Assessment Tools for Feedback and Entrustment Decisions in the Clinical Workplace: A Systematic Review

Chantal C.M.A. Duijn ■ Emma J. van Dijk ■ Mira Mandoki ■ Harold G.J. Bok ■ Olle Th.J. ten Cate

ABSTRACT

BACKGROUND: Entrustable Professional Activities (EPAs) combine feedback and evaluation with a permission to act under a specified level of supervision and the possibility to schedule learners for clinical service. This literature review aims to identify workplace-based assessment tools that indicate progression toward unsupervised practice, suitable for entrustment decisions and feedback to learners. **METHODS:** A systematic search was performed in the PubMed, Embase, ERIC, and PsycINFO databases. Based on title/abstract and full text, articles were selected using predetermined inclusion and exclusion criteria. Information on workplace-based assessment tools was extracted using data coding sheets. The methodological quality of studies was assessed using the medical education research study quality instrument (MERSQI). **RESULTS:** The search yielded 6,371 articles (180 were evaluated in full text). In total, 80 articles were included, identifying 67 assessment tools. Only a few studies explicitly mentioned assessment tools used as a resource for entrustment decisions. Validity evidence was frequently reported, and the MERSQI score was 10.0 on average. **CONCLUSIONS:** Many workplace-based assessment tools were identified that potentially support learners with feedback on their development and support supervisors with providing feedback. As expected, only few articles referred to entrustment decisions. Nevertheless, the existing tools or the principals could be used for entrustment decisions, supervision level, or autonomy.

Key words: assessment, clinical science, communications, competency/competencies, educational methods, medicine, problem-based learning, veterinary teaching hospital

INTRODUCTION

Evaluating learners' clinical competence and skills is challenging for educators. Better patient care, caused by better acquisition of clinical skills by learners is associated with enhanced supervision in competency-based veterinary and medical education.[1] Supervision and clinical education are closely linked. Learner improvement in expertise depends not only on adequate teaching, but also on accurate and detailed assessment and feedback, which are manifestations of supervision.[2] The use of entrustable professional activities (EPAs) can help to improve learner assessment within competency-based education through entrustment decision making.[3,4]

The shift to competency-based education has challenged educators in veterinary medicine and medicine[5–7] to develop new methods of teaching and assessing clinical competence. The ultimate purpose of assessment is typically not to know how learners have acted in the past, but to predict how they will act in the near future.[8] A summative entrustment decision for an EPA should lead to acknowledged transfer of responsibilities to a learner from then on, which can be viewed as a certification or a license to act unsupervised.[4,8] Such a summative entrustment decision requires a grounded entrustment process, based on information collected in assessment forms over time.

The emergent concept of EPAs[4,9,10] has stimulated an interest in tools indicating whether learners have the ability to provide care unsupervised—that is, to be trusted with the tasks of the profession.[11] Entrustment decisions combine feedback and evaluation with permission to act under a specified level of supervision and the possibility to schedule learners for clinical duties (e.g., performing physical examination, venipuncture or even performing surgery without the supervisor in the same room).[4,12]

Clinical learning environments in medical education are in many respects similar to those in veterinary education: learners participate and learn in authentic clinical settings with patients, under (in)direct supervision.[13] To optimize both entrustment decision making and feedback to learners in veterinary and medical education to prepare them for execution of critical tasks, it is useful to explore the assessment tools in the workplace that are most informative for feedback and decision making.

In a 2009 systematic review by Kogan et al.,[14] direct observation tools to assess medical learners' clinical skills with actual patients were identified and summarized.[13] Although many tools were identified, validity evidence appeared to be scarce. Furthermore, the educational outcome of these tools appeared to be poorly described in the

available literature. To our knowledge, a systematic review of clinical assessment tools suitable for making entrustment decisions in veterinary and medical education, including indirect assessment or assessment without actual patients, has not yet been performed. This literature review aims to expand the current literature by describing workplace-based assessment tools that are indicators for progression toward unsupervised practice and clinical competence, suitable for feedback and entrustment decision making.

METHODS

A research protocol was prepared using methodology recommended by the Best Evidence Medical Education (BEME) collaboration.[15] The research proposal was discussed within the research team, and minor alterations were made until consensus was achieved.

Review Question

The primary research question for this systematic review was, "What assessment tools are described in literature that can indicate progression toward unsupervised clinical practice, and are suitable for entrustment decisions and feedback to learners in medical and veterinary clinical education?"

Framework

In a recent comprehensive article on entrustable professional activities (EPAs), a preliminary overview of workplace-based assessment tools and approaches to entrustment decisions distinguished the following six categories of assessment tools:[4]

1. Knowledge tests (such as written); not applicable for this review
2. Simulation tests (such as with standardized patients [SPs] and technology)
3. Short-practice observations (such as Mini-Clinical Evaluation Exercises [Mini-CEX])
4. Long-practice observation (such as Multi-Source Feedback ([MSF])
5. Case-based discussion (such as exploring adaptive competence)
6. Product evaluation (such as medical records)

This framework was used to categorize the results of our literature review.

Search Strategy

A systematic literature search was conducted on December 21, 2015, using the PubMed, Embase, Education Resource Information Center (ERIC), and PsycINFO electronic databases. The search query was developed by the research team with assistance of a health sciences librarian. A pilot search, including an initial electronic search followed by manual cross-reference checking and reviewing of 10 articles, was performed by the first and second author (CD, EvD) to evaluate and refine the search query and research protocol. The final search comprised terms relating to veterinary and medical education, clinical workplace, feedback, and entrustment decisions. The search included all studies that were written in the English language and were published after January 1, 2000. The latter was decided because it was expected that the relevant assessment tools for current education have been incorporated in the more recent tools and to limit the number of results. The complete search strategies for the different databases are presented in Table S3 (Tables S3–S7 are available at https://utpjournals.press/doi/suppl/10.3138/jvme.0917-123r.).

Study Selection

All identified studies were collected and organized in the RefWorks reference manager database. Titles and abstracts were screened by the first and second author, discrepancies were reconciled by discussion. When an abstract was insufficient to determine study eligibility, the article was included for full-text review. All included articles were subsequently reviewed in full text by both the first and second author independently. Title and abstract screening and full-text reviews were based on predetermined inclusion criteria (Figure 1).

Moreover, only studies with an empirical study design describing an assessment tool that had been evaluated in the clinical workplace were included. Hesitations and disagreements about inclusion were discussed to achieve consensus. Subsequently, the reference lists of all included articles were screened to identify potentially relevant articles that had not been identified with the literature search. Reference lists of newly included articles were then screened until no more additional references were found.

Data Extraction and Assessment of Methodological Quality

To extract and code data from the included studies, all articles were independently reviewed by at least two of the four reviewing authors (EvD, CvD, HB, MM). Extracted information was recorded in data coding sheets that were developed specifically to correspond to the research question. Data coding sheets were collected to compare results between reviewers. Disagreements were discussed or evaluated by a third reviewer until consensus was achieved. Extracted information included name and description of the evaluated assessment tool, study design and study context, expected learning outcomes and reported validity evidence. Expected learning outcomes included skills the assessment tool intended to measure, whether the tool was used to provide feedback or formative assessment in a qualitative or quantitative way, and whether the assessment tool was used to make entrustment decisions or to measure the required level of supervision.

The following four sources of validity evidence[16,17] were evaluated:

- Content validity: the relationship between the test's content and the construct it is intended to measure
- Internal structure: acceptable reliability and factor structure, for example reported as measures of internal consistency, generalizability, inter-rater reliability
- Relationships to other variables: the relationship between scores and other variables relevant to the measured construct
- Response process: the relationship between thought processes of individual participants and the intended construct

Furthermore, the methodological quality of all studies was assessed using the medical education research study quality instrument (MERSQI). MERSQI[18] is a checklist

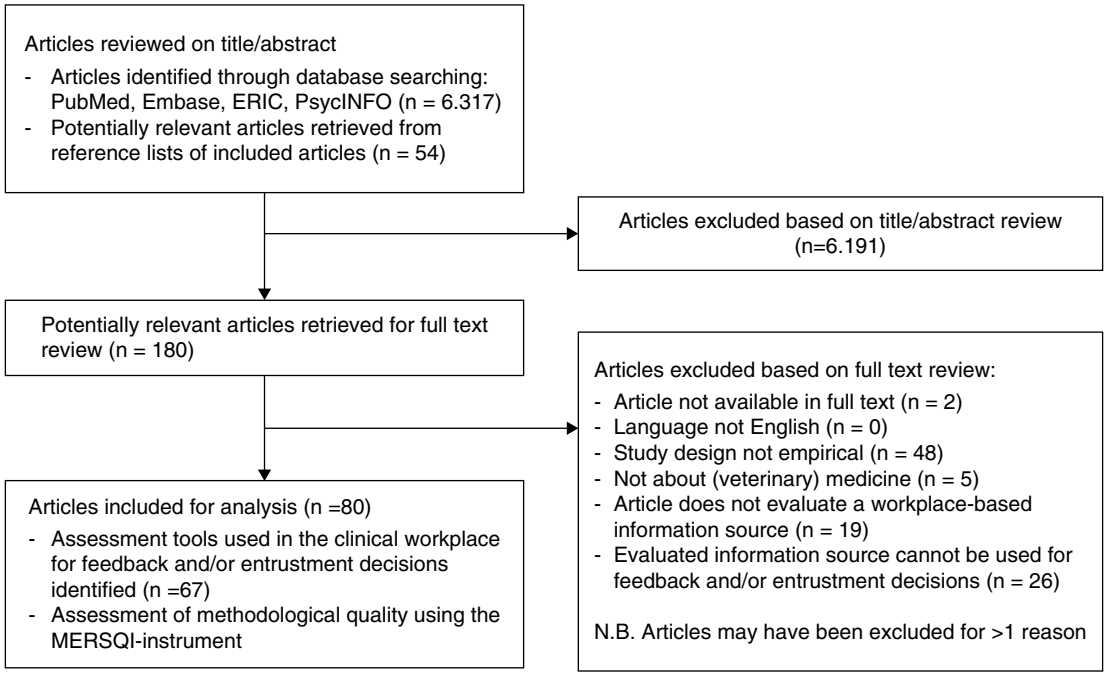

Figure 1: Flowchart of the search strategy

tool to evaluate the quality of medical education research (experimental, quasi-experimental, and observational studies) and cross-sectional studies. It includes 10 items, reflecting 6 domains of study quality: study design, sampling, type of data (subjective or objective), validity, data analysis, and outcomes. All items were scored on ordinal scales and were summed up to determine a total MERSQI score. The maximum score for each domain was 3, producing a maximum possible MERSQI score of 18 (potential range of 5 to 18). The total MERSQI score was calculated as the percentage of total achievable points (accounting for 'not applicable' responses) and then adjusted to a standard denominator of 18 to allow for comparison of scores across studies. To analyze and interpret the results, data were summarized in the described previously six categories.

RESULTS

The search yielded 6,371 articles; after title and abstract review for potential relevance, 180 articles were selected for full-text review. Two articles could not be retrieved in full text despite requests sent to the authors, and were subsequently excluded from the review. Eventually, 80 articles were found to meet the inclusion criteria, describing 67 unique tools in total. The flowchart in Figure 1 summarizes the overall review process.

Quality of Studies Included

All studies included had a MERSQI score of 6.0 or higher (N = 80, range 6 to 12.5) from a theoretical range of 5 to 18.[18] The majority of the studies had a MERSQI score of 9.0 or higher and the average MERSQI score was 10.01 with a standard deviation of 1.36 (average in theory 9.95, SD 2.34.[18]

Description of Articles and Tools

Most included studies had a prospective observational design (72 studies, 90.0%), 6 studies (7.5%) had a retrospective observational design and 2 studies (2.5%) were questionnaires. Fifty-three studies (66.2%) were conducted with postgraduates and 27 studies (33.8%) with under-graduate students. Almost all studies evaluated workplace-based assessment tools in medical education (77 studies, 96.2%), only 3 studies (3.8%) were conducted in veterinary education.

Of the 67 assessment tools that were identified, 43 were categorized as short-practice observations (64.2%), of which 16 tools were used to assess procedural or surgical skills and performance. Fourteen assessment tools (20.9%) were categorized as long-practice observations, 6 (9.0%) as product evaluations, 3 (4.5%) as simulation tests, and 1 (1.5%) as case-based discussion. No written or electronic knowledge tests were identified that were relevant to the research question. Details about the assessment tools are presented in Table 1 and Table 2. In Table S4A to Table S7 the assessment tools are discussed per category according to the previously described framework. Each category is explained in the following paragraphs.

Short-Practice Observations

A short-practice observation (SPO) usually takes 5 to 15 minutes and is focused on work in practice. It is documented with a judgment on paper or electronically, includes feedback afterwards and is meant to be conducted multiple times. An SPO can be applied to patient consultations, but can also be applied to many other snapshots of clinical activities.[19]

Table 1: Overview of workplace-based assessment tools identified in the literature

- Short-practice observations (Table S4A)
- Mini-Clinical Evaluation Exercise (Mini-CEX)[20–38]
- Handoff Mini-Clinical Evaluation Exercise (Handoff Mini-CEX)[39]
- Professionalism Mini-Evaluation Exercise (P-MEX)[40]
- Electronic problem-specific Clinical Evaluation Exercise (eCEX)[41]
- Just In Time Medicine (JIT) Clinical Evaluation Exercise[42]
- Ophthalmic Clinical Evaluation Exercise[43,44]
- Teamwork Mini-Clinical Evaluation Exercise[45]
- Personal Digital Assistant (PDA)-based Mini-Clinical Evaluation Exercise[46]
- Outpatient Clinic Evaluation (OPC)[57]
- Systematically Observed Medical Encounters[90]
- Mini-assessed Clinical Encounter[48]
- Assessment of Clinical Expertise[48]
- Case Presentation (CP)[48]
- Minicard[67]
- Direct Observation Clinical Encounter Examination (DOCEE)[84]
- Acute Care Assessment Tool (ACAT)[56]
- Direct Observation of Clinical Skills (DOCS)[68]
- Structured Clinical Observation (SCO)[96]
- Unnamed – Analysis of videotaped patient interactions[52]
- Davis Observation Code (DOC)[63]
- Case Conference Assessment Tool (cCAT)[71]
- Standardized Direct Observation Tool (SDOT)[73]
- Unnamed scoring form – global rating form[72]
- Unnamed – group assessment tool[91]
- Unnamed – one minute mentor[92]
- Long Case[77,78]
- Video Communication Assessment and Feedback (Video-CAF)[80]

Procedural short-practice observations (Table S4B)

- Procedure-Based Assessment (PBA)[49,86,95]
- Direct Observation of Procedural Skills (DOPS)[28,31,37,38,47,48]
- Direct Observation of Procedural Skills in Surgery (SDOPS)[95]
- Objective Structured Assessment of Technical Skills (OSATS)[85,86]
- Non-Technical Skills for Surgeons (NOTSS)[86]
- Structured Assessment of Microsurgical Skills (SAMS)[87]
- Hopkins Assessment of Surgical Competency (HASC)[88]
- Procedure Feedback Form (PFF)[98]
- Global Rating Index for Technical Skills (GRITS)[89]
- Zwisch Scale[50]
- Global Evaluative Assessment of Robotic Skills (GEARS)[83]
- Global Operative Assessment of Laparoscopic Skills[77,82]
- **Unnamed assessment tool to measure competence in video-assisted thoracoscopic surgery**[70]
- Operative Performance Rating Scale (OPRS)[51]
- Unnamed scoring form – checklist for task-specific items[72]

Long-practice observations (Table S5)

- Multi-Source Feedback (MSF)[38,52]
- Sheffield Peer Review Assessment Tool (SPRAT) (based on MSF)[53]
- Team Observation Tool (based on MSF)[54]
- Ward Evaluations (WE)[57]
- Preceptor Evaluations (PE)[57]
- Mini-Peer Assessment Tool[48]
- Patient Satisfaction Questionnaires (PSQ)[48]
- Patient Survey (PS)[56]
- **In-training Assessment (ITA)**[93]
- **Clinical Work Sampling (CWS) for In-Training Evaluation**[76]
- **McGill Electronic Evaluation Portfolio (MEEP)**[97]
- **Consultation And Relational Empathy (CARE)**[52]
- **Specifically, monthly performance evaluations**[63]
- **Ambulatory Care Competency Assessment Tool**[81]

Case-based discussion (Table S5)

- Case-based discussion[48,56]

Simulation tests (Table S6)

- Objective Structured Clinical Examination (OSCE)[57–65]
- Unnamed – based on OSCE[66]
- Integrated Procedural Performance Instrument[69]

Product evaluation (Table S7)

- Audit Assessment[56]
- Criterion Audit[53]
- Assessment of referral letters[52]
- Significant Event Analysis (SEA)[52]
- Clinical Learning Electronic Portfolio System[94]
- Vanderbilt Learning Portfolio (VLP)[75]

Table 2: Learner qualities assessed by tools reported in MERSQI ≥9.5 studies

Learner qualities/skills/ competencies evaluated	Simulation*	SPO†	Procedural SPO‡	LPO§	CBD¶	Product evaluation**
Interviewing skill	•	•		•		
Examination skill	•	•		•		
Therapeutic skill	•	•				
Diagnostic reasoning, decision making, clinical judgment	•	•		•	•	
Skill in communication, counseling, obtaining consent	•	•	•	•		
Team work (multi-disciplinary, multi-professional)	•			•		
Professionalism	•	•		•		
Knowledge	•					
Organization		•				
Collaboration		•				
Empathy				•		
Situation awareness			•			
Leadership			•			
Procedural skill	•		•			
Surgical skills and performance (micro, robotic, laparoscopic)			•			
Thoracoscopy skills			•			
Preoperative management	•					
Postoperative management	•					
Patient admission				•		
Data gathering		•				
Medical record keeping						•
Handover skill		•				
Creating referral letters						•
Oral case presentation		•				
Ambulatory care		•		•		
Time management				•		
Chairing case conference		•				
Critical appraisal				•		
Audit						•

MERSQI = medical education research study quality instrument; SPO = short-practice observation; LPO = long-practice observation; CBD = case-based discussion

* Simulation: OSCE and similar simulations

† Short-practice observation: Mini-CEX; P-MEX; Team Mini-CEX; MiniCard; DOPS; cCat and other tools

‡ Procedural SPO: mostly short-practice observation in the operating room

§ Long-practice observation: cumulative evaluations after training, including multi-source feedback tools

¶ Case-based discussion: with qualitative and quantitative reports

** Product evaluation: including portfolio evaluation

Forty-three unique assessment tools were identified in the short-practice observations category (Table S4A). The most evaluated workplace-based assessment tool was the Mini-Clinical Evaluation Exercise (Mini-CEX),[20–38] with several adjusted versions.[39–46] The most frequently assessed skills were interviewing, examination, communication, professionalism, organization, overall competence and various task-specific skills. Most feedback was quantitative (a numbered score), but qualitative written and oral feedback were also often described. Two studies evaluated the use of the Mini-CEX to make entrustment decisions by assessing the required level of supervision and learner independence/autonomy.[31,36]

Sixteen assessment tools specifically assessed procedural or surgical skills and competence Table S4B. Direct Observation of Procedural Skills (DOPS) was the most commonly studied assessment tool.[28,31,37,38,47,48] Procedural skills, communication, professionalism and operative performance were commonly assessed competencies. Several articles described assessment tools that assessed the required level of supervision, learner autonomy, or the amount of guidance or overall operative performance, each of which could be used as a resource to make entrustment decisions.[49–51]

Long-Practice Observations

Long-practice observations pertain to observed behavior over a longer period and usually focus on other qualities than those observed in SPOs. Observers are asked in advance to observe over a specified period (a long shift up to several weeks), allowing them to be alert when they see or interact with the learner.

Fourteen assessment tools were categorized as long-practice observations (Table S5). The MSF was described twice[38,52] and two modified versions of the MSF were evaluated.[53,54] The long-practice observations assess several competences, such as professionalism, clinical skills, overall performance, patient management and empathy. None of the articles evaluated an assessment tool in the context of making entrustment decisions.

Case-Based Discussion

A case-based discussion (CBD) is a short oral discussion with the learner on knowledge and clinical reasoning (10–15 min) after a clinical encounter or procedure.[55] It is prompted using two types of questions: (1) What have you done and why; what was your reasoning; and, possibly scenario-based; (2) What would you have done differently if this patient had shown X or Y; if there would have been an unexpected finding or if you would have encountered complication Z.

CBD was evaluated in two studies,[48,56] no additional assessment tools based on CBD were identified (Table S5). Three different skills were assessed with CBD, namely clinical competence,[48] clinical reasoning and clinical decision making.[56] The possibility of using CBD for entrustment decisions was not described. Feedback or formative assessment, both quantitative[48,56] and qualitative,[56] was provided in written format.

Simulation Tests

Skills testing in a simulated and standardized environment involve Objective Structured Clinical Evaluations (OSCEs) and similar examinations with low or high-fidelity equipment and with or without SPs.

Three simulation tests were identified (Table S6). The OSCE was the most studied,[57–65] with an adapted version for an orthopedic OSCE.[66] Simulation tests were used to assess a variety of skills, including interviewing, examination, procedural, communication, clinical decision-making and professionalism skills. Feedback was provided in both quantitative and qualitative ways. None of the assessment tools were explicitly used to make entrustment decisions or to determine the required level of supervision.

Product Evaluation

Products may include discharge summaries and letters, medication prescriptions and other entries into the electronic health record, presentations and case-reports.

Six assessment tools based on product evaluation were identified (Table S7). Both postgraduates and undergraduates in medical education were assessed on their medical records, on the quality of practice changes for patient care and on the quality of referral letters. None of the tools were used as a resource for making entrustment decisions.

Validity Evidence

Content validity was reported for 64 assessment tools (68.7%).[a] Descriptions of content selection usually comprised expert opinions or literature reviews evaluating if the educational competencies were an adequate reflection of the construct intended to be measured. Validity evidence for internal structure was described for 53 assessment tools (79.1%).[b] Most reported measures were generalizability,[c] internal consistency[d] and inter-rater reliability.[e] However, inter-item correlations,[f] item-total correlations,[77] variance,[g] factor analyses[22,26,40,53,74,88] and test-retest reliability[52] measures were also provided by numerous articles. Internal consistency was usually acceptable to high (Cronbach's $\alpha \geq 0.70$).

Information on relationships to other variables was provided for 44 assessment tools (65.7%).[h] Learner level scores of participant subgroups correlated with their level of experience[i] and concurrent validity (correlation between scores and outcomes from other, validated, assessment tools)[j] were the most reported measures. Several studies evaluated if confounding factors could be identified that influenced scores, such as race, gender and ethnicity.[k] Few studies provided measures of predictive validity (a correlation between scores and future assessments).[34,68,75]

Validity support in the form of evaluation of response process was described for 40 assessment tools (59.7%)[l] and typically involved training sessions to explain the use of assessment tools to raters or pilot studies to evaluate the clarity and practicality of the assessment tool.

DISCUSSION

This systematic literature review aimed to provide an overview of workplace-based assessment tools that have been used for feedback and summative entrustment decisions

in veterinary and medical education. A comprehensive literature search resulted in 80 articles describing 67 assessment tools. Various assessment tools, sufficient to provide learners with feedback or formative assessments, were identified. However, only few studies evaluated workplace-based assessment tools as a resource to support entrustment decision making. That is not surprising, as the terminology of entrustment decisions in clinical education has only recently been used in the literature. Nevertheless, a high number of studies was identified within the short-practice observations category that evaluated entrustment decision making in the operative or procedural setting. Kogan et al.[14] concluded in their systematic review on direct observation tools that few assessment tools had been profoundly evaluated and tested.[14] This systematic review showed literature increasingly providing validity evidence for many of the identified assessment tools. Some commonly used assessment tools, such as the Mini-CEX,[20–23,25–33,35,36,37] DOPS,[28,31,37,48] and the OSCE, [57–64] had validity evidence from multiple studies. Three different simulation tests were identified,[57–66,69] including OSCE.[57–65] Entrustment decisions pertain to actual practice, and it might confuse to include studies using simulations. However, in cases in which a program fails to provide authentic learning opportunities, or when certain clinical activities occur infrequently, simulation could be a reasonable alternative.

Technology in various forms, such as smartphone applications, online assessment forms and electronic portfolios, was incorporated in numerous assessment tools. We expect that technology will be increasingly used in the future for assessment of learners in the clinical workplace and that many developments in this area are yet to be made.

The set of qualities, skills and competencies (Table 1) as reported in literature might not be comprehensive, since it might be altered to use the tool for a different evaluation purpose. Conversely, many tools may be adequately suitable for learner qualities, but were not reported with that purpose. For example, case-based discussions and long-practice observations such as MSF are certainly suitable for evaluation of collaboration skill, while that was not explicitly reported in the included studies. The list can be regarded as a report of the minimum applicability of the assessment tools, where applicability in practice is likely to be wider. Only in a limited number of SPO tools references to entrustment decisions, required supervision and autonomy granted based on evaluation tools were included.[31,36,49,86]

In short, much is known about workplace-based assessment tools in the clinical workplace to provide learners with feedback or formative assessment, including validity evidence for a variety of available assessment tools. However, few studies have researched the application of these assessment tools for the purpose of making decisions about the required level of supervision. Future studies are needed to explore which assessment tools may be most feasible and valid for entrustment decision making.

A summative entrustment decision cannot be made based on a single information source at a single moment of evaluation, aligning with current thinking about programmatic assessment.[99] Earlier studies have investigated the qualities needed to trust a learner performing a critical task unsupervised.[100–103] To enable sufficient trust to perform a task unsupervised, learners need to demonstrate these qualities. This cannot be credited to a single quality, but requires a multitude of factors and multiple assessment moments.[3,102] Many workplace-based assessment tools were identified that potentially can support learners with feedback on their development and can support supervisors with providing feedback. As expected, few studies have researched the application of these assessment tools for the purpose of making decisions about the required level of supervision. This area of interest has only recently received attention in the literature.

A strength of this study is that more than 6000 articles were initially identified for review. The research group was international and comprised researchers from both medical and veterinary education, resulting in input from various viewpoints. The methodology applied to conduct this systematic review was based on BEME guidelines[15] and included both a pilot literature search and a pilot review, resulting in a more valid search and data extraction method. We also added MERSQI quality evaluation scores.

However, this review should be considered in the light of its limitations. First, many cross-references were identified, indicating that additional, relevant articles might have been missed. Because of the broad research question, a primary literature search had yielded more than 150,000 articles. As a result, more specific search terms had to be created. However, manual screening of the reference lists of newly included articles only yielded three additional articles. When these articles were subsequently screened for cross-references, no additional articles were identified, indicating that we expect that only few articles were missed. Second, publication bias may be of concern. It is conceivable that articles evaluating assessment tools with poor validity outcomes have not been published. The reported validity evidence for an assessment tool may subsequently be an overestimation of quality.

Third, a potential limitation might be the difference between veterinary and medical education. This review included studies in the fields of both veterinary and medical education because the learning environments are very similar: learners participate and learn in authentic clinical settings with patients under direct or indirect supervision. However, only a limited number of studies evaluated assessment tools used in veterinary education. The identified assessment tools might not be applicable in both fields of education in an identical way. Our findings suggest several next steps in workplace-based assessment to improve the quality of feedback and the use of entrustment decisions in veterinary and medical education. Furthermore, we cannot exclude that literature from dentistry, pharmacy or nursing would have resulted in different assessment tools.

Although quite some assessment tools were used for implementation in multispecialty settings, making them useful for quite some different clerkships, it could be necessary to adapt tools to make them effective in different contexts. Further research should investigate factors for

successful implementation of various assessment tools in different curricula. This might lead to valid indicators of progression toward unsupervised practice and might enhance the learning environment and patient care.

Lastly, based on the assigned MERSQI-scores we conclude that most included studies were of reasonable methodological quality.[18,104] Some studies had limitations, such as a single-center design, low response rates and lack of validity evidence. Most of the studies scored poorly on the MERSQI components 'criteria outcomes' and 'type of data', because assessment was usually performed by study participants and outcomes were usually knowledge, skills, perceptions or opinions.

CONCLUSIONS

This review identified assessment tools used in the clinical workplace that can support learners with feedback on their development and that can support supervisors with making entrustment decisions. Although many assessment tools were described for providing feedback or formative assessment, little is known about the application of these assessment tools for the purpose of making summative entrustment decisions, assessing the required level of supervision or autonomy. Nevertheless, the several existing assessment tools have the opportunity to support the entrustment decision process.

ACKNOWLEDGMENTS

We wish to thank Bianca Kramer, librarian at University Medical Center Utrecht, for her assistance in defining the search strategy. Furthermore, we would like to thank Dr. Carrie Chen for reviewing the final version of this manuscript.

CONFLICT OF INTEREST

The authors report no conflicts of interest. Only the authors are responsible for the purpose and writing of this article. Initially, the manuscript was based on a report drafted for the Workplace-based e-Assessment Technology for Competency-based Higher Multi-professional Education (WATCHME) project, a multi-professional, multi-institutional, multi-country project that has received funding from the European Union's Seventh Framework Program for research, technological development and demonstration, under grant agreement 619349.

NOTES

a References 20,22,23,25–28,30,32,35–40,43–45,47,49,53,54,56–58,60–62,64–89
b References 20–23,25–33,35,36,38–40,42,44–46,48–54,56–62,65,66,70,71,73–80,83–94
c References 20–22,26,29–32,35,36,38–40,45,48,49,52,56,57,61,62,78,84,86,90,93
d References 22,23,25,28,29,31,35,39,44–46,51,52,58,60,61,65,66,70,71,73,74,76,77,79,83,84,86–89,92
e References 27,30,39,42,44,50,52,54,62,65,66,71,75,76,80,83,91,92,94
f References 33,35,36,45,46,49–51,54,66,73,74,76,77,83–85,88
g References 20,21,31–33,35,38,40,45,49,76,79,80
h References 20,22,23,25,26,28,29,31,33,34,36–38,41,49–53,56–58,60,63,64,66–68,70,71,73,77,79,80,82–90,92,94,95
i References 20,22,29,31,37,49–51,53,60,66,67,70,77,80,82,83,85,87–89,95
j References 23,24,26,28,29,31,34,42,48,50,57,58,63,64,66,68,79,84,86,90,92
k References 29,32,34,38,49,52,53,71,73,80,83
l References 21,26,28,30,34,39–41,44–46,48,49,52,53,56,60–62,65,67–71,77,80,83,86,92,96,97

REFERENCES

1 Duffy FD, Gordon GH, Whelan G, et al.; Participants in the American Academy on Physician and Patient's Conference on Education and Evaluation of Competence in Communication and Interpersonal Skills. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. Acad Med. 2004;79(6):495–507. https://doi.org/10.1097/00001888-200406000-00002. Medline:15165967

2 Dornan T, Tan N, Boshuizen H, et al. How and what do medical students learn in clerkships? Experience based learning (ExBL). Adv Health Sci Educ Theory Pract. 2014;19(5):721–49. https://doi.org/10.1007/s10459-014-9501-0. Medline:24638146

3 Hauer K. Evaluating clinical trainees in the workplace: on supervision, trust and the role of competency committees [dissertation]. Utrecht, The Netherlands: Utrecht University; 2015.

4 ten Cate O, Chen HC, Hoff RG, et al. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. Med Teach. 2015;37(11):983–1002. https://doi.org/10.3109/0142159X.2015.1060308. Medline:26172347

5 Bok HG, Jaarsma DA, Teunissen PW, et al. Development and validation of a competency framework for veterinarians. J Vet Med Educ. 2011;38(3):262–9. https://doi.org/10.3138/jvme.38.3.262. Medline:22023978

6 Carraccio C, Wolfsthal SD, Englander R, et al. Shifting paradigms: from Flexner to competencies. Acad Med. 2002;77(5):361–7. https://doi.org/10.1097/00001888-200205000-00003. Medline:12010689

7 Carraccio C, Englander R, Van Melle E, et al.; International Competency-Based Medical Education Collaborators. Advancing competency-based medical education: a charter for clinician-educators. Acad Med. 2016;91(5):645–9. https://doi.org/10.1097/ACM.0000000000001048. Medline:26675189

8 ten Cate O, Hart D, Ankel F, et al.; International Competency-Based Medical Education Collaborators. Entrustment decision making in clinical training. Acad Med. 2016;91(2):191–8. https://doi.org/10.1097/ACM.0000000000001044. Medline:26630606

9 ten Cate O. Entrustability of professional activities and competency-based training. Med Educ. 2005;39(12):1176–7. https://doi.org/10.1111/j.1365-2929.2005.02341.x. Medline:16313574

10 ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? Acad Med. 2007;82(6):542–7.

https://doi.org/10.1097/ACM.0b013e31805559c7. Medline:17525536
11 Rekman J, Gofton W, Dudek N, et al. Entrustability scales: outlining their usefulness for competency-based clinical assessment. Acad Med. 2016;91(2):186–90. https://doi.org/10.1097/ACM.0000000000001045. Medline:26630609
12 ten Cate O, Hart D, Ankel F, et al.; International Competency-Based Medical Education Collaborators. Entrustment decision making in clinical training. Acad Med. 2016;91(2):191–8. https://doi.org/10.1097/ACM.0000000000001044. Medline:26630606
13 Duijn CCMA, Welink LS, Mandoki M, et al. Am I ready for it? Students' perceptions of meaningful feedback on entrustable professional activities. Perspect Med Educ. 2017;6(4):256–64. https://doi.org/10.1007/s40037-017-0361-1. Medline:28577253
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(12):1316–26. https://doi.org/10.1001/jama.2009.1365. Medline:19773567
15 Hammick M, Dornan T, Steinert Y. Conducting a best evidence systematic review. Part 1: From idea to data coding. BEME Guide No. 13. Med Teach. 2010;32(1):3–15. https://doi.org/10.3109/01421590903414245. Medline:20095769
16 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;119(2):166.e7–e16. https://doi.org/10.1016/j.amjmed.2005.10.036. Medline:16443422
17 Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ. 2003;37(9):830–7. https://doi.org/10.1046/j.1365-2923.2003.01594.x. Medline:14506816
18 Reed DA, Cook DA, Beckman TJ, et al. Association between funding and quality of published medical education research. JAMA. 2007;298(9):1002–9. https://doi.org/10.1001/jama.298.9.1002. Medline:17785645
19 ten Cate TJ, Fluit CRMG. Richtlijn korte praktijkbeoordeling [Guideline for short practice observations]. Neth J Med Educ. 2011;29(5):s105–34. https://doi.org/10.1007/s12507-010-0101-6. Dutch.
20 Alves de Lima A, Barrero C, Baratta S, et al. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. Med Teach. 2007;29(8):785–90. https://doi.org/10.1080/01421590701352261. Medline:17917984
21 Alves de Lima A, Conde D, Costabel J, et al. A laboratory study on the reliability estimations of the mini-CEX. Adv Health Sci Educ Theory Pract. 2013;18(1):5–13. https://doi.org/10.1007/s10459-011-9343-y. Medline:22193944
22 Cook DA, Beckman TJ, Mandrekar JN, et al. Internal structure of mini-CEX scores for internal medicine residents: factor analysis and generalizability. Adv Health Sci Educ Theory Pract. 2010;15(5):633–45. https://doi.org/10.1007/s10459-010-9224-9. Medline:21120648
23 Durning SJ, Cation LJ, Markert RJ, et al. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med. 2002;77(9):900–4. https://doi.org/10.1097/00001888-200209000-00020. Medline:12228088
24 Goel A, Singh T. The usefulness of Mini Clinical Evaluation Exercise as a learning tool in different pediatric clinical settings. Int J Appl Basic Med Res. 2015;5(Suppl 1):S32–4. https://doi.org/10.4103/2229-516X.162266. Medline:26380207
25 Hatala R, Ainslie M, Kassen BO, et al. Assessing the mini-Clinical Evaluation Exercise in comparison to a national specialty examination. Med Educ. 2006;40(10):950–6. https://doi.org/10.1111/j.1365-2929.2006.02566.x. Medline:16987184
26 Hill F, Kendall K, Galbraith K, et al. Implementing the undergraduate mini-CEX: a tailored approach at Southampton University. Med Educ. 2009;43(4):326–34. https://doi.org/10.1111/j.1365-2923.2008.03275.x. Medline:19335574
27 Holmboe ES, Huot S, Chung J, et al. Construct validity of the miniclinical evaluation exercise (miniCEX). Acad Med. 2003;78(8):826–30. https://doi.org/10.1097/00001888-200308000-00018. Medline:12915378
28 Karanth K.L., Kanagasabai S., Kader IS, Najimuddin M., Upali MD, De S. Structured program for final-year undergraduate students to improve clinical skills to prepare for effective patient management [Internet]. Internet J Gynecol Obstet. 2015;19(1). Available from: http://ispub.com/IJGO/19/1/28355.
29 Kogan JR, Bellini LM, Shea JA. Feasibility, reliability, and validity of the mini-clinical evaluation exercise (mCEX) in a medicine core clerkship. Acad Med. 2003;78(10 Suppl):S33–5. https://doi.org/10.1097/00001888-200310001-00011. Medline:14557089
30 Margolis MJ, Clauser BE, Cuddy MM, et al. Use of the mini-clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: a validity study. Acad Med. 2006;81(10 Suppl):S56–60. https://doi.org/10.1097/01.ACM.0000236514.53194.f4. Medline:17001137
31 Moonen-van Loon JM, Overeem K, Donkers HH, et al. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. Adv Health Sci Educ Theory Pract. 2013;18(5):1087–102. https://doi.org/10.1007/s10459-013-9450-z. Medline:23494202
32 Nair BR, Alexander HG, McGrath BP, et al. The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. Med J Aust. 2008;189(3):159–61. Medline:18673104
33 Norcini JJ, Blank LL, Duffy FD, et al. The mini-CEX: a method for assessing clinical skills. Ann Intern Med. 2003;138(6):476–81. https://doi.org/10.7326/0003-4819-138-6-200303180-00012. Medline:12639081
34 Rogausch A, Beyeler C, Montagne S, et al. The influence of students' prior clinical skills and context characteristics on mini-CEX scores in clerkships–a

multilevel analysis. BMC Med Educ. 2015;15(1):208. https://doi.org/10.1186/s12909-015-0490-3. Medline:26608836
35 Weller JM, Jolly B, Misur MP, et al. Mini-clinical evaluation exercise in anaesthesia training. Br J Anaesth. 2009;102(5):633–41. https://doi.org/10.1093/bja/aep055. Medline:19336537
36 Weller JM, Misur M, Nicolson S, et al. Can I leave the theatre? A key to more reliable workplace-based assessment. Br J Anaesth. 2014;112(6):1083–91. https://doi.org/10.1093/bja/aeu052. Medline:24638231
37 Wiles CM, Dawson K, Hughes TA, et al. Clinical skills evaluation of trainees in a neurology department. Clin Med (Lond). 2007;7(4):365–9. https://doi.org/10.7861/clinmedicine.7-4-365. Medline:17882854
38 Wilkinson JR, Crossley JG, Wragg A, et al. Implementing workplace-based assessment across the medical specialties in the United Kingdom. Med Educ. 2008;42(4):364–73. https://doi.org/10.1111/j.1365-2923.2008.03010.x. Medline:18338989
39 Arora VM, Berhie S, Horwitz LI, et al. Using standardized videos to validate a measure of handoff quality: the handoff mini-clinical examination exercise. J Hosp Med. 2014;9(7):441–6. https://doi.org/10.1002/jhm.2185. Medline:24665068
40 Cruess R, McIlroy JH, Cruess S, et al. The professionalism mini-evaluation exercise: A preliminary investigation. Acad Med. 2006;81(10 Suppl):S74–8. https://doi.org/10.1097/00001888-200610001-00019. Medline:17001141
41 Ferenchick GS, Foreback J, Towfiq B, et al. The implementation of a mobile problem-specific electronic CEX for assessing directly observed student-patient encounters [Internet]. Med Educ Online. 2010;15(1):Article 4276. https://doi.org/10.3402/meo.v15i0.4276. Medline:20174616
42 Ferenchick GS, Solomon D. Using cloud-based mobile technology for assessment of competencies among medical students. Peer J. 2013;1:e164. https://doi.org/10.7717/peerj.164. Medline:24109549
43 Golnik KC, Goldenhar LM, Gittinger JW Jr, et al. The ophthalmic clinical evaluation exercise (OCEX). Ophthalmology. 2004;111(7):1271–4. https://doi.org/10.1016/j.ophtha.2004.04.014. Medline:15234125
44 Golnik KC, Goldenhar L. The ophthalmic clinical evaluation exercise: reliability determination. Ophthalmology. 2005;112(10):1649–54. https://doi.org/10.1016/j.ophtha.2005.06.006. Medline:16111754
45 Olupeliyawa AM, O'Sullivan AJ, Hughes C, et al. The Teamwork Mini-Clinical Evaluation Exercise (T-MEX): a workplace-based assessment focusing on collaborative competencies in health care. Acad Med. 2014;89(2):359–65. https://doi.org/10.1097/ACM.0000000000000115. Medline:24362380
46 Torre DM, Simpson DE, Elnicki DM, et al. Feasibility, reliability and user satisfaction with a PDA-based mini-CEX to evaluate the clinical skills of third-year medical students. Teach Learn Med. 2007;19(3):271–7. https://doi.org/10.1080/10401330701366622. Medline:17594223
47 Bindal N, Goodyear H, Bindal T, et al. DOPS assessment: a study to evaluate the experience and opinions of trainees and assessors. Med Teach. 2013;35(6):e1230–4. https://doi.org/10.3109/0142159X.2012.746447. Medline:23627359
48 Brittlebank A, Archer J, Longson D, et al. Workplace-based assessments in psychiatry: evaluation of a whole assessment system. Acad Psychiatry. 2013;37(5):301–7. https://doi.org/10.1176/appi.ap.11110198. Medline:23771296
49 Marriott J, Purdie H, Crossley J, et al. Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. Br J Surg. 2011;98(3):450–7. https://doi.org/10.1002/bjs.7342. Medline:21254025
50 George BC, Teitelbaum EN, Meyerson SL, et al. Reliability, validity, and feasibility of the Zwisch scale for the assessment of intraoperative performance. J Surg Educ. 2014;71(6):e90–6. https://doi.org/10.1016/j.jsurg.2014.06.018. Medline:25192794
51 Larson JL, Williams RG, Ketchum J, et al. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. Surgery. 2005;138(4):640–9. https://doi.org/10.1016/j.surg.2005.07.017. Medline:16269292
52 Murphy DJ, Bruce DA, Mercer SW, et al. The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. Adv Health Sci Educ Theory Pract. 2009;14(2):219–32. https://doi.org/10.1007/s10459-008-9104-8. Medline:18306052
53 Archer J, McGraw M, Davies H. Assuring validity of multisource feedback in a national programme. Arch Dis Child. 2010;95(5):330–5. https://doi.org/10.1136/adc.2008.146209. Medline:20457700
54 Wood L, Wall D, Bullock A, et al. 'Team observation': a six-year study of the development and use of multi-source feedback (360-degree assessment) in obstetrics and gynaecology training in the UK. Med Teach. 2006;28(7):e177–84. https://doi.org/10.1080/01421590600834260. Medline:17594543
55 Setna Z, Jha V, Boursicot KA, et al. Evaluating the utility of workplace-based assessment tools for speciality training. Best Pract Res Clin Obstet Gynaecol. 2010;24(6):767–82. https://doi.org/10.1016/j.bpobgyn.2010.04.003. Medline:20598644
56 Johnson G, Booth J, Crossley J, et al. Assessing trainees in the workplace: results of a pilot study. Clin Med (Lond). 2011;11(1):48–53. https://doi.org/10.7861/clinmedicine.11-1-48. Medline:21404785
57 Auewarakul C, Downing SM, Jaturatamrong U, et al. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. Med Educ. 2005;39(3):276–83. https://doi.org/10.1111/j.1365-2929.2005.02090.x. Medline:15733163
58 Bark H, Cohen R. Use of an objective, structured clinical examination as a component of the final-year examination in small animal internal medicine and surgery. J Am Vet Med Assoc. 2002;221(9):1262–5.

https://doi.org/10.2460/javma.2002.221.1262. Medline:12418690
59 Bateman K, Menzies P, Sandals D, et al. Objective structured clinical examinations (OSCEs) as a summative evaluation tool in a ruminant health management rotation for final-year DVM students. J Vet Med Educ. 2008;35(3):382–8. https://doi.org/10.3138/jvme.35.3.382. Medline:19066355
60 Dwyer T, Theodoropoulos JS, Herold J, et al. Assessing competence of orthopaedic residents: the reliability and validity of an objective structured clinical examination after a sports medicine rotation. J Bone Joint Surg Am. 2013;95(22):e177. https://doi.org/10.2106/JBJS.M.00148. Medline:24257677
61 Hecker K, Read EK, Vallevand A, et al. Assessment of first-year veterinary students' clinical skills using objective structured clinical examinations. J Vet Med Educ. 2010;37(4):395–402. https://doi.org/10.3138/jvme.37.4.395. Medline:21135408
62 Lukas RV, Adesoye T, Smith S, et al. Student assessment by objective structured examination in a neurology clerkship. Neurology. 2012;79(7):681–5. https://doi.org/10.1212/WNL.0b013e3182648ba1. Medline:22855865
63 Nuovo J, Bertakis KD, Azari R. Assessing resident's knowledge and communication skills using four different evaluation tools. Med Educ. 2006;40(7):630–6. https://doi.org/10.1111/j.1365-2929.2006.02506.x. Medline:16836535
64 Varkey P, Natt N. The Objective Structured Clinical Examination as an educational tool in patient safety. Jt Comm J Qual Patient Saf. 2007;33(1):48–53. https://doi.org/10.1016/S1553-7250(07)33006-7. Medline:17283941
65 Wilkinson TJ, Newble DI, Wilson PD, et al. Development of a three-centre simultaneous objective structured clinical examination. Med Educ. 2000;34(10):798–807. https://doi.org/10.1046/j.1365-2923.2000.00669.x. Medline:11012928
66 Dwyer T, Wadey V, Archibald D, et al. Cognitive and psychomotor entrustable professional activities: Can simulators help assess competency in trainees? Clin Orthop Relat Res. 2016;474(4):926–34. https://doi.org/10.1007/s11999-015-4553-x. Medline:26394640
67 Donato AA, Park YS, George DL, et al. Validity and feasibility of the minicard direct observation tool in 1 training program. J Grad Med Educ. 2015;7(2):225–9. https://doi.org/10.4300/JGME-D-14-00532.1. Medline:26221439
68 Kang Y, Bardes CL, Gerber LM, et al. Pilot of direct observation of clinical skills (DOCS) in a medicine clerkship: feasibility and relationship to clinical performance measures. Med Educ Online. 2009;14(9):Article 4519. https://www.tandfonline.com/doi/abs/10.3402/meo.v14i.4519. Medline:20165523
69 Kneebone R, Nestel D, Yadollahi F, et al. Assessing procedural skills in context: exploring the feasibility of an Integrated Procedural Performance Instrument (IPPI). Med Educ. 2006;40(11):1105–14. https://doi.org/10.1111/j.1365-2929.2006.02612.x. Medline:17054620
70 Konge L, Lehnert P, Hansen HJ, et al. Reliable and valid assessment of performance in thoracoscopy. Surg Endosc. 2012;26(6):1624–8. https://doi.org/10.1007/s00464-011-2081-7. Medline:22179467
71 O'Connor RJ, Playford ED. The case Conference Assessment Tool (cCAT): a new workplace-based assessment. Clin Med (Lond). 2014;14(4):386–90. https://doi.org/10.7861/clinmedicine.14-4-386. Medline:25099840
72 Ringsted C, Østergaard D, Ravn L, et al. A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. Med Teach. 2003;25(6):654–8. https://doi.org/10.1080/01421590310001605642. Medline:15369915
73 Shayne P, Gallahue F, Rinnert S, et al, and the CORD SDOT Study Group. Reliability of a core competency checklist assessment in the emergency department: the Standardized Direct Observation Assessment Tool. Acad Emerg Med. 2006;13(7):727–32. https://doi.org/10.1111/j.1553-2712.2006.tb01711.x. Medline:16636361
74 Silber CG, Nasca TJ, Paskin DL, et al. Do global rating forms enable program directors to assess the ACGME competencies? Acad Med. 2004;79(6):549–56. https://journals.lww.com/academicmedicine/Fulltext/2004/06000/Do_Global_Rating_Forms_Enable_Program_Directors_to.10.aspx. Medline:15165974
75 Spickard A III, Ridinger H, Wrenn J, et al. Automatic scoring of medical students' clinical notes to monitor learning in the workplace. Med Teach. 2014;36(1):68–72. https://doi.org/10.3109/0142159X.2013.849801. Medline:24195470
76 Turnbull J, MacFadyen J, van Barneveld C, et al. Clinical work sampling: a new approach to the problem of in-training evaluation. J Gen Intern Med. 2000;15(8):556–61. https://doi.org/10.1046/j.1525-1497.2000.06099.x. Medline:10940147
77 Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg. 2005;190(1):107–13. https://www.sciencedirect.com/science/article/pii/S0002961005003478. Medline:15972181
78 Wass V, Jones R, Van der Vleuten C. Standardized or real patients to test clinical competence? The long case revisited. Med Educ. 2001;35(4):321–5. https://doi.org/10.1046/j.1365-2923.2001.00928.x. Medline:11318993
79 Wilkinson TJ, Campbell PJ, Judd SJ. Reliability of the long case. Med Educ. 2008;42(9):887–93. https://doi.org/10.1111/j.1365-2923.2008.03129.x. Medline:18715486
80 Wouda JC, van de Wiel HB. The effects of self-assessment and supervisor feedback on residents' patient-education competency using videoed outpatient consultations. Patient Educ Couns. 2014;97(1):59–66. https://doi.org/10.1016/j.pec.2014.05.023. Medline:24993839

81 Wendling AL. Assessing resident competency in an outpatient setting. Fam Med. 2004;36(3):178–84. Medline:14999574
82 Gumbs AA, Hogle NJ, Fowler DL. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. J Am Coll Surg. 2007;204(2):308–13. https://doi.org/10.1016/j.jamcollsurg.2006.11.010. Medline:17254935
83 Goh AC, Goldfarb DW, Sander JC, et al. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. J Urol. 2012;187(1):247–52. https://doi.org/10.1016/j.juro.2011.09.032. Medline:22099993
84 Hamdy H, Prasad K, Williams R, et al. Reliability and validity of the direct observation clinical encounter examination (DOCEE). Med Educ. 2003;37(3):205–12. https://doi.org/10.1046/j.1365-2923.2003.01438.x. Medline:12603759
85 Beard JD, Choksy S, Khan S; Vascular Society of Great Britain and Ireland. Assessment of operative competence during carotid endarterectomy. Br J Surg. 2007;94(6):726–30. https://doi.org/10.1002/bjs.5689. Medline:17315174
86 Beard JD, Marriott J, Purdie H, et al. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. Health Technol Assess. 2011;15(1):i–xxi. https://doi.org/10.3310/hta15010. Medline:21227024
87 Chan W, Niranjan N, Ramakrishnan V. Structured assessment of microsurgery skills in the clinical setting. J Plast Reconstr Aesthet Surg. 2010;63(8):1329–34. https://doi.org/10.1016/j.bjps.2009.06.024. Medline:19625227
88 Chou B, Bowen CW, Handa VL. Evaluating the competency of gynecology residents in the operating room: validation of a new assessment tool. Am J Obstet Gynecol. 2008;199(5):571.e1–5. https://www.sciencedirect.com/science/article/pii/S000293780800714X. Medline:18799153
89 Doyle JD, Webber EM, Sidhu RS. A universal global rating scale for the evaluation of technical skills in the operating room. Am J Surg. 2007;193(5):551–5. https://doi.org/10.1016/j.amjsurg.2007.02.003. Medline:17434353
90 Bergus GR, Woodhead JC, Kreiter CD. Using systematically observed clinical encounters (SOCEs) to assess medical students' skills in clinical settings. Adv Med Educ Pract. 2010;1:67–73. https://doi.org/10.2147/AMEP.S12962. Medline:23745065
91 Thomas MR, Beckman TJ, Mauck KF, et al. Group assessments of resident physicians improve reliability and decrease halo error. J Gen Intern Med. 2011;26(7):759–64. https://doi.org/10.1007/s11606-011-1670-4. Medline:21369769
92 Topps D, Evans RJ, Thistlethwaite JE, et al. The one minute mentor: a pilot study assessing medical students' and residents' professional behaviours through recordings of clinical preceptors' immediate feedback. Educ Health (Abingdon). 2009;22(1):189. Medline:19953438
93 Daelmans HE, van der Hem-Stokroos HH, Hoogenboom RJ, et al. Feasibility and reliability of an in-training assessment programme in an undergraduate clerkship. Med Educ. 2004;38(12):1270–7. https://doi.org/10.1111/j.1365-2929.2004.02019.x. Medline:15566538
94 Spickard A III, Gigante J, Stein G, et al. Automatic capture of student notes to augment mentor feedback and student performance on patient write-ups. J Gen Intern Med. 2008;23(7):979–84. https://doi.org/10.1007/s11606-008-0608-y. Medline:18612728
95 Ali AS, Bussey M, O'Flynn KJ, et al. Quality of feedback using workplace based assessments in urological training. Br J Med Surg Urol. 2012;5(1):39–43. https://doi.org/10.1016/j.bjmsu.2011.10.001.
96 Lane JL, Gottlieb RP. Structured clinical observations: a method to teach clinical skills with limited time and financial resources. Pediatrics. 2000;105(4 Pt 2):973–7. Medline:10742358
97 Duque G, Finkelstein A, Roberts A, et al, and the Members of the Division of Geriatric Medicine, McGill University. Learning while evaluating: the use of an electronic evaluation portfolio in a geriatric medicine clerkship. BMC Med Educ. 2006;6(1):4. https://doi.org/10.1186/1472-6920-6-4. Medline:16409640
98 Cook MR, Watters JM, Barton JS, et al. A flexible postoperative debriefing process can effectively provide formative resident feedback. J Am Coll Surg. 2015;220(5):959–67. https://doi.org/10.1016/j.jamcollsurg.2014.12.048. Medline:25907872
99 Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment: AMEE Guide No. 57. Med Teach. 2011;33(10):783–97. https://doi.org/10.3109/0142159X.2011.611022. Medline:21942477
100 Sterkenburg A, Barach P, Kalkman C, et al. When do supervising physicians decide to entrust residents with unsupervised tasks? Acad Med. 2010;85(9):1408–17. https://doi.org/10.1097/ACM.0b013e3181eab0ec. Medline:20736669
101 Kennedy TJ, Regehr G, Baker GR, et al. Point-of-care assessment of medical trainee competence for independent clinical work. Acad Med. 2008;83(10 Suppl):S89–92. https://doi.org/10.1097/ACM.0b013e318183c8b7. Medline:18820510
102 Wijnen-Meijer M, van der Schaaf M, Nillesen K, et al. Essential facets of competence that enable trust in graduates: a delphi study among physician educators in the Netherlands. J Grad Med Educ. 2013;5(1):46–53. https://doi.org/10.4300/JGME-D-11-00324.1. Medline:24404226
103 Choo KJ, Arora VM, Barach P, et al. How do supervising physicians decide to entrust residents with unsupervised tasks? A qualitative analysis. J Hosp Med. 2014;9(3):169–75. https://doi.org/10.1002/jhm.2150. Medline:24446195
104 Reed DA, Beckman TJ, Wright SM, et al. Predictive validity evidence for medical education research study quality instrument scores: quality of submissions to JGIM's Medical Education Special Issue. J Gen Intern Med. 2008;23(7):903–7. https://doi.org/10.1007/s11606-008-0664-3. Medline:18612715

AUTHOR INFORMATION

Chantal C.M.A. Duijn, MSc, is a large-animal veterinarian and a PhD student in veterinary medical education at the Centre for Quality Improvement in Veterinary Education, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands. Email: c.c.m.a.duijn@uu.nl.

Emma J. van Dijk, MD, MSc, is an ANIOS (medical doctor not in training) at the St Antonius Hospital, PO Box 2500, 3430 EM, Nieuwegein.

Mira Mandoki, PhD, is an Associate Professor at the Department of Pathology and Forensic Veterinary Medicine, Szent István University, P.O. Box 2, H-1400 Budapest, Hungary.

Harold G.J Bok, DVM, is a veterinarian and Assistant Professor at the Centre for Quality Improvement in Veterinary Education, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 1, 3584 CL Utrecht, The Netherlands. His interests include workplace-based learning and assessment, feedback, competency development, and (veterinary) educational research.

Olle Th.J. ten Cate, PhD, is a Professor of Medical Education and Director of the Center for Research and Development of Education at University Medical Center, Utrecht, The Netherlands; he is also an Adjunct Professor in the Department of Medicine, University of California, San Francisco, USA.