

Supplemental Digital Appendix 1 (pg. 1): Search Strategies for Questionnaire-Based Tools for Physicians' Clinical, Teaching, and Research Performance Assessment^a

Supplemental Digital Appendix 2 (pg. 4): Description of the 118 Studies on Questionnaire-Based Assessment Tools for Physicians' Clinical and Teaching Performance Included in a Systematic Analysis of the Literature Published 1966 – October 2016^a

Supplemental Digital Appendix 3 (pg. 22): Validity Evidence Scores of the 15 Questionnaire-Based Assessment Tools on Physicians' Clinical Performance from 46 Studies and 38 Questionnaire-Based Assessment Tools on Physicians' Teaching Performance From 72 Studies Included in a Systematic Analysis of the Literature Published 1966 – October 2016^a

Supplemental Digital Appendix 1

Search Strategies for Questionnaire-Based Tools for Physicians' Clinical, Teaching, and Research Performance Assessment^a

	Clinical performance	Teaching Performance	Research performance
Topic of the study	("Employee Performance Appraisal"[Mesh] OR "Peer Review, Health Care"[Mesh] OR "Health Care Surveys"[Mesh] OR "Peer Review"[Mesh] OR "Feedback"[Mesh] OR "Self-Assessment"[Mesh] OR "Patient Satisfaction"[Mesh] OR "Employee Performance Appraisal"[OT] OR "Health Care Surveys"[OT] OR "Peer Review"[OT] OR "Feedback"[OT] OR "Self-Assessment"[OT] OR "Patient Satisfaction"[OT]) AND (assess*[tiab] OR evaluat*[tiab] OR recertification[tiab] OR improve[tiab] OR measure[tiab]) AND	("Faculty, Medical"[Mesh] OR "Education, Medical"[Mesh] OR "Teaching"[Mesh] OR (teaching[ti] OR teacher*[ti])) AND (assess*[tiab] OR evaluat*[tiab] OR recertification[tiab] OR improve[tiab] OR measure[tiab]) AND	("Research"[Mesh] OR "Research Personnel"[Mesh] OR "Research"[OT] OR "Research Personnel"[OT]) AND (assess*[tiab] OR evaluat*[tiab] OR recertification[tiab] OR improve[tiab] OR measure[tiab]) AND

What will be evaluated	("Clinical competence"[Mesh] OR "Professional Competence/standards"[Mesh] OR performance[tiab] OR skills[tiab] OR qualities[tiab] OR competenc*[tiab] OR practice*[tiab]) AND	(teaching[tiab] OR "vocational training"[tiab] OR "educational framework"[tiab] OR "resident evaluations"[tiab]) AND	((research*[tiab] AND (skills[tiab] OR performance[tiab] OR practices[tiab] OR competence[tiab])) OR (scholar*[tiab] AND (skills[tiab] OR performance[tiab] OR practices[tiab] OR competence[tiab])) OR (scien*[tiab] AND (skills[tiab] OR performance[tiab] OR practices[tiab] OR competence[tiab]))) AND
What type of tool	("Surveys and Questionnaires"[Mesh] OR questionnaire*[tiab] OR survey*[tiab] OR rating*[tiab] OR method[tiab] OR measure[tiab] OR system[tiab] OR instrument[tiab] OR battery[tiab] OR scale[tiab] OR inventory[tiab] OR test[tiab] OR score[tiab] OR scorecard[tiab]) AND	("Surveys and Questionnaires"[Mesh] OR questionnaire*[tiab] OR survey*[tiab] OR rating*[tiab] OR method[tiab] OR measure[tiab] OR system[tiab] OR instrument[tiab] OR battery[tiab] OR scale[tiab] OR inventory[tiab] OR test[tiab] OR score[tiab] OR scorecard[tiab]) AND	("Surveys and Questionnaires"[Mesh] OR questionnaire*[tiab] OR survey*[tiab] OR rating*[tiab] OR method[tiab] OR measure[tiab] OR system[tiab] OR instrument[tiab] OR battery[tiab] OR scale[tiab] OR inventory[tiab] OR test[tiab] OR score[tiab] OR scorecard[tiab]) AND
Analyses	("Validation Studies"[pt] OR ("Clinical Competence/standards"[Mesh] AND "Employee Performance Appraisal"[Mesh]) OR valid*[tiab] OR reliab*[tiab] OR psychometric[tiab] OR factor analys*[tiab] or internal consistency[tiab] OR "Reproducibility of Results"[Mesh]) AND	("Validation Studies"[pt] OR valid*[tiab] OR reliab*[tiab] OR psychometric[tiab] OR factor analys*[tiab] or internal consistency[tiab] OR "Reproducibility of Results"[Mesh]) AND	("Validation Studies"[pt] OR valid*[tiab] OR reliab*[tiab] OR psychometric[tiab] OR factor analys*[tiab] or internal consistency[tiab] OR "Reproducibility of Results"[Mesh]) AND

Whom will be evaluated	(“Physicians”[Mesh] OR physicians[tiab] OR physician[tiab] OR doctors[tiab] OR doctor[tiab] OR clinician*[tiab] OR GP[tiab] OR general practitioner[tiab] OR general practitioners[tiab] OR hospitalist*[tiab] OR anesthesiologist*[tiab] OR anaesthesiologist*[tiab] OR gynecologist*[tiab] OR gynaecologist*[tiab] OR surgeon*[tiab] OR pediatrician*[tiab] OR radiologist*[tiab] OR neurologist*[tiab] OR psychiatrist*[tiab] OR surgical[tiab])	(“Physicians”[Mesh] OR physicians[tiab] OR physician[tiab] OR doctors[tiab] OR doctor[tiab] OR clinician*[tiab] OR GP[tiab] OR general practitioner[tiab] OR general practitioners[tiab] OR hospitalist*[tiab] OR anesthesiologist*[tiab] OR anaesthesiologist*[tiab] OR gynecologist*[tiab] OR gynaecologist*[tiab] OR surgeon*[tiab] OR pediatrician*[tiab] OR radiologist*[tiab] OR neurologist*[tiab] OR psychiatrist*[tiab] OR surgical[tiab] OR faculty[tiab] OR teacher*[tiab] OR educator*[tiab] OR instructor*[tiab] or physician*[tiab] OR trainer*[tiab] OR attending*[tiab] OR doctor*[tiab] OR resident*[tiab] OR supervisor*[tiab])	(“Physicians”[Mesh] OR physicians[tiab] OR physician[tiab] OR doctors[tiab] OR doctor[tiab] OR clinician*[tiab] OR GP[tiab] OR general practitioner[tiab] OR general practitioners[tiab] OR hospitalist*[tiab] OR anesthesiologist*[tiab] OR anaesthesiologist*[tiab] OR gynecologist*[tiab] OR gynaecologist*[tiab] OR surgeon*[tiab] OR pediatrician*[tiab] OR radiologist*[tiab] OR neurologist*[tiab] OR psychiatrist*[tiab] OR surgical[tiab] OR "physician-scientist" OR "clinical- investigator"[tiab])
------------------------------	--	--	--

^aThe search has been conducted in the following databases: PubMed, ERIC, PsycINFO and Web of Sciences.

Supplemental Digital Appendix 2

Description of the 118 Studies on Questionnaire-Based Assessment Tools for Physicians' Clinical and Teaching Performance Included in a Systematic Analysis of the Literature Published 1966 – October 2016^a

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
Al Ansari 2016 ⁶³	BDF	1	EM	30	269 colleague evaluations	BH	39 items; 5 point Likert scale	n/a, 4.3 min	Paper-based mail	12
Archer 2011 ⁵²	SPRAT	1	IM S GP	68	626 assessors	GB	25 items; 6 point Likert scale, 13 items; 5 point scale	n/a	n/a	n/a
Bhogal 2012 ⁵⁶	QBT1	1	n/a	22	Evaluated each other	US	7 items; 5 point Likert scale	n/a	n/a	n/a
Campbell 2010 ⁵⁰	CFEP360	n/a	Pc	179	2421 colleagues, 8474 patients	GB	18 colleague items, 14 patient items; n/a	n/a	n/a	n/a
Campbell 2008 ³⁹	GMC CQ	18	multiple	309	13754 patients 4269 colleague	GB	16 patient items, 27 colleague items; 5 point Likert scale, 2 patient items,	n/a	n/a	22 patients, 8 colleagues

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
Campbell 2011 ⁵³	GMC CQ	11	n/a	1065	17031 colleagues, 30333 patients	GB	1 colleague item; binary scale 18 colleague items, 9 patient items; 5 point scale	n/a	n/a	22 patients, 8 colleagues
Carline 1989 ¹⁹	ABIM PAR	multiple	IM	255	1249 colleague evaluations	US	9 items; 6 point Likert scale	n/a	Paper-based mail	12 total score, for individual items varied from 10 to 32
Crossley 2008 ⁴⁰	SPRAT	1	RA	107	577 colleagues	GB	25 items; 6 point Likert scale	n/a	Paper-based mail	n/a
Elwyn 2005 ³²	Adapted ABIM PAR	1	GP	113	1271 colleagues	GB	10 items; 9 point Likert scale,	n/a	n/a	15
Fidler 1999 ²⁵	CPSA-PAR	n/a	FM GP	220	4302 colleague evaluations	CA	26 items, 23 items, 21 items, 17 items; 5 point Likert scale	200 Canadian dollars, n/a	n/a	n/a
Hall 1999 ²⁶	CPSA-PAR	n/a	FM GP	308	4302 colleague	CA	26 items, 23	\$200	n/a	n/a

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
			OG IM P		evaluations		items, 21 items, 17 items; 5 point Likert scale	per physician, n/a		
Hess 2009 ⁴⁶	CRP-PIM	n/a	IM	803	12212 colleagues	n/a	13 items; 6 point Likert scale	n/a	Paper-based mail or person, e-mail	>10 referring physicians
Hill 2012 ⁵⁷	GMC CQ	2	GP	12	n/a	GB	n/a; 5 point Likert scale	n/a	n/a	20
Lelliot 2008 ⁴¹	ACP 360	n/a	Ps	347	4422 colleagues 6657 patients	GB	17 patient items, 57 colleague items; 6 point Likert scale	n/a	Web- and Paper-based	13 colleagues, 25 patients
Lipner 2002 ²⁷	ABIM PAR	n/a	IM	356	3560 colleagues	US	11 items; 9 point Likert scale	n/a, 8 min	Telephone survey	>10
Lockyer 2003 ²⁸	CPSA-PAR	n/a	GS Vs Ns Op U O ENT Ps OG	144	n/a	CA	31 items, 17 items; 5 point Likert scale	n/a	n/a	8 colleagues, 8 coworkers, 25 patients
Lockyer 2004 ³¹	CPSA-PAR	n/a	IM P Ps	304	2306 colleagues	CA	36 items; 5 point Likert	n/a	n/a	7.6

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
							scale			
Lockyer 2006a ³⁵	CPSA-PAR	n/a	A	186	2822 colleagues, 3135 patients	CA	11 patient items, 19 coworker items, 29 peer items; 5 point Likert scale	n/a	n/a	8 colleagues, 8 coworkers, 25 patients
Lockyer 2006b ³⁶	CPSA-PAR	n/a	EM	187	2850 colleagues, 4039 patients	CA	16 patient items, 20 coworker items, 30 colleague items; 5 point Likert scale	n/a	n/a	8 colleagues, 8 coworkers, 25 patients
Lockyer 2008 ⁴²	CPSA-PAR	n/a	R	190	6838 colleagues	CA	38 peer items, 29 referral items, 20 coworker items; 5 point Likert scale	n/a	n/a	8 colleagues, 8 coworkers, 25 patients
Lockyer 2009 ⁴⁷	CPSA-PAR	n/a	Pa	101	2210 colleagues	CA	39 peer items, 30 referral items, 22 coworker items; 5 point Likert scale	n/a	n/a	8 peers, 8 referrals, 8 coworkers

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
Mackillop 2011a ⁵⁴	GMC GQ	n/a	GP	205	2789 colleague evaluations	GB	21 items; 4 point Likert scale	n/a	Email	>15
Mackillop 2011b ³	GMC GQ	n/a	A EM GP OG Op P Pa Ps R S	977	12540 colleague evaluations	GB	10 items; 4 point Likert scale	5 min, n/a	Electronic ally	>12
Overeem 2012a ⁵⁸	IFMS	26	S IM	146	1758 colleagues, 1960 patients	NL	33 peer items, 22 coworker items, 22 patient items; 9 point Likert scale	n/a	n/a	5 peers, 5 coworkers, 11 patients
Overeem 2012b ⁵⁹	IFMS	26	D C Pd IM Ps N P A R lab GS U O OG Op ENT	238	n/a	NL	n/a	n/a	n/a	n/a
Ramsey 1989 ²⁰	ABIM PAR	n/a	IM	259	n/a	US	n/a; 9 point Likert scale	n/a	Paper-based mail	n/a
Ramsey 1993 ²¹	ABIM PAR	n/a	IM	314	n/a	US	n/a; 9 point Likert scale	n/a	Paper-based mail	>11

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
Ramsey 1996 ²³	ABIM PAR	11	IM	228	3005 colleague evaluations	US	11 items; 9 point Likert scale	n/a	Paper-based mail	>10
Richards 2009 ⁴⁸	GMC CQ	n/a	AC PC	309	1636 colleagues	GB	17 items; 5 point Likert scale, 1 item; binary scale	n/a	Paper-based mail, email	8
Rosenbaum 2005 ³³	ABMS/ACGME Faculty Peer Ratings	1	FM	21	n/a	US	19 items; 10 point Likert scale	n/a	n/a	n/a
Sargeant 2003 ²⁹	CPSA-PAR	n/a	FM	142	1876 colleagues	CA	31 items, 17 items; 5 point Likert scale	n/a	n/a	8 colleagues, 8 coworkers, 25 patients
Sargeant 2005 ³⁴	CPSA-PAR	n/a	FM	15	n/a	CA	26 items, 23 items, 21 items, 17 items; 5 point Likert scale	n/a	n/a	8 colleagues, 8 coworkers, 25 patients
Sargeant 2007 ³⁸	CPSA-PAR	n/a	FM C IM D EM OG O S	23	n/a	CA	n/a	n/a	n/a	n/a

Supplemental digital content for van der Meulen MW, Smirnova A, Heeneman S, oude Egbrink MGA, van der Vleuten CPM, Lombarts KMJMH. Exploring validity evidence associated with questionnaire-based tools for assessing the professional performance of physicians: A systematic review. Acad Med.

Sargeant 2008 ⁴³	CPSA-PAR	n/a	FM	n/a	n/a colleagues, n/a patients	CA	n/a	n/a	n/a	8 colleagues
--------------------------------	----------	-----	----	-----	---------------------------------	----	-----	-----	-----	-----------------

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
										, 8 coworkers, 25 patients
Sargeant 2009 ⁴⁹	CPSA-PAR	n/a	FM	28	n/a	CA	n/a	n/a	n/a	n/a
Sargeant 2011 ⁵⁵	CPSA-PAR	n/a	FM	28	n/a	CA	n/a	n/a	n/a	n/a
Shepherd 2010 ⁵¹	QBT2	10	GP	176	n/a	GB	n/a	n/a	email	n/a
Vinod 2013 ⁶¹	CPSA-PAR	1	On	7	55 patients, 123 colleagues	AU	n/a	n/a, 900 Australian dollars	Paper-based mail	10 patients, 10 coworkers, 10 referrals
Violato 1997 ²⁴	CPSA-PAR	n/a	FM	28	734 patients, 673 colleagues	CA	26 items, 23 items, 21 items, 17 items; 5 point Likert scale	n/a, 200 Canadian dollars	Paper-based	6
Violato 2003 ³⁰	CPSA-PAR	n/a	Vs OG Ps ENT O GS Ths Ns Op U S	201	2859 colleagues 4185 patients	CA	34 colleague items, 19 coworker items; n/a	n/a	n/a	8 colleagues, 8 coworkers, 25 patients
Violato 2006 ³⁷	CPSA-PAR	n/a	P	100	2341 patients, 1522	CA	40 patient	n/a	n/a	8

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
					colleagues		items, 22 coworker items, 38 colleague items; 5 point Likert scale			colleagues, 8 coworkers, 25 patients
Violato 2008a ⁴⁴	CPSA-PAR	n/a	FM GP	250	n/a	CA	31 items, 17 items, 40 patient items; 5 point Likert scale	n/a	n/a	8 colleagues, 8 coworkers, 25 patients
Violato 2008b ⁴⁵	CPSA-PAR	n/a	Ps	101	2456 patient evaluations 1508 colleague evaluations	CA	40 patient items, 22 coworker items, 38 colleague items; 5 point Likert scale	n/a	Paper-based mail	8 colleagues, 8 coworkers, 25 patients
Warner 2015 ⁶²	MOCA	1	A	46	732 colleague evaluations	US	n/a	n/a	Web-based	45 patients, 10 colleagues
Wenrich 1993 ²²	ABIM PAR	175	IM	232	1877 colleagues	US	13 items, n/a items; 9 point Likert scale	n/a	Paper-based mail	10-15
Wright 2012 ⁶⁰	GMC CQ	10	multiple	1057	17012 colleague evaluations	GB	18 items; 5 point Likert scale, 1 item;	n/a	Paper- or web-based	34 patients, 15

Author, year & reference	Instrument	No. of insti- tutions	Spe- cialty	No. of physi- cians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and dura- tion	Platform	No. of assessors needed
							binary scale			colleagues
Teaching Performance Tools										
Afonso 2005 ⁸²	QBT1	1	IM	30	n/a residents and students	US	18 items; 5- point Likert scale	n/a	Paper- based	n/a
Arah 2011 ¹⁰³	SETQ	16	IM C N P R RT CG Pa NM PR Ps	494	403 residents	NL	23 core items, 2 global ratings; 5-point Likert scale	n/a	Web- based	4
Arah 2012 ¹⁰⁸	SETQ	20	N/A	962	690 residents	NL	22 core items; 5 point-Likert scale	n/a	Web- based	5
Archer 2013 ¹¹³	LDMES	1	AC MH PC	665	3587 resident evaluations	GB	18 items, 1 global rating; 6- point scale	n/a	Web- based	n/a
Backeris 2013 ¹¹⁴	QBT2	1	A	133	n/a	US	13 items; 9 point scale	n/a	Electronic ally	n/a
Baker 2010 ⁹⁸	QBT3	1	A	197	194 residents	US	7 items; 10 point Likert scale	n/a	Paper- based	2
Beckman 2005 ⁸³	MTE	1	IM	60	n/a residents	US	14 items; 5 point scale	n/a	Electronic ally	n/a
Beckman 2006 ⁸⁵	MTE	1	IM C	126	n/a residents	US	14 items; 5 point scale	n/a	Electronic ally	n/a

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
Beckman 2010 ⁹⁹	MTE	1	IM	356	209 residents	US	16 items; 5 point scale	n/a	Electronically	n/a
Bierer 2007 ⁸⁹	CTE	1	IM S P A R Pa	872	n/a residents medical students	US	15 items, 1 global rating; 5-point Likert scale	n/a	Web-based	1 to 6
Boerebach 2012 ¹⁰⁹	SETQ	16	S Ns O Op ENT P su U	302	269 residents	NL	26 items, 2 global ratings; 5-point Likert scale	n/a	Web-based	n/a
Boerebach 2016 ¹²⁸	SETQ	46	Multiple	2835	2021 trainees	NL	20 items; 5-point Likert scale	n/a	Web-based	n/a
Coletti 2010 ¹⁰⁰	RMS	1	EM	31	27 residents	US	18 items; 9-point scale	n/a	Electronically	n/a
Copeland 2000 ⁷³	(CC) CTEI	1	A IM Pa P R S	711	n/a	US	15 items; 5 point scale	n/a	n/a	n/a
Da Dalt 2015 ¹²⁵	TAQ	1	P	26	51 residents	IT	8 items; 5 point scale	n/a	Web-based	n/a
De Groot 2003 ⁷⁶	QBT4	1	Ps	289	1765 resident evaluations	CA	7 items; 5 point scale	n/a	n/a	n/a
De Oliveira 2008 ⁹⁴	QBT5	4	A	38	18 residents	n/a	9 items; 4 point scale	n/a	Web-based	n/a
Dexter 2016 ¹²⁹	QBT5	1	A	76	14585 resident evaluations	US	9 items; 4 point scale	n/a	Web-based	n/a

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
Donner-Banzhoff 2003 ⁷⁷	QBT6	n/a	GP	n/a	101 registrars	DE	43 items; n/a scale	n/a	n/a	n/a
Egbe 2012 ¹¹⁰	QBT7	n/a	n/a	31	128 trainees, 115 fellow trainers	GB	25 items on a 4 point scale	n/a	Web- and paper-based	12
Fluit 2012 ¹¹¹	EFFECT	4	P Pd S	117	106 residents	NL	58 items; 5-point Likert-scale	n/a, <10 min	Web-based	n/a
Fluit 2013 ¹¹⁵	EFFECT	1	n/a	24	237 residents evaluations	NL	58 items; 5-point Likert scale	n/a, <10 min	Web-based	n/a
Hindman 2013 ¹¹⁶	QBT5	1	A	49	39 residents	US	9 items; 4 point scale	n/a	Web-based	15
Huete 2016 ¹³⁰	MEDUC-RX32	1	R	18	28 residents	CL	32 items; 7 point Likert scale	n/a	n/a	8
Kelly 2007 ⁹⁰	EDTS	1	EM	31	36 residents	US	7 items; 10 point Likert scale	n/a	n/a	n/a
Kikukawa 2014 ¹¹⁹	QBT8	1	IM P Es Bs U OG En D N Ic	12	10 residents 5 educational experts	JP	25 items; 6 point scale	n/a	n/a	n/a
Kripalani 2004 ⁷⁹	CTE	1	IM	63	423 medical students and	US	25 items; 6 point scale	n/a, 20 min	n/a	n/a

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
					housestaff					
Lases 2014 ¹²⁰	SETQ	17	S G	302	204 residents	NL	20 items; 5 point scale	n/a	Web-based	n/a
Lee 2014 ¹²¹	MTE	1	IM	123	n/a residents	US	17 items; n/a	n/a	n/a	n/a
Litzelman 1999 ⁷²	SFDP	1	IM P	36	45 residents	US	26 items; n/a	n/a	Paper-based	n/a
Logio 2011 ¹⁰⁴	HTPE	1	IM P	241	886 resident evaluations	US	10 items; 5 point Likert scale	n/a	n/a	n/a
Lombarts 2009 ⁹⁶	SETQ	1	A	36	30 residents	NL	24 items; 2 global ratings; 5 point Likert scale	n/a	Web-based	4
Lombarts 2010 ¹⁰¹	SETQ	15	IM C Ga Chm N R Rt P Gs A Ns Ps Op OG PR CG Pa O ENT	662	407 residents	NL	22 items; 5 point Likert scale	n/a	Web-based	n/a
Lombarts 2014 ¹²²	SETQ	17	n/a	502	451 residents	NL	22 items; 5 point Likert scale	n/a	Web-based	n/a

Author, year & reference	Instrument	No. of insti- tutions	Spe- cialty	No. of physi- cians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and dura- tion	Platform	No. of assessors needed
Lombarts 2016 ¹³¹	SETQ (Smart)	n/a	A	247	240 residents	AT, DK, DE, NL, SE, GB	25 items; 7 point Likert scale	n/a	Web- based	n/a
Maker 2004 ⁸⁰	QBT9	1	S	44	39 residents	US	9 items; 3 point scale	n/a	n/a	n/a
Maker 2006 ⁸⁶	QBT9	1	S	42	40 residents	US	9 items; 3 point scale	n/a	n/a	n/a
McLeod 1991 ⁶⁶	CTE		IM	24	n/a	CA	25 items; 6 point scale	n/a	Paper- based mail	n/a
McOwen 2007a ⁹²	QBT10	1	n/a	399	436 residents	US	7 items; 5 point scale, 5 items; 2 point scale	n/a	Web- based	>4
McOwen 2007b ⁹¹	QBT10	18	n/a	720	516 residents	US	9 items; 5 point scale	n/a	Web- based	n/a
Metz 1996 ⁶⁴	QBT11	1	IM	23	215 students, 162 residents	US	8 items; 5 point scale	n/a	Paper- based	>5
Mintz 2015 ¹²⁶	SFDP	1	IM	n/a	119 medical students	CA	25 items; 5 point Likert scale	n/a	n/a	n/a
Mourad 2006 ⁸⁷	TES	Mu- lti- cen	IM	40	677 resident, intern, medical student evaluations	CA	15 items; 5 point scale	n/a	n/a	n/a

Author, year & reference	Instrument	No. of insti- tutions	Spe- cialty	No. of physi- cians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and dura- tion	Platform	No. of assessors needed
		tre	C He Id Re A Cm EM FM IM OG P Ro R S							
Nation 2011 ¹⁰⁵	CTAI	1		170	14 clinical clerks, 229 residents, 53 fellows, 21 n/a	CA	19 items; 5 point scale	n/a	n/a	n/a
Ramsbottom- Lucier 1994 ⁶⁹	CTAF	5	IM	29	639 resident evaluations	US	8 items; 6 point scale	n/a	Paper- based mail	>10
Risucci 1992 ⁶⁸	QBT12	n/a	S	62 in '88, 64 in '89	23 in '88, 24 in '89 residents	n/a	10 items; 5 point scale	n/a	Paper- based mail	n/a
Robinson 2015 ¹²⁷	QBT13	2	IM	18	32 medical students	US	10 items; 5 point Likert scale	n/a	n/a	n/a
Scheepers 2014 ¹²³	SETQ	18	25 specialti es, 7 surgical	622	560 residents	NL	21 items; 5 point Likert scale	n/a	Web- based	n/a
Scheepers 2016 ¹³²	SETQ	18	n/a	636	549 residents	NL	23 items; 5 point Likert scale	n/a	Web- based	n/a
Schönrock- Adema 2012 ¹¹²	CTEI	n/a	n/a	n/a	112 residents	NL	15 items; 5 point scale	n/a	n/a	n/a

Author, year & reference	Instrument	No. of institutions	Specialty	No. of physicians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and duration	Platform	No. of assessors needed
Schum 1996 ⁷⁰	QBT14	1	P	44	n/a	US	10 items; 7 point scale	n/a	Paper-based	n/a
Shea 2002 ⁷⁵	QBT15	1	IM	132	163 students, 219 residents	US	10 items; 4 points scale, 5 items; 5 point scale	n/a	Web-based	n/a
Shea 2009 ⁹⁷	QBT15	1	n/a	1210	18012 trainees evaluations	US	9 items; 5 point scale	n/a	Web-based	n/a
Silber 2006 ⁸⁸	QBT16	1	IM S	11	89 residents, 1 program director	US	23 items; 5 point scale	n/a	n/a	n/a
Smith 2004 ⁸¹	QBT17	1	IM	99	145 residents	US	32 items; 5 point Likert scale	n/a, 10 min	Paper-based mail	>8
Solomon 1997 ⁷¹	QBT18	1	IM	147	1570 clerk evaluations	US	13 items; 4 point Likert scale	n/a	Paper-based	n/a
Stalmeijer 2008 ⁹⁵	MTCQ	n/a	-	-	10 educationalists, 16 doctors, 12 medical students	NL	27 items; 5 point scale	n/a, 5 min	Paper-based	n/a
Stalmeijer 2010 ¹⁰²	MTCQ	2	IM S P OG N D ENT Op Ps	291	1315 medical students evaluations	NL	24 items; 5 point scale	n/a, 5 min	Paper-based	n/a
Steiner 2000 ⁷⁴	ER scale	3	EM	29	18 residents	CA	4 items; 5 point Likert scale	n/a	Paper-based	n/a
Steiner 2003 ⁷⁸	ER scale	7	EM	115	562 residents	CA	4 items; 5 point	n/a	Paper-	n/a

Author, year & reference	Instrument	No. of insti- tutions	Spe- cialty	No. of physi- cians	No. and type assessors	Study origin	No. and type of items	Feasibility Tool		
								Costs and dura- tion	Platform	No. of assessors needed
							Likert scale		based	
Steiner 2005 ⁸⁴	ER Scale	7	EM	115	562 residents	CA	4 items; 5 point Likert scale	n/a	Paper- based	n/a
Tiberius 1989 ⁶⁵	QBT19	1	IM	n/a	n/a	CA	52 items; 7 point scale	n/a	n/a	n/a
Tortolani 1991 ⁶⁷	QBT12	n/a	S	62	47 residents	US	10 items; 5 point Likert scale	n/a	Paper- based	n/a
Van der Hem- Stokroos 2005 ¹³⁵	CTEI	1	S	51	n/a	NL	15 items; 5 point Likert scale	n/a	Paper- based	n/a
Van der Leeuw 2011 ¹⁰⁶	SETQ	9	OG	99	77 residents	NL	26 items; 5 point Likert scale, 2 global ratings	n/a	Web- based	n/a
Van der Leeuw 2013 ¹¹⁷	SETQ	6	IM N OG ENT P R S	24	n/a	NL	20-25 items; 5 point scale	n/a	Web- based	n/a

Van der Leeuw 2013 ^{b118}	SETQ	20	IM P D On Ps R A Pa S Os U OG Op ENT ThS Vs Ns	917	659 residents	NL	23-30 specialty-specific questions; 5 point Likert scale	n/a	Web-based	n/a
Van der Leeuw 2016 ¹³³	SETQ	16	n/a	992, 847	901, 816, 862 residents	NL	22 items, 2-8 specialty specific; 5 point	n/a	Web-based	n/a
Wingo 2016 ¹³⁴	MTE	1	IM	107	542 residents	US	14 items; 5 point scale	n/a	n/a	n/a
Young 2014 ¹²⁴	PACT	1	IM P FM Ps S OG	567	178 clerks	CA	18 items; 5 point scale	n/a	Web-based	>2
Zibrowski 2011 ¹⁰⁷	CTEI & SETOC	1	IM P	223	3034 resident evaluations	CA	15 items; 7 point scale	n/a	Web- & Paper-based	n/a
Zuberi 2007 ⁹³	SETOC	1	S IM S Op ENT FM Os P OG	87	224 clerks	CA & PK	15 items; 7 point scale	n/a	n/a	n/a

Abbreviations: n/a, not available. *Instruments* BDF = Bahrain Defense Force, SPRAT = The Sheffield Peer Rating Assessment Tool, CFEP360 = n/a, GMC CQ = General Medical Council patient and colleague questionnaires, ABIM PAR = American Board of Internal Medicine Peer Assessment Review, Adapted ABIM PAR = Adapted American Board of Internal Medicine Peer Assessment Review, CPSA-PAR = College of Physicians and Surgeons of Alberta Physician Achievement Review, CRP-PIM = Communication with Referring Physicians Practice Improvement Module, ACP 360 = n/a, GMC GQ = General Medical Council generic questionnaire, IFMS = Individueel Functioneren Medisch Specialisten, ABMS/ACGME Faculty Peer Ratings = American Board of Medical Specialties/Accreditation Council for Graduate Medical Education Faculty Peer Ratings, MOCA = Maintenance of Certification in Anesthesiology Program patient and peer surveys, CTAF = Clinical Teaching Assessment Form, CTE = Clinical Teaching Effectiveness, CTEI & (CC) CTEI = (Cleveland Clinic's) Clinical Teaching Effectiveness Instrument, EDTS = Emergency Department Teaching Survey, EFFECT = Evaluation and Feedback For Effective Clinical Teaching, ER scale =

Supplemental digital content for van der Meulen MW, Smirnova A, Heeneman S, oude Egbrink MGA, van der Vleuten CPM, Lombarts KMJM. Exploring validity evidence associated with questionnaire-based tools for assessing the professional performance of physicians: A systematic review. Acad Med.

Emergency Rotation Scale, HTPE = Humanistic Teaching Practices Effectiveness Questionnaire = LDMES = London Deanery 'MSF for Educational Supervisors', MEDUC-RX32 = Medicina Universidad Católica—Radiology 32 items, MTCQ = Maastricht Clinical Teaching Questionnaire, MTE = Mayo Teaching Effectiveness, PACT = Professionalism Assessment of Clinical Teachers, RMS = Residency Management Suite, SETOC = Student Evaluation of Teaching in Outpatient Clinics, SETQ & SETQ (Smart) = System for Evaluation of Teaching Qualities, SFDP = Stanford Faculty Development Program, TAQ = Tutor Assessment Questionnaire, TES = Teaching Effectiveness Scores, *Specialty* A = Anesthesiology, En = Endocrinology, Id = Infectious Diseases, Op = Ophthalmology, R = Radiology, Bs = Brain Surgery, ENT = Ear Nose Throat, IM = Internal Medicine, P = Pediatrics, Re = Respiriology, AC = Acute Care, ES = Emergency Surgery, MH = Mental Health, Pa = Pathology, Ro = Radiation Oncology, C = Cardiology, FM = Family Medicine, N = Neurology, PC = Primary Care, RT = Radiotherapy, CG = Clinical Genetics, G = Gynecology, NM = Nuclear Medicine, Pd = Pulmonary Diseases, S = Surgery, ChM = Chest medicine, Ga = Gastroenterology, NS = Neurosurgery, PR = Physical Rehabilitation, TS = Thoracic Surgery, CM = Community Medicine, He = Hematology, O = Orthopedics, Ps = Psychiatry, U = Urology, D = Dermatology, Ic = Infection Control, OG = Obstetrics & Gynecology, Psu = Plastic Surgery, Vs = Vascular Surgery. *Study origin* AT = Austria, AU = Australia, BH = Bahrain, CA = Canada, CL = Chile, DE = Germany, DK = Denmark, SE = Sweden, GB = United Kingdom, IT = Italy, JP = Japan, NL = The Netherlands, PK = Pakistan, US = United States

“The feasibility of the tools was determined by examining the “Costs and duration”, “Platform” and “No. of assessors needed” which implied respectively, how much the questionnaire-based tool costs to use, how long it would take assessors to fill out the questionnaire, how the questionnaire-based tool was administered, and how many assessors were needed to achieve reliable scores.

Supplemental Digital Appendix 3

Validity Evidence Scores of the 15 Questionnaire -Based Assessment Tools on Physicians' Clinical Performance from 46 Studies and 38 Questionnaire-Based Assessment Tools on Physicians' Teaching Performance From 72 Studies Included in a Systematic Analysis of the Literature Published 1966 – October 2016^a

Instrument & References	Scoring			Generalization		Extrapolation		Implications	
	Items	Raters	Scores	Reliability	Generalizability	Constructs	Performance	Intended	Unintended
ACP 360 ⁴¹	2	1	2	3	2	2	2	0	0
BDF ⁶³	3	2	1	3	3	2	0	0	0
CRP-PIM ⁴⁶	0	1	1	0	3	2	2	0	2
ABIM PAR ^{19-23,27}	3	2	2	0	3	3	2	1	0
ABIM Par Adapted ³²	2	1	1	0	0	0	0	0	0
ABMS/ACGME FPR ³³	3	1	1	2	0	2	0	0	0
QBT2 ⁵¹	3	0	0	0	0	0	0	1	0
MOCA ⁶²	0	1	1	0	0	0	0	2	0
GMC M ^{3,54}	3	2	1	0	3	0	0	0	0
IFMS ^{58,59}	3	1	1	3	0	2	2	2	0
GMC	3	1	1	3	2	2	3	2	2

Instrument & References	Scoring			Generalization		Extrapolation		Implications	
	Items	Raters	Scores	Reliability	Generalizability	Constructs	Performance	Intended	Unintended
CQ ^{39,48,53,57,60}									
CFEP 360 ⁵⁰	0	1	1	3	3	2	2	0	0
CPSA-PAR ^{24-26,28-31,34-38,42-45,47,49,55,61}	3	1	1	3	3	3	0	2	0
SPRAT ^{40,52}	3	2	3	0	0	0	2	0	2
QBT1 ⁵⁶	3	2	0	0	0	0	2	0	2
Teaching									
MTE ^{83,85,99,121,134}	3	0	0	3	0	2	3	2	2
CTAF ⁶⁹	3	2	0	2	0	0	0	0	0
CTE ^{66,79,89}	3	2	1	0	0	0	2	0	0
CTEI ^{73,107,112,135}	3	2	2	3	3	2	2	0	0
EFFECT ^{111,115}	2	1	0	2	0	3	0	2	0
EDTS ⁹⁰	2	0	0	0	0	0	3	0	0
ER SCALE ^{74,78,84}	0	2	0	3	0	0	3	2	0
HTPE ¹⁰⁴	0	0	1	3	0	2	3	0	0
MEDUC-	0	0	1	3	3	2	3	0	0

Instrument & References	Scoring			Generalization		Extrapolation		Implications	
	Items	Raters	Scores	Reliability	Generalizability	Constructs	Performance	Intended	Unintended
RX32 ¹³⁰									
MCTQ ^{95,102}	3	1	0	3	3	3	0	0	0
QBT5 ^{94,116,129}	3	2	1	3	3	2	3	0	0
PACT ¹²⁴	3	1	0	3	3	2	3	0	0
RMS ¹⁰⁰	2	0	1	2	0	2	0	0	0
SETOC ^{93,107}	3	1	0	3	3	2	3	0	0
SETQ ^{96,101,103,106,108,109,117,118,120,122,123,128,131-133}	3	2	1	3	3	3	3	2	0
SFDP ^{72,126}	2	1	0	3	0	3	0	0	0
TAQ ¹²⁵	0	1	0	3	0	0	0	0	0
TES ⁸⁷	2	0	0	3	0	0	3	0	0
LDMES ¹¹³	3	2	1	3	3	3	2	2	2
QBT1 ⁸²	2	2	0	3	0	0	0	0	0
QBT2 ¹¹⁴	0	0	0	0	0	0	3	2	0
QBT3 ⁹⁸	0	1	0	3	0	0	3	2	0
QBT4 ⁷⁶	0	0	0	0	0	0	2	0	2
QBT6 ⁷⁷	3	0	0	0	0	0	0	0	0

Instrument & References	Scoring			Generalization		Extrapolation		Implications	
	Items	Raters	Scores	Reliability	Generalizability	Constructs	Performance	Intended	Unintended
QBT7 ¹¹⁰	0	1	0	0	0	0	0	0	0
QBT8 ¹¹⁹	3	0	0	0	0	0	0	0	0
QBT9 ^{80,86}	1	1	0	0	0	0	0	2	0
QBT10 ^{91,92}	1	1	1	3	3	2	3	0	2
QBT11 ⁶⁴	0	1	0	0	0	0	2	0	0
QBT12 ^{67,68}	0	1	0	3	0	2	0	0	0
QBT13 ¹²⁷	0	1	1	0	0	0	3	0	0
QBT14 ⁷⁰	1	0	0	0	0	0	0	2	0
QBT15 ^{75,97}	0	1	2	0	2	0	3	0	0
QBT16 ⁸⁸	3	1	0	0	0	0	0	0	0
QBT17 ⁸¹	3	1	0	3	0	0	3	0	0
QBT18 ⁷¹	0	1	1	0	3	0	0	0	0
QBT19 ⁶⁵	3	0	0	0	0	2	0	2	0
CTAI ¹⁰⁵	3	2	0	3	2	2	0	0	0

^aFor certain tools multiple studies were found and scores in this table were based on these multiple studies, with duplicate evidence only counted once.

Abbreviations: *Instruments* BDF = Bahrain Defense Force, SPRAT = The Sheffield Peer Rating Assessment Tool, CFEP360 = n/a, GMC CQ = General Medical Council patient and colleague questionnaires, ABIM PAR = American Board of Internal Medicine Peer Assessment Review, Adapted ABIM PAR = Adapted American Board of Internal Medicine Peer Assessment Review, CPSA-PAR = College of Physicians and

Surgeons of Albert Physician Achievement Review, CRP-PIM = Communication with Referring Physicians Practice Improvement Module, ACP 360 = n/a, GMC GQ = General Medical Council generic questionnaire, IFMS = Individueel Functioneren Medisch Specialisten, ABMS/ACGME Faculty Peer Ratings = American Board of Medical Specialties/Accreditation Council for Graduate Medical Education Faculty Peer Ratings, MOCA = Maintenance of Certification in Anesthesiology Program patient and peer surveys, CTAF = Clinical Teaching Assessment Form, CTE = Clinical Teaching Effectiveness, CTEI & (CC) CTEI = (Cleveland Clinic's) Clinical Teaching Effectiveness Instrument, EDTS = Emergency Department Teaching Survey, EFFECT = Evaluation and Feedback For Effective Clinical Teaching, ER scale = Emergency Rotation Scale, HTPE = Humanistic Teaching Practices Effectiveness Questionnaire = LDMES = London Deanery 'MSF for Educational Supervisors', MEDUC-RX32 = Medicina Universidad Católica—Radiology 32 items, MTCQ = Maastricht Clinical Teaching Questionnaire, MTE = Mayo Teaching Effectiveness, PACT = Professionalism Assessment of Clinical Teachers, RMS = Residency Management Suite, SETOC = Student Evaluation of Teaching in Outpatient Clinics, SETQ & SETQ (Smart) = System for Evaluation of Teaching Qualities, SFDP = Stanford Faculty Development Program, TAQ = Tutor Assessment Questionnaire, TES = Teaching Effectiveness Scores