

Appendix 1. Characteristics of 34 assessment tools in 48 studies

Tool	Item evaluation	Assessed knowledge, skills, and behaviour (Canadian Safety Competencies domains*)	Source	Study aim	Setting	Learners	Main findings
MEDICAL							
The Miller level "knows"							
A knowledge, skills, and attitudes questionnaire	28 items (5 MCQ, and 18 items of 5 point scale)	Medical error reporting, physician's attitude toward medical errors, tolerance of uncertainty, reaction to uncertainty, institutional safety culture (1, 6)	Madigosky WS, 2006 ²³	To evaluate patient safety and medical fallibility curriculum	Single US medical school	92 second-year medical students	Improvement in self-reported knowledge, and attitudes changed after training
			Hall LW, 2010 ²⁴	To evaluate patient safety training	Single US medical school	146 third-year medical students (81 study group, 65 control group)	Training increased student comfort in identifying the cause of an error.
Unnamed online test	26 items, MCQ	Patient safety, error prevention, system theory, United States health care system (1, 4)	Kerfoot BP, 2007 ²⁵	To evaluate a web-based training program	US web-based program	276 medical students (years 2 and 3) and 217 residents (PGY1 to 5)	Participants' test scores improved after completing the first module.
Unnamed questionnaire	14 items, MCQ	Error prevention, system theory (1, 4)	Kerfoot BP, 2007 ²⁶	To determine the current patient safety knowledge	7 US residencies and 2 US medical schools	268 second and third-year medical students and 372 medical residents (PGY 1 to 5)	Participants correctly answered a mean 58.4%. Patient safety knowledge levels varied significantly by participants' characteristics.
Unnamed questionnaire	MCQ	RCA analysis: defining violation errors, barriers to error, contributory factors, types of error (1)	Wallace LM, 2009 ²⁷	To evaluate a training program	UK national training program	374 health professionals immediately after the training, 350 participants 6 months after the training.	The RCA training enhanced knowledge of RCA.
The Miller level "knows how"							
Computer-based case management of USMLE Step 3	Free text entry: 25 cases and 11 cases	Patient management, errors in decision making (4)	Harik P, 2009 ²⁸	To examine potentially dangerous actions	US license examination	25,283 medical students	More than 20% of examinees ordered at least one dangerous action.
			Floreck LM, 2002 ²⁹	To examine potentially dangerous actions	US license examination	About 7,000 medical students	Examinees performed at least one non-indicated action in more than 40% of cases.

The Miller level "shows how"							
Unnamed checklist	10 items	Safe handover (4)	Klamen DL, 2009 ³⁰	To develop a training program	Single US medical school	69 second-year medical students	More than 80% of students could safely handover
Competencies assessed at multiple levels							
8 station OSCE	Checklist and written examinations (MCQ, short answer, scored based on original checklist)	Quality measurement, RCA, evidence-based medicine, insurance systems, team collaboration (only assess team competence), prescription errors, Nolan's model, negotiation (1, 2, 3, 4, 6)	Varkey P, 2008 ³¹	To develop an assessment tool	Single US quality improvement elective	9 medical residents (preventive medicine and endocrinology)	All faculty members agreed that the OSCE was realistic and capable of providing accurate assessments.
			Gupta P, 2009 ³²	To develop an assessment tool	Single US quality improvement elective	9 medical residents (preventive medicine and endocrinology)	Pilot study. All faculty members agreed that they were able to accurately assess competency in RCA using the tool.
6 station OSCE	Not mentioned	Interview: error case, health behaviour change case; simulations; video: teamwork, co-ordination of care; chart: adverse event case, deficient care case; written examination: behavioural skills, medication safety, system approach (1, 2, 3, 4, 6)	Singh R, 2009 ³³	To develop an assessment tool to evaluate a training program	US residency programs (patient safety curriculum)	59 medical residents (PGY1 to 3, 47 study group, 12 control group, family medicine)	Safety training increased systems thinking and inculcation of a culture of safety.
10 station OSCE	Checklist and written examinations (MCQ)	Accurate data gathering, accurate written communication, patient-centered communication, timely/effective patient care, team functioning/inter-professional communication, infection prevention (2, 3, 4, 6)	Wagner DP, 2009 ³⁴	To develop an assessment tool	64 US residency programs (patient safety curriculum)	235 medical residents (PGY1, multispecialty)	Significant performance deficits were noted. Trainee and administrator evaluation of the experience was positive.

ANESTHESIA

The Miller level "shows how"

The Anesthetists' Non-Technical Skills System (ANTS)	4 categories, 15 elements, 4 point scale	Non-technical skills: situation awareness, task management, team working, decision making (2, 3, 4)	Zausig YA, 2009 ³⁵	To compare non-technical skills and medical management skills after training	2 German university hospitals and 5 community hospitals	42 physicians (anesthesia)	A single session of training did not improve non-technical skills when compared with anesthesiologists who only received medical management training.
			Savoldelli GL, 2006 ³⁶	To evaluate the effect of debriefing	Single Canadian University	42 medical residents (PGY 1, 2, and 4, anesthesia)	Participants' non-technical skills did not improve in the control group, whereas the provision of oral feedback, either assisted or not assisted with videotape review, resulted in significant improvement ($P < 0.005$).
			Yee B, 2005 ³⁷	To evaluate a training program	Single Canadian University	20 medical residents (PGY 2 and 4, anesthesia)	A significant improvement in the non-technical skills of residents was demonstrated from their first to second session and from their first to third session (both $P < 0.005$).
			Welke TM, 2009 ³⁸	To evaluate a training program	Single Canadian University	30 medical residents (PGY1 and 2, anesthesia)	Posttest ($P = 0.009$) and retention ($P < 0.001$) performances of non-technical skills were significantly improved in the standardized multimedia instruction group compared with pretest.
			Müller MP, 2009 ³⁹	To evaluate a training program	5 ICUs at a single German university hospital	32 physicians (anesthesia)	ANTS scores as well as clinical performances were significantly better in the post-intervention scenario.
			Fletcher G, 2003 ⁴⁰	To develop an assessment tool	17 UK hospitals	50 physicians (anesthesia)	The results showed that the skills are observable and can be rated with acceptable levels of agreement and accuracy.
			Bruppacher HR, 2010 ⁴¹ (the "does" level)	To compare training programs	Single Canadian hospital	20 medical residents (PGY 4 or higher, anesthesia)	The simulation group scored significantly higher than the seminar group at both posttest and retention test.

Rating scale	3 categories, 5 point scale	Behaviour (non-technical skills), medical management, overall performance (2)	Weller JM, 2008 ⁴²	To develop an assessment tool	All anesthesia trainees in a region of New Zealand	20 medical residents (anesthesia)	Scores for behaviour were less dependent on the specific clinical context than the scores for overall performance and medical management.
Competencies assessed at multiple levels							
4 station: 1)Structured interview, 2)Portfolio review, 3)Presentation , 4)Simulation	Structured interview, portfolio, presentation, and simulation with checklist	Personal skills (communication, organization, working under pressure, situational awareness and decision making, team working), content of portfolio, global rating (2, 3, 4, 5)	Gale TC, 2010 ⁴³	To develop a selection centre for recruitment	Candidates attending interview for anesthesia training posts in the South West Peninsula Deanery in UK	224 candidates (anesthesia)	Candidates and assessors demonstrated strong approval of the selection centre with more than 70% of ratings 'good' or 'excellent'.
SURGERY							
The Miller level "shows how"							
The Non-technical Skills for Surgeons (NOTSS)	5 categories, 14 elements, 4 point scale	Non-technical skills: situation awareness, decision making, task management, leadership, communication and teamwork (2, 3, 4)	Yule S, 2006 ⁴⁴	To develop an assessment tool	UK	27 consultant surgeons	Five categories of non-technical skills were identified.
The Non-Technical Skills System (NOTECHS)	5 categories, 19 elements, 5 or 6 point scale	Non-technical skills: communication and interaction, situation awareness and vigilance, co-operation and team skills, leadership and managerial skills, decision making (revised NOTECH version) (2, 3, 4)	Black SA, 2010 ⁴⁵	To assess surgical competence in a simulated operating theatre	UK simulated operating theatre	30 physicians (surgery)	There was a significant difference in non-technical skill with ascending grade for both scenarios (P < 0.001). There was a highly significant correlation between technical and non-technical performance in both scenarios.
			Moorthy K, 2006 ⁴⁶	To evaluate a training program	UK simulated operating theatre	20 medical residents (surgery, 10 senior and 10 junior trainees)	While the senior trainees scored higher than the juniors for technical skills (P=0.001), there were no differences in human factors skills.

			Moorthy K, 2005 ⁴⁷	To develop a simulated environment	UK simulated operating theatre	26 surgical trainees (surgery, trainees were divided into 3 groups according to their level of experience)	While technical skills ratings discriminated between surgeons according to experience (P=0.002), there were no differences in terms of the checklist and teamwork skills (P=0.70).
			Undre S, 2007 ⁴⁸	To develop a training program	UK simulated operating theatre	20 teams (N=80, trainee surgeon, anesthetist, operating department practitioner, scrub nurse)	Leadership and decision making were scored lower than the other three non-technical skills (communication, team skills, and vigilance).
			Powers KA, 2008 ⁴⁹	To develop a simulation model	US simulated operating theatre	5 FLS-certified surgeon experts and 5 non-FLS certified novices	There was a significant difference between the two groups in their overall technical and non-technical skills.
The PAR matrix	18 skill-goals (the y-axis is divided into six levels and the x-axis into three columns: unsatisfactory, competent, good)	Crew resource management skills, surgical technical skills (ergonomics) (2, 3, 5)	Papaspyros SC, 2009 ⁵⁰	To develop an assessment tool	Single UK hospital	14 medical residents (surgery)	The format of the PAR model is such that it allows trainer and trainee to objectively assess progress, identify deficiencies and strengths and formulate an appropriate plan of action.
Observational Teamwork Assessment for Surgery (OTAS)	5 elements, 7 point scale	Teamwork: shared monitoring, communication, co-operation, co-ordination, shared leadership (2)	Wetzel CM, 2010 ⁵¹	To investigate the effect of stress on performance	UK simulated operating theatre	30 physicians (surgery)	Coping significantly enhanced non-technical skills (beta, 0.302; 0.117-1.624, P= 0.03).

EMERGENCY/RESUSCITATION							
The Miller level "shows how"							
Stanford University for crisis management behaviours	11 items, 5 point scale	Orientation to case, inquiry/assertion, communication, feedback, leadership, group climate, anticipation/planning, workload distribution, vigilance and re-evaluation, overall leadership (1, 2, 3, 5)	Wallin CJ, 2007 ⁵²	To evaluate a training program	Swedish simulated emergency environment	15 medical students	Nine of 10 observed team skills improved significantly in response to practice, in parallel with a global rating of team skills.
A behaviour grading sheet	3 categories, 3 point scale (poor/fail, adequate, excellent)	Crisis management skills: teamwork, decision making, situation awareness (2, 4)	Knudson MM, 2008 ⁵³	To evaluate a training program	US simulated trauma emergency environment	18 medical residents (PGY2 to 3, surgery)	Simulation-based trained residents received higher overall scores and higher scores for crisis management skills compared with the didactic lecture-trained group, which was most evident in the scores received for the teamwork category (p=0.04).
Non-technical scorecard at simulated septic shock scenario	7 items, 5 point scale	Non-technical skills: anticipation and planning, communication, leadership, information transfer, task distribution, communication content, information use (2, 3, 4)	Williams JB, 2009 ⁵⁴	To assess reliability of real-time vs. videotaped evaluation	US simulated environment	26 medical residents (emergency and transitional)	Nland-Altman plot analysis of both conditions revealed substantial agreement between the real-time and videotaped review scores by reviewers.
A behaviourally anchored team skill rating scale	5 categories, 5 point scale	Assertiveness, decision making, leadership, communication, situation assessment (2, 4)	Wright MC, 2009 ⁵⁵	To evaluate the relationship between teamwork skills and clinical performance	US simulated environment	35 first-year medical students	There was moderate to high correlation between observer ratings of team skill and checklist-based measures of team performance.
Modified ANTS and Anti-Air Teamwork Observation Measure (ATOM)	7 categories	Modified ANTS: co-ordination, information exchange, use of authority, assessing capabilities, supporting behaviours; ATOM: shared mental model (2, 4)	Westli HK, 2010 ⁵⁶	To evaluate teamwork skills	Norwegian hospitals	27 trauma teams (139 clinicians, including surgeons, anesthesiologists, nurses, and radiographers, emergency)	Specific teamwork skills and behavioural markers were associated with indicators of good team performance.

The Ottawa Crisis Resource Management Global Rating Scale (Ottawa GRS) and The Ottawa Crew Resource Management checklist (Ottawa CRM checklist)	The Ottawa GRS: 5 categories, 7 point scale; the Ottawa CRM checklist: 5 categories, 12 items, 2 point scale ⁴²	Crew resource management skills: leadership, problem solving, situation awareness, resource utilization, communication (2, 3, 4)	Kim J, 2009 ⁵⁷	To compare assessment tools	Canadian university tertiary hospital	60 medical residents (PGY1 and 2, multispecialty)	Users indicated a strong preference for the Ottawa GRS given ease of scoring, presence of an overall score, and the potential for formative evaluation.
			Kim J, 2006 ⁵⁸	To develop an assessment tool	Canadian university tertiary hospital	60 first and third-year medical residents (multispecialty)	Construct validity was measured on the basis of content validity, response process, internal structure, and response to other variables.
Teamwork Behaviour Frequencies Form	Frequency (number of behaviours per minute)	Information sharing, inquiry, assertion, intentions shared and evaluation of plans, vigilance, workload (2, 3, 4, 5)	Thomas EJ, 2007 ⁵⁹	To evaluate a training program	US neonatal resuscitation program	51 medical residents (multispecialty)	The interns in the neonatal resuscitation program with team training exhibited more frequent team behaviours than interns in the control group.
Observational Skill-based Clinical Assessment tool for Resuscitation (OSCAR)	6 categories	Communication, co-operation, co-ordination, monitoring, leadership, decision making (2)	Walker S, 2010 ⁶⁰	To develop an assessment tool	UK	Multidisciplinary experts to assess content validity, 2 clinician observers used (resuscitation)	OSCAR is a psychometrically robust (reliable, content- and face- valid) tool for the assessment of team working skills in cardiac arrest events.
The Miller level "knows"							
Unnamed scale	36 items, 5 point scale	Team building, shared decision making in emergency situations (2)	Haller G, 2008 ⁶¹	To evaluate a training program	Swiss labour and delivery units of a university-affiliated hospital	239 multidisciplinary (midwives, nurses, physicians and technicians)	Except for seven items, the 36-item survey testing participants' learning demonstrated a significant change (P<0.05) towards better knowledge of teamwork and shared decision making after the training program.

NURSING							
The Miller level "knows"							
Quality Improvement Knowledge, Skills, and Attitudes (QulSKA)	73 items, (26 items: MCQ; 6 items: true-false; 31 items: 4 or 6 point scale)	QSEN competencies (quality improvement, safety, evidence-based practice, teamwork, patient-centered care, informatics) (1, 2, 3, 4)	Dycus P, 2009 ⁶²	To develop an assessment tool	US and Latin American nurses practicing at a hospital and affiliate partner sites	37 nurses (pediatric nursing)	The QulSKA inter-item correlation coefficient was 0.839 (P = .001). The mean knowledge score (based on 100) was 69.2 +/- 11.3.
The patient safety attitudes, skills, and knowledge scale (PS-ASK)	26 items, 5 point scale	Patient safety, patient safety practices relevant to nursing, evidence-based advanced practice nursing care (1, 4)	Schnall R, 2008 ⁶³	To develop an assessment tool	US nursing school	285 nursing students	The final instrument consists of 26 items and three separate scales: attitudes, skills, and knowledge.
The Miller level "knows how"							
Essay	N/A	Unit-based quality improvement, meaning of sentinel event, interdisciplinary team (1, 2, 6)	Murray ME, 2010 ⁶⁴	To describe patient safety education in nursing schools	US	-	The application assignment gives students an empowering experience of implementing a process change and measuring results.
Situation awareness check list	17 items	Situation awareness (4)	Cooper S, 2010 ⁶⁵	To examine the ability of nursing students in a simulated environment	Australian nursing school	112 nursing students	The mean situation awareness score across both scenarios was 59% (range 38-82%).
The Miller level "shows how"							
Clinical Simulation Evaluation Tool (CSET)	15 items, numeric scale	Safety, basic assessment, prioritization, problem-focused assessment, ensuring intervention, delegation, communication (3, 4)	Radhakrishnan K, 2006 ⁶⁶	To examine clinical performance	US simulation center	12 senior second degree BSN nursing students	Students who practiced with the simulation had significantly higher scores than the control group on patient identification and on assessing vital signs.
Knowledge, Skills, Attitudes criterion (Checklist) : QSEN	16 items, dichotomous scale	QSEN Competencies (quality improvement, safety, evidence-based practice, teamwork, patient-centered care, informatics) (1, 2, 3, 4)	Ironside PM, 2009 ⁶⁷	To evaluate a training program	US simulation center	67 nursing students from 8 universities (acute care nursing)	The study showed that students' safety competencies improved significantly from the first to the second simulation.

Unnamed checklist	6 items	Patient-centered care of QSEN (1, 2, 3, 4, 6)	McKeon LM, 2009 ⁶⁸	To evaluate a training program	US university	53 nursing students (34 computer-based simulation, 31 traditional simulation)	Students in both intervention computer-based simulation and traditional simulation achieved similar competencies in patient-centered care.
Clinical Performance Evaluation Tool	6 categories, 3 choices (satisfactory, needs improvement, unsatisfactory)	Patient-centered care, teamwork and collaboration, evidence-based nursing, quality improvement, safety, informatics (2, 4)	Walsh T, 2010 ⁶⁹	To describe the development of an assessment tool	US core group of undergraduate faculty	Nurse	The tool more accurately appraised clinical capabilities by focusing on quality and safety in health care. Further research is needed for evaluation of reliability and validity.
The Miller level "does"							
Unnamed checklist	44 items, 6 point scale	Patient comfort and safety, time management, medication administration (4)	Fisher M, 2000 ⁷⁰	To develop an assessment tool	Australian tertiary and referral health service	30 nurses	The competency assessment tool has face validity and is internally consistent.

Abbreviations: MCQ=Multiple Choice Question; QSEN=Quality and Safety Education for Nurses

*Canadian Safety Competencies: 1) Contribute to a culture of patient safety, 2) Work in teams for patient safety, 3) Communicate effectively for patient safety, 4) Manage safety risks, 5) Optimize human and environmental factors, 6) Recognize, respond to and disclose adverse events