r=0.85
Test-retest reliability/intra-rater reliability	-	-	-	-
Inter-method (or parallel forms) reliability /concurrent validity	-	MiniCEX vs ABIM subcomponents r=0.7-0.81 MiniCEX vs ITE clinical judgment r=0.6, overall clinical competence r=0.5	Mini-CEX overall clinical competence rating vs the structured oral component, the bedside station subscore and the written examination of the RCPSC IM exam r=0.73, 0.67 and 0.72	-
Feasibility	-	12-14 encounters	-	Up to 20 ratings only by attending physicians and nurses
Predictive validity	-	-	-	-
Risk of bias	Low	Low	Unclear	Unclear

ACGME, Accreditation Council for Graduate Medical Education; MiniCEX, Mini Clinical Evaluation Exercise; ABIM, American Board of Internal Medicine; MEF, monthly evaluation form; ITE, In-Training Examination; RCPSC-IM, Royal College of Physicians and Surgeons of Canada Comprehensive Examination in Internal Medicine; MSF, multisource feedback

Online Supplementary Table S3 (continued): Summary of articles from other specialties evaluating mini-clinical examination and multisource feedback

Author, country, year [ref]	Davis JD, USA, 2002 [42]	Musick DW, USA, 2003 [40]	Massagli TL, USA, 2007 [39]	Murphy DJ, UK, 2009 [33]
Competence assessed	Clinical competence, interpersonal skills, and overall assessment	Overall competence	The ACGME 6 general competencies	Clinical skills
Tool	MSF	MSF	MSF	CA, MSF, PSQs, ARL, SEA, Videos
Comparator tool	none	none	none	Each other
Sample size	16	18	56	171
Workplace based or simulation based	Workplace based	Workplace based	Workplace based	Workplace based
Specialty	Obstetrics and Gynecology	Physical Medicine and rehabilitation	Physical Medicine and rehabilitation	General practice
Current practice or research purpose	Research purpose	Research purpose	Research purpose	Current practice
Control population	none	none	none	none
Study duration	1 year	4 years	3 years	-
Internal consistency (Cronbach alpha unless otherwise stated)	ICC nurses on CC 0.23, IS) 0.45, OA 0.35. ICC attending physicians CC 0.66, IS 0.84, OA 0.79. ICC Peers on CC 0.78, IS 0.90, OA 0.85	0.99	0.77-0.90	CA 0.64, ARL 0.31, SEA 0.25, Videos 0.44 MSF clinical raters 0.80, MSF non clinical raters 0.81, PSQs 0.82
Inter-rater reliability	No correlation between attending physicians and self-evaluations and between nurses and attending physicians Attending physicians vs peers r=0.72-0.86	No significant differences between ratings	Medical students ratings are significantly higher than other raters	CA 0.22, MSF clinical raters 0.30, MSF non clinical raters 0.31, PSQs 0.07, ARL 0.22, SEA 0.25, Videos 0.21
Test-retest reliability/intra-rater reliability	-	-	-	MSF clinical raters 0.29, MSF non clinical raters 0.34, ARL 0.08, Videos 0.20 SEA 0.80
Inter-method (or parallel forms) reliability /concurrent validity	-	-	-	-
Feasibility	-	-	5 ratings from research nurses, 4 from other staff and 23 from medical students. With other raters not possible to feasibly achieve the target.	MSF and PSQs
Predictive validity	-	-	-	-
Risk of bias	Unclear	Unclear	Unclear	Low

MSF, multisource feedback; CA, Criterion audit; PSQs, Patient satisfaction questionnaires; ARL, Assessment of referral letters; SEA, Significant event analysis; ABIM, American Board of Internal Medicine; ICC, intraclass correlation coefficient; CC, clinical competence; IS, interpersonal skills

Online Supplementary Table S4: Summary of articles from other specialties evaluating written exams and the script concordance test and corresponding references

Author, country, year	Norcini JJ, USA, 1984	Veloski JJ, USA, 1999	Gagnon R, France, 2005	Bland AC, USA, 2005
Competence assessed	Knowledge and clinical skills	Knowledge and clinical skills	Clinical reasoning	Clinical reasoning
Tool	Written exam (PMP, MCQs, A-, M-, X-types)	Written exam (UnQ)	SCT	SCT
Comparator tool	Program director scoring	MCQ	none	none
Sample size	8000	14174	80	62
Workplace based or simulation based	Simulation based	Simulation based	Simulation based	Simulation based
Specialty	Internal medicine	General practice	General practice	Nephrology
Current practice or research purpose	Current practice	Current practice	Research purpose	Research purpose
Control population	none	35 board certified family physicians	38 GP	15 medical students, 8 nephrologists
Study duration	3 years	2 years	1 year	-
Internal consistency (Cronbach alpha unless otherwise stated)	Moisier formula: Knowledge 0.76-0.83; synthesis 0.69-0.78; Clinical judgement 0.69-0.74; PMPs 0.72-0.75; Exam composite 0.91-0.92	0.74	0.76	0.68-0.78
Inter-rater reliability	-	-	The higher the number of panel members (at least 15-20) the higher the correlation with a full panel	The results were virtually identical by rotating the raters
Test-retest reliability/intra-rater reliability	-	-	-	-
Inter-method (or parallel forms) reliability /concurrent validity	No significant correlation between written exam and program director scoring	UnQ vs MCQ r=0.73	-	-
Feasibility	-	-	-	-
Predictive validity	-	-	-	-
Risk of bias	Unclear	Low	High	Unclear

PMP, Linear patient management problems; MCQ, Multiple choice questionnaires; A-types, one best answer questions; M-types, matching questions; X-types, multiple true/false questions, UnQ, Open-ended uncued test; SCT, Script concordance test; GP, general practitioner

References

1. Norcini JJ, Swanson DB, Grosso LJ et al. A comparison of knowledge, synthesis, and clinical judgment. Multiple-choice questions in the assessment of physician competence. *Eval Health Prof* 1984;7(4):485-499
2. Veloski JJ, Rabinowitz HK, Robeson MR et al. Patients don't present with five choices: an alternative to multiple-choice tests in assessing physicians' competence. *Acad Med* 1999;74(5):539-546
3. Gagnon R, Charlin B, Coletti M et al. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Med Educ* 2005;39(3):284-291
4. Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med* 2005;80(4):395-399

