



Competency-Based Veterinary Education: Toolkit



CBVE: Toolkit — Introduction

Competency-based Veterinary Education (CBVE) focuses on learner-centered education with outcomes-based assessments. One of the most commonly asked questions about CBVE is: What evaluative or assessment tools should be used in various circumstances?

In the full version of the CBVE Toolkit (see: <https://cbve.org/assessment-toolkit>), collated informative summaries and references for 21 evaluative tools are provided. Each tool is described with pertinent

domains and competencies listed. Examples, documented uses, evidence for efficacy (pros and cons where available), and selected references are provided for each tool. When available, links to examples in use are provided.

For this publication, the abbreviated CBVE Toolkit includes only descriptions and pros/cons of each assessment tool.

Notes

CBVE: Toolkit – Summary Table

	COMPETENCIES ASSESSED		
	Clinical Knowledge “What You Know”	Critical Reasoning “How You Think”	
	Individual Animal Care, Animal Population Care, & Public Health	Clinical Reasoning & Decision-Making	Gathering & Evaluating Information
Domains of Competence	2, 3, 4	1	1, 2, 5
Entrustable Professional Activities (EPAs)	2, 3, 4, 6, 7, 8	2, 3, 4, 6, 7	2, 3
ASSESSMENT TYPE			
Written & Oral Examinations			
Multiple Choice Questions (MCQ)	X	X	
Extended Matching Questions (EMQ)	X	X	
Fill in the Blank (FITB)	X	X	
Short Answer Questions (SAQ)	X	X	
Essay Questions	X	X	
Script Concordance Testing (SCT)		X	
Oral Examinations	X	X	
Chart Stimulated Recall Examination (CSR)	X	X	
Key Features Examination		X	
Practical (Skills) Examinations			
In-Training Evaluation Report (ITER)	X	X	X
Case-Based Discussion	X	X	
Direct Observation of Procedural Skills (DOPS)			
Objective Structured Clinical Examination (OSCE)			
Clinical Evaluation Exercise (CEX)			
Mini-Clinical Evaluation Exercise (Mini-CEX)			
Longitudinal Evaluation of Performance (LEP)	X	X	X
360° Evaluations (Multi-Source Feedback)			
Portfolios	X	X	X
Entrustment-supervision scales	X	X	X
Student Assignments			
Capstone Assignments			X
Case Logs	X	X	X

This table reconciles the CBVE Toolkit assessments with domains and entrustable professional activities (EPAs). The reader can search for any of the nine domains of competence and the eight EPAs, and in that same column can then identify different assessment tools that are suggested to address those domains and EPAs. The individual tools suggested for assessment of each competency are described in greater detail in the CBVE Toolkit.

	Technical Skills “What You Can Do”	Professional Identity “How You Interact”		
	Medical, Surgical, & Anesthetic Procedures	Written Communication	Verbal Communication	Collegiality & Teamwork
	1, 2, 6, 7	5, 8, 9	5, 8, 9	6
	3, 6, 7	1, 3, 4, 5, 8	1, 3, 4, 5, 8	4, 7
		X		
		X		
		X		
		X	X	
			X	
			X	
		X	X	
	X	X	X	X
			X	
	X			
	X		X	
	X			
	X			
	X			
	X	X	X	
				X
		X		
	X	X	X	X
		X		
		X		

CBVE: Toolkit — Assessment Tools

WRITTEN & ORAL EXAMINATIONS:

Multiple Choice Questions

DESCRIPTION:

Multiple choice questions (MCQs) are commonly written as one-best-answer items. One-best-answer MCQs have a stem that is followed by a series of response options. The response options include one correct answer and a series of “distractors” that are incorrect. True-false items are another MCQ type. True-false MCQs have a lead-in question and a series of response options where the number of “true” responses varies from one to all of the set of responses.

PROS & CONS:

Positive Aspects/Pros:

- Well-written MCQs can produce high item discrimination and reliability, and so are commonly used for high-stakes licensing veterinary and medical licensing exams.
- MCQs can be used to broadly sample curricular content in assessment blueprinting.
- Depending on how they are targeted and written, MCQs can assess several levels of Bloom’s hierarchy of learning objectives in the cognitive domain. This enables assessment of a variety of foundational and clinical knowledge and problem-solving skills using fact-oriented or scenario-oriented assessment.
- Test-takers are generally familiar with the MCQ format.
- MCQs can be efficiently graded using computer-based testing and student performance statistics readily evaluated.

Negative Aspects/Cons:

- MCQs are not well suited to evaluating performance of procedural skills or performance in the workplace.
- Other forms of assessment are more effective for evaluating the psychomotor or attitudinal domains of Bloom’s taxonomy.
- Training and time are required to write high quality MCQs without flaws.

Extended Matching Questions

DESCRIPTION:

Extended matching questions (EMQs) are a type of selected-response (multiple choice) question. They are similar in format to simple one-best-answer multiple choice questions but differ in that they involve large potential option sets, with multiple question stems per item. There are two varieties of extended matching questions, including one-best-answer and pick-N-options (in which there are more than one correct answer) formats.

PROS & CONS:

As of 2021, no systematic reviews or meta-analyses of EMQs were found in the literature. However, available studies (Bhakta et al., 2005; Buellens et al., 2005; Swanson et al., 2008) and the prevalent use of EMQs in high-stakes professionally designed examinations suggest that EMQs demonstrate good psychometric properties and are considered valuable items in valid assessments of abilities such as clinical reasoning.

Positive Aspects/Pros:

- EMQs provide superior item discrimination when compared to one-best-answer selected-response questions.
- Like well-written one-best-answer selected-response questions, EMQs can produce high item discrimination and reliability when compared to many other item formats.
- EMQs are well suited to measuring conceptual knowledge, principles, and problem solving (e.g. diagnostic and clinical reasoning).

Negative Aspects/Cons:

- EMQs are not well suited to evaluation in workplace-based settings, measurement of psychomotor skills, memorization of verbal information, or evaluation of procedures.
- Training is required to write effective questions.
- Learners who are not familiar with the format may find it confusing.

CBVE: Toolkit — Assessment Tools

Fill in the Blank Questions

DESCRIPTION:

Fill in the blank (FITB) questions are a type of assessment best used when testing for one or several explicitly correct responses. Fill in the blank questions are best designed using clear, explicit instructions to specify the answer format and acceptable variation. They should be clearly worded using correct and neutral grammar to avoid extraneous clues and so that students understand the nature of information being requested. They should be phrased so that the answer is brief and specific. The blank to be completed should be at or near the end of the question. When a numeric response is required, it should specify the degree of precision expected or units of measurement. For best practices in scoring, the question should be phrased so that there is only one answer, or a limited range of possible answers and use pre-established scoring rubrics.

PROS & CONS:

Positive Aspects/Pros:

- Assesses cognitive domain for assessing who, what, where, and when information
- Limited to knowledge and comprehension questions
- Reduced rote memorization by not using direct quotes in the stem of the question
- Easy to administer
- Possible to identify questions on the entire curriculum
- Requires students to fill in the important term or phrase
- Promotes more in-depth study to recall answers
- Quicker for students to complete than multiple choice (consider having to read through all options prior to making an informed response)
- More comprehensive and reliable than essay questions
- Provides diagnostic information when looking at types of errors
- Improved reliability if there are structured marking schemes, clear outline answers, and independent double scoring used
- Less likely for scores to be influenced by guessing

Negative Aspects/Cons:

- Encourages rote memorization
- Unable to assess skills and attitude
- Inability to measure linguistic skill and power of expression
- Inability to measure higher mental faculties (e.g., logic)
- Inability to measure insight and foresight
- Handwriting and spelling skills may influence assessment.
- Difficult to write items that have only one clear answer
- Not suitable for item analysis
- Subjective scoring takes more time and is more difficult.

Short Answer Questions

DESCRIPTION:

Although some conflate short answer questions with fill in the blank (FITB) questions, the education literature is fairly clear that these are separate question types that assess different levels of thinking and learning. Short answer questions are open-ended questions requiring students to respond with a brief written answer, generally a paragraph or less but often restricted to one to three sentences. Because they require students to recall information rather than select from a list of potential responses, short answer questions are better for testing higher order thinking skills and separating out those students who have achieved deep learning from those who have obtained only surface level learning. In general, students should earn more credit for these types of questions as they are more demanding than those requiring recognition of answers (MCQs, FITB, T/F).

PROS & CONS:

Positive Aspects/Pros:

- Relies on recall vs recognition
- Differentiates deep vs surface learning of students
- Easier to write well than MCQs, FITB
- Good for testing higher order thinking skills

Negative Aspects/Cons:

- Less efficient than MCQ, FITB exams
- Cannot cover as much breadth of material due to length
- Poor validity and reliability unless multiple graders
- Can be demotivating to students

CBVE: Toolkit — Assessment Tools

Essay Questions

DESCRIPTION:

Essay questions allow for different or original responses or patterns of responses and elicits responses that must consist of more than one sentence. Essay questions also provide learners with an indication of the types of thinking and content to use in responding to the essay question and require learners to compose rather than select a response option. These types of questions require subjective judgment by a competent specialist to assess the accuracy and quality of responses and where double marking might be recommended to improve reliability.

PROS & CONS:

Positive Aspects/Pros:

- Can assess higher-order or critical thinking skills
- Can evaluate student thinking and reasoning
- Provides authentic experience closer to real practice
- Written feedback possible (+) but also time consuming (-)
- May provide practice to improve poor or unpolished writing

NEGATIVE ASPECTS/CONS:

- Assesses a limited sample of the range of content
- Difficult and time consuming to grade
- Reliability often low as sampling across content tends to be low, unless a large number of essays are used
- Labor intensive scoring
- The word “essay” can be confusing to students in their interpretation of the format.
- Not recommended for high-stakes assessment
- Techniques to detect plagiarism should be considered.

Script Concordance (SCT)

DESCRIPTION:

The SCT is a written test designed to evaluate examinees’ ability to interpret clinical information in ill-defined situations and then compares their judgment to experts. The test is based on illness script theory. Examinees are presented with a brief case vignette, followed by an initial hypothesis. The second step presents new information that may affect the likelihood of the hypothesis. In the third step, examinees are asked to indicate the effect of the new information on the original hypothesis. There is no single correct answer. Instead, scores are determined based on the proportion of experienced clinicians selecting a particular answer (Lubarsky et al., 2013; Ramaekers, 2010).

PROS & CONS:

Positive aspects/pros:

- Studied in many health professions
- Assesses a specific domain of clinical reasoning – ability to interpret medical information under ill-defined conditions
- Good construct validity
- Compares examinees to clinicians
- Written test – easy to administer, 60-90 minutes
- Requires students to apply their knowledge

Negative aspects/cons:

- Items difficult to construct
- Minimum of 25 cases with 3 items each for optimal reliability
- Requires a panel of experts (10-15) to review/score proposed questions
- Not useful for evaluating content-area knowledge
- Considerable concerns about test validity:
 - Process validity (concerns about whether examinees share the same view of constructs under examination as experts)
 - Aggregate scoring of SCT may not be valid.
 - Potential bias against examinees who select extreme options.

CBVE: Toolkit — Assessment Tools

Oral Examination

DESCRIPTION:

The oral examination, also referred to as *viva voce*, is the oldest form of examination (Stray, 2001). It is comprised of the examiner(s) verbally asking questions to which the examinee provides a verbal reply. Oral examinations may be either unstructured or structured. Structured oral examinations utilize a standard scenario, questions, and scoring methodology, whereas unstructured ones do not. Unstructured oral examinations in medical training were standard for many years but were discontinued by the National Board of Medical Examiners in 1963 when data demonstrated poor correlation (0.25) between examiners (Hodges, 2006).

When compared to unstructured oral examinations, structured oral examinations demonstrate improved reliability and inter-rater reliability (Anastakis et al., 1991; Jefferies et al., 2011). Structured oral examinations also had greater correlation to multiple-choice questions and objective structured clinical examinations when compared to unstructured oral examinations (Anastakis et al., 1991). Reliability is further increased by using more than one oral examination per examinee, use of several examiners when multiple oral examinations are given, standardization of questions, scoring with a rubric, and training the examiners (Daelmans et al., 2001; Davis and Karunathilake, 2005; Touchie et al., 2010; Wakeford et al., 1995).

Oral examinations can be constructed to evaluate critical thinking, reasoning skills, and higher order cognitive skills. However, studies evaluating oral examinations found that they frequently test at a lower cognitive level (i.e., knowledge/recall of information) rather than at the higher levels of understand, apply, analyze, synthesize, and evaluate (Davis and Karunathilake, 2005).

Unstructured oral examinations can be utilized as a formative assessment.

Structured oral examinations can be utilized either as formative or summative assessments.

PROS & CONS:

Positive Aspects/Pros:

- Evidence for efficacy
- Elucidate knowledge, critical thinking, and reasoning skills, and may allow evaluator to form subjective impressions of other characteristics (attitudes, values, beliefs)
- Allows for immediate feedback

Negative Aspects/Cons:

- Poor reliability and validity with unstructured or poorly structured oral exams
- Time consuming – can be further compounded if multiple examiners and multiple exams are utilized
- Can be intimidating for students
- Risk of evaluator bias (sex, race, age, language proficiency)

Chart-Stimulated Recall

DESCRIPTION:

Chart-stimulated recall (CSR) is similar to case-based discussion. It is a hybrid format with elements of oral examination and case-based discussion. Both the evaluator and the examinee are provided with the chart of a patient previously examined/treated by the examinee. The specific case can either be selected by the examinee (self-selected) or may be chosen by the examiner. Both parties independently review the medical record prior to meeting. This process gives the examinee a chance to re-familiarize themselves with the case. The examiner may or may not have any previous association with the case. The examinee is then interviewed regarding case specifics with the goal of determining the process and reasonings for clinical decision making, which are often not included in the chart audit. This tool can be used either as a summative or formative method of assessment. If the goal is to utilize this assessment in a summative manner, training of evaluators and guidance with case selection are strongly recommended.

PROS & CONS:

Positive Aspects/Pros:

- Evidence for efficacy (Norman et al., 1993; Cunningham et al., 1997)
- Helps elucidate critical thinking and reasoning skills
- Provides opportunity for quality feedback and mentorship
- Encourages reflective practice

CBVE: Toolkit — Assessment Tools

- High perceived value (Holt and Sofair, 2017)
- Can be used for both formative and summative assessment

Negative Aspects/Cons:

- Time consuming
- Can be intimidating to student
- Risk of evaluator bias (sex, race, age, language proficiency)
- Poorer recall for evaluation of cases further in the past
- Varying degree of case difficulty
- Use in summative assessment recommends training of evaluators with guidance in case selection

Key Features Examination

DESCRIPTION:

Assessments including key feature items help evaluate clinical decision-making. A key feature is defined as a critical step in the resolution of a problem, where two corollaries are important to include in the general definition of a key feature: “1) it focuses on a step in which examinees are most likely to make errors in the resolution of a problem, and 2) it is a difficult aspect of the identification and management of the problem in practice” (Page et al., 1995).

A number of formats for these items are described in the literature, but most common are the short menu (a variation of “Pick N”) and write-in formats (Nayer et al., 2018). One key feature case will commonly contain a problem scenario, with an average of 2-3 questions, and allows sequential pieces of clinical information to be provided between questions (Farmer and Page, 2005). By focusing on the most challenging decisions and actions in each case, examinations using key feature items may contain many short, focused cases and increase the number of cases per testing time, resulting in better content representation for the domain assessed. Furthermore, key feature items can provide improved item discrimination by focusing on the most important diagnostic features of a problem and reducing the impact of other kinds of knowledge on test scores.

PROS & CONS:

Positive Aspects/Pros:

- Key feature questions assess clinical decision making. They do not assess knowledge retention and instead assess decision-making based on synthesis and evaluation of information in Bloom’s cognitive taxonomy.
- By focusing on the critical steps required for successful resolution of a clinical problem, a key feature case can reliably and validly assess decision-making skills in a particular area using as few as 2-3 items per case vignette (Norman et al., 2006).
- Oral and written examinations tend to overly reward thoroughness (i.e., the more good things an examinee does, the higher the score). However, it has been shown that thoroughness is a poor predictor of performance (Elstein, 1978) and is indicative of novice behavior.
- Scoring that rewards only key decisions contributes to more reliable and valid test scores (Bordage and Page, 2018b).

Negative Aspects/Cons:

- Key feature questions do not assess clinical reasoning; they measure the outcome of the clinical reasoning process and not the process itself.
- As key feature items also do not assess knowledge retention, they must be used with caution in preclinical courses, by making sure that students have the requisite support and clinical background to handle the question type.
- Key feature questions are not designed or well suited to evaluating performance of procedural skills or for assessment in the psychomotor or attitudinal domains of Bloom’s taxonomy.
- Training and time are required to write high quality and valid key features items.

CBVE: Toolkit — Assessment Tools

PRACTICAL (SKILLS) EXAMINATIONS

In-Training Evaluation Report (ITER)

DESCRIPTION:

Historically, ITERs are the most commonly used evaluation form in the veterinary medicine clinical training environment (and the one with the least published about it). They are often used in a summative manner to provide a learner with feedback during or following a learning experience (typically a clinical rotation). Scoring in an ITER usually includes rating on a Likert or numerical scale and qualitative comments. ITERs are often heavily focused on professionalism, work ethic, knowledge, and communication. Written comments may be most helpful in identifying struggling learners but are often criticized for being vague and hard to interpret. ITERs are also referred to in the literature as clinical performance reports, performance assessment forms, clinical performance progress reports, or end-of-clinical rotation reports.

PROS & CONS:

Positive aspects/pros:

- Used broadly across the health professions training settings
- Recent research from medicine has focused on completing ITERs more effectively, especially the qualitative comments.
- Can assess the quality of the ITER using the Completed Clinical Evaluation Report Rating (CCERR)

Negative aspects/cons:

- Evidence suggests that the final assessment (i.e., pass versus fail) marked on the ITER is not always consistent with the evaluator's judgment of a trainee's performance, especially for poorly performing residents.
- May be difficult to recall events that occurred earlier in the training period, resulting in criticism of this method as focusing on the more recent events only
- Some negative perception from faculty related to the amount of time needed to complete these forms

Case-Based Discussion

DESCRIPTION:

This evaluation is a formal discussion between a student and clinician/professor about a case for which a student has had direct responsibility. The discussion includes all case records. The instructor asks questions to determine the student's depth of understanding, decision-making and clinical judgment. The instructor should be determining the quality of all aspects of the student's case management skills (e.g., record-keeping, client communications). The student is offered the chance to explain their decision-making throughout the discussion. A consistent rubric should be used and discussed with the student, followed by a short feedback session to help the student improve on the next case. This tool is used primarily for formative assessment (versus summative).

PROS & CONS:

Positive Aspects/Pros:

- Evidence for efficacy
- Lower stakes for student
- Helps elucidate critical thinking and reasoning skills
- Provides opportunity for quality feedback and mentorship

Negative Aspects/Cons:

- Time consuming
- Can be intimidating to student
- Not valid/reliable for summative assessments

CBVE: Toolkit — Assessment Tools

Direct Observation of Procedural Skills (DOPS)

DESCRIPTION:

This evaluation format is specifically designed to assess practical skills in a workplace setting. A trainee is observed and scored by an assessor while performing a routine practical procedure during normal clinical work. A standardized DOPS form is used to score the technique. Based on studies in medical education, for any specific skill the trainee must pass a number of repeated assessments; typically six, though more recent studies suggest fewer (three) may be needed to be signed off as competent at that skill with a reasonable level of reliability.

PROS & CONS:

Positive Aspects/Pros:

- High authenticity
- Evidence for validity and reliability in specific settings e.g., gastroscopy in medical education (Siau et al., 2020)
- Multiple assessments of the same skill can be evaluated using standardized form.
- Valuable opportunity for formative feedback

Negative Aspects/Cons:

- Time consuming to administer
- Requires the availability of a dedicated observer for an entire clinical encounter
- Multiple observations over time are needed for reliability.

Objective Structured Clinical Examination (OSCE)

DESCRIPTION:

OSCEs have been used in medical education over the past four decades and are now widely accepted in health professions education to assess hands-on technical skills or communication skills. They consist of a timed circuit of multiple mini stations with different skills or tasks being assessed in each station. OSCEs use a standardized form for grading – a binary checklist or global rating scale (GRS). Assessments are at the level of “Shows” on Miller’s Pyramid of Clinical Competence. The pass mark or minimum performance level (MPL) is set in advance using standard setting techniques such as modified Angoff, Ebel, or borderline regression.

PROS & CONS:

Positive Aspects/Pros:

- Considered gold standard for assessment of technical and communication skills across the health professions training settings outside of the clinical workplace
- Widely used in veterinary medicine for clinical skills and communication skills training – goes by many names such as OSPEs (objective structured practical exams), OSPVEs (objective structured practical veterinary exams)
- Near-peer assessment has been used.
- Generalizability theory can help determine where the source of variation between student performance comes from. Ideally the only source of variation would be from the students’ ability, but often there are many factors to consider – different animals, different raters, different sites, or different days, for instance. The more variables that can be controlled for, the better.
- Piloting stations before the OSCE can improve reliability.
- Assessors should be trained in advance and repeat rater volunteers should have their training refreshed on a regular interval. Consistency amongst raters is important regarding what elements of student performance are critical and this should be based upon how the skill was taught in the skills center.

Negative Aspects/Cons:

- Can be anxiety-inducing for novice learners
- Poorly designed OSCEs can have low reliability so quality assurance and review of evaluation of reliability are essential as part of the examination process. Cronbach’s alpha is a commonly used method of internal consistency (reliability assessment).

Clinical Evaluation Exercise (CEX)

DESCRIPTION:

The CEX is used to evaluate a student’s (or resident’s) clinical skills in a workplace setting. In the predecessor “long case evaluation,” students took a history and performed a physical examination, and then reported their findings to one or more supervising evaluators who questioned the student about the case in order to perform an evaluation (Norcini, 2001). The evaluation was therefore indirect in that the student self-reported to the examiner(s) and the examiner(s) did not directly observe the student interacting with the patient. Ultimately, long case evaluation inter-examiner reliability proved poor (Wilson et al., 1969). The CEX was developed as a tool where

CBVE: Toolkit — Assessment Tools

the examiner directly observes the student's history taking and physical examination, completes a standardized evaluation, and provides immediate feedback to the student.

PROS & CONS:

Positive Aspects/Pros:

- Evidence for efficacy: Veterinary educational efficacy remains unproven but this observational technique is used often in veterinary teaching hospitals. The potentially cumbersome length and impracticality have led to the development of the mini-CEX to improve efficiency.
- Evaluations can be quick: 15-20 minutes for a specific technique or procedure, but that focused evaluation could be considered a mini-CEX instead of a typically lengthier CEX.
- Feedback is immediate for the student.
- Scoring of the CEX is more standardized than the previous long case evaluation, so the CEX should be more consistent between students and between evaluators.
- Inter-rater scoring can be made more reliable with more than one rater simultaneously observing the same patient interaction, and with standardization of evaluation rubrics.

Negative Aspects/Cons:

- Reliability of the CEX has been criticized (Durning et al., 2002). Different evaluators may score the same student differently.
- The original CEX was developed to be 2 hours in duration as part of a standardized certification examination, making it impractical in most clinical business settings (Searle 2008).

Mini-Clinical Evaluation Exercise (mini-CEX)

DESCRIPTION:

The mini-CEX is a method used to evaluate learners engaged in an authentic clinical encounter in a workplace-based setting. It is patterned after the original, longer, CEX. Originally designed to measure "a focused history and physical examination" in the training of medical residents (Norcini et al. 1995 p. 795)", this modality is now employed to assess history taking, physical examination skills, communication skills, clinical judgment, professionalism, organization/ efficiency, and overall clinical care (Norcini and Burch, 2007), and is employed in a variety of health professions including nursing, midwifery, dentistry,

and veterinary medicine (Lorwad et al. 2017). In a typical mini-CEX encounter, the evaluator observes the examinee for approximately 20 minutes conducting a task or series of tasks in an authentic clinical setting. The evaluator then provides oral feedback, as well as a completed evaluation form. Students are likely to be evaluated using multiple mini-CEX encounters over time in order to increase reliability and document change/improvement.

PROS & CONS:

Positive Aspects/Pros:

- Evidence for efficacy: The mini-CEX has been studied in a variety of workplace-based settings, and there is evidence for its effectiveness and feasibility across a variety of disciplines in medical sciences education (see cited literature, below). The mini-CEX has been studied less in veterinary medical education contexts than in other medical education settings, but available research suggests that it performs similarly in veterinary education settings to other medical education settings. Weijs, Coe, and Hecker (2015) found that students and instructors found mini-CEX to be beneficial for learning and assessment, and Bok and colleagues (2018) found mini-CEXs to be a valuable component of their validated programmatic assessment approach.
- As a global observation tool, the mini-CEX is suitable for evaluating learners' overall ability in broad areas of performance such as "medical interviewing", "physical examination," and "professionalism." The mini-CEX is a familiar and proven tool in medical education, and multiple exemplars exist in the literature and in common use.

Negative Aspects/Cons:

- The mini-CEX does not specifically provide information at the level of subcompetencies, so inferences regarding proficiency at the level of subcompetencies must be provided through comments, or can be inferred from scores assigned at the broader competency level.
- Like other workplace-based assessment tools, the mini-CEX is time-consuming to administer, requiring the availability of a dedicated observer for an entire clinical encounter, and multiple observations over time.

CBVE: Toolkit — Assessment Tools

Longitudinal Evaluation of Performance (LEP)

DESCRIPTION:

The Longitudinal Evaluation of Performance (LEP) is a method that was adapted from mini-CEX where learners are observed in a clinic setting and the evaluator makes judgments regarding learner performance across several broad categories using standardized assessment forms. The primary difference with LEP is that serial observations are made over time to monitor learner progression towards achieving competence and the individual evaluations serve as formative feedback for the learners. The observations can be driven by the learner or the evaluator, with the former allowing learners to take greater responsibility for their own professional development. The use of multiple different evaluators to provide learners with feedback in LEP is considered advantageous to guard against bias and situations where there are professional relationship problems between the learner and evaluator.

PROS & CONS:

Positive Aspects/Pros:

- Evaluation forms can be flexibly adapted to assess a variety of different technical skills and professional skills.
- Allows monitoring of learner progression over time so that problems with performance can be identified earlier and remediation measures put in place to ensure learners have the opportunity to achieve competence
- Having feedback from multiple evaluators can provide a more holistic view of the student and guard against potential biases arising from problems in the professional relationship between a learner and evaluator.

Negative Aspects/Cons:

- Although each assessment form is generally quick to complete, it may generate a time burden for learners and evaluators if there are a large number of clinical events to assess.
- If students are not required to complete evaluation forms for all clinical events, they may tend to select events for evaluation that align with their strengths and avoid those with the potential to highlight their weaknesses.
- Requires a system for collating results from the evaluation form to monitor learner progress
- Requires a dedicated staff person and/or team to track the evaluations and develop remediation plans.

360° Evaluations (Multi-Source Feedback)

DESCRIPTION:

360° evaluations have traditionally been used to assess how individuals perform in the workplace environment by soliciting anonymous feedback on their behaviors and outcomes from people who are knowledgeable about their work. This group of evaluators includes raters who are hierarchically above, at the same level, and below the person being evaluated as this is thought to provide a more balanced assessment of performance than traditional top-down supervisor driven feedback. In a veterinary teaching hospital setting, potential raters could include line managers, mentors, clinicians, residents, interns, nurses, animal care assistants, receptionists, support staff, administrators, students, clients, and alumni. While 360° evaluations are often more traditionally used to evaluate faculty performance, there is potential for adapting them for use in student assessment.

Ratings from the 360° evaluations are also often compared against self-ratings to assess how well an individual can reflect on their own performance. The assessment surveys should be designed to take no longer than 5-10 minutes to complete given that raters may be asked to provide feedback on many other individuals in their work environment. Each 360° evaluation needs to include ratings from approximately 8-12 individuals in order to be effective.

CBVE: Toolkit — Assessment Tools

PROS & CONS:

Positive Aspects/Pros:

- Good for evaluating professionalism, communication, and interpersonal skills. Rubrics can easily be adapted to collect information on different attributes of performance
- Having feedback from multiple individuals with different professional relationships to the person being evaluated can provide a more holistic view of performance and is less prone to positive bias (“halo effect”) and negative bias (“millstone effect”).
- It can provide an anonymous means for individuals to provide feedback on their colleagues, particularly since many faculty have never been trained to give effective feedback and are uncomfortable discussing performance issues with students or residents.

Negative Aspects/Cons:

- It can be difficult to get faculty to make positive behavioral changes in response to feedback received through 360° evaluations (Lockyer et al., 2003).
- Collecting feedback can be time and resource intensive. There is often a need to purchase or subscribe to specialized software, which can be expensive for smaller programs.
- Some individuals experience strong negative emotional reactions to receiving negative feedback, particularly if they have higher self-ratings of their own performance (Sargeant et al., 2008; van der Meulen et al., 2021).

Portfolios

DESCRIPTION:

A portfolio is a cumulative body of work demonstrating a student’s learning and achievements. Each individual piece or the entire body of work may be assessed as a demonstration of the cumulative learning for a course, semester, year, or program. The process of putting the portfolio together and receiving feedback on the individual pieces serves as a great formative assessment tool because it provides multiple opportunities for student-instructor interaction. Incorporation of checklists and/or specific tasks to be mastered are helpful to students and instructors, with suggestions for types of documentation to be included that would demonstrate attempts at and final mastery of skills also recommended.

Content may be paper-based, electronic (e-portfolio), or a mixture, and may include materials selected by the student with or without guidance from the instructor. Materials should be diverse and might include written assignments, instructor feedback, case write-ups (including SOAP notes, documentation of client communications), links to videos of client/peer interactions, resumes/CVs, budgets, and the like.

PROS & CONS:

Positive Aspects/Pros:

- Thorough
- Longitudinal
- Requires a variety of skills and student reflection
- Actively promotes metacognition
- Can be formative and summative

Negative Aspects/Cons:

- Time-consuming
- Can be challenging to grade
- Requires excellent rubrics and/or multiple graders

EPAs and Entrustment-Supervision Scales

DESCRIPTION:

Historically, workplace-based assessments in health professions education have focused on evaluating learner proficiency, with scales typically employing anchors referencing specific behaviors (e.g. “Efficient time/motion but some unnecessary moves” on a surgery “Time and Motion” scale) or norm-based standards (e.g. “meets expectations” or “competent”). The CBVE Model incorporates an additional scale type, referred to by ten Cate et al. (2020) as ‘Entrustment-Supervision (ES)’ scales, which reference the extent to which the supervisor has confidence in the learner’s ability to complete a task without assistance. Such scales are not stand-alone assessment tools but can be used with practical assessment tools such as mini-CEXs or ITERs. A number of ES scales exist. They most commonly allow the rater to indicate one of the following: 1. How much supervision was required for the learner to accomplish the task (e.g., what the supervisor had to do), 2. How independently the learner could accomplish the task, or 3. How much supervision the rater believes would be necessary the next time the student attempts the task.

CBVE: Toolkit — Assessment Tools

PROS & CONS:

Positive Aspects/Pros:

- EPAs are currently employed across health professions – nursing, dentistry, pharmacy, physiotherapy, and medicine.
- EPAs are observable in process and measurable in outcome.
- Assesses at the “does” or “is” level of Miller’s pyramid of clinical competence (ten Cate et al., 2021)
- EPAs require application of knowledge, skills, and attitudes acquired in training which creates observable examples of competencies being performed.
- EPAs draw on multiple competencies but not necessarily equally. EPAs can be mapped against a competency framework to show how competencies are assessed.
- Can be used summatively or formatively. Summative decisions should be made on multiple sources of information.

Negative Aspects/Cons:

- Became very popular in a short time so perhaps there is less evidence to date for support than other assessment methods. Mostly descriptive publications to date.
- EPAs can sometimes be confused with competencies. Learners can possess competencies, knowledge, or skills (all abilities that the learner brings to the role) but a learner cannot possess an EPA (this is the professional work to be done).
- Entrustment, trust, and competence are also not readily distinguished (Melvin et al., 2020)
- Suitable for helping with assessment of workplace-based activities, but current EPAs do not include evaluation of all competencies in the CBVE framework.

STUDENT ASSIGNMENTS

Capstone Assignments

DESCRIPTION:

Capstone experiences, also called culminating experiences or transition experiences, are significant, summative, academic exercises positioned at milestone moments in curricula. The intent of a capstone from the student perspective can be to integrate, summarize, analyze, and critically reflect on what they have learned thus far in their training. From a curricular perspective, a capstone can be used to prepare students for and determine student readiness for subsequent phases in their professional training, to assess program outcomes or Entrustable Professional Activities (EPAs), and to develop student professional identity and employability. Generally, satisfactory completion of a capstone is a requirement for curricular progression or even for degree completion.

To encourage students to synthesize, integrate, analyze, and reflect, capstone experiences can focus on service-learning, workplace-based experiences, collaboration, research, international work, or even creation of new knowledge such as inventions. These experiences typically culminate in a formal write-up of some kind, be it a research essay, thesis, reflective essay, or even business proposal. Alternatively, capstones can also take the form of high stakes barrier assessments. Regardless of the format of the written assessment, the content challenges students to grapple with the complex and diverse challenges that are encountered at a high level of content understanding.

CBVE: Toolkit — Assessment Tools

PROS & CONS:

Positive Aspects/Pros:

- Certain capstones are well-validated (high stakes written barrier assessments, EPA assessments)
- Develops professional identity
- Helps students transition to the next phase in their training (e.g., clinics, employment)
- Helps elucidate critical thinking and reasoning skills
- Provides opportunity for quality feedback and mentorship
- Encourages reflective practice, integration of content, and critical analysis
- High perceived value
- Can be used for curricular outcomes assessment

Negative Aspects/Cons:

- Validated methods of assessment may be limited or nonexistent for certain project-based assignments or research essays, which may subject examinees to subjective biases in their evaluations.
- Assessments of this nature must be associated with appropriate scaffolding, support, training, and feedback so that students understand expectations for performance and evaluation; this may be challenging in the case of international experiences or humanities-based experiences for which students may have limited experience, support, or examples in other aspects of their training.
- Summative, high-stakes nature of the experiences presents barriers to progression for students.
- Resource-intensive and time-intensive to administer and assess

Case Logs

DESCRIPTION:

Case logs, also known as procedure logs, are a method of depicting patient activity including number of patients treated, procedures observed, and procedures performed. Maintenance of case logs have been implemented in medical and veterinary educational models as an easily applicable method of measuring individual clinical experiences of patient encounters and procedures. Case logs may be utilized in all

levels of medical training, but the most common application is in advanced training programs for house officers. Common practices for maintaining case logs include hand-written logs, electronic spreadsheets, electronic health record generated reports, web-based platforms, and more recently, artificial intelligence tools. Guidelines for including and describing cases are necessary to reduce variability in reporting. Clinical information such as case number, date of procedure, patient signalment, patient identifier, diagnostic procedures, length of procedure, primary and secondary diagnoses, role of trainee, and location of procedure should be included. Further reflection of the procedure including what skills were utilized, what was learned, what went well, and how the experience could be improved may also be included. Use of medical coding systems may be employed for data entry of these categories. Institutional programs typically have access to the system content, which allows for institutional oversight of patient care encounters and procedures. Audit of case logs may inform institutions of high-yield clinical rotations or be used to compare experiences to learner needs and outcomes.

PROS & CONS:

Positive Aspects/Pros:

- Low cost
- Provides measure of volume and breadth of experience
- Provides rapid objective data of experiences
- Encourages reflective practices
- May correlate with confidence in practice or comfort with procedures

Negative Aspects/Cons:

- Time consuming
- Increased clerical duties
- Relies on self-reporting, which may not accurately reflect experiences (Salazar et al., 2014)
- Does not correlate to trainees' medical knowledge, skill, or clinical acumen (Neumayer et al., 1998; Greenburg and Getson, 1999).
- Coding may be inconsistent (Balla et al., 2016).

Notes