AAVMC
CBVE
Competency-Based
Veterinary Education

Competency-Based Veterinary Education:

Toolkit

Version April 2024

Authors

- Jonathan H. Foreman, DVM, MS, DACVIM-LAIM, Professor, Associate Dean for Academic and Student Affairs, University of Illinois College of Veterinary Medicine
- Michelle Coleman, DVM, PhD, DACVIM-LAIM, Associate Professor, University of Georgia College of Veterinary Medicine
- Jared A. Danielson, PhD, Professor, Senior Associate Dean of Academic and Student Programs, Iowa State University College of Veterinary Medicine
- Katherine Fogelberg, DVM, PhD, Professor, Associate Dean for Professional Programs, Virginia-Maryland Regional College of Veterinary Medicine
- Jody Frost, PT, DPT, PhD, FAPTA, FNAP, Immediate Past President, National Academies of Practice, Education Consultant and Facilitator
- Carolyn Gates, BSc, VMD, PhD, MRCVS, SFHEA, Associate Professor, Massey University School of Veterinary Science
- Ariana Hinckley-Boltax, DVM, PhD, Assistant Professor, Tufts University Cummings School of Veterinary Medicine
- Jennifer L. Hodgson, BVSc, GradCertEdStud(HigherEd), PhD, DACVM, Professor, Virginia-Maryland Regional College of Veterinary Medicine
- Shane Lyon, DVM, MS, DACVIM-SAIM, Clinical Associate Professor, Clinical Skills Coordinator, Kansas State University College of Veterinary Medicine
- Susan M. Matthew, PhD, BVSc(Hons), BSc(Vet)(Hons), GradCertEdStud(HigherEd), Professor, Associate Chair of Veterinary Medical Education, Washington State University College of Veterinary Medicine
- Emma Read, DVM, MVSc, DACVS, Professor, Associate Dean for Professional Programs, The Ohio State University College of Veterinary Medicine
- Regina Schoenfeld-Tacher, PhD, Professor, North Carolina State University College of Veterinary Medicine

Introduction

Competency Based Veterinary Education (CBVE) represents an adoption of the medical education competency model which 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 we use in various circumstances? To that end, here we have collated informative summaries and references for several evaluative tools. Each tool is described and has pertinent domains and competencies listed. Examples, documented uses, evidence for efficacy (pro and con where available), and selected references are provided for each tool. When available, links to examples in use are provided.

Each tool could employ a variety of scales, including conventional scales related to levels of performance (excellent to poor) or competence (developing to proficient). The CBVE approach has contributed an additional scale type, referred to by ten Cate et al. (2020) as “entrustment - supervision 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 assessment tools but can be used with practical assessment tools such as mini-CEXs, ITERs, or others. Because entrustment-supervision scales are still novel, we have included a description of them in the “practical examinations” section of this document.

Tools

Assessment tools in the toolkit include:

Written (selected and constructed response) examinations:

- Multiple choice questions (MCQ)
- Extended matching questions (EMQ)
- Fill in the blank (FITB) questions
- Short answer questions
- Essay questions
- Script concordance (SCT)

Practical examinations:

- Oral examination
- Chart-stimulated recall (CSR)
- Key features examination
- In training evaluation report (ITER)
- Case-based discussion
- 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)
- 360° evaluations (multiple sources)
- Portfolios
- Entrustment-supervision scales
- Capstone assignment
- Case logs

Table of Contents

Table of Contents 3

Summary Table 8

Multiple Choice Questions 9

Description: 9

Suggested CBVE Domains and Competencies Assessed: 9

Documented Uses and Examples: 9

Pros and Cons: 9

References: 10

Extended Matching Questions 11

Description: 11

Suggested CBVE Domains and Competencies Assessed: 11

Documented Uses and Examples: 11

Pros and Cons: 12

References: 12

Fill in the Blank Questions 14

Description: 14

Suggested CBVE Competencies and Domains Assessed: 14

Documented Uses and Examples: 14

Pros and Cons: 14

References: 15

Short Answer Questions 16

Description: 16

Suggested CBVE Domains and Competencies Assessed: 16

Documented Uses and Examples: 16

Pros and Cons: 16

References: 17
Essay Questions 18
Description: 18
Suggested CBVE Domains and Competencies Assessed: 18
Documented Uses and Examples: 18
Pros and Cons: 18
References: 19
Script Concordance (SCT) 20
Description: 20
Suggested CBVE Domains and Competencies Assessed: 20
Documented Uses and Examples: 20
Pros and Cons: 20
References: 21
Oral Examination 23
Description: 23
Suggested CBVE Domains and Competencies Assessed: 23
Documented Uses and Examples: 24
Pros and Cons: 24
References: 24
Chart-Stimulated Recall 26
Description: 26
Suggested CBVE Domains and Competencies Assessed: 26
Documented Uses and Examples: 26
Pros and Cons: 26
References: 27
Key Features Examination 28
Description: 28
CBVE Domain/Competency Assessed: 28
Documented Uses and Examples: 28
Pros and Cons: 28
References: 29
In Training Evaluation Report (ITER) 31

Description: 31
Suggested CBVE Domain and Competency Assessed: 31
Documented Uses and Examples:....31
Pros and Cons:....31
References:....32
Case-Based Discussion....33
Description: 33
Suggested CBVE Domains and Competencies Assessed:....33
Documented Uses and Examples:....33
Pros and Cons:....33
References:....34
Direct Observation of Procedural Skills (DOPS)....35
Description: 35
Suggested CBVE Domains and Competencies Assessed:....35
Documented Uses and Examples:....35
Pros and Cons:....35
References:....36
Objective Structured Clinical Examination (OSCE) 38
Description: 38
CBVE Domain/Competency Assessed: 38
Documented Uses and Examples:....38
Pros and Cons:....39
References:....40
Clinical Evaluation Exercise (CEX)....41
Description: 41
Suggested CBVE Domains and Competencies Assessed:....41
Documented Uses and Examples:....41
Pros and Cons:....41
References:....42
Mini-Clinical Evaluation Exercise (mini-CEX) 43
Description: 43
Suggested CBVE Domains and Competencies Assessed:....43

Documented Uses and Examples:....43
Pros and Cons:....43
References:....44
Longitudinal Evaluation of Performance (LEP)....46
Description:....46
Suggested CBVE Domains and Competencies Assessed:....46
Documented Uses and Examples:....46
Pros and Cons:....46
References:....47
360° Evaluations (Multi-Source Feedback)....49
Description:....49
Suggested CBVE Domains and Competencies:....49
Documented Uses and Examples:....49
Pros and Cons:....50
References:....51
Portfolios....53
Description:....53
Suggested CBVE Domains and Competencies Assessed:....53
Documented Uses and Examples:....53
Pros and Cons:....54
References:....54
Entrustment-Supervision Scales....56
Description:....56
Suggested CBVE Domains and Competencies Assessed:....56
Documented Uses and Examples:....56
Pros and Cons:....57
References:....58
Capstone Assignments....59
Description:....59
Suggested CBVE Domains and Competencies Assessed:....59
Documented Uses and Examples:....60
Pros and Cons:....60

References: 61
Case Logs 63
Description: 63
Suggested CBVE Domains and Competencies Assessed: 63
Documented Uses and Examples: 63
Pros and Cons: 64
References: 64

Summary Table

	Competencies Assessed						
	Clinical Knowledge "What You Know"	Critical Reasoning "How You Think"		Technical Skills "What You Can Do"	Professional Identity "How You Interact"		
	Individual Animal Care, Animal Population Care, & Public Health	Clinical Reasoning & Decision-Making	Gathering & Evaluating Informa-tion	Medical, Surgical, & Anesthetic Procedures	Written Communi-cation	Verbal Communi-cation	Collegiality & Teamwork
Domains of competence	2, 3, 4	1	1, 2, 5	6, 7	5, 8, 9	5, 8, 9	6
Entrustable professional activities (EPAs)	2, 3, 4, 6, 7, 8	2, 3, 4, 6, 7	2, 3	3, 6, 7	1, 3, 4, 5, 8	1, 3, 4, 5, 8	4, 7
Assessment Type							
Written & Oral Examinations							
Multiple Choice Questions (MCQ)	X	X					
Extended Matching Questions (EMQ)	X	X					
Fill in the blank (FITB)	X	X			X		
Short Answer Questions (SAQ)	X	X			X		
Essay Questions	X	X			X		
Script Concordance Testing (SCT)		X			X	X	
Oral Examinations	X	X				X	
Chart Stimulated Recall Examination (CSR)	X	X				X	
Key Features Examination		X			X	X	
Practical (Skills)Examinations							
In Training Evaluation Report (ITER)	X	X	X	X	X	X	X
Case-Based Discussion	X	X				X	
Direct Observation of Procedural Skills (DOPS)				X			
Objective Structured Clinical Examination (OSCE)				X		X	
Clinical Evaluation Exercise (CEX)				X			
Mini-Clinical Evaluation Exercise (Mini-CEX)				X			
Longitudinal Evaluation of Performance (LEP)	X	X	X	X	X	X	
360° Evaluations (Multi-Source Feedback)							X
Portfolios	X	X	X		X		
Entrustment-supervision scales	X	X	X	X	X	X	X
Student Assignments							
Capstone Assignments			X		X		
Case Logs	X	X	X		X		

Multiple Choice Questions

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.2
- 3.1-3.3
- 4.1-4.2
- 5.3
- 8.1-8.3
- 9.1 - 9.2

Documented Uses and Examples:

The National Board of Medical Examiners (NBME) Item-Writing Guide on constructing written test questions for the health sciences provides detailed guidelines for and examples of well-written MCQs. The guide can be downloaded from: https://www.nbme.org/publications/item-writing-manual.html.

Haladyna et al. (2002) provides a structured taxonomy of item-writing guidelines.

MCQs are used commonly for formative and summative assessment in veterinary and medical school curricula. MCQs are used in the North American Veterinary Licensing Examination (NAVLE), and National Board of Medical Examiners (NBME) student progress and licensure examinations.

Pros and 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 examinations. MCQs can be used broadly to 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 use 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.

References:

Cook AK, Lidbury JA, Creevy KE, Heseltine JC, Marsilio S, Catchpole B, Whittlestone KD. Multiple-choice questions in small animal medicine: An analysis of cognitive level and structural reliability, and the impact of the characteristics on student performance. J Vet Med Educ 47(4):497-505, 2020.

Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. Appl Measurem Educ 15(3):309-334, 2002.

National Board of Medical Examiners. NBME Item-Writing Guide: Constructing Written Test Questions for the Health Sciences. 6th Edition, February 2021.

Royal K, Dorman D. Comparing item performance on three- versus four-option multiple choice questions in a veterinary toxicology course. Vet Sci 5(2):55, 2018.

Extended Matching Questions

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.2
- 3.1-3.3
- 4.1-4.2
- 5.3
- 8.1-8.3
- 9.1-9.2

Note: Any application in which multiple choice questions are appropriate is also appropriate for extended matching. Well-written selected-response questions can measure a variety of relevant outcomes including conceptual knowledge, principles, and problem solving (diagnostic and clinical reasoning).

Documented Uses and Examples:

A good source of information regarding EMQs is Chapters 6 and 7 of Case and Swanson's "Constructing Written Test Questions for the Basic and Clinical Sciences (3rd Edition). This resource is available for free download: https://www.researchgate.net/publication/242759434_Constructing_Written_Test_Questions_For_the_Basic_and_Clinical_Sciences

EMQs have been used in high stakes examinations including US Medical Board examinations (Case and Swanson 1993), Obstetrics and Gynecology board examinations (Duthie et al., 2006), and Psychiatry board examinations (Samuels, 2006).

Pros and 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.

References:

Beullens J, Struyf E, Van Damme B. Do extended matching multiple-choice questions measure clinical reasoning? Med Educ 39: 410-417, 2005. https://doi.org/10.1111/j.1365-2929.2005.02089.x

Bhakta B, Tennant A, Horton M et al. Using item response theory to explore the psychometric properties of extended matching questions examination in undergraduate medical education. BMC Med Educ 5:9, 2005. https://doi.org/10.1186/1472-6920-5-9

Burton JL. How to write and how to answer EMQs. Obst Gynaec Reprod Med 19(12):359-361, 2009.

Case SM, Swanson DB. Extended-matching items: A practical alternative to free-response questions. Teach Learn Med 5(2):107-115, 1993. https://DOI:10.1080/10401339309539601

Duthie S, Hodges P, Ramsay I, Reid W. EMQs: a new component of the MRCOG Part 2 exam. Obstet Gynaec 8(3):181-185, 2006.

Samuels A. Extended matching questions and the Royal Australian and New Zealand College of Psychiatrists written examination: an overview. Australasian Psych 14(1):63-66, 2006.

Swanson DB, Holtzman KZ, Allbee K. Measurement characteristics of content-parallel single-best-answer and extended-matching questions in relation to number and source of options. Acad Med 83(10):S21-S24, 2008. https://doi:10.1097/ACM.0b013e318183e5bb

Fill in the Blank Questions

Return to Table of Contents

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.

Suggested CBVE Competencies and Domains Assessed:

- 1.1-1.7
- 2.1-2.2
- 3.1-3.3
- 4.1-4.2
- 5.3
- 8.1-8.3
- 9.1 - 9.2

Documented Uses and Examples:

Example: A patient is receiving dopamine at 12 ml/hr. The solution hanging indicates "D5W 250 ml/200 mg dopamine." The patient weighs 70 lbs. What dosage (mcg/kg/min) is the patient receiving? _____

European College of Veterinary Pathologists. Example Exam Questions. 2021. https://www.ecvpath.org/example-exam-questions/

Pros and 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
- Although evaluation of short answer questions is more objective than essay questions, short answer questions are neither fully valid, reliable, nor objective.
- 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.

References:

Jacobs LC. How to write better tests: A handbook for improving test construction skills. Evaluation Services and Testing, Indiana University-Bloomington. https://www.uky.edu/Ag/CLD/CETL/files/f09workshop/IU%20How%20to%20write%20better%20tests.pdf

Park University, Faculty Development website, Writing Test Items. http://captain.park.edu/facultydevelopment/writing_test_items.htm

Schuwirth L. ABC of learning and teaching in medicine: Written assessment. Brit Med J 326(73-90): 643-645, 2003.

Schuwirth L, van der Vleuten C. Different written assessment methods: what can be said about their strengths and weaknesses? Med Educ 38(9):974-979, 2004.

Short Answer Questions

Return to Table of Contents

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).

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.2
- 3.1-3.3
- 4.1, 4.2
- 5.3
- 7.3
- 8.1, 8.2
- 9.1, 9.2

Documented Uses and Examples:

https://unmc.instructure.com/courses/6155/pages/short-answer-and-essay-answer-questions

https://www.csu.edu.au/division/learning-teaching/assessments/assessment-types/exams/short-answer-questions

Pros and 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

References:

Haynie WJ. Effects of multiple-choice and short-answer tests on delayed retention learning. J Technol Educ 6(1): Fall 1994. https://doi.org/10.21061/jte.v6i1.a.3

Rademakers J, ten Cate T, Bar P. Progress testing with short answer questions. Med Teach 27(7):578-582, 2005.

Schuwirth LWT and van der Vleuten CPM. ABC of teaching and learning in medicine: Written assessment. BMJ 326(7390):643-645, 2003. https://www.jstor.org/stable/25453994

Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: What can be said about their strengths and weaknesses? Med Educ 38: 974-979, 2004.

Stunden A, Jefferies D. The effectiveness of short answers test papers in evaluating academic nursing programs: A review of the literature. Nurse Educ Pract 33:94-101, 2018. https://doi.org/10.1016/j.nepr.2018.09.004

van Hoeij MJW, Haarhuis JCM, Wierstra RFA, van Beukelen P. Developing a classification tool based on Bloom's taxonomy to assess the cognitive level of short essay questions. J Vet Med Educ 31(3): 261-267, 2004.

Essay Questions

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.2
- 3.1-3.3
- 4.1, 4.2
- 5.2, 5.3
- 7.3
- 8.1, 8.2
- 9.1, 9.2, 9.3

Documented Uses and Examples:

European College of Veterinary Pathologists. Example Exam Questions. 2021. https://www.ecvpath.org/example-exam-questions/

Pros and 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.

References:

Christian CM, Bothell TW, Sudweeks RR, Wood B. Preparing effective essay questions: A Self-Directed Workbook for Educators. New Forums Press, 2002. https://testing.byu.edu/handbooks/WritingEffectiveEssayQuestions.pdf

Schuwirth L, van der Vleuten C. Different written assessment methods: What can be said about their strengths and weaknesses? Med Educ 38(9):974-979, 2004.

Schuwirth, L. ABC of learning and teaching in medicine: Written assessment. Brit Med J 326(73-90):643-645, 2003.

Veterinary Medicine Essays: Examples, Topics, Titles, & Outlines (paperdue.com)

Script Concordance (SCT)

Return to Table of Contents

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).

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.6

Documented Uses and Examples:

There are limited examples showing the use of SCTs in veterinary education. The first reference illustrates the use of an SCT to assess the clinical reasoning of practicing food-animal clinicians. Highlights of this example include a description of the process used to create the scoring key and validate the SCT (Dufour et al., 2012).

A more recent publication illustrates the use of a modified SCT (3 response options instead of 5) as a teaching tool, instead of a true assessment (Tayce and Saunders, 2021).

Pros and Cons:

Positive aspects/pros (Carriere et al., 2009; Charlin et al., 2000; Lubarsky et al., 2011; Nouh et al., 2012; Subra et al., 2017):

- 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 (Lineberry et al., 2013; Lineberry et al., 2019; Steinberg et al., 2020):

- 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 examines 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.

References:

Carrière B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: Validity evidence for a script concordance test. Annals of Emerg Med 53(5):647-652, 2009.

Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. Teach Learn in Med 12(4):189-195, 2000.

Dufour S, Latour S, Chicoine Y, Fecteau G, Forget S, Moreau J, Trépanier A. Use of the script concordance approach to evaluate clinical reasoning in food-ruminant practitioners. J Vet Med Educ 39(3):267-275, 2012. https://jvme.utpjournals.press/doi/full/10.3138/jvme.0112-13R

Lineberry M, Kreiter CD, Bordage G. Threats to validity in the use and interpretation of script concordance test scores. Med Educ 47(12):1175-1183, 2013.

Lineberry M, Hornos E, Pleguezuelos E, Mella J, Brailovsky C, Bordage G. Experts' responses in script concordance tests: A response process validity investigation. Med Educ 53(7):710-722, 2019.

Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CP. Script concordance testing: A review of published validity evidence. Med Educ 45(4):329-338, 2011. https://doi-org.prox.lib.ncsu.edu/10.1111/j.1365-2923.2010.03863.x

Lubarsky S, Dory V, Duggan P, Gagnon R, Charlin B. Script concordance testing: From theory to practice: AMEE Guide No. 75. Med Teach 35:184-193, 2013.

Nouh T, Boutros M, Gagnon R, Reid S, Leslie K, Pace D, Pitt D, Walker R, Schiller D, MacLean A, Hameed M. The script concordance test as a measure of clinical reasoning: A national validation study. Am J Surg 203(4):530-534, 2012.

Ramaekers S, Kremer W, Pilot A, Beukelen PV, Keulen HV. Assessment of competence in clinical reasoning and decision-making under uncertainty: The script concordance test method.

Assessm & Eval in Higher Educ 35(6):661-673, 2010. https://doi: 10.1080/02602938.2010.500103

Steinberg E, Cowan E, Lin MP, Sielicki A, Warrington S. Assessment of emergency medicine residents' clinical reasoning: Validation of a script concordance test. Western J Emerg Med 21(4):978, 2020.

Subra J, Chicoulaa B, Stillmunkès A, Mesthé P, Oustric S, Rougé Bugat ME. Reliability and validity of the script concordance test for postgraduate students of general practice. Eur J Gen Pract 23(1):209-214, 2017.

Tayce JD, Saunders AB. The use of a modified script concordance test in clinical rounds to foster and assess clinical reasoning skills. J Vet Med Educ, 2021 Nov 16:e20210090. https://jvme.utpjournals.press/doi/epub/10.3138/jvme-2021-0090

Oral Examination

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.2
- 3.1-3.3
- 4.1, 4.2
- 5.1, 5.2
- 6.4
- 7.1 - 7.3
- 8.1, 8.2
- 9.1

Documented Uses and Examples:

Oral examinations are typically a component of the M.S. or Ph.D. thesis defense process. These are generally unstructured in nature as they are not comprised of standard questions for all candidates and scored using a standardized method, but rather are reminiscent of the original *viva voce* utilized in the beginnings of the academe.

Oral examinations are utilized in some specialty board examinations within the American Board of Physician Specialties (diagnostic radiology, obstetrics and gynecology, radiation oncology, disaster medicine, surgery, emergency medicine, orthopedics, anesthesiology, family medicine obstetrics, and psychiatry).

Pros and 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)

References:

Anastakis DJ, Cohen R, Reznick RK. The structured oral examination as a method for assessing surgical residents. Am J Surg 162(1):67-70, 1991.

Daelmans HE, Scherpbier AJ, Vleuten CPvd, Donker AJ. Reliability of clinical oral examinations re-examined. Med Teach 23(4):422-4, 2001.

Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. Med Teach 27(4):294-7, 2005.

Hodges BD. The objective structured clinical examination: three decades of development. J Vet Med Educ 33(4):571-7, 2006.

Jefferies A, Simmons B, Ng E, Skidmore M. Assessment of multiple physician competencies in postgraduate training: utility of the structured oral examination. Adv Health Sci Educ 16(5):569-77, 2011.

Stray C. The shift from oral to written examination: Cambridge and Oxford 1700–1900. Assessment in Education: Principles, Policy & Practice 8(1):33-50, 2001.

Touchie C, Humphrey-Murto S, Ainslie M, Myers K, Wood TJ. Two models of raters in a structured oral examination: does it make a difference? Adv Health Sci Educ 15(1):97-108, 2010.

Wakeford R, Southgate L, Wass V. Improving oral examinations: selecting, training, and monitoring examiners for the MRCGP. BMJ 311(7010):931-5, 1995.

Chart-Stimulated Recall

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.1, 2.2
- 3.1, 3.2
- 4.1, 4.2
- 5.3
- 6.1-6.3
- 7.1-7.4
- 8.2
- 9.1-9.3

Documented Uses and Examples:

CSR was first described for the re-certification of emergency room physicians and is heavily utilized in the United Kingdom and Canada to evaluate practicing physicians. CSR is also described in post-graduate medical education programs like residency training.
An example of a chart-stimulated recall worksheet from the Accreditation Council for Graduate Medical Education can be found at:
https://www.acgme.org/globalassets/430_chartstimulatedrecall.pdf

Pros and Cons:

Positive Aspects/Pros:

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

References:

Cunnington JP, Hanna E, Turnhbull J, Kaigas TB, Norman GR. Defensible assessment of the competency of the practicing physician. Acad Med: J Assoc Am Med Coll 72(1):9-12, 1997.

Goulet F, Jacques A, Gagnon R, Racette P, Sieber W. Assessment of family physicians' performance using patient charts: interrater reliability and concordance with chart-stimulated recall interview. Eval Health Prof 30(4):376-392, 2007.

Holt SR, Sofair AN. Resident and faculty perceptions of chart-stimulated recall. Southern Med J 110(2):142-146, 2017.

Norman GR, Davis DA, Lamb S, Hanna E, Caulford P, Kaigas T. Competency assessment of primary care physicians as part of a peer review program. J Am Med Assoc 270(9):1046-1051, 1993.

Reddy ST, Endo J, Gupta S, Tekian A, Park YS. A case for caution: Chart-stimulated recall. J Grad Med Educ 7(4):531-535, 2015.

Sinnott C, Kelly MA, Bradley CP. A scoping review of the potential for chart stimulated recall as a clinical research method. BMC Health Serv Res 17(1):583, 2017.

Key Features Examination

Return to Table of Contents

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.

CBVE Domain/Competency Assessed:

- 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7
- 2.2
- 3.1, 3.2, 3.3

Documented Uses and Examples:

Key feature questions have been used at different levels of medical education and practice including undergraduate and graduate education, licensing examinations, and medical practice (Farmer and Hinchy, 2005; Fischer et al., 2005; Leung et al., 2016; Bronander et al., 2015). Since its inception, the accumulated validity evidence for the key features approach supports the decision-making construct measured and its use to assess clinical decision-making skills at all levels of training and practice and with various types of examination formats (Bordage and Page, 2018b).

Pros and 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). Further, 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. Therefore, 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.

References:

Bordage G, Page G. Key features to assess clinical decisions. Med Teach 40:1195-1196, 2018a.

Bordage G, Page G. The key-features approach to assess clinical decisions: Validity evidence to date. Adv in Health Sci Educ 23:1005-1036, 2018b.

Bronander KA, Lang VJ, Nixon LJ, Harrell HE, Kovach R, Hingle S, Berman N. How we developed and piloted an electronic key features examination for the internal medicine clerkship based on a US national curriculum. Med Teach 37:807-812, 2015.

Elstein AS, Shulman LS, Sprafka SA. Medical Problem Solving. Cambridge: Harvard University Press, 1978.

Farmer EA, Hinchy J. Assessing general practice clinical decision-making skills: A key features approach. Austral Family Phys 34:1059-1061, 2005.

Farmer EA, Page GA. Practical guide to assessing clinical decision-making skills using key features approach. Med Educ 39:1188-1194, 2005.

Fischer MR, Kopp V, Holzer M, Ruderich F, Jünger J. A modified electronic key features examination for undergraduate medical students: Validation threats and opportunities. Med Teach 27:450-455, 2005.

Hyrnchak P, Takahashi SG, Nayer M. Key-features questions for assessment of clinical reasoning: A literature review. Med Educ 48:870-883, 2014.

Leung F-H, Herold J, Iglar K. Family medicine mandatory assessment of progress: Results of a pilot administration of a family medicine competency-based in-training examination. Can Family Phys 62:e263-c267, 2016.

Nayer M, Takahashi SG, Hrynchak P. Twelve tips for developing key-feature questions (KFQ) for effective assessment of clinical reasoning. Med Teach 40:1116-1122, 2018.

Norman G, Bordage G, Page G, Keane D. How specific is case specificity? Med Educ 40:618-623, 2006.

Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. Acad Med 70(3):194-201, 2019.

In Training Evaluation Report (ITER)

Return to Table of Contents

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.

Suggested CBVE Domain and Competency Assessed:

- All CBVE domains and competencies can be assessed using this format.

Documented Uses and Examples:

Read EK, Brown A, Maxey C, Hecker KG. Comparing entrustment and competence: An exploratory look at performance-relevant information in the final year of a veterinary program. J Vet Med Educ 48(5):652-572, 2021. https://doi:10.3138/jvme-2019-0128

Weijs CA, Coe JB, Hecker KG. Final-year students' and clinical instructors' experience of workplace-based assessments used in a small-animal primary-veterinary-care clinical rotation. J Vet Med Educ 42(4):382-392, 2015. https://doi:10.3138/jvme.1214-123R1

Pros and 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

References:

Bismil R, Dudek NL, Wood TJ. In-training evaluations: Developing an automated screening tool to measure report quality. Med Educ 48(7):724-732, 2014. https://doi:10.1111/medu.12490

Ginsburg S, Eva K, Regehr G. Do in-training evaluation reports deserve their bad reputations? A study of the reliability and predictive ability of ITER scores and narrative comments. Acad Med 88(10):1539-1544, 2013. https://doi:10.1097/ACM.0b013e3182a36c3d

Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: Faculty interpretations of narrative evaluation comments. Med Educ 49(3):296-306, 2015. https://doi:10.1111/medu.12637

Ginsburg S, van der Vleuten CPM, Eva KW, Lingard L. Cracking the code: Resident's interpretation of written assessment comments. Med Educ 51(4):401-410, 2017. https://doi:10.1111/medu.13158

Hatala R, Sawatsky AP, Dudek N, Ginsburg S, Cook DA. Using in-training evaluation report (ITER) qualitative comments to assess medical students and residents: A systematic review. Acad Med 92(6):868-879, 2017. https://doi:10.1097/ACM.0000000000001506

Kassam A, Donnon T, Rigby I. Validity and reliability of an in-training evaluation report to measure the CanMEDS roles in emergency medicine residents. Can J Emerg Med 16(2):144-150, 2014. https://doi:10.2310/8000.2013.130958

Read EK, Brown A, Maxey C, Hecker KG. Comparing entrustment and competence: An exploratory look at performance-relevant information in the final year of a veterinary program. J Vet Med Educ 48(5):652-572, 2021. https://doi:10.3138/jvme-2019-0128

Weijs CA, Coe JB, Hecker KG. Final-year students' and clinical instructors' experience of workplace-based assessments used in a small-animal primary-veterinary-care clinical rotation. J Vet Med Educ 42(4):382-392, 2015. https://doi:10.3138/jvme.1214-123R1

Case-Based Discussion

Return to Table of Contents

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).

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.1, 2.2
- 3.1, 3.2
- 4.1, 4.2
- 5.1, 5.3
- 7.2, 7.3, 7.5
- 8.2, 8.3

Documented Uses and Examples:

https://www.youtube.com/watch?v=14b8yZy-6Zo

https://www.youtube.com/watch?v=8LtC5AqnV9M

Pros and Cons:

Positive Aspects/Pros:

- Evidence for efficacy
    - Cunningham JPW, Hanna E, Turnbull J, Kaigas TB, Norman GR. Defensible assessment of the competency of the practicing physician. Acad Med 72(1):9-12, 1997.
    - Jyothirmayi, R. Case-based discussion: Assessment tool or teaching aid? Clin Oncol 24:649-653, 2012.

- 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

References:

A rubric commonly used in the UK can be found at: Intercollegiate Surgical Curriculum Website: http://www.iscp.ac.uk/Assessment/WBA/CBD.aspx)

*information paraphrased from Baillie and Rhind: A guide to assessment methods in veterinary medicine, version 1.1 (September 2008)

Direct Observation of Procedural Skills (DOPS)

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1, 1.7
- 2.1
- 5.1

Documented Uses and Examples:

Siau K, Crossley J, Dunckley P, Johnson G, Feeney M, Hawkes ND, Beales ILP, Joint Advisory Group on Gastrointestinal Endoscopy (JAG). Direct observation of procedural skills (DOPS) assessment in diagnostic gastroscopy: Nationwide evidence of validity and competency development during training. Surg Endosc 34(1):105-114, 2020. https://doi:10.1007/s00464-019-06737-7

Note: similar articles by these authors for:

- Pediatric gastroscopy (https://doi:10.1097/MPG.0000000000002089)
- Colonoscopy (https://doi:10.14309/ajg.0000000000000426)
- Sigmoidoscopy (https://doi:10.15403/jgld.2014.1121.281.nov)

Specific example for veterinary education: Magnier K, Dale V, Pead M. Workplace-based assessment instruments in the health sciences. J Vet Med Educ 39(4):389-395, 2012.

Pros and 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.

References:

Dabir S, Hoseinzadeh M, Mosaffa F, Hosseini B, Dahi M, Vosoughian M, Moshari M, Tabashi S, Dabbagh A. The effect of repeated direct observation of procedural skills (R-DOPS) assessment method on the clinical skills of anesthesiology residents. Anesth Pain Med 11(1):e111074, 2021. doi: 10.5812/aapm.111074

Erfani K, Ebadi FA. Direct observation of procedural skills (DOPS) evaluation method: Systematic review of evidence. Med J Islam Repub Iran 32:45, 2018. https://doi:10.14196/mjiri.32.45

Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. J Am Med Assoc 302(12):1316-1326, 2009. https://doi:10.1001/jama.2009.1365

Lorwald AC, Lahner FM, Nouns ZM, Berendok C, Norcini J, Greif R, Huwenediek S. The educational impact of mini-clinical evaluation exercise (Mini-CEX) and direct observation of procedural skills (DOPS) and its association with implementation: A systematic review and meta-analysis. PLoS ONE 13(6):e0198009, 2018. https://doi:10.1371/journal.pone.0198009

Magnier K, Dale V, Pead M. Workplace-based assessment instruments in the health sciences. J Vet Med Educ 39(4):389-395, 2012.

Magnier K, Pead M. Performance and workplace-based assessment. In: Veterinary Medical Education: A Practical Guide. Hodgson J, Pelzer J (eds), Wiley Blackwell, Iowa, pp 255-272, 2017.

McLeod R, Mires G, Ker J. Direct observed procedural skills assessment in the undergraduate setting. Clin Teach 9(4):228—232, 2012.

Naeem N. Validity, reliability, feasibility, acceptability and educational impact of direct observation of procedural skills (DOPS). J Coll Phys Surg Pak 23(1):77-82, 2013. PMID 23286629.

Schuwirth L. ABC of learning and teaching in medicine: Written assessment. British Med J 326(73-90):643-645, 2003.

Schuwirth L, van der Vleuten C. Different written assessment methods: what can be said about their strengths and weaknesses? Med Educ 38(9):974-979, 2004.

Siau K, Crossley J, Dunckley P, Johnson G, Feeney M, Hawkes ND, Beales ILP, Joint Advisory Group on Gastrointestinal Endoscopy (JAG). Direct observation of procedural skills (DOPS) assessment in diagnostic gastroscopy: Nationwide evidence of validity and competency development during training. Surg Endosc 34(1):105-114, 2020. https://doi:10.1007/s00464-019-06737-7

Wilkinson J, Crossley J, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. Med Educ 42(4):364-373, 2008.

Wragg A, Wade W, Fuller G, Cowan G, Mills P. Assessing the performance of specialist registrars. Clin Med 3(2):131-134, 2003.

Objective Structured Clinical Examination (OSCE)

Return to Table of Contents

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.

CBVE Domain/Competency Assessed:

- 1.1-1.7
- 2.1-2.2
- 3.1, 3.2
- 4.1, 4.2
- 5.1-5.3

Documented Uses and Examples:

Annandale A, Fosgate GT, Bok H, Holm DE. Ability of a bovine transrectal palpation objective structured clinical examination to predict veterinary students' pregnancy diagnosis accuracy. Vet Rec 185(6):1-9, 2019. https://doi:10.1136/vr.105022 . Epub 2019 Jun 7.

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 221(9):1262-1265, 2006. https://doi:10.2460/javma.2002.221.1262

Davis MH, Ponnamperuma GG, McAleer S, Dale VHM. The objective structured clinical examination (OSCE) as a determinant of veterinary clinical skills. J Vet Med Educ 33(4):578-587, 2006. https://doi:10.3138/jvme.33.4.578

Hecker K, Read EK, Vallevand A, Krebs G, Donszelmann D, Muelling CKW, Freeman SL. Assessment of first-year veterinary students' clinical skills using objective structured clinical examinations. J Vet Med Educ 37(4):395-402, 2010. https://doi:10.3138/jvme.37.4.395

Hecker KG, Adams CL, Coe JB. Assessment of first-year veterinary students' communication skills using an objective structured clinical examination: The importance of context. J Vet Med Educ 39(3):304-310, 2012. https://doi:10.3138/jvme.0312.022R

Hunt JA, Anderson S. Remote assessment of veterinary clinical skills courses during the COVID-19 pandemic. Mar 3;e20200084, 2021. doi: 10.3138/jvme-2020-0084.

Read EK, Bell C, Rhind S, Hecker KG. The use of global rating scales for OSCEs in veterinary medicine. Plos ONE 10(3):e0121000, 2015. https://doi:10.1371/journal.pone.0121000

Tan J-Y, Ma IWY, Hunt JA, Kwong GPS, Farrell R, Bell C, Read EK. Video recording in veterinary medicine OSCEs: Feasibility and inter-rater agreement between live performance examiners and video recording reviewing examiners. J Vet Med Educ 48(4):485-491, 2021. https://doi:10.3138/jvme-2019-0142

Pros and 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 structures 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).

References:

Dunne K, Moffett J, Loughran ST, Duggan V, Campion DP. Evaluation of a coaching workshop for the management of veterinary nursing students' OSCE-associated test anxiety. Irish Vet J 71:15, 2018. https://doi:10.1186/s13620-018-0127-z

Hecker K, Violato C. Validity, reliability, and defensibility of assessments in veterinary education. J Vet Med Educ 36:271-275, 2009.

Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The objective structured clinical examination (OSCE): AMEE guide no. 81. part II: Organisation & administration. Med Teach 35**:**e1447-e1463, 2013.

Lee CB, Madrazo L, Khan U, Thangarasa T, McConnell M, Khamisa K. A student-initiated objective structured clinical examination as a sustainable cost-effective learning experience. Med Educ Online 23**:** 1440111, 2018.

May S, Head S. Assessment of technical skills: best practices. J Vet Med Educ 37(3):258-265, 2010. doi: 10.3138/jvme.37.3.258.

McKinley DW, Norcini JJ. How to set standards on performance-based examinations: AMEE guide no. 85. Med Teach 36**:**97-110, 2014.

Miller G. The assessment of clinical skill/competence/performance. Acad Med 65**:**63-67, 1990.

Pell G, Fuller R, Homer M, Roberts T. How to measure the quality of the OSCE: a review of metrics–AMEE guide no. 49. Med Teach 32**:**802-811, 2010.

Royal KD, Hecker KG. Rater errors in clinical performance assessments. J Vet Med Educ 43(1):5-8, 2016. https://doi:10.3138/jvme.0715-112R

Clinical Evaluation Exercise (CEX)

Return to Table of Contents

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

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.1-2.2
- 3.1-3.3
- 4.1-4.2
- 5.1-5.3
- 6.1, 6.2, 6.4
- 7.1-7.2, 7.4
- 8.2, 8.3
- 9.1-9.3

Documented Uses and Examples:

https://www.iscp.ac.uk/curriculum/surgical/assessment_cex.aspx

Pros and 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).

References:

Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med 77:900-904, 2002.

Norcini JJ. The validity of long cases. Med Educ 35:720-721, 2001.

Searle GF. Is CEX good for psychiatry? An evaluation of workplace-based assessment. Psych Bull 32:271-273, 2008.

Wilson GM, Lever R, Harden RM, et al. Examination of clinical examiners. The Lancet 293:37-40, 1969.

Mini-Clinical Evaluation Exercise (mini-CEX)

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.1-2.2
- 3.1-3.3
- 4.1-4.2
- 5.1-5.2
- 6.1-6.2, 6.4
- 8.1-8.3
- 9.2-9.3

Documented Uses and Examples:

https://www.abim.org/Media/qlvp1fhb/mini-cex.pdf

Ansari AA, Ali SK, Donnon T. The construct and criterion validity of the mini-CEX. Acad Med 88(3):413-420, 2013. https://doi:10.1097/ACM.0b013e318280a953

Pros and 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 work-place 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.

References:

Ansari AA, Ali SK, Donnon T. The construct and criterion validity of the mini-CEX. Acad Med 88(3):413-420, 2013. https://doi:10.1097/ACM.0b013e318280a953

Bok HGJ, de Jong LH, O'Neill T, et al. Validity evidence for programmatic assessment in competency-based education. Perspect Med Educ 7:362–372, 2018. https://doi.org/10.1007/s40037-018-0481-2

Hejri SM, Jalili M, Masoomi R, Shirazi M, Nedjat S, Norcini J. The utility of mini-clinical evaluation exercise in undergraduate and postgraduate medical education: A BEME review: BEME Guide No. 59. Med Teach 42(2):125-142, 2000. https://doi:10.1080/0142159X.2019.1652732

Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees: A systematic review. J Am Med Assoc 302(12):1316–1326, 2009. https://doi:10.1001/jama.2009.1365

Lörwald AC, Lahner FM, Nouns ZM, Berendonk C, Norcini J, et al. The educational impact of mini-clinical evaluation exercise (mini-CEX) and direct observation of procedural skills (DOPS)

and its association with implementation: A systematic review and meta-analysis. PLoS ONE 13(6):e0198009, 2018. https://doi.org/10.1371/journal.pone.0198009

Pelgrim EAM, Kramer AWM, Mokkink HGA, et al. In-training assessment using direct observation of single-patient encounters: A literature review. Adv in Health Sci Educ 16, 131–142, 2011. https://doi.org/10.1007/s10459-010-9235-6

Weijs CA, Coe JB, Hecker KG. Final-year students' and clinical instructors' experience of workplace-based assessments used in a small-animal primary-veterinary-care clinical rotation. J Vet Med Educ 42(4):382-392, 2015. https://doi:10.3138/jvme.1214-123R1

Longitudinal Evaluation of Performance (LEP)

Return to Table of Contents

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 judgements 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 competency and the individual evaluations serve as formative feedback for the learners. The observations can be driven by the learner or the assessor 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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.1-2.2
- 3.1-3.3
- 4.1-4.2
- 5.1-5.2
- 6.1-6.2, 6.4
- 8.1-8.3
- 9.2-9.3

Documented Uses and Examples:

There are few examples in the literature documenting the use of LEP in health education. The following is an example of an evaluation form that was utilized in the LEP for the Scottish Dental Assessment Programme:

Prescott L, McKinlay P, Rennie J. The development of an assessment system for dental vocational training and general professional training: A Scottish approach. British Dental J 190(1):41-44, 2001. https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2923.2002.01099.x#f1

Pros and 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.

References:

Dickie JDR. Longitudinal clinical assessment of undergraduate dental students: Building an argument for its validity: University of Glasgow, 2021.

Newton WP, Rode K, O'Neill T, Fain R, Baxley E. Longitudinal assessment: Where we are and why it is important. Am Board Family Med 2019.

Prescott-Clements L, Van Der Vleuten CP, Schuwirth LW, Hurst Y, Rennie JS. Evidence for validity within workplace assessment: the Longitudinal Evaluation of Performance (LEP). Medical Education. 2008;42(5):488-95.

Prescott L, McKinlay P, Rennie J. The development of an assessment system for dental vocational training and general professional training: a Scottish approach. British dental journal. 2001;190(1):41-4.

Prescott L, Norcini J, McKinlay P, Rennie J. Facing the challenges of competency-based assessment of postgraduate dental training: Longitudinal Evaluation of Performance (LEP). Medical education. 2002;36(1):92-7.

Schumacher DJ, West DC, Schwartz A, Li S-T, Millstein L, Griego EC, et al. Longitudinal assessment of resident performance using entrustable professional activities. JAMA network open. 2020;3(1):e1919316-e.

360° Evaluations (Multi-Source Feedback)

Return to Table of Contents

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.

Berk and colleagues (2009) identified three main purposes for 360° evaluations in a clinical teaching setting including:

- Formative decisions and feedback about teaching improvement
- Summative decisions and feedback for merit pay and contract renewal
- Formative decisions and feedback about professional behaviors in the academic setting.

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.

Suggested CBVE Domains and Competencies:

- 5.1-5.3
- 6.1-6.4
- 7.1-7.5

Documented Uses and Examples:

Implementing 360° evaluations for clinical teaching faculty generally requires separate surveys for the 3 broad categories of raters:

- Co-workers/colleagues/administrators
- Clients/patients

- Students/alumni

Several studies have described the development and validation of surveys for specific disciplines within medicine:

- A systematic review of the quality and utility of observer-based instruments for assessing medical professionalism (Lelliott et al., 2008)
- Assessing the practicing physician using patient surveys: a systematic review of instruments and feedback methods (Evans et al., 2007)
- Developing specialty-specific multisource feedback tools (Bindels et al., 2019)
- Evaluation of physicians' professional performance: An iterative development and validation study of multisource feedback instruments (Overeem et al., 2012)
- Validation of the INCEPT: A multisource feedback tool for capturing different perspectives on physicians' professional performance (van der Meulen et al., 2017)

Pros and 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).

References:

Archer J, McGraw M, Davies H. Republished paper: Assuring validity of multisource feedback in a national programme. Postgrad Med J 86:526-531, 2010.

Berk RA. Using the 360 multisource feedback model to evaluate teaching and professionalism. Med Teach 31:1073-1080, 2009.

Bindels E, Boerebach B, van der Meulen M, Donkers J, van den Goor M, Scherpbier A, Lombarts K, Heeneman S. A new multisource feedback tool for evaluating the performance of specialty-specific physician groups: Validity of the group monitor instrument. J Contin Educ Health Prof 39:168-177, 2019.

Donnon T, Al Ansari A, Al Alawi S, Violato C. The reliability, validity, and feasibility of multisource feedback physician assessment: A systematic review. Acad Med 89:511-516, 2014.

Evans R, Elwyn G, Edwards A. Review of instruments for peer assessment of physicians. BMJ 328:1240, 2004.

Evans RG, Edwards A, Evans S, Elwyn B, Elwyn G. Assessing the practising physician using patient surveys: A systematic review of instruments and feedback methods. Family Pract 24:117-127, 2007.

Kwan YH, Png K, Phang JK, Leung YY, Goh H, Seah Y, Thumboo J, Ng APSC, Fong W, Lie D. A systematic review of the quality and utility of observer-based instruments for assessing medical professionalism. J Grad Med Educ 10:629-38, 2018

Lelliott P, Williams R, Mears A, Andiappan M, Owen H, Reading P, Coyle N, Hunter S. Questionnaires for 360-degree assessment of consultant psychiatrists: Development and psychometric properties. Brit J Psych 193:156-160, 2008.

Lockyer J, Violato C, Fidler H. Likelihood of change: A study assessing surgeon use of multisource feedback data. Teach Learn Med 15:168-174, 2003.

Lockyer JM, Violato C, Fidler HM. What multisource feedback factors influence physician self-assessments? A five-year longitudinal study. Acad Med 82:S77-S80, 2007.

Overeem K, Wollersheim HC, Arah OA, Cruijsberg JK, Grol RP, Lombarts KM. Evaluation of physicians' professional performance: An iterative development and validation study of multisource feedback instruments. BMC Health Services Res 12:1-11, 2012.

Riveros R, Kimatian S, Castro P, Dhumak V, Honar H, Mascha EJ, Sessler DI. Multisource feedback in professionalism for anesthesia residents. J Clin Anesth 34:32-40, 2016.

Rodgers KG, Manifold C. 360-degree feedback: possibilities for assessment of the ACGME core competencies for emergency medicine residents. Acad Emerg Med 9:1300-1304, 2002.

Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. Adv Health Sci Educ 13:275-288, 2008.

van der Meulen MW, Boerebach BC, Smirnova A, Heeneman S, oude Egbrink MG, van der Vleuten CP, Arah OA, Lombarts KM. Validation of the INCEPT: A multisource feedback tool for capturing different perspectives on physicians' professional performance. J Cont Educ Health Prof 37:9-18, 2017.

van der Meulen MW, Arah OA, Heeneman S, oude Egbrink MG, van der Vleuten CP, Lombarts KM. When feedback backfires: Influences of negative discrepancies between physicians' self and assessors' scores on their subsequent multisource feedback ratings. J Cont Educ Health Prof 41:94-103, 2021.

Violato C, Lockyer J, Fidler H. Multisource feedback: A method of assessing surgical practice. BMJ 326:546-548, 2003.

Wood L, Hassell A, Whitehouse A, Bullock A, Wall D. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. Med Teach 28:e185-e91, 2006a.

Wood L, Wall D, Bullock A, Hassell A, Whitehouse A, Campbell I. ‘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 28:e177-e84, 2006b.

Wright C, Richards SH, Hill JJ, Roberts MJ, Norman GR, Greco M, Taylor MR, Campbell JL. Multisource feedback in evaluating the performance of doctors: The example of the UK General Medical Council patient and colleague questionnaires. Acad Med 87, 1668-78, 2012.

Portfolios

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.2
- 3.1-3.3
- 4.1. 4.2
- 5.3
- 6.1, 6.3, 6.4
- 7.2, 7.3, 7.4
- 8.1-8.3
- 9.1-9.3

Documented Uses and Examples:

Numerous examples can be found at: https://uwaterloo.ca/centre-for-teaching-excellence/resources/integrative-learning/eportfolios/examples-student-eportfolios

https://jvme.utpjournals.press/doi/full/10.3138/jvme.0113-016R

https://jvme.utpjournals.press/doi/full/10.3138/jvme.0917-128r1

https://www.squarespace.com/websites/create-a-portfolio/?channel=pnb&subchannel=go&campaign=pnb-go-us-en-verticals_portfolio_tier1-e&subcampaign=(portfolio_portfolio-examples_e)&&utm_source=google&utm_medium=pnb&utm_campaign=pnb-go-us-en-verticals_portfolio_tier1-e&utm_term=portfolio%20examples&gclid=Cj0KCQjw7MGJBhD-ARIsAMZ0eesrSG8AXB6PhpytVP1wi1dFoU7EoESDHePmiz8ud7Qer88y_iFPK18aAmq_EALw_wcB&gclsrc=aw.ds

Pros and 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

References:

Ciesielkiewicz M. The use of e-portfolios in higher education: From the students' perspective. Issues Educ Res 29(3):649-667, 2019.

Davis MH, Ponnamperuna GG. Portfolio assessment. J Vet Med Educ 32(3):279-284, 2005.

Driessen EW, van Tartwijk J, Overeem K, Vermunt JD, van der Vleuten CPM. Conditions for successful reflective use of portfolios in undergraduate medical education. Med Educ 39:1230-1235, 2005.

Handel M, Wimmer B, Ziegler A. E-portfolio use and its effect on exam performance – a field study. Studies Higher Educ 2018 https://doi.org/10.1080/03075079.2018.1510388

Miller R, Morgaine W. The benefits of e-portfolios for students and faculty in their own words. Peer Review, 2009, published by the Association of American Colleges and Universities. https://www.aacu.org/publications-research/periodicals/benefits-e-portfolios-students-and-faculty-their-own-words

Pool AO, Govaerts MJ, Jaarsma DADC, Driessen EW. From aggregation to interpretation: How assessors judge complex data in a competency-based portfolio. Adv Health Sci Educ 23:275-287, 2018.

Entrustment-Supervision Scales

Return to Table of Contents

Description:

An Entrustable Professional Activity (EPA) is a unit of professional practice that can be entrusted to a learner, once the student has demonstrated the necessary level of competence to perform the activity with minimal supervision. EPAs should feature legitimate activities that are part of everyday veterinary practice and are those activities that can be completed only in a clinical setting or context. The main purpose of EPAs is to operationalize competency-based education through safe engagement of a learner in an actual practice setting. They link progressive development of the trainee alongside progressive autonomy from the supervisor. Terminology is critical to describing EPAs and entrustment correctly. A number of different entrustment scales exist for assessing EPAs.

- Schumacher DJ, ten Cate O, Damodaran A, Richardson D, Hamstra SJ, Ross S, Hodgson J, Touchie C, Molgaard L, Gofton W, Carraccio C & on behalf of the ICBME Collaborators (2021) Clarifying essential terminology in entrustment. Med Teach 43(7): 737-744, 2021. https://doi:10.1080/0142159X.2021.1924365
- ten Cate O, Schwartz A, Chen HC. Assessing trainees and making entrustment decisions: On the nature and use of entrustment-supervision scales. Acad Med 95:1662–1669, 2020. https://doi:10.1097/ACM.0000000000003427

Suggested CBVE Domains and Competencies Assessed:

CBVE features 8 EPAs in booklet 2.

Documented Uses and Examples:

A veterinary degree should guarantee safety and readiness for individual practice but the total scope of assessing that is more than one evaluator can do alone. EPAs are snapshots that can be summed to form a complete picture – programmatic assessment (ten Cate, 2013; ten Cate et al., 2015).

In the health professions, there are five main levels of entrustment scale that have been described (ten Cate et al., 2021, Entrustment decision making: Extending Miller's pyramid. https://doi:10.1097/ACM.0000000000003800):

- Level 1: the learner is allowed to be present and observe, not to enact an EPA.
- Level 2: the learner is allowed to execute the EPA with direct, pro-active supervision, present in the room.
- Level 3: the learner is allowed to carry out the EPA without a supervisor in the room, but quickly available if needed, i.e. with indirect, reactive, supervision.
- Level 4: the learner is allowed to work unsupervised.

- Level 5: the learner is allowed to provide supervision to more junior learners.

Veterinary examples include:

- Molgaard LK, Chaney KP, Bok HGJ, Read EK, Hodgson JL, Salisbury SK, Rush BR, Ilkiw JE, May SA, Danielson JA, Frost JS, Matthew SM. Development of core entrustable professional activities linked to a competency-based veterinary education framework. Med Teach 41(12):1404-1410, 2019. https://doi:10.1080/0142159X.2019.1643834
- Duijn CCMA, ten Cate O, Kremer WDJ, Bok HGJ. The development of entrustable professional activities for competency-based veterinary education in farm animal health. J Vet Med Educ 46(2):218-224, 2019. https://doi:10.3138/jvme.0617-073r
- Salisbury SK, Rush BR, Ilkiw JE, Matthew SM, Chaney KP, Molgaard LK, May SA, Bok HGJ, Hodgson JL, Frost JS, Read EK. Collaborative development of core entrustable professional activities for veterinary education. J Vet Med Educ 47(5):607-618, 2020. https://doi:10.3138/jvme.2019-0090
- Favier RP, Godijn M, Bok HGJ. Identifying entrustable professional activities for surgical skills training in companion animal health. Vet Rec 186(4):122, 2020. https://doi:10.1136/vr.105386
- Favier RP, ten Cate O, Duijn C, Bok HGJ. Bridging the gap between undergraduate veterinary training and veterinary practice with entrustable professional activities. J Vet Med Educ 48(2):136-138, 2021. https://doi:10.3138/jvme.2019-0051

Pros and 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 work-place based activities, but current EPAs do not include evaluation of all competencies in the CBVE framework.

References:

Bok HGJ, de Jong LH, O'Neill TO, Maxey C, Hecker KG. Validity evidence for programmatic assessment in competency-based education. Perspect Med Educ 7:362-372, 2018. https://doi.org/10.1007/s40037-018-0481-2

Duijn C, Bok H, ten Cate O, Kremer W. Qualified but not yet fully competent: Perceptions of recent graduates on their day-one skills. Vet Rec 186(7):216, 2020. https://doi:10.1136/vr.105329

Melvin L, Rassos J, Stroud L, Ginsburg S. Tensions in assessment: The realities of entrustment in internal medicine. Acad Med 95:609-615, 2020. https://doi:10.1097/ACM.0000000000002991

Read EK, Brown A, Maxey C, Hecker KG. Comparing entrustment and competence: An exploratory look at performance-relevant information in the final year of a veterinary program. J Vet Med Educ Mar 4:e20190128, 2021. https://doi:10.3138/jvme-2019-0128

Shorey S, Lau TC, Lay ST, Ang E. Entrustable professional activities in health care education: A scoping review. 53(8):766-777, 2019. https://doi:10.1111/medu.13879

ten Cate O, Taylor DR. The recommended description of an entrusted professional activity: AMEE Guide No.140. Med Teach, 2020. https://doi:10.1080/0142159X.2020.1838465

Violato C, Cullen MJ, Englander R, Murray KE, Hobday PM, Borman-Shoap E, Ersan O. Validity evidence for assessing entrustable professional activities during undergraduate medical education. Acad Med 96(7S):S70-S75, 2021. https://doi:10.1097/ACM.000000000000409

Capstone Assignments

Return to Table of Contents

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.

Suggested CBVE Domains and Competencies Assessed:

Due to wide variety of capstone experiences, a variety of CBVE domain(s) may be assessed. For this reason, two examples are listed below for suggested CBVE domains and competencies, based on two different capstone experiences.

Example 1: Workplace-based experience: Assessment of selected Entrustable Professional Activities

- 1.1-1.7
- 2.1-2.2
- 5.1-5.3
- 7.2

Example 2: Research experience: Analysis of adverse health outcomes

- 4.2
- 5.3
- 8.3
- 9.1-9.3

Documented Uses and Examples:

Capstones are pervasive in both undergraduate and graduate coursework as well as across disciplines, from the humanities to the STEM fields. To engage students in critical reflection about the meaning of professional medical practice, the Northeastern Ohio Universities College of Medicine instituted a month-long course rooted in the humanities, through media such as poetry readings, short fiction, essays, and films. The course culminated in critical reflective writing, a "personal oath statement," and a "letter to a 3rd year student" (Wear and Zarconi, 2006).

With an emphasis on creativity, leadership, global citizenship, and diversity of professional perspectives, final year occupational therapy students from Australia engaged in a project-based global health internship, in which the students partnered with community health organizations in India and Vietnam to address a community health need. Students in the program were challenged by the high expectations and high levels of freedom afforded to them, as well as the development of cultural competency and cultural humility (Fortune et al., 2019). Other models of global health capstones exist at other schools as well.

In the Global Medicine Program at the University of Illinois Chicago College of Medicine, student participate in a longitudinal, four-year capstone project, that incorporates specific deliverables each year of the curriculum (Chamberlain et al., 2020).

A mastery learning capstone course at the Feinberg School of Medicine at Northwestern University focused on deliberate practice with feedback for three selected AAMC EPAs. While there was varied competence across the EPAs on baseline evaluation, following the capstone, all 130 students met or exceeded minimum standards on two EPAs, and 84% of students met or exceeded minimum standards on the third EPA (Saltzman et al., 2019).

In the context of high-stakes assessments as capstone experiences, Colorado State University College of Veterinary Medicine and Biomedical Sciences created a three-part capstone examination in their curriculum, modeled on the USMLE step examinations. The written/multiple choice assessments are delivered either online or in class at the start of second year, the start of third year, and just before clinics (Avery et al., 2020).

In the University of Illinois College of Veterinary Medicine, capstone or milestone written and OSCE examinations are given in March of both the second and third years of the curriculum (Foreman et al., 2017). Similarly, the faculty at the Cummings School of Veterinary Medicine at Tufts University are developing a Comprehensive Milestone Assessment to be administered at the end of the second year, which will consist of both written and practical components.

Pros and 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

References:

Avery AC, Dowers KL, West AB, Graham BJ, Hellyer P, Avery PR, Ballweber LR, Hassel DM, Oaks JF, Frye MA. Student, faculty, and program outcomes associated with capstone examinations administered to veterinary students at Colorado State University. J Am Vet Med Assoc 257(2):165-175, 2020.

Chamberlain S, Gonzalez N, Dobiesz V, Edison M, Lin J, Weine S. A global health capstone: an innovative educational approach in a competency-based curriculum for medical students. BMC Med Educ. 20(1):159, 2020.

Foreman JH, Morin DE, Graves TK, Mitchell MA, Zuckermann FA, Whiteley HE. Veterinary curriculum transformation at the University of Illinois, 2006-2016. J Vet Med Educ 44(3):471-479, 2017.

Fortune T, Borkovic S, Bhopti A, Somoza R, Nhan HC, Rangwala S. Transformative learning through international project-based learning in the global south: Applying a students-as-partners lens to a “high-impact” capstone. J Stud International Educ 23(1):49-65, 2019.

Reid KJ, Chiavaroli NG, Bilszta JLC. Assessing a capstone research project in medical training: Examiner consistency using generic versus domain-specific rubrics. J Med Educ Curricul Developm 9:23821205221081813, 2022.

Salzman DH, McGaghie WC, Caprio TW, Hufmeyer KK, Issa N, Cohen ER, Wayne DB. A mastery learning capstone course to teach and assess components of three entrustable professional activities to graduating medical students. Teach Learn Med 31(2):186-194, 2019.

Wear D, Zarconi J. A humanities-based capstone course in medical education: An affirming and difficult look back. J Learning through Arts 2(1):1-14, 2006.

Case Logs

Return to Table of Contents

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 allow 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.

Suggested CBVE Domains and Competencies Assessed:

- 1.1-1.7
- 2.1-2.2
- 3.1-3.3
- 4.1
- 7.2-7.4

Documented Uses and Examples:

Even though there has been some implementation, there is a paucity of peer-reviewed evaluation of the application and validity of case logs in veterinary medicine. Most case logging is documented in the human medical profession, especially in procedure-based specialties such as surgery, obstetrics-gynecology, and neurological surgery. Evaluation of cases logs by institutional programs may provide insight into the trainees' experiences, deficiencies, and needs.

Pros and 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 (Fronza et al., 2012; Suwonabol et al., 2009).

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).

References:

Balla F, Garwe T, Motghare P, Stamile T, Kim J, Mahnken H, Lees J. Evaluating coding accuracy in general surgery residents' Accreditation Council for Graduate Medical Education procedural case logs. J Surg Educ 73:59-63, 2016.

Fronza JS, Prystowsky JP, Darosa D, Fryer JP. Surgical residents' perception of competence and relevance of the clinical curriculum to future practice. J Surg Educ 69(6): 792-797, 2012.

Gisondi MA, Regan L, Branzetti J, Hopson L. More learners, finite resources, and the changing landscape of procedural training at the bedside. Acad Med 93(5):699-704, 2018.

Greenberg LW, Getson P. Clinical experiences of medical students in a pediatric clerkship and performance on the NBME subject test: Are they related? Med Teach 21:420-423, 1999.

Neumayer L, McNamara RM, Dayton M, Kim B. Does volume of patients seen in an outpatient setting impact test scores? Am J Sug 175:511-514, 1998.

Salazar D, Schiff A, Mitchell E, Hopkinson W. Variability in Accreditation Council for Graduate Medical Education Resident Case Log System practices among orthopedic surgery residents. J Bone Joint Surg Am 96(3):e22, 2014.

Suwanabol PA, McDonald R, Foley E, Weber SM. Is surgical resident comfort level associated with experience? J Surg Res 156(2):240-244, 2009.

Thanawala R, Jesneck J, Shelton J, Rhee R, Seymour N. Overcoming system factors in case logging with artificial intelligence tools 79:1024-1030, 2022.