

In-training assessment using direct observation of single-patient encounters: a literature review

E.A.M. Pelgrim, A.W.M. Kramer, H. Mokkink. L. van den Elsen, R.P.T.M. Grol, C.P.M. van der Vleuten

Appendices and references

Appendix 1 Characteristics.....	2
Appendix 2 Characteristics.....	9
Appendix 3 Feasibility.....	15
Appendix 4 Reliability	18
Appendix 5 Validity	21
Appendix 6 Effect on trainee's learning.....	24
References	26

Appendix 1 Characteristics

	setting	year of (first) publication	formative/summative	trainee (T) and rater (R)	encounters	competences	description of behaviours?	scale	criteria for allocation of marks	frame of reference
mini-CEX	(internal) medicine residency, Philadelphia, USA. Also used in USA, Canada, UK, Argentina international medical graduates of Australia	1995	both	T: resident R: attending, senior resident	inpatient, outpatient, ED, other (including admission, discharge)	1. medical interviewing 2. physical examination 3. humanistic qualities / professionalism 4. clinical judgement 5. counselling 6. organization / efficiency 7. overall clinical competence	yes (total of 26, specified per competency) (www.abim.org/pdf/paper-tools/minicex.pdf)	9-point scale: 1-3 = unsatisfactory 4-6 = marginal/satisfactory 7-9 = superior plus: "not observed"	specified by ABIM for resident assessment unsatisfactory = consistently falls short of reasonable expectations marginal = in general meets some expectations but occasionally falls short satisfactory = always meets and occasionally exceeds reasonable expectations superior = far exceeds reasonable expectations	only 4 on 9-point scale is specified as: marginal; conveys expectation that with remediation resident will meet certification standards otherwise not specified
mini-CEX in clerkship	(internal) medicine clerkship, Canada, USA	2002	both	T: student R: attending, resident	inpatient and outpatient	1. medical interviewing 2. physical examination 3. humanistic qualities / professionalism 4. clinical judgement 5. counselling 6. organization / efficiency 7. overall clinical competence	no (not mentioned)	9-point scale: 1-3 = unsatisfactory 4-6 = marginal/satisfactory 7-9 = superior plus: "not observed"	not specified	not specified
OCEX	ophthalmology residency, USA	2004	formative	T: resident R: attending	new patients (n= 3 - 4 per year)	assessment of: - patient care - professionalism - interpersonal skills - medical knowledge categories on form:	description of 33 subitems, some are behaviours and some regard	4-point scale: 1 = does not meet expectations 2 = meets some expectations	allocation of score 1 to 4 is specified for each subitem in a scoring rubric (<i>Golnik et al., 2004, table 1</i>)	not specified

						1.interview skills 2.physical exam 3.interpersonal skills / professionalism 4.case presentation	specification of the category-subject	3 = meets all expectations 4 = exceeds expectations plus: "not applicable"		
palliative care CEX	internal medicine residency, university of Pittsburgh, USA	2005	formative	T: (1 st year) resident R: one of the clinicians of "palliative care service" (physicians, nurse, medical ethicist and social worker)	giving bad news to or discussing wishes regarding "code status" with seriously ill patient (n = 1)	communication skills in giving bad news and discussing code status	yes (total of 18, see: <i>Han et al., 2005 table 3</i>)	not specified	not specified	not specified
P-MEX	clerkship in: internal medicine, general surgery, paediatrics, psychiatry and obstetrics / gynaecology, McGill university, Montreal, Canada	2006	both	T: 3 rd -4 th year student R: faculty and senior resident (multiple raters per student, number not specified)	situations where student's behaviour can be observed (e.g. patient encounters, small group sessions, sign-out rounds)	professional behaviours: 1. doctor-patient relationship skills 2. reflective skills 3. time management 4. interprofessional relationship skills	yes (total of 24, specified per competency, see: <i>Cruess et al., 2006, table 1</i>).	4-point scale: 1 = unacceptable 2 = below expectations 3 = met expectations 4 = exceeded expectations plus: "not observed or not applicable"	not specified	not specified
CBA rheumatology	rheumatology residency, West Midlands, UK	2006	formative	T: resident R: consultant in rheumatology (n = 1-2 per trainee)	new and follow-up outpatients with variety of rheumatology problems (n = 4 - 7 in a 3-hour clinic)	1. history taking 2. examination 3. diagnostic skills 4. management plan 5. communicative skills 6. letter dictation 7. overall impression	no	3-point scale: 1 = does not reach standard 2 = borderline 3 = good	allocation of score 1, 2 and 3 are specified (<i>Dowson & Hassell, 2006, table 3</i>).	not specified
SCO	paediatric clerkship, Jefferson Medical College, Philadelphia, USA	2000	formative	T: 3 rd year student R: general paediatrician (multiple raters, number not specified)	part of paediatric patient encounters (max. 3 min of observation, not specified whether in-patient or out-patient)	- history taking or - physical examination or - information giving	yes (total of 47, specified per competence) - history taking: total of 22 items - physical exam: total of 8 items - information	not specified	not specified	not specified

							giving: total of 21 items (see: Lane & Gottlieb, 2000, table 1,2,3)			
PEAF	surgical residency (general surgery and trauma rotations), Michigan State university, USA	2005	formative	T: PGY I and II resident R: faculty	new outpatient (scheduled during teaching patient hour)	<ol style="list-style-type: none"> 1. punctuality / introduction / describes role 2. quality / appropriateness of history taking skills 3. open ended questions / listening / interviewing skills 4. quality of physical exam / attention to patient anatomy 5. knowledge of test results / images 6. ability to formulate & discuss assessment / classic presentation / differential diagnosis 7. knowledge of evidence-based data 8. decision making skills / knowledge of steps, orders / plan 9. ability to educate others / communicate with team 10. quality / timeliness of data entry / dictation <p>plus: "overall score for level of training"</p>	no	<p>for each item a score on 11-point scale: 0-100%, increments of 10%</p> <p>and 8-letter scale for "overall score for level of training": A+ = best resident A = superb B+ = exceeds expectations B = solid C+ = below expectations C = marginal D = poor performance F = failure</p>	not specified	<p>for the 11-point scale: novice, beginner, advanced, expert, mastery</p> <p>PGYI is expected to perform at novice or beginner level (0-30%); PGYII at advanced level; senior resident at 80-100%</p> <p>8-letter scale: score compares resident to other residents of same level</p>
GRFA	anaesthesiology residency, Denmark	2003	summative <i>(NB, used in research setting)</i>	T: resident R: consultant in anaesthesiology	<ul style="list-style-type: none"> - insertion of epidural catheter - pre-op consult - emergency induction of anaesthesia - round on patient in ICU <p><i>(NB, research purposes, not described for</i></p>	<ol style="list-style-type: none"> 1. patient communication 2. clinical and technical skills 3. knowledge 	no	<p>5-point scale:</p> <ol style="list-style-type: none"> 1 = clear fail 2 = borderline fail 3 = borderline pass 4 = clear pass 5 = excellent <p>plus: an overall dichotomous scoring of pass/fail</p>	not specified	mastery (all elements of performance must be correct)

					<i>which purposes it was intended in practice)</i>					
WRF	internal medicine clerkship, university of Ottawa, Canada	2000	potentially formative	T: 3 rd year student R: attending	inpatient discharge	1. therapeutic strategies 2. communication skills 3. consultation skills 4. management skills 5. interpersonal behaviours 6. continued learning skills 7. health advocacy skills plus: a global rating of overall performance	no	5-point scale: 1 = unsatisfactory 2 = ? 3 = ? 4 = ? 5 = excellent	not specified	not specified
CPB	family practice residency, USA	2002	not specified	T: resident R: attending	outpatients in family practice	1. history taking and physical examination 2. interpersonal skills 3. clinical problem solving	no	likert scale, not further specified	not specified	not specified
CEE	emergency medicine residency, Atlanta, USA	2002	formative	T: resident R: attending	"typical" patient encounter in ED	categories on form: - emergency stabilization - data-gathering-history - data-gathering-physical examination - case presentation - diagnosis and plan for diagnostic studies and medical care according to authors the following competences are assessed: - patient care - practice-based learning - interpersonal and communication skills - professionalism	yes, total of 27, specified per category. Items are scored, not category as a whole	3-point scale: 1 = below expected 2 = meets expected 3 = outstanding	1= falls short of reasonable expectations 2 = always meets and occasionally exceeds expectations 3 = far exceeds expectations for level of training	level of training

						- system-based practices				
CSAF	psychiatric clerkship, McMaster university, Hamilton, Canada	1984	formative	T: student R: clinical supervisor	new psychiatric patients (inpatient or outpatient not specified)	- interviewing - history taking - mental status examination - doctor-patient-relationship - problem formulation - treatment plan	yes, total of 17, specified per competence (see: <i>Links et al, 1984, table 1</i>)	7-point scale not further specified	1= non or few features were demonstrated 2-6 = not specified 7 = criterion performance	criterion or ideal performance
SDOT	emergency medicine residency, USA	2006	not specified	T: resident R: attending	patient encounters in emergency department	1st assessment on form: 4 categories: - data gathering - synthesis / diagnosis - management - disposition 2nd assessment on form 5 competencies: - patient care - medical knowledge - interpersonal and communication skills - professionalism - system-based practice plus: overall clinical competence	yes, total of 26, specified per competence. These 26 items are separately scored	1st assessment and overall clinical competence: 3-point scale: 1 = needs improvement 2 = meets expectations 3 = above expected plus: not assessed 2nd assessment: 5-point scale: score 1-5 in combination with - needs improvement, - meets expectations, - above expectations" (NB, it is not completely clear which score combines with which description)	extensive (7-page) description of behavioural anchors for each item (www.cordtests.org)	level of training
ECS	general practice residency, West Australia	1997	formative	T: resident R: experienced GP	new and follow-up patients in general practice	communication skills: 14 behaviours are to be scored (<i>Nyman & Sheridan, 1997, figure 1</i>)	no further specification	5-point scale: 1 = strongly disagree 2 = disagree 3 = unsure	not specified	not specified

								4 = agree 5 = strongly agree plus: "not relevant"		
LEP	dentists, dental vocational training, Scotland	2002	formative (can contribute towards a summative decision)	T: postgraduate dental trainee R: clinicians, advisers, nurses, other members of the clinical team	almost any clinical situation across dentistry	1. examination & consultations skills 2. clinical judgment & diagnosis 3. technical ability & manual dexterity 4. communication skills 5. professionalism 6. knowledge 7. organisation	no further specification	9-point scale: 1 - 3 = need improvement 4 - 6 = satisfactory 7 - 9 = superior	not specified	completion of training
MiniCard	internal medicine, 4 different residency programmes in the USA	2008	probably both	T: resident R: internal medicine faculty member	problem commonly seen in internal medicine	four sections: 1. history 2. physical 3. presentation of plan 4. counselling each section consists of three domains: a. interpersonal communication b. medical knowledge c. professional domains	3 to 11 prompts given for each domain	4-point scale: 1 = excellent 2 = good 3 = marginal 4 = poor	each rating category has adjectival and behavioural anchors specific to the domain	level of training for medical knowledge, mastery level for interpersonal communication and professionalism
CEC	A: surgery clerkship, USA <hr/> B: internal medicine clerkship, Canada	1999	A + B: formative	A + B: T: 3th year medical student R: faculty or PGY resident	A + B: inpatient and outpatient, new and follow-up	A + B: 1. history taking 2. physical examination 3. professional behaviour 4. technical skills 5. case presentation 6. problem formulation (diagnosis) 7. problem formulation (therapy) 8. other	A + B: no further specification	A: 6-point scale 1 = unsatisfactory 2 = below level of average 3 = at level of average 4 = above level of average 5 = outstanding above level of average 6 = at level of intern <hr/> B: 5-point scale: 1 = unsatisfactory 2 = below level of average 3 = at level of	A + B: not specified	A + B: norm-referenced (not further specified)

								average 4 = above level of average 5 = at level of intern		
BFA	medicine clerkship at the university of Cape Town, South Africa	2006	formative	T: 4 th year medical student R: clinician educators, all specialist physicians with 5 years teaching experience	'blinded' patient encounters: student examines inpatient without access to their clinical records	1.basic clinical skills a. information b. report 2.diagnostic reasoning skills 3. knowledge a. investigations b. plan	descriptions given on the form	9 - point scale: 1 - 3 = poor 4 - 6 = adequate 7 - 9 = good	not specified	not specified

Appendix 2 Characteristics

	type of feedback	additional information on feedback	way of training raters	instruction of trainees	form displayed in article?	instrument retrievable on internet?
mini-CEX	<p>quantitative: yes</p> <p>qualitative: - verbal: yes - written: yes (space on form for comments)</p>	<p>- 80% of sessions lead to recommendations for improvement</p> <p>- in 61% learner interaction was promoted</p> <p>- in 34% self-assessment occurred</p> <p>- in 10% action plans were developed (in 3 out of 11 sessions this was generated by the resident).</p>	<p>raters were instructed on goals of mini-CEX (including importance of feedback to promote improvement of clinical skills) and an instructional ABIM-handout was provided.</p> <p>-in the study of Cook et al., 2008 intervention of an intensive workshop was studied. The workshop included rater error training, performance dimension training, behavioural observation training, frame of reference training using lecture, video and facilitated discussion. Interrater reliability was higher after the workshop than before.</p> <p>- <i>In other studies on mini-CEX: no training or not specified.</i></p> <p>- <i>A written instruction for raters can be found on the ABIM homepage (www.abim.org/pdf/paper-tools/attending.pdf)</i></p>	<p>- instruction for trainees is not specified in articles except for the research on the workshop in the article of Cook et al., 2008.</p> <p>- <i>a written instruction for trainees can be found on the ABIM homepage (www.abim.org/pdf/paper-tools/resident.pdf)</i></p>	yes	<p>yes: guideline, form, written instruction for raters and trainees.</p> <p>http://www.abim.org/pdf/paper-tools/minicex.pdf</p> <p>http://www.abim.org/pdf/paper-tools/attending.pdf</p> <p>http://www.abim.org/pdf/paper-tools/resident.pdf</p>
mini-CEX in clerkship PDA-based	<p>quantitative: yes</p> <p>qualitative: - verbal: not specified (probably after observation) - written: not specified</p> <p>quantitative: yes</p> <p>qualitative: - verbal: not specified - written: yes</p>	- feedback was provided in 96%	<p>raters were oriented to rationale and use of mini-CEX during a meeting. Outpatient faculty were also mailed an orientation package. There was no formal training to use the form.</p> <p>no formal training. Before start of rotation, raters received orientation package and e-mail in which purpose of mini-CEX was explained, key features of the PDA-based form were discussed and instructions were</p>	<p>instruction was limited to number of forms to be handed in.</p> <p>trainees were oriented to form and procedure by clerkship director during clerkship orientation session. Purpose of form and rationale for each item</p>	<p>no (<i>mini-CEX form is used with some adjustments</i>)</p> <p>yes</p>	no

			provided on how to structure the exercise and complete the form. There was no detailed guidance on how to deliver feedback.	was reviewed in detail. Trainees were instructed to assist raters in the use of the PDA-form if necessary.		
OCEX	quantitative: yes qualitative: - verbal: probably, but not specified in the article - written: yes (space on form for comments)	resident is to receive immediate feedback.	raters received a form, scoring rubric and a 3 minute explanatory introduction video-CD. (NB: research purposes))	not specified (<i>studies did not regard resident use of OCEX</i>)	yes	http://www.acgme.org/acWebsite/RR_C_240/240_OCEX.pdf
palliative care CEX	quantitative: probably, however not specified qualitative: - verbal: yes - written: not specified	feedback included a review of both residents' and observers' impressions of the strengths and weaknesses of the discussion with the patient.	raters were individually trained (by authors) on all aspects of the study protocol, procedures for conducting the palliative care CEX and proper scoring of the instrument to encourage consistency and reliability.	raters provided a review of the circumstances, goals and strategies for the discussion with the patient.	no	no
P-MEX	quantitative: yes qualitative: - verbal: not specified - written: yes: space on form for comments	rater is expected to give timely feedback to student.	instructions on using the form were given to all raters. not further specified	not specified	no	no <i>only a presentation on NVMO-website with example of P-MEX</i> http://www.nvmo.nl/portals/0/congres2006/Steinert.pdf
CBA rheumatology	quantitative: yes qualitative: - verbal: not specifically mentioned, however "constructive feedback was given during and/or at the end of the assessment" - written: yes (space on form for comments on each item); 1 form for each encounter in the session and 1 summary form at the end of the session	-	raters were given information regarding aims, format and scoring criteria for the assessment. Scoring criteria were shown at least one month before assessment. not further specified, no training sessions	trainees were given information regarding aims, format and scoring criteria for the assessment. Scoring criteria were shown at least one month before assessment. not further specified, no training sessions	yes	no

SCO	<p>quantitative: no</p> <p>qualitative: - verbal: yes - written: ?</p>	<p>observation was followed by immediate feedback, outside patient room, lasting maximally 2 minutes. The short observation would limit the number of feedback points, so that feedback did not become overwhelming for the students and would not take more than a few minutes for the rater to complete. Since there were repeated observations, students would be given a chance to incorporate what they had been taught, feel that they were gaining mastery and improving their clinical skills.</p> <p>raters were instructed to write down exactly what trainees said and did, to give focussed and effective feedback.</p> <p>if 1 or 2 major feedback points are identified in the first few seconds of observation, the SCO is ended.</p>	<p>two-hour workshop, consisting of:</p> <ul style="list-style-type: none"> - discussion of rationale for SCO program - overview of the SCO methodology - overview of basic tenets of effective feedback - discussion of procedures - practice in observing and using SCO forms by looking at 2-3-minute video vignettes of students interacting with patients - role play to practise giving feedback 	<p>students were oriented to SCO program at start of 6-week clerkship and again just before starting the SCO program (a 2-week block). The skill guidelines were reviewed and discussed, procedures were discussed and it was emphasized that it was a teaching experience and not an evaluation.</p>	no	no
PEAF	<p>quantitative: yes</p> <p>qualitative: - verbal: ? : "comments were invited on each case reviewed" - written: yes, space on form for comments / improvements</p>	-	<p>raters were trained on use of the evaluation form.</p> <p>not further specified</p>	not specified	yes	no
GRFA	<p>quantitative: yes (in wording)</p> <p>qualitative: - verbal: no - written: ? : participants were invited to give written comments, not specified if this included feedback, or only comments about the instrument</p>	<p>participants were invited to give written comments, not specified if this included feedback, or only comments about the instrument. The instrument was used in a study situation with video-taped encounters.</p>	<p>the instrument was used in a study situation. Raters were instructed to:</p> <ul style="list-style-type: none"> - use their professional expertise and judgement in evaluation of resident's performance - make an overall decision regarding pass or fail according to assessment protocol (criterion for passing was mastery) 	<p>not applicable because of the study situation where a standardized resident was rated.</p>	yes	no

WRF	quantitative: yes qualitative: - verbal: yes - written: yes	formative feedback on each performance was accommodated by a comment section on the form.	two-hour workshop, followed by monthly communications reminding raters of the basis of the project and the need to complete evaluations.	students were oriented at beginning of rotation and subsequently met with research associate on weekly basis to discuss issues or questions related to the project.	no	no
CPB	quantitative: yes qualitative: - verbal: ? - written: yes	specificity of feedback improved significantly after intervention as described in the next cell, (pre-intervention score mean of 5.25/9, post-intervention mean score 7.44/9; $p < .01$) timeliness and immediacy of feedback also improved dramatically (4.43/9 pre-intervention, 7.81/9 post-intervention, $p < .01$)	2-hour course in which: - basic outline of evaluation was reviewed - criteria for the standardized assessment and completion of the CPB were discussed. the format of the course contained the analysis of 5-to-10-minute videotaped vignettes of outstanding, satisfactory and unsatisfactory resident performances.	not mentioned	no	no
CEE	quantitative: yes (in wording) qualitative: - verbal: yes - written: yes (space on form for comments on each category)	rater is to provide immediate feedback on encounter in oral and written form.	no / not mentioned	no / not mentioned	yes	no
CSAF	quantitative: yes qualitative: - verbal: yes - written: ?	following the encounter, there was a 30-min feedback session, during which the student presented a problem-formulation and management plan.	no/ not mentioned	no / not mentioned	no ("further details about the CSAF can be obtained from the first author")	no

SDOT	quantitative: yes qualitative: - verbal: ? - written: yes (space on form for summary comments and comments for each competence)	-	for reliability study: raters were briefed on the purpose of SDOT	not specified for practice (not applicable for reliability study, because standardized residents were used)	no, although authors refer to website where SDOT is available	yes: www.cordtetsts.org
ECS	quantitative: yes qualitative: - verbal: ? - written: no	-	detailed set of written guidelines was provided for raters	not mentioned	yes	no
LEP	quantitative: yes qualitative: - verbal: ? - written: yes	-	all participants were trained (not further specified)	all participants were trained (not further specified)	yes	trainers guide http://www.nes.scot.nhs.uk/dentistry/dvt_dft_htvt/assessment/trainers/document/LEPAtrainersguide1.pdf
MiniCard	quantitative: yes qualitative: - verbal: yes - written: yes	qualitative feedback is written in the research design, but given verbal in real-world situations.	1-hour training session with videotaped description of the tool, followed by a practice session.	not applicable (study situation where a standardized resident was rated)	yes	no
CEC	quantitative: yes qualitative: - verbal: yes - written: yes	-	the encounter cards were presented at a number of forums, including departmental meetings, academic half-day for internal medicine residents and an educational faculty retreat (Hatla & Norman, 1999) evaluators were instructed to rate performance on a global scale and give verbal and written feedback. A letter explained the system (Paukert et al., 2002).	students were informed and CEC's were explained (not further specified)	yes	no
BFA	quantitative: yes qualitative: - verbal: probably - written: yes	-	clinician educators who volunteered to participate in the study attended 2 workshops in which the principles and purpose of formative assessment were discussed and the structured feedback form was designed.	not mentioned	yes	no

Abbreviations used in Appendix 1 and 2

ABIM = American Board of Internal Medicine

ED = emergency department

PGY = postgraduate year

pre-op = preoperative

ICU = intensive care unit

Appendix 3 Feasibility

	Feasibility determined	How was this performed	Outcomes
mini-CEX	yes	<p>completion rate (Alves de Lima et al., 2007; Durning et al., 2002; Wilkinson et al., 2008)</p> <p>survey (Nair et al., 2008)</p> <p>focus group (Malhorta, 2008)</p>	<p>low completion rate highlights feasibility problems, lack of time was the main factor preventing completion (Wilkinson et al., 2008)</p> <p>completion rate of 96.43% (Durning et al., 2002)</p> <p>15% of residents were evaluated four or more times during the study period. Feasibility was defined by a minimum of 50% of the residents obtained at least four mini-CEXs. Feasibility in this study was poor (Alves de Lima et al., 2007)</p> <p>never or only occasionally experienced difficulties. Most examiners (14/18) were satisfied or very satisfied (Nair et al., 2008).</p> <p>the assessment first reflects anxiety. Repeated mini-CEX experiences made the resident learner more confident in controlling and enforcing the positive educational effect (Malhorta, 2008)</p>
mini-CEX in clerkship	yes	<p>completion rate (Kogan et al., 2002 & Kogan et al., 2003)</p> <p>satisfaction ratings (Kogan et al., 2002)</p>	<p>completion rate: mean of 7.3 mini-CEX, 9 were asked (Kogan et al., 2002)</p> <p>completion rate: 89% of target (Kogan et al., 2003)</p> <p>satisfaction rate: faculty/residents: 7.2 students: 6.8 (Kogan et al., 2002)</p>
PDA-based mini-CEX	yes	<p>completion rate (Torre et al., 2007)</p> <p>satisfaction rating (Torre et al., 2007)</p>	<p>completion rate: 100% = 354 forms -feedback provided in 96%</p> <p>satisfaction rate: student satisfaction 8.0 resident satisfaction 8.1 faculty satisfaction 7.4</p> <p>researchers conclude that the PDA-based mini-CEX is feasible</p>
OCEX	no	-	-
palliative care CEX	yes	resident questionnaires (telephone survey and written questionnaire) (Han et al., 2005)	most residents rated the Palliative Care CEX very highly (>4 on five-point scale) among several dimensions: -educational value, -overall quality of the experience, -preceptor's effectiveness in creating a positive learning experience, -improvement in comfort with

			discussions, -importance of formal instruction in end-of-life communication, -value of preceptor feedback). the study provides support for the feasibility in this setting.
P-MEX	yes	demographic and contextual data were computed (Crues et al., 2006)	-211 forms, 74 students, 47 evaluators -context of evaluation varied greatly researchers conclude that this study suggests a feasible format for evaluating professionalism in clerkship training
CBA rheumatology	yes	consultant and resident questionnaires (Dowson & Hassell, 2006)	12/13 of consultants and 12/12 residents were positive about the process researchers conclude that this is a worthwhile exercise for evaluating important clinical and consultations skills
SCO	yes	student and faculty questionnaires (Lane & Gottlieb, 2000) observation rate (Lane & Gottlieb, 2000)	the SCO program was highly regarded as a clinical skills teaching tool by both students and faculty it maintains the number of observations. researchers conclude that the program was feasible.
PEAF	no	-	-
GRFA	yes	participants indicate the appropriateness of the scoring form they had used on a 5-point scale and with written comments (Ringsted et al., 2003)	clinicians found checklists with task-specific items more appropriate than rating forms with general dimensions of competence
WRF	yes	completion rate (Turnbull et al., 2000)	mean: 8.1 WRFs submitted per student 23% response rate while response rates were low, the minimum number needed to achieve adequate level of reliability is exceeded. Thereby meeting requirement for feasibility.
CPB	yes	questionnaires - satisfaction levels of residents and receptors (Ross, 2002)	higher rate of documented direct observation and feedback than previously reported the specificity of feedback improved significantly (5.25/9 – 7.44/9; p<.01) timeliness and immediacy of feedback improved dramatically (4.43/9 – 7.81/9; p< .01) residents and faculty were highly supportive of continuing CPB as part of the evaluation during continuity clinic.
CEE	no	-	-
CSAF	yes	compliance of students and supervisors (Links et al., 1984)	students and staff complied with the exercise. particularly when the students were given more responsibility for completion of the exercise.
SDOT	no	-	-
ECS	no	-	-

LEP	yes	average time spent on each assessment (Prescott-Clements et al., 2008) rater questionnaires (Presscott-Clements et al., 2008)	average time spent is 22 minutes for observation and 8 minutes for giving feedback 65% founds implementation easy (only 11% disagreed on this) 63% thinks the assessment had sufficient flexibility
MiniCard	no	-	-
CEC	yes	return rate (Hatala & Norman, 1999)	64% participated. mean of 7.9 encounters with CEC during rotation
BFA	yes	staff questionnaires (Burch et al 2006)	most participating clinical educators agreed that the assessment could be satisfactorily integrated into bedside teaching sessions

Appendix 4 Reliability

	Reliability determined?	Type(s) of reliability	How was this performed	Outcomes
mini-CEX	yes	<p>(1)generalizability (Alves de Lima et al., 2007 ; Margolis et al., 2006; Nair et al, 2008; Wilkinson et al., 2008)</p> <p>(2)reproducibility (Norcini et al., 1995)</p> <p>(3)internal consistency (Durning et al., 2002 ; Hatala et al., 2006)</p> <p>(4)inter-encouter reliability (Hatala et al., 2006)</p> <p>(5)interrater reliability (Cook & Beckman, 2008 ; Cook et al., 2008)</p>	<p>(1)generalizability analysis (Margolis et al., 2006; Nair et al, 2008; Wilkinson et al., 2008) or SEM benchmark and D-coefficient (Alves de Lima et al., 2007)</p> <p>(2)reproducibiliy coefficients (ratios of universe score variance to total variance)</p> <p>(3)cronbach's α</p> <p>(4)cronbach's α</p> <p>(5)ICC</p>	<p>(1)eight different assessors observing at least two encounters (Wilkinson et al., 2008).</p> <p>one rater 10 cases, reproducibility .39. ten raters 1 case, reproducibility .83 (Margolis et al., 2006)</p> <p>G coefficient for eight encounters was 0.88, suggesting reliability of .90 for 10 encounters (Nair et al, 2008)</p> <p>using the SEM benchmark of .26, a minimum of 10 evaluations was necessary to produce a minimal reliable inference. This corresponds to a D-coefficient of .44 (Alves de Lima et al., 2007)</p> <p>(2)between 12 and 14 encounters are required to reach a reproducibility of .80 (Norcini et al., 1995).</p> <p>(3)$\alpha = .90$ for 162 forms (Durning et al., 2002) & $\alpha = .79$ (Hatala et al., 2006)</p> <p>(4)inter-encouter reliability $\alpha = .74$ (Hatala et al., 2006).</p> <p>(5)interrater reliability on five-point scale (research format) was .40, on nine-point scale (original m-CEX) .43. the interrater reliability of mini-CEX was similar for 5-point and 9-point rating scale (Cook & Beckman, 2008)</p> <p>interrater reliability was higher after intensive workshop intervention than before the intervention, (ICC .34 vs. ICC .18) but standard error of measurement was similar for both periods (Cook et al., 2008).</p>
mini-CEX	yes	generalizability (Kogan et al., 2003)	generalizability analysis	reproducibility for 4 raters is.62. for 6 raters, .71. For 8

in clerkship				raters .77.
OCEX	yes	(1)internal consistency (Golnik & Goldenhar, 2005) (2)interrater reliability (Golnik & Goldenhar, 2005)	(1)cronbach's α (2)analyzing the distribution of ratings across each category for each item. with an acceptable distribution of having at least 85% of ratings occur in 2 consecutive categories	(1)coefficient α for OCEX as a whole = .81 interview skills = .65 examination = .27 interpersonal skills/ professionalism = .73 ease presentation = .70 (2)94% of the items has at least 85% of the ratings occur in 2 consecutive categories
palliative care CEX	no	-	-	-
P-MEX	yes	generalizability (Cruess et al., 2006)	generalizability analysis (GENOVA)	between 10 and 12 encounters are required to obtain a reproducibility coefficient of 0.80.
CBA rheumatology	no	-	-	-
SCO	no	-	-	-
PEAF	no	-	-	-
GRFA	yes	interrater reliability (Ringsted et al., 2003)	agreement of pass/fail decisions	poor agreement among clinicians regarding pass/fail decisions reliability could be enhanced by having several different assessors and several observations. and familiarity with the concept, to train the clinicians as assessors.
WRF	yes	within method reliability/ generalizability (to determine the number of assessments necessary) (Turnbull et al., 2000)	spearman brown prophecy formula	spearman brown = 3.2 forms necessary to achieve within method reliability (reliability of .70).
CPB	no	-	-	-
CEE	no	-	-	-
CSAF	yes	(1)interrater reliability (Links et al., 1984) (2)test- retest reliability (Links et al., 1984)	(1)review of a single videotape of a student performance by 12 clinical supervisors. (satisfactory agreement: agreement plus or minus one on the 7-point scale by greater than 60% of observers). number of actual agreement over the number of possible agreements.	(1)percentage of agreement varies between 52% and 89%. overall percentage was 71%. in 13 of 17 cases agreement met arbitrary standard of 60% (2)significant correlation in 16 of 17 categories (sig. < 0.05) (N=38).

			(2)correlation coefficient between the pre-exercise and post-exercise (5-week interval)	
SDOT	yes	(1)interrater reliability (Shayne et al., 2006) (2)internal consistency (Shayne et al., 2006)	(1)video presentations of 2 simulated resident-patient encounters. video 1: average resident performance, video 2: several significant weaknesses. they (N=82) watched alone and could watch twice before completing the form. interrater reliability on each item and of each combined competency score. (2)cronbach's α is calculated and intraclass correlation coefficients.	(1) four of the 26 items were less reliable. overall reliability may be improved by modifying these 4 portions. multivariate analysis found no differences in rating by faculty when examined by experience, academic title, site, of previous use of the SDOT. (2) α = .93. if four items were removed α = .95 ICC = .74. if four items were removed α = .81 SDOT appears to be reliable according to the researchers.
ECS	no	-	-	-
LEP	no	-	-	-
MiniCard	yes	interrater reliability (Donato et al., 2008)	fleiss' kappa	K = .520 – moderate agreement
CEC	yes	(1)generalizability (Richards et al., 2007) (2)inter-encounter reliability (Hatala & Norman, 1999)	(1)ICC (2)intraclass correlations(ICC)	(1) ICC .58 on 20 encounters, 7 evaluators. ICC .69 on 18 encounters, 12 evaluators (2)inter-encounter reliability of .32 for single encounter. reliability of .79 overall for 8 encounters per student
BFA	no	-	-	-

Appendix 5 Validity

	Validity determined?	Type(s) of validity	How was this performed	Outcomes
mini-CEX	yes	<p>(1) criterion validity (Hatala et al., 2006)</p> <p>(2) criterion validity (Durning et al., 2002)</p> <p>(3) construct validity (Alves de Lima et al., 2007; Holmboe et al., 2003)</p> <p>(4) test of halo (Cook et al., 2008)</p> <p>(5) face validity (Wilkinson et al., 2008)</p> <p>(6) accuracy (Cook & Beckman, 2008)</p>	<p>(1) pearson product-moment correlation between mini-CEX and RCSPC-IM</p> <p>(2) comparing mCEX mean scores with widely used methods for evaluation of residents' clinical competence. analysis of variance (ANOVA)</p> <p>(3) variance analysis with ANOVA (video analysis on three clinical skills on three different performance levels)</p> <p>ability of the mini-CEX to discriminate between pre-existing levels of expertise. significant mean differences between different years of training</p> <p>(4) test of halo</p> <p>(5) questionnaires</p> <p>(6) scores compared to scripted competence level</p>	<p>(1) correlations: -overall competence (mCEX) vs. structured oral (RCSPC-IM) = .73 -overall competence (mCEX) vs. bedside station sub score (RCSPC-IM) = .67 -overall competence (mCEX) vs. written examination (RCSPC-IM) = .72</p> <p>(2) strong correlations between mCEX scores and corresponding ABIM MEF scores and ITE scores.</p> <p>(3) for 3 clinical skills faculty participants were able to discriminate among performance levels. differences were statistically significant, however, the range in ratings among the participants for each videotape was wide (Holmboe et al., 2003). the mini-CEX discriminates between pre-existing levels of global competency between residents: first year residents 7.19, second 7.51, third 7.76, fourth 8.16. these differences are statistically significant (P=.0008) (Alves de Lima et al., 2007).</p> <p>(4) values did not vary significantly between groups or testing periods, nor was the interaction significant.</p> <p>(5) participants felt that the assessment methods were fair. with each method, more senior trainees received significant higher scores</p> <p>(6) nine-point scales appear to provide more accurate scores</p>
Mini-CEX in clerkship	yes	concurrent validity / criterion validity (Kogan et al., 2003)	relationships with other assessments and an increase in scores with increasing clinical experience	significant correlations between mean mCEX scores and exam scores/ write-ups/ inpatient, outpatient and final course grades. students who received 'honors' on their exam had higher mean mCEX scores than those who received 'pass'. mean mCEX ratings increased over the course of the year.
OCEX	yes	<p>(1) construct validity (Golnik & Goldenhar, 2005)</p> <p>(2) content validity (Golnik et al., 2004)</p>	<p>(1) not mentioned</p> <p>(2 & 3) modifying the OCEX in ways suggested by a panel of experts</p>	<p>(1) they concluded good validity</p> <p>in the subscales 'examination' and 'interview skills' they perhaps should not necessarily hang together. it might not be indicating a poorly functioning scale per se. the OCEX shows validity.</p>

		(3)face validity (Golnik et al., 2004)		(2 & 3)OCEX has face and content validity. it can be used to assess a resident's patient care skills, medical knowledge, and interpersonal skills
palliative care CEX	no	-	-	-
P-MEX	yes	(1)content validity (Cruess et al., 2006) (2)construct validity (Cruess et al., 2006)	(1)process of item generation, and processes compared to other groups in North America (2)factor analysis	(1)extensive process of item generation, and the results of this process were 'triangulated' with similar processes conducted by other groups in North America. researchers conclude that this study suggests content validity (2)the 24 original items cluster into identifiable factors or facets of the construct. this is evidence for construct validity.
CBA rheumatology	yes	face validity (Dowson & Hassell, 2006)	not mentioned	researchers conclude that it is clear that it will test the skills used in routine outpatient consultations, i.e. it has face validity researchers mentioned that if looked at in isolation, there are weaknesses in terms of the validity. the assessment described should be viewed as one of a selection of assessments utilized.
SCO	no	-	-	-
PEAF	no	-	-	-
GRFA	no	-	-	-
WRF	yes	(1)content validity (Turnbull et al., 2000) (2)concurrent validity (Turnbull et al., 2000)	(1)process of internal and external expert reviews (2)examining the correlations of scores on the different measures	(1)the form did reflect the necessary domains of practice (is considered to have content validity). the high item-total correlation suggests that evaluators may not be reflecting these competencies. (2)correlation between WRF and Admission Rating Form is moderate (.47). correlation between WRF and Multidisciplinary Team Rating Forms were lower (.37 and .26 (history and physical forms)). modest correlations between different forms provides further supportive evidence of validity
CPB	no	-	-	-
CEE	no	-	-	-
CSAF	yes	construct validity (Links et al., 1984)	improvement in skills from the pre- to post- direct observation exercise was measured. different real patients, (in most cases) same supervisor. paired t-test (sig. < .0025)	9 of 17 significant on .0025 level. 13 of 17 significant on .05 level.
SDOT	no	-	-	-
ECS	no	-	-	-
LEP	yes	construct validity	LEP results of 2 cohorts and rater questionnaires	-appropriate range of assessment

		(Prescott-Clements et al., 2008)		-increase in ratings throughout the year
MiniCard	yes	accuracy (Donato et al., 2008)	comparison of script	the new format is more accurate than the mini-CEX form
CEC	yes	criterion validity (Richards et al., 2007)	pearson correlation coefficient	significant positive correlation on CEC with 'clinical performance ratings' 'NBME' and 'final grades'. no positive correlation on CEC with 'OSCE scores'.
BFA	no	-	-	-

Appendix 6 Effect on trainee's learning

	Effect on trainee's learning determined?	How was this performed	Outcomes
mini-CEX	yes	(1)questionnaires (Cook et al., 2008) (2)focus groups (Malhorta et al., 2008)	(1) although participants' confidence in using the mini-CEX improved, beliefs about the accuracy and usefulness of the mini-CEX changed little (2) repeated mini-CEX experiences made the resident learner more confident in controlling and enforcing the positive educational effect
mini-CEX in clerkship	no	-	-
OCEX	no	-	-
palliative care CEX	yes	questionnaires (telephone survey and written questionnaire) (Han et al., 2005)	after the intervention there is observed significant improvement in self-rated competence and confidence in giving bad news ($p < .05$) researchers concluded that the study provides support for the educational value of the form.
P-MEX	yes	semi-structured interviews (Cruess et al., 2006)	-useful in promoting self-reflection -awareness of the importance of professionalism in daily encounters -identifying behaviours consistent with professionalism -teaching about these subjects matters.
CBA rheumatology	no	-	-
SCO	yes	questionnaires with faculty members (Lane & Gottlieb, 2000)	-value of SCO to student education, 4 out of 5 on a 5-point scale -researchers conclude that they have evidence of students incorporating feedback suggestions into future encounters either by actually observing them or by the student telling that they had done so. a quantitative evaluation of the evaluation of the educational impact of the program needs to be undertaken.
PEAF	no	-	-

GRFA	no	-	-
WRF	no	-	-
CPB	no	-	-
CEE	no	-	-
CSAF	yes	<p>students (N=19) completed a ranking of the components of the clerkship. The components were ranked in terms of their contribution to the development of knowledge, skills and attitude (Links et al., 1984)</p> <p>a comparison was made of the quality of written evaluation forms completed prior to introduction (N = 23) and after its introduction (N=35). rated on a 5-pointscale (Links et al., 1984)</p>	<p>students rated the direct observation exercise as the second most valuable component in terms of assisting skill acquisition.</p> <p>quality rating before introduction 2.8 (range 2 to 4). quality rating after introduction 3.6 (range 2 to 5). (sign. 0.05).</p>
SDOT	yes	opinion of SDOT task force (Shayne et al., 2006)	in the opinion of the SDOT task force, specific observations of resident behaviour and actions provide more substantive feedback for residents to understand and modify their performance.
ECS	no	-	-
LEP	yes	rater questionnaire (Prescott-Clements et al., 2008)	in first cohort 79% of raters founds positive response from trainees on feedback, in second cohort 86%
MiniCard	no	-	-
CEC	yes	end of clerkship evaluations of intervention-group compared with historical control-group (Paukert et al., 2002)	improved satisfaction with feedback and improved understanding of history taking, physical examination and decision making skills
BFA	yes	student questionnaires (Burch et al., 2006)	<p>95,6% recognised the learning value of BFA</p> <p>70% acknowledged the informative, advisory and motivational role of feedback</p> <p>71,9% increased in preparatory reading</p>

References

- [1] Alves de Lima, A., Barrero, C., Baratta, S., Castillo Costa, Y., Bortman, G., Carabajales, J., Conde, D., Galli, A., Degrange, G. & Vleuten, C., van der (2007). "Validity, reliability, feasibility and satisfaction of the mini-clinical evaluation exercise (mini-CEX) for cardiology residency training". *Medical Teacher*, vol. 29, pp. 785 - 790.
- [2] Anderson, C.I., Jentz, A.B., Harkema, J.M., Kareti, L.R., Apeltgren, K.N. & Slomski, C.A. (2005). "Assessing the competencies in general surgery residency training". *The American Journal of Surgery*, vol. 189, pp. 288 - 292.
- [3] Burch, V.C., Seggie, J.L. & Gary, N.E. (2006). "Formative assessment promotes learning in undergraduate clinical clerkships". *South African Medical Journal*, vol. 96, pp. 430 - 433.
- [4] Cook, D.A. & Beckman, T.J. (2008). "Does scale length matter? A comparison of nine- versus five-point rating scales for the mini-CEX". [Electronic version]. *Advances in Health Science Education*.
- [5] Cook, D.A., Dupras, D.M., Beckman, T.J., Thomas, K.G. & Pankratz, S. (2008). "Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial". *Journal of General Internal Medicine*, vol. 24, pp. 74 - 79.
- [6] Cruess, R., McIlroy, J.H., Cruess, S., Ginsburg, S. & Steinert, Y. (2006). "The professionalism mini-evaluation exercise: a preliminary investigation". *Academic Medicine*, vol. 81, pp. S74 - S78.
- [7] Donato, A.A., Pangaro, L., Smith, C., Rencic, J., Diaz, Y., Mensinger, J. & Holmboe, E. (2008). "Evaluation of a novel assessment form for observing medical residents: a randomised, controlled trial". *Medical Education*, vol. 42, pp. 1234 - 1242.
- [8] Dowson, C. & Hassell, A. (2006). "Competence-based assessment of specialist registrars: evaluation of a new assessment of out-patient consultations". *Rheumatology*, vol. 45, pp. 459 - 464.
- [9] Durning, S.J., Cation, L.J., Markert, R.J. & Pangaro, L.N. (2002). "Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training". *Academic Medicine*, vol. 77, pp. 900 - 904.
- [10] Golnik, K.C. & Goldenhar, L. (2005). "The ophthalmic clinical evaluation exercise: reliability determination". *Ophthalmology*, vol. 112, pp. 1649 - 1654.
- [11] Golnik, K.C., Goldenhar, L.M., Gittinger, J.W. & Lustbader, J.M. (2004). "The ophthalmic clinical evaluation exercise (OCEX)". *Ophthalmology*, vol. 111, pp. 1271 - 1274.
- [12] Han, P.K., Keranen, L.B., Lescisin, D.A. & Arnold, R.M. (2005). "The palliative care clinical evaluation exercise (CEX): an experience-based intervention for teaching end-of-life communication skills". *Academic Medicine*, vol. 80, pp. 669 - 676.
- [13] Hatala, R., Ainslie, M., Kassen, B.O., Mackie, I. & Roberts, J.M. (2006). "Assessing the mini-clinical evaluation exercise in comparison to a national specialty examination". *Medical Education*, vol. 40, pp. 950 - 956.
- [14] Hatala, R. & Norman, G.R. (1999). "In-training evaluation during an internal medicine clerkship". *Academic Medicine*, vol. 74, pp. S118 - S120.
- [15] Holmboe, E.S., Huot, S., Chung, J., Norcini, J. & Hawkins, R.E. (2003). "Construct validity of the mini clinical evaluation exercise (miniCEX)". *Academic Medicine*, vol. 78, pp. 826 - 830.
- [16] Kogan, J.R., Bellini, L.M. & Shea, J.A. (2002). "Implementation of the mini-CEX to evaluate medical students' clinical skills". *Academic Medicine*, vol. 77, pp. 1156 - 1157.
- [17] Kogan, J.R., Bellini, L.M. & Shea, J.A. (2003). "Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship". *Academic Medicine*, vol. 78, pp. S33 - S35.
- [18] Kogan, J.R. & Hauer, K.E. (2006). "Brief report: Use of the mini-clinical evaluation exercise in internal medicine core clerkships". *Journal of General Internal Medicine*, vol. 21, pp. 501 - 502.
- [19] Kogan, J.R., Holmboe, E.S. & Hauer, K.E. (2009). "Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review". *JAMA*, vol. 302, no. 12, pp. 1316-1326.
- [20] Lane, J.L. & Gottlieb, R.P. (2000). "Structured clinical observations: a method to teach clinical skills with limited time and financial resources". *Pediatrics*, vol. 105, pp. 973 - 977.
- [21] Links, P.S., Colton, T. & Norman, G.R. (1984). "Evaluating a direct observation exercise in a psychiatric clerkship". *Medical Education*, vol. 18, pp. 46 - 51.
- [22] Malhotra, S., Hatala, R. & Courneya, C.A. (2008). "Internal medicine residents' perceptions of the mini-clinical evaluation exercise". *Medical Teacher*, vol. 30, pp. 414 - 419.

- [23] Margolis, M.J., Clauser, B.E., Cuddy, M.M., Ciccone, A., Mee, J., Harik, P. & Hawkins, R.E. (2006). "Use of the mini-clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: a validity study". *Academic Medicine*, vol. 81, pp. S56 - S60.
- [24] Nair, B.R., Alexander, H.G., McGrath, B.P., Parvathy, M.S., Kilsby, E.C., Wenzel, J., Frank, I.B., Pachev, G.S. (2008). "The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates". *The Medical Journal of Australia*, vol. 189, pp. 159 - 161.
- [25] Norcini, J.J., Blank, L.L., Arnold G.K. & Kimball, H.R. (1995). "The mini-CEX (clinical evaluation exercise): a preliminary investigation". *Annals of Internal Medicine*, vol. 123, pp. 795 - 799.
- [26] Norcini, J.J., Blank, L.L., Arnold, G.K. & Kimball, H.R. (1997). "Examiner differences in the mini-CEX". *Advances in Health Science Education*, vol. 2, pp. 27 - 33.
- [27] Norcini, J.J., Blank, L.L., Duffy, F.D. & Fortna, G.S. (2003). "The mini-CEX: a method for assessing clinical skills". *Annals of Internal Medicine* vol.138, pp. 476 - 481.
- [28] Norcini, J. & Burch, V. (2007). "Workplace-based assessment as an educational tool: AMEE Guide No. 31". *Medical Teacher*, vol. 29, pp. 855 - 871.
- [29] Nyman, K.C. & Sheridan, B. (1997). "Evaluation of consulting skills of trainee general practitioners". *Australian Family Physician*, vol. 26, pp. S28 - S35.
- [30] Paukert, J.L., Richards, M.L. & Olney, C. (2002). "An encounter card system for increasing feedback to students". *The American Journal of Surgery*, vol. 183, pp. 300 - 304.
- [31] Prescott, L.E., Norcini, J.J., McKinlay, P. & Rennie, J.S. (2002). "Facing the challenges of competency-based assessment of postgraduate dental training: Longitudinal Evaluation of Performance (LEP)". *Medical Education*, vol. 36, pp. 92 - 97.
- [32] Prescott-Clements, L., Vleuten, C.P.M., van der, Schuwirth, L.W.T., Hurst, Y. & Rennie, J.S. (2008). "Evidence for validity within workplace assessment; the Longitudinal Evaluation of Performance (LEP)". *Medical Education*, vol. 42, pp. 488 - 495.
- [33] Richards, M.L., Paukert, J.L., Downing, S.M. & Bordage, G. (2007). "Reliability and usefulness of clinical encounter cards for a third-year surgical clerkship". *Journal of Surgical Research*, vol. 140, pp. 139 - 148.
- [34] Ringsted, C., Ostergaard, D., Ravn, L., Pedersen, J.A., Berlac, P.A. & Vleuten, van der C.P.M. (2003). "A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills". *Medical Teacher*, vol. 25, pp. 654 - 658.
- [35] Ross, R. (2002). "A clinical-performance biopsy instrument". *Academic Medicine*, vol. 77, p.268.
- [36] Shayne, P., Gallahue, F., Rinnert, S., Anderson, C.L., Hern, G. & Katz E. (2006). "Reliability of a core competency checklist assessment in the emergency department: the standardized direct observation assessment tool". *Academic Emergency Medicine*, vol. 13, pp. 727 - 732.
- [37] Shayne, P., Heilpern, K., Ander, D. & Palmer-Smith, V. (2002). "Protected clinical teaching time and a bedside clinical evaluation instrument in an emergency medicine training program". *Academic Emergency Medicine*, vol. 9, pp. 1342 - 1349.
- [38] Torre, D.M., Simpson, D.E., Elnicki, D.M., Sebastian, J.L. & Holmboe, E.S. (2007). "Feasibility, reliability and user satisfaction with a PDA-based mini-CEX to evaluate the clinical skills of third-year medical students". *Teaching and Learning in Medicine*, vol. 19, pp. 271 - 277.
- [39] Turnbull, J., MacFadyen, J., Barneveld, van, C. & Norman, G. (2000). "Clinical work sampling. A new approach to the problem of in-training evaluation". *Journal of General Internal Medicine*, vol. 15, pp. 556 - 561.
- [40] Vleuten, van der, C.P.M. (1996). "The assessment of professional competence: developments, research and practical implications". *Advances in Health Sciences Education*, vol. 1, pp. 41 - 67.
- [41] Wass, V. & Vleuten, van der, C. (2004). "The long case". *Medical Education*, vol. 38, pp. 1176 - 1180.
- [42] Wass, V., Vleuten, van der, C., Shatzer, J. & Jones, R. (2001). "Assessment of clinical competence". *Lancet*, vol. 357, pp. 945 - 949.
- [43] Wilkinson, J.R., Crossley, J.G.M., Wragg, A., Mills, P., Cowan, G. & Wade, W. (2008). "Implementing workplace-based assessment across the medical specialties in the United Kingdom". *Medical Education*, vol. 42, pp. 364 - 373.
- [44] <http://www.abim.org/pdf/paper-tools/attendig.pdf> American Board of Internal Medicine (ABIM), document: "Attending physicians, your role in evaluating clinical competence". [Retrieved at January 7, 2009]
- [45] <http://www.abim.org/pdf/paper-tools/minicex.pdf> American Board of Internal Medicine (ABIM), mini-CEX brochure on internet. [Retrieved at January 7, 2009]

- [46] <http://www.abim.org/pdf/paper-tools/resident.pdf> American Board of Internal Medicine (ABIM). document: "Residents: evaluating your clinical competence". [Retrieved at January 7, 2009]
- [47] http://www.acgme.org/acWebsite/RRC_240/240_OCEX.pdf [Retrieved at November 2008]
- [48] <http://www.cordtest.org> [Retrieved at November 2008]
- [49] <http://www.nvmo.nl/portals/0/congres2006/Steinert.pdf> [Retrieved at November 2008]