

Assessing the patient safety competencies of healthcare professionals: a systematic review

Ayako Okuyama,^{1,2} Kartinie Martowirono,¹ Bart Bijnen^{1,3}

► An additional appendix is published online only. To view this file please visit the journal online (<http://qualitysafety.bmj.com>).

¹Foreest Medical School, Medical Center Alkmaar, Alkmaar, The Netherlands

²Graduate School of Medicine, Osaka University, Osaka, Japan

³VU University Medical Center, Institute for Education and Training, Amsterdam, The Netherlands

Correspondence to

Ayako Okuyama, 1–7 Yamadaoka, Suitashi, Osaka, Japan 565-0871; aokuyama-ky@umin.ac.jp

Accepted 21 July 2011
Published Online First
31 August 2011

ABSTRACT

Background: Patient safety training of healthcare professionals is a new area of education. Assessment of the pertinent competencies should be a part of this education. This review aims to identify the available assessment tools for different patient safety domains and evaluate them according to Miller's four competency levels.

Methods: The authors searched PubMed, MEDLINE, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Web of Science, psycINFO and the Education Resource Information Center (ERIC) from the start of each database to December 2010 for English-language articles that evaluated or described tools for the assessment of the safety competencies of individual medical and/or nursing professionals. Reports on the assessment of technical, clinical, medication and disclosure skills were excluded.

Results: Thirty-four assessment tools in 48 studies were identified: 20 tools for medical professionals, nine tools for nursing professionals, and five tools for both medical and nursing professionals. Twenty of these tools assessed the two highest Miller levels ('shows how' and 'does') and four tools were directed at multiple levels. Most of the tools that aimed at the higher levels assessed the skills of working in teams (17 tools), risk management (15 tools), and communication (11 tools). Internal structure (reliability, 22 tools) and content validity (14 tools) when described were found to be moderate. Only a small number of tools addressed the relationship between the tool itself and (1) other assessments (concurrent, predictive validity, eight tools), and (2) educational outcomes (seven tools).

Conclusions: There are many tools designed to assess the safety competencies of healthcare professionals. However, a reliable and valid toolbox for summative testing that covers all patient safety domains at Miller's four competency levels cannot yet be constructed. Many tools, however, are useful for formative feedback.

INTRODUCTION

Patient safety is nowadays increasingly recognised as a key dimension of quality care, and is thus progressively being integrated into the education of healthcare professionals.^{1–4} The Institute of Medicine, renowned for its 2000 publication *To Err is Human*,⁵ stated in a 2003 report that 'all health professionals should be educated to deliver patient-centred care as members of an interdisciplinary team, emphasising evidence-based practice, quality improvement approaches, and informatics'.⁶ Professional bodies and the WHO have since endorsed a patient safety competencies framework for healthcare professionals to enhance local patient safety training programs.^{6–9}

Assessment is an essential component of competency-based education,^{10–12} and should be used for giving feedback as well as for summative examination. While a significant number of studies have been carried out to assess the competency of healthcare professionals, none of the previous systematic reviews on professional competence have focused on the domains relevant to patient safety.^{13 14} Competence is context dependent because it interrelates the ability of the healthcare professional, the task at hand, the ecology of the working environment, and the clinical contexts in which the task is performed.¹⁵ Accordingly, competence can only be assessed in the work place, but trainees must know what is required in order to provide safe care, and know how to use the knowledge they have accumulated. Thus, ideally, there would be a toolbox available that enrolls all of the elements of the defined patient safety competencies throughout the curriculum, assessing each of these at that level of Miller that is fitting to the stage of



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://qualitysafety.bmj.com/site/about/unlocked.xhtml>

training reached.¹⁶ This review therefore aims to identify and evaluate the available tools for the assessment of the safety competencies of physicians, residents, medical students, nurses, and nursing students in a hospital setting. Four review questions are discussed:

1. What types of tool, focused on knowledge, skills, or behaviour, are available for the assessment of the individual safety competencies?
2. What types of content are measured by these assessment tools?
3. To what extent are these assessment tools reliable and valid?
4. Which safety domains⁷ are being covered?

METHODS

Data sources

Relevant English language articles published from the start of each database to December 2010 were sourced using PubMed, MEDLINE, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Web of Science, psycINFO, and the Education Resource Information Center (ERIC). Combinations of search terms were used, relating to assessment (educational measurement, teaching, assessment, curriculum, education professional), competence (professional competence, competence, competencies, educational status, ability, skills), and patient safety (medical error, risk management, safety, patient safety, error, errors).

An additional keyword search was also conducted using PubMed. Again, combinations of search terms were employed, relating to assessment (OSCE, peer assessment, oral examination, essay, portfolio, CEX, mini-CEX, non-technical skill) and patient safety (safety, patient safety, error, errors). The Medical Subject Headings (MeSH) were used when available.

Hand searches were also conducted of relevant journals on medical/nursing education and patient safety (*Medical Education*, *Academic Medicine*, *Medical Teacher*, *Journal of Nursing Education*, *Journal of Patient Safety*, *The Joint Commission Journal on Quality and Patient Safety*, *BMJ Quality & Safety*). Furthermore, the referenced articles listed in each of the selected publications were examined, the abstracts of relevant congresses were screened, and several patient safety and educational experts were consulted.

Selection of articles

An article was selected only if it fulfilled all of the following criteria: the subjects of the study were physicians, residents, medical students, nurses, or nursing students; the article described the tools for the assessment of the individual safety competencies, focusing on knowledge, skills or behaviour; and the content of the

assessment covered the domains of the Canadian Safety Competencies Framework.⁷ The Canadian Patient Safety Institute finalised this inter-professional safety competencies framework in 2008.⁷ The framework consists of six core competency domains: 1. Contributing to a culture of patient safety. 2. Working in teams for patient safety. 3. Communicating effectively for patient safety. 4. Managing safety risks. 5. Optimising human and environmental factors. 6. Recognising, responding to, and disclosing adverse events.

An article was excluded if it focused on the assessment of procedural or technical skills (eg, surgical skills, time and motion instrument handling), skills of medication safety (eg, drug prescribing skills, drug administration skills), general clinical skills, diagnostic skills, or open disclosure skills.

Data extraction

Two authors (AO, KM) reviewed the titles and abstracts of citations generated by the search to assess their eligibility for further review based on the selection criteria, and chose relevant articles for possible inclusion. They, supported by the third author (BB), then reviewed all of the selected articles and decided which to include in this study. The standard Best Evidence in Medical Education coding sheet¹⁷ was modified to focus on relevant parameters (country, single/multi-institution, speciality, trainee level, use for formative/summative evaluation) and tool characteristics (assessment methods, assessed competencies, reliability, validity, outcomes). One author (AO) abstracted data using this modified coding sheet, and then all of the authors reviewed the abstractions to ensure completeness and accuracy. Differences in data abstraction were resolved by consensus.

Information on the reliability, validity and outcomes of the assessment tools described in the selected articles was also extracted. Messick's unitary concept, 'construct validity', was used: the degree to which a score can be interpreted as representing the intended underlying construct.^{14 18–20} Validity evidence was evaluated for five areas:

1. Content (face) validity—the relationship between a tool's content and the construct it intends to measure.
2. The response process—the relationship between the intended construct and the thought processes of subjects or observers. This provides evidence that raters have been properly trained (faculty development).
3. Internal structure (reliability)—looks at internal consistency, test–retest reliability, agreement (inter-rater reliability), and generalisability.
4. Relationship with other variables (concurrent, predictive validity)—the correlation of scores with those

from other assessments or outcomes, and the differences between scores of learner subgroups.

5. Outcomes (educational)—the consequences of assessment.

A modified version of Kirkpatrick's hierarchy was used to evaluate the outcomes of tool implementation.^{14 21} Outcome levels which were abstracted included: 1) participation, in other words the learners' and observers' views on the tool and its implementation; 2) self-assessed modification of learner and observer knowledge, skills and behaviour; 3) transfer of learning, in other words an objectively measured change in learner or observer knowledge or skills; and 4) results in terms of a change in organisational delivery or quality of patient care. Information regarding the cost of tool development and implementation was also extracted.

Data synthesis

Following identification of the patient safety domains covered by the assessment tools, the tools were categorised according to Miller's four competency levels: 'knows', 'knows how', 'shows how', and 'does'.¹⁶ For each of these Miller competency levels, the trainee should be able to demonstrate the ability to imitate or replicate a protocol, apply principles in a familiar situation, adapt principles to new situations, and associate new knowledge with previously learned principles.^{16 22}

All data included in this report were previously published and publicly available. Hence, our study did not require submission to the local institutional review board for ethical approval.

RESULT

Search results and article overview

The initial literature search identified 4773 citations (figure 1). Of these articles, 209 were filtered for detailed review to determine whether they met the inclusion criteria. Following the title and abstract review by two of the authors (KM, AO), the value of Cohen's κ was calculated to be 0.91. Forty-six articles met our inclusion criteria, one article was retrieved from the references of a selected article, and one article was retrieved from a hand search. In total, 48 articles were selected, and 34 unique assessment tools were identified (table 1). More detailed information on the selected studies is available in the online appendix.^{23–70}

Forty-five of our selected articles (94%) were published after 2005. The years 2009 (16 articles, 33%) and 2010 (10 articles, 21%) yielded the highest numbers of publications relevant to this review. Almost half of the selected articles (22 articles, 46%) came from the USA, and of the remaining 26, 14 (29%) came from Europe, and 6 (13%) from Canada. Thirty-four of the selected

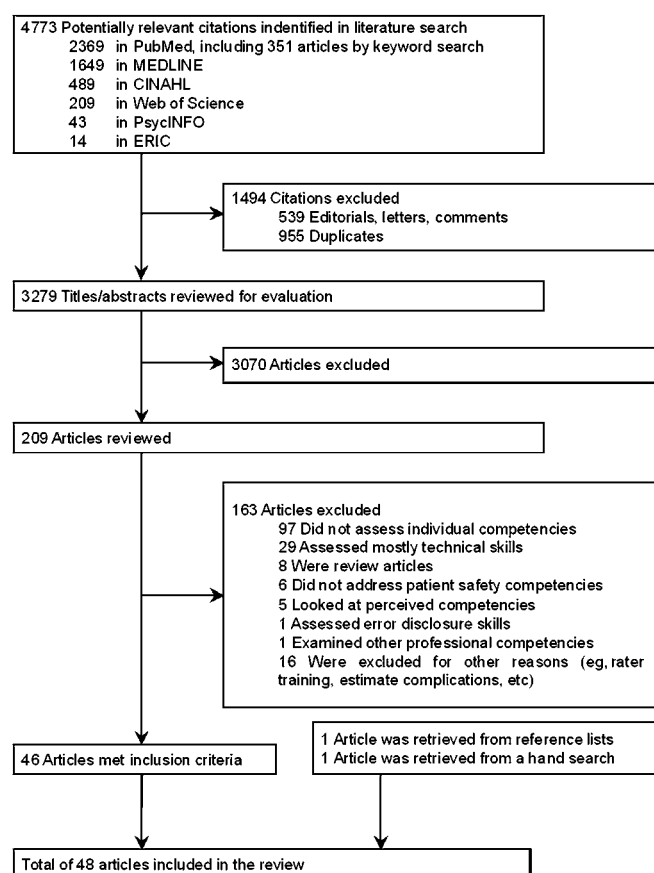


Figure 1 Literature search and study selection process.

studies (71%) were conducted within single institutions. Just one of our selected articles (2.1%) refers to the implementation cost of the competency assessment.³¹

Of our 34 identified assessment tools, 12 (35%) failed to adequately explain the purpose of competency assessment. Seventeen tools (50%) in 20 studies were used for the summative evaluation of training programs.^{23–25 27–29 31–33 50 52 53 59 61 62 64 66–69} Three tools (8.8%) were employed for the formative evaluation of trainees.^{30 34 70} Two tools (5.9%) were utilised for both formative and summative evaluation purposes.^{35–41 45–49}

Description of tools

Of the 34 assessment tools identified in our 48 selected studies, seven tools (21%) measured the trainees' knowledge of patient safety (the 'knows' level) (table 2). Three tools (8.8%) assessed the trainees' applied knowledge using case management (the 'knows how' level). Nineteen tools (56%) were used in simulations, with or without standardised patients, to evaluate the trainees' performance (the 'shows how' level). Four tools (12%) assessed the trainees' competencies at multiple levels (the 'knows', 'knows how', and 'shows how' levels); three of these tools were 'Objective Structured Clinical Examinations' (OSCEs). Two tools (5.9%) implemented the direct observation of trainees with

Table 1 Characteristics of 48 studies describing 34 tools for patient safety competencies

Characteristics	N (%)
Location	
United States	22 (46)
Canada	6 (13)
Europe	14 (29)
Australia/New Zealand	3 (6.3)
Other	3 (6.3)
Single/multi-institution	
Single institution	34 (71)
Multi-institution	11 (23)
Other	3 (6.3)
Publication year	
2000–2004	3 (6.3)
2005	2 (4.2)
2006	5 (10)
2007	6 (13)
2008	6 (13)
2009	16 (33)
2010	10 (21)
Speciality	
Anaesthesiology	9 (19)
Surgery	9 (19)
Emergency medicine	2 (4.2)
Family medicine/general practice	1 (2.1)
Multispecialty*	10 (21)
Nursing	9 (19)
Not specified	8 (17)
Learners	
Medical students	8 (17)
Residents/fellows	17 (35)
Physicians	6 (13)
Combination of medical students, residents, and physicians	3 (6.3)
Nurses/nursing students	9 (19)
Multidisciplinary (both physicians and nurses)	5 (10)
Study design	
Randomised controlled trial	9 (19)
Prospective cohort, historical control, or pre-post	9 (19)
Prospective cohort without baseline	11 (23)
Cross-sectional	6 (13)
Other†	13 (27)
Institutional review board approval	32 (67)
Cost mentioned	1 (2.1)

*Multiple specialities or disciplines included within a single study.

†Includes descriptive studies and studies that did not report a specific statement of study design.

actual patients: one evaluated the non-technical skills of anaesthesia residents in the operating theatre,⁴¹ and the other assessed nurses' patient safety handling at inpatients' bedsides (the 'does' level).⁷⁰

Twenty of our 34 identified tools (59%) assessed the competencies of medical professionals. Of the remaining 14 tools, nine (26%) assessed the competencies of nursing professionals, and five (15%) tested both medical professionals and nurses.^{27 48 49 56 60 61}

Assessed content of the Canadian safety competencies

Referring back to the six aforementioned domains of the Canadian Safety Competencies Framework, 25 of our tools (74%) assessed the 'managing safety risks' competencies, and 24 tools (71%) measured the 'working in teams for patient safety' competencies. Approximately half of the tools (16 tools, 47%) assessed the 'communicating effectively for patient safety' competencies, and 12 tools (35%) evaluated the 'contributing to a culture of patient safety' competencies. Six tools (18%) assessed the 'recognizing, responding to, and disclosing adverse events' competencies, while only four tools (12%) measured the 'optimizing human and environmental factors' competencies. The three competency domains 'managing safety risks' (medical 15 tools, nursing eight tools), 'working in teams for patient safety' (medical 15 tools, nursing five tools), and 'communicating effectively for patient safety' (medical 11 tools, nursing four tools) were assessed by a number of tools for both physicians and nurses (figure 2). The 'contributing to a culture of patient safety' domain was evaluated for nurses by five of a total of nine tools (56%), and for physicians by only six of 20 tools (30%).

The majority of tools at the lower levels ('knows' and 'knows how') (seven of a total of 10 tools, 70%) assessed the 'contributing to a culture of patient safety' domain. In contrast, most of the tools at the higher levels ('shows how' and 'does') assessed the 'working in teams for patient safety' (17 of a total of 20 tools, 85%), 'managing safety risks' (15 of a total of 20 tools, 75%), and 'communicating effectively for patient safety' (11 of 20 tools, 55%) domains. Four tools used at multiple levels assessed all of the domains except the 'optimizing human and environmental factors' domain.

Validity evidence

The issue of content validity was raised in the descriptions of 14 tools (41%). According to the response process, rater-training was described for nine tools (26%), usually only once and briefly. The matter of internal structure was mentioned for 22 tools (65%). The majority of our tools have acceptable reliability. Correlations between evaluated scores and other assessment variables were described for eight tools (24%).

The safety competency results obtained using our assessment tools were compared with technical skills,^{35 45 55 56} work place performance,⁴³ and written examination scores,^{28 31 65} whether tests were failed or passed.²⁸ Our safety competency results show a high ($r=0.65-0.80$)^{35 45 55} or modest ($r=0.34$)⁵⁶ correlation with technical skills. Conversely, they exhibit a low ($r=0.13$)^{32 68} or modest ($r=0.25$)^{32 35 68} correlation with final written examination scores. Performance scores were also compared across training levels and trainee

Table 2 Description of 34 tools in 48 studies for individual trainee safety competency assessment

Tool*	Miller's level†	Safety competencies‡						Validity evidence§		Internal structure	Relationships to other variables	Outcomes¶
								Content validity	Response process			
Medical	1	✓				✓	No	No	No	Pre/post intervention	No	
A knowledge, skills, and attitudes questionnaire ^{23 24}	1	✓					Yes	No	IC, TRR	Learner level	1	
Unnamed online test ²⁵	1	✓		✓			Yes	No	IC, TRR	Learner level	No	
Unnamed questionnaire ^{26 **}	1	✓					No	No	No	No	No	
Unnamed questionnaire ²⁷	1	✓				✓	No	No	Reliability coefficient, inter-case reliability	Concurrent ²⁸	No	
Computer-based case management of USMLE Step 3 ^{28 29}	2						No	No	No	No	No	
USMLE checklist ³⁰	3		✓	✓		✓	No	No	IC, IRR, inter-station correlation	Concurrent validity	No	
8-station OSCE ^{31 32}	5	✓	✓	✓	✓	✓	Yes	Yes	IC, IRR, inter-station correlation	Concurrent validity	1 ³²	
6-station OSCE ³³	5	✓		✓	✓	✓	No	No	coefficients, G	Learner level	No	
10-station OSCE ³⁴	5		✓	✓	✓	✓	No	No	No	No	1	
Anaesthesia												
Anesthetists' Non-technical Skills System ^{35–41}	3/4	✓	✓	✓	✓	✓	Yes	Yes	IC, IRR, IRA, rater reference rater difference	Concurrent validity, ³⁵ learner level, ^{36 41} pre/post intervention ^{37–39}	1 ⁷¹	
Rating scale ⁴²	3	✓					No	Yes	G	No	No	
4-station ⁴³	5	✓	✓	✓	✓	✓	Yes	Yes	IC, IRA	Predictive validity	No	
Surgery												
Non-technical Skills for Surgeons ⁴⁴	3	✓	✓	✓	✓	✓	Yes	No	IRR	No	No	
Non-technical Skills System ^{45–49 §§}	3	✓	✓	✓	✓	✓	Yes	Yes	IC, IRR, IRA	Concurrent validity, ^{45 47} learner level ^{45 48 49}	1 ^{45 46 48 49}	
PAR matrix ⁵⁰	3	✓	✓	✓		✓	No	No	No	No	1	
Observational teamwork assessment for surgery ⁵¹	3	✓	✓				Yes	Yes	No	No	No	
Emergency/resuscitation												
Stanford University for Crisis Management Behaviours ⁵²	3	✓	✓	✓		✓	No	No	IRR	Pre/post intervention	No	
A behaviour grading sheet ⁵³	3	✓			✓		No	No	IRR	Learner level	No	
Non-technical scorecard ⁵⁴	3	✓	✓	✓	✓		No	No	IC, IRR	No	No	
A behaviourally anchored team skill rating scale ⁵⁵	3	✓	✓	✓	✓		No	No	IRA	Concurrent validity	No	
Modified ANTS and anti-air teamwork observation measure ^{56 †† §§}	3	✓	✓	✓	✓		Yes	Yes	IRR	Concurrent validity	No	

Continued

Continued

Table 2 Continued

Tool*	Miller's level†	Safety competencies‡						Validity evidence§		Internal structure	Relationships to other variables		Outcomes¶
		1	2	3	4	5	6	Content validity	Response process				
		✓	✓	✓	✓	✓	✓	Yes	Yes	IC, IRR	Learner level	No	
Ottawa Crisis Resource Management Global Rating Scale ^{57 58}	3	✓	✓	✓	✓	✓	✓	Yes	Yes	IC, IRR	Learner level	No	No
Ottawa Crew Resource Management Checklist ⁵⁸	3	✓	✓	✓	✓	✓	✓	Yes	Yes	IRR	Learner level	No	No
Teamwork behaviour frequencies form ⁵⁹	3	✓	✓	✓	✓	✓	✓	Yes	No	IC, IRR, inter-observer reliability	No	No	No
Observational Skill-based Clinical Assessment Tool for Resuscitation ^{60 †† §§}	1	✓	✓	✓	✓	✓	✓	No	No	No	Pre/post intervention	No	No
Unnamed scale ^{61 †† §§}	1	✓	✓	✓	✓	✓	✓	Yes	No	IC	Learner level	No	No
Quality Improvement Knowledge, Skills, and Attitudes ⁶²	1	✓	✓	✓	✓	✓	✓	Yes	No	IC	No	No	No
Patient Safety Attitudes, Skills, and Knowledge Scale ^{63 ††}	1	✓	✓	✓	✓	✓	✓	Yes	No	IC	No	No	No
Essay ⁶⁴	2	✓	✓	✓	✓	✓	✓	No	No	No	Concurrent validity	No	No
Situation Awareness Check list ⁶⁵	2	✓	✓	✓	✓	✓	✓	No	No	No difference in scores by two raters tested	No	No	No
Clinical Simulation Evaluation Tool ^{66 ††}	3	✓	✓	✓	✓	✓	✓	No	No	No	No	No	No
Knowledge, skills, attitudes criterion ⁶⁷	3	✓	✓	✓	✓	✓	✓	No	No	IC	Pre/post intervention	No	No
Unnamed checklist ⁶⁸	3	✓	✓	✓	✓	✓	✓	No	No	No	No	No	No
Clinical Performance Evaluation Tool ⁶⁹	3	✓	✓	✓	✓	✓	✓	No	No	No	No	No	No
Unnamed checklist ⁷⁰	4	✓	✓	✓	✓	✓	✓	No	No	IC, IRR	No	No	1

*Tools were labelled as unnamed if the tool was not named in the study.

†Miller's levels: 1=Knows; 2=Knows how; 3=Shows how; 4=Does; 5=Competencies assessed at multiple levels.

‡Canadian Safety Competency domains: 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=Optimise human and environmental factors; 6=Recognise, respond to, and disclose adverse events.

§Refers to whether each validity component was evaluated. Boldface denotes an acceptable level of validity.

¶Outcomes were evaluated using the lowest level (1)=participation; learners' or observers' views on the tool or its implementation) of a modified Kirkpatrick hierarchy.

**Article authors reported that this is not appropriate for determining individual competency standards because the tool was designed for program improvement.

††Conference report or pilot study.

§§Seventeen items come from 'the Knowledge, Skills, and Attitudes Questionnaire'.²⁵

§§§These tools assessed the competencies of both medical and nursing professionals.

ANTS, Anaesthetists' Non-Technical Skills System; G, generalisability; IC, internal consistency; IRR, inter-rater agreement; IRR, inter-rater reliability; OSCE, Objective Structured Clinical Examination; TRR, test retest reliability.

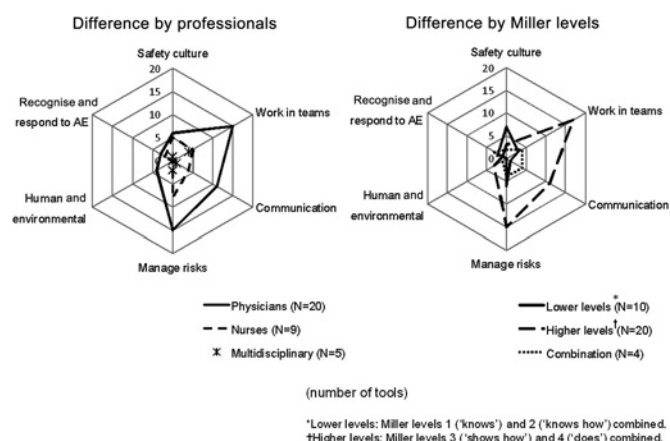


Figure 2 Number of tools available for assessing each of the Canadian Safety Competency domains.

characteristics for 13 of our assessment tools (38%); 11 tools in 17 studies produced scores that increased with training level,^{23–26 33 37–39 45 48 49 52 57–59 61 67} while for three tools in four studies this trend was not observed.^{36 41 53 62} Of 19 tools which were used for the summative assessment, six evaluated both content and internal structure.^{25 38 39 43 46 59 62} Validity evidence for another five tools was not reported.^{27 50 64 66 69}

Trainees and observers reported their experiences as being generally positive (Kirkpatrick level 1) for seven tools in 10 studies (21%).^{25 31 34 45 47–50 70 71}

DISCUSSION

Developing reliable and valid tools for the assessment of safety competencies is a challenging task.²² In the past few years the number of publications on this subject has increased, and several assessment tools are now available.

When assessing safety competencies, instructors should understand the limitations of different approaches and use a range of methods that fit the specific learning situations of Miller's four competency levels.⁷² Of the 34 tools identified in this review, 10 can be used to assess the knowledge of key patient safety concepts at Miller's two lower levels, such as adverse events, just culture, and system thinking. Four^{25 26 62 63} of these 10 tools, which are of proven content validity and internal reliability, are valuable for the evaluation of the cognitive knowledge and reasoning of both medical and nursing students, particularly after a course in basic patient safety science. The remaining six of the 10 tools have limited content validity and reliability. Written examinations such as multiple choice questions, however, tend to measure the potential to perform rather than actual work performance, and more advanced trainees should be assessed through direct observation in simulation scenarios or real practice for the higher competency levels.⁷³ We found 20 tools for

the examination of the 'shows how' and 'does' levels in the domains of 'managing safety risks', 'working in teams for patient safety', and/or 'communicating effectively for patient safety'. Of these 20 tools found for the assessment of the competencies of medical and nursing professionals, three^{56–59} have a proven appropriate content validity and internal reliability, and are suitable for use in an emergency situation. Two tools^{44–49} out of the 20 are ideal for use in surgery, and a further two are suitable for the assessment of anaesthesiologists in training.^{35–41 43}

Each tool of course has its strengths and limitations, and so complementing methods must be sought in order to overcome these weaknesses.^{22 72} For instance, no tool covered all the patient safety competency domains, and therefore tools should be combined or alternated.^{31 33 34} Unfortunately, the various tools cover the domains unevenly, and more work has to be done to investigate tools in the important domains of 'contributing to safety culture', 'optimizing human and environmental factors', and the sub domain of 'recognizing, responding to, and disclosing adverse events'.

It should further be noted that the described tools were mainly applied in specific situations, such as in the operating theatre, during the provision of anaesthesia, and for crisis management, and only a few tools have been developed for the assessment of competencies in daily practice in other workplaces. For example, the Anaesthetists' Non-Technical Skills System (ANTS) and the Non-Technical Skills System (NOTECHS) are available for the assessment of safety competencies in their respective clinical situations (the higher competency levels). The importance of context in measuring the higher competency levels means that such tools cannot simply be copied and used in other working environments. Each discipline or speciality must adapt the tools to its own requirements.

Another issue for discussion is the limited proven reliability and validity of the described tools. The majority of the tools assessed internal structure and content validity. A modest inter-rater reliability was reported for some tools, such as the ANTS and the Non-Technical Skills for Surgeons. Sufficient rater-training is a requirement for summative assessment,^{44 74 75} and yet in the majority of the studies rater-training was either minimally described or not described at all.

The correlation of assessed competency scores with other types of assessment and educational outcomes was evaluated in only a few studies. Four tools showed a strong correlation with technical skills.^{35 45 55 56} However, for one of these tools, the ANTS, it was pointed out that there was an imperfect distinction between non-technical skills and pure medical knowledge in some of the elements of the ANTS scale.³⁶ One review has stated

that “improving patient outcomes as a result of patient safety education represents a particularly daunting task, given that intensive, large-scale quality improvement efforts often fail to demonstrate improvements in health outcomes”.⁷⁶ This clearly explains why evaluating educational outcomes can be so difficult.

Finally, assessment tools should also have an impact on future learning and practice, be acceptable to learners and faculty, and costs must be considered.⁷⁷ Although data are scant, these subjects are important for the compliance of the involved staff and trainees and the feasibility of implementing the assessment tools.

The assessment of safety competencies is a new field of education, and it is clearly difficult to develop reliable and valid assessment tools. We would like to express our appreciation for the pioneering work of the researchers in this field, as several useful tools could be identified. Similarly, this review has its own limitations. First, our search strategy for article selection focused on competency rather than knowledge; we searched for articles using the terms ‘competence’, ‘ability’, and ‘skill’, and if articles did not use these specific terms, we might have unfairly omitted them from our study. Hence our search methodology may have excluded articles which described or evaluated patient safety knowledge. We tried to cover this by searching the reference lists of the quoted articles, hand searching relevant journals on medical/nursing education and patient safety, engaging in discussions with experts, and screening abstract books of patient safety conferences. Second, we did not focus on technical skills and medical knowledge. Thus tools used principally to assess technical skills, for example, the Objective Structured Assessment of Technical Skill,⁷⁸ were excluded, even if they implicitly included some elements of non-technical skills. Making the non-technical skills more explicit in such studies would, in our opinion, enhance patient safety awareness. Finally, studies which assessed team competencies such as the Oxford NOTECHS scale were also excluded, as our focus was the individual learner.⁷⁹ Therefore, our results may not show the whole picture of the tools for the assessment of safety competencies. Team training is undoubtedly important to patient safety, as is the evaluation of the functioning of a team.

CONCLUSION

This review has identified many tools for the assessment of safety competencies of healthcare professionals at each of Miller’s competency levels. The reliability and validity of these tools, where reported, are however modest at best. The tools do not cover the whole spectrum of patient safety competencies on all of Miller’s competency levels, and they often suit only specific working situations.

New research should focus more on the domains of ‘contributing to a culture of patient safety’ and ‘optimizing human and environmental factors’. Appropriate use of such tools can motivate a trainees’ learning and at least enhance patient safety awareness.

Acknowledgements We would like to thank the medical librarians, Mr Otten at VU Amsterdam University Library, Medical Library in the Netherlands and Mr Swa at Osaka University Life Sciences Library in Japan for assistance with the literature search. We would like to thank Dr van Luijk, associate professor of Medical Education, VU University Medical Center and Institute for Education and Training, Amsterdam, the Netherlands for his critical reading of the manuscript. These individuals did not receive compensation for their efforts beyond their usual salary.

Funding Supported in part by the Stichting Kwaliteitsgelden Medisch Specialisten in the Netherlands (SKMS, Quality foundation of the Dutch Medical Specialists). Martowirono is supported by a scholarship for an advanced researcher from the SKMS.

Competing interests None.

Contributors Okuyama had full access to all of the data in the study and takes full responsibility for the integrity of the data and the accuracy of data analysis. Study concept: Okuyama, Martowirono, Bijnen; acquisition of data: Okuyama; analysis and interpretation of data: Okuyama; drafting the manuscript: Okuyama; critical revision of the manuscript for intellectual content: Okuyama, Martowirono, Bijnen; statistical analysis: Okuyama; study supervision: Bijnen.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Flanagan B, Nestel D, Joseph M. Making patient safety the focus: crisis resource management in the undergraduate curriculum. *Med Educ* 2004;38:56–66.
2. Mayer D, Klamen DL, Gunderson A, *et al*. Designing a patient safety undergraduate medical curriculum: the Telluride Interdisciplinary Roundtable experience. *Teach Learn Med* 2009;21:52–8.
3. Wakefield A, Attree M, Braidman I, *et al*. Patient safety: do nursing and medical curricula address this theme? *Nurse Educ Today* 2005;25:333–40.
4. Smith EL, Cronenwett L, Sherwood G. Current assessments of quality and safety education in nursing. *Nurs Outlook* 2007;55:132–7.
5. Kohn LT, Corrigan J, Donaldson MS, eds. *To err is human: building a safer health system*. 1st edn. Washington, DC: National Academy Press, 2000.
6. Institute of Medicine (US). *Health professions education: a bridge to quality*. http://books.nap.edu/openbook.php?record_id=10681 (accessed 19 Apr 2011).
7. Canadian Patient Safety Institute (Canada). *The safety competencies: enhancing patient safety across the health professions*. <http://www.patientsafetyinstitute.ca/English/toolsResources/safetyCompetencies/Pages/default.aspx> (accessed 22 Aug 2011).
8. Walton MM, Shaw T, Barnett S, *et al*. Developing a national patient safety education framework for Australia. *Qual Saf Health Care* 2006;15:437–42.
9. WHO. *The WHO patient safety curriculum guide for medical schools*. <http://www.who.int/patientsafety/education/curriculum/download/en/index.html> (accessed 19 Apr 2011).
10. Carraccio C, Wolfsthal SD, Englander R, *et al*. Shifting paradigms: from Flexner to competencies. *Acad Med* 2002;77:361–7.
11. Shumway JM, Harden RM; Association for Medical Education in Europe. AMEE Guide No. 25: the assessment of learning outcomes for the competent and reflective physician. *Med Teach* 2003;25:569–84.
12. Crossley J, Humphris G, Jolly B. Assessing health professionals. *Med Educ* 2002;36:800–4.
13. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA* 2002;287:226–35.
14. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees. A systematic review. *JAMA* 2009;302:1316–26.
15. Klass D. Reevaluation of clinical competency. *Am J Phys Med Rehabil* 2000;79:481–6.

16. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65(9 Suppl):S63–7.
17. Best Evidence Medical Education Collaboration. *Best evidence in medical education*. <http://www2.warwick.ac.uk/fac/med/beme> (accessed 19 Apr 2011).
18. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol* 1995;50:741–9.
19. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006;119:166.
20. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ* 2003;37:830–7.
21. Kirkpatrick D. Evaluation of training. In: Craig R, Mittel I, eds. *Training and development handbook*. New York, NY: McGraw-Hill, 1967: 87–112.
22. Epstein RM. Assessment in medical education. *N Engl J Med* 2007;356:387–96.
23. Madigosky WS, Headrick LA, Nelson K, et al. Changing and sustaining medical students' knowledge, skills, and attitudes about patient safety and medical fallibility. *Acad Med* 2006;81:94–101.
24. Hall LW, Scott SD, Cox KR, et al. Effectiveness of patient safety training in equipping medical students to recognise safety hazards and propose robust interventions. *Qual Saf Health Care* 2010;19:3–8.
25. Kerfoot BP, Conlin PR, Trivison T, et al. Web-based education in systems-based practice: a randomized trial. *Arch Intern Med* 2007;167:361–6.
26. Kerfoot BP, Conlin PR, Trivison T, et al. Patient safety knowledge and its determinants in medical trainees. *J Gen Intern Med* 2007;22:1150–4.
27. Wallace LM, Spurgeon P, Adams S, et al. Survey evaluation of the National Patient Safety Agency's Root Cause Analysis training programme in England and Wales: knowledge, beliefs and reported practices. *Qual Saf Health Care* 2009;18:288–91.
28. Harik P, Cuddy MM, O'Donovan S, et al. Assessing potentially dangerous medical actions with the computer-based case simulation portion of the USMLE step 3 examination. *Acad Med* 2009;84(10 Suppl):S79–82.
29. Floreck LM, Guernsey MJ, Clyman SG, et al. Examinee performance on computer-based case simulations as part of the USMLE step 3 examination: are examinees ordering dangerous actions? *Acad Med* 2002;77(10 Suppl):S77–9.
30. Klamen DL, Reynolds KL, Yale B, et al. Students learning handovers in a simulated in-patient unit. *Med Educ* 2009;43:1097–8.
31. Varkey P, Natt N, Lesnick T, et al. Validity evidence for an OSCE to assess competency in systems-based practice and practice-based learning and improvement: a preliminary investigation. *Acad Med* 2008;83:775–80.
32. Gupta P, Varkey P. Developing a tool for assessing competency in root cause analysis. *Jt Comm J Qual Patient Saf* 2009;35:36–42.
33. Singh R, Singh A, Fish R, et al. A patient safety objective structured clinical examination. *J Patient Saf* 2009;5:55–60.
34. Wagner DP, Hoppe RB, Lee CP. The patient safety OSCE for PGY-1 residents: a centralized response to the challenge of culture change. *Teach Learn Med* 2009;21:8–14.
35. Zausig YA, Grube C, Boeker-Blum T, et al. Inefficacy of simulator-based training on anaesthesiologists' non-technical skills. *Acta Anaesthesiol Scand* 2009;53:611–19.
36. Savoldelli GL, Naik VN, Park J, et al. Value of debriefing during simulated crisis management: oral versus video-assisted oral feedback. *Anesthesiology* 2006;105:279–85.
37. Yee B, Naik VN, Joo HS, et al. Nontechnical skills in anesthesia crisis management with repeated exposure to simulation-based education. *Anesthesiology* 2005;103:241–8.
38. Welke TM, LeBlanc VR, Savoldelli GL, et al. Personalized oral debriefing versus standardized multimedia instruction after patient crisis simulation. *Anesth Analg* 2009;109:183–9.
39. Müller MP, Hänsel M, Fichtner A, et al. Excellence in performance and stress reduction during two different full scale simulator training courses: a pilot study. *Resuscitation* 2009;80:919–24.
40. Fletcher G, Flin R, McGeorge P, et al. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth* 2003;90:580–8.
41. Bruppacher HR, Alam SK, LeBlanc VR, et al. Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery. *Anesthesiology* 2010;112:985–92.
42. Weller JM, Jolly B, Robinson B. Generalizability of behavioral skills in simulated anesthetic emergencies. *Anaesth Intensive Care* 2008;36:185–9.
43. Gale TC, Roberts MJ, Sice PJ, et al. Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. *Br J Anaesth* 2010;105:603–9.
44. Yule S, Flin R, Paterson-Brown S, et al. Development of a rating system for surgeons' non-technical skills. *Med Educ* 2006;40:1098–104.
45. Black SA, Nestel DF, Kneebone RL, et al. Assessment of surgical competence at carotid endarterectomy under local anesthesia in a simulated operating theatre. *Br J Surg* 2010;97:511–16.
46. Moorthy K, Munz Y, Forrest D, et al. Surgical crisis management skills training and assessment: a stimulation-based approach to enhancing operating room performance. *Ann Surg* 2006;244:139–47.
47. Moorthy K, Munz Y, Adams S, et al. A human factors analysis of technical and team skills among surgical trainees during procedural simulations in a simulated operating theatre. *Ann Surg* 2005;242:631–9.
48. Undre S, Koutantji M, Sevdalis N, et al. Multidisciplinary crisis simulations: the way forward for training surgical teams. *World J Surg* 2007;31:1843–53.
49. Powers KA, Rehrig ST, Irias N, et al. Simulated laparoscopic operating room crisis: an approach to enhance the surgical team performance. *Surg Endosc* 2008;22:885–900.
50. Papaspyros SC, Javangula KC, O'Regan DJ. Surgical training in the 48-h week: a novel simulation and educational tool. From amateur golfer to professional pilot. *Eur J Cardiothorac Surg* 2009;36:511–15.
51. Wetzel CM, Black SA, Hanna GB, et al. The effects of stress and coping on surgical performance during simulations. *Ann Surg* 2010;251:171–6.
52. Wallin CJ, Meurling L, Hedman L, et al. Target-focused medical emergency team training using a human patient simulator: effects on behaviour and attitude. *Med Educ* 2007;41:173–80.
53. Knudson MM, Khaw L, Bullard MK, et al. Trauma training in simulation: translating skills from SIM time to real time. *J Trauma* 2008;64:255–63.
54. Williams JB, McDonough MA, Hilliard MW, et al. Intermethod reliability of real-time versus delayed videotaped evaluation of a high-fidelity medical simulation septic shock scenario. *Acad Emerg Med* 2009;16:887–93.
55. Wright MC, Phillips-Bute BG, Petrusa ER, et al. Assessing teamwork in medical education and practice: relating behavioural teamwork ratings and clinical performance. *Med Teach* 2009;31:30–8.
56. Westli HK, Johnsen BH, Eid J, et al. Teamwork skills, shared mental models, and performance in simulated trauma teams: an independent group design. *Scand J Trauma Resusc Emerg Med* 2010;18:47.
57. Kim J, Neilipovitz D, Cardinal P, et al. A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated emergencies (abbreviated as 'CRM simulator study IB'). *Simul Healthc* 2009;4:6–16.
58. Kim J, Neilipovitz D, Cardinal P, et al. A pilot study using high-fidelity simulation to formally evaluate performance in the resuscitation of critically ill patients: the University of Ottawa Critical Care Medicine, High-Fidelity Simulation, and Crisis Resource Management I Study. *Crit Care Med* 2006;34:2167–74.
59. Thomas EJ, Taggart B, Crandell S, et al. Teaching teamwork during the Neonatal Resuscitation Program: a randomized trial. *J Perinatol* 2007;27:409–14.
60. Walker S, Lambden S, McKay A, et al. Development, reliability, and content validation of the observational skill-based clinical assessment for resuscitation (OSCAR). *23rd Annual Meeting of the European Society of Intensive Care Medicine*; 9–13 October 2010, Springer, 2010;36:S328.
61. Haller G, Garnerin P, Morales MA, et al. Effect of crew resource management training in a multidisciplinary obstetrical setting. *Int J Qual Health Care* 2008;20:254–63.
62. Dycus P, McKeon L. Using QSEN to measure quality and safety knowledge, skills, and attitudes of experienced paediatric oncology nurses: an international study. *Qual Manag Health Care* 2009;18:202–8.
63. Schnell R, Stone P, Currie L, et al. Development of a self-report instrument to measure patient safety attitudes, skills, and knowledge. *J Nurs Scholarsh* 2008;40:391–4.
64. Murray ME, Douglas S, Girdley D, et al. Teaching quality improvement. *J Nurs Educ* 2010;49:466–9.
65. Cooper S, Kinsman L, Buyck P, et al. Managing the deteriorating patient in a simulated environment: nursing students' knowledge, skill and situation awareness. *J Clin Nurs* 2010;19:2309–18.
66. Radhakrishnan K, Roche JP, Cunningham H. Measuring clinical practice parameters with human patient simulation: a pilot study. *Int J Nurs Educ Scholarsh* 2007;4:Article8.

67. Ironside PM, Jeffries PR, Martin A. Fostering patient safety competencies using multiple-patient simulation experiences. *Nurs Outlook* 2009;57:332–7.
68. McKeon LM, Norris T, Cardell B, *et al.* Developing patient-centered care competencies among prelicensure nursing students using simulation. *J Nurs Educ* 2009;48:711–15.
69. Walsh T, Jairath N, Paterson MA, *et al.* Quality and safety education for nurses clinical evaluation tool. *J Nurs Educ* 2010;49:517–22.
70. Fisher M, Parolin M. The reliability of measuring nursing clinical performance using a competency based assessment tool: a pilot study. *Collegian* 2000;7:21–7.
71. Patey RE, Fletcher G, Flin R, *et al.* Usability of the ANTS system. *Anesth Analg* 2004;98:S33.
72. Galbraith RM, Holtman MC, Clyman SG. Use of assessment to reinforce patient safety as a habit. *Qual Saf Health Care* 2006;15 Suppl 1:i30–3.
73. Norcini JJ. Current perspectives in assessment for assessing practice performance at work. *Med Educ* 2005;39:880–9.
74. Yule S, Rowley D, Flin R, *et al.* Experience matters: comparing novice and expert ratings of non-technical skills using the NOTSS system. *ANZ J Surg* 2009;79:154–60.
75. Graham J, Hocking G, Giles E. Anaesthesia non-technical skills: can anaesthetists be trained to reliably use this behavioural marker system in 1 day? *Br J Anaesth* 2010;104:440–5.
76. Wong BM, Etchells EE, Kuper A, *et al.* Teaching quality improvement and patient safety to trainees: a systematic review. *Acad Med* 2010;85:1425–39.
77. Vleuten CP. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996;1:41–67.
78. Martin JA, Regehr G, Reznick R, *et al.* Objective Structured Assessment of Technical Skills (OSATS) for surgical residents. *Br J Surg* 1997;84:273–8.
79. Mishara A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Qual Saf Health Care* 2009;18:104–8.