

Chapter 7

Assessment of Clinical Reasoning Using the CBCR Test

Olle ten Cate

A CBCR course for preclinical students should be completed with some form of student evaluation, as any course in a medical curriculum should be concluded with a valid decision about the extent to which every student has reached the objectives of the course. Many students proceed through education with one predominant, returning question in mind: *what must I do to pass this course and its examination?* This guides their efforts in learning, and as the saying goes, “assessment drives learning” or “students rather learn what you *inspect* than what you *expect*.” This is not something to be disappointed about, but a fact of university life that should be understood and respected. After all, it is the school that sets these rules. However, the lesson to be learned is that assessment should be aligned with the educational objectives in such way that what we inspect is exactly what we expect (Biggs 1996). If we test students on clinical reasoning skills, the examination should rather not consist of four-item multiple-choice or true/false questions that focus on factual knowledge. That would drive students in the direction of rehearsing the systems approach to biomedical knowledge that, as we have shown, is in contrast with the patient-oriented approach in CBCR sessions. Assessment needs to focus more specifically on clinical reasoning skill, acknowledging however the importance of biomedical knowledge for clinical reasoning. That should include questions like: What, at this stage, is a likely differential diagnosis? What findings would you expect to find with physical examination if hypothesis X were true? Which laboratory tests of the ordering list would you check for this patient at this stage? In other words, multiple options from a larger array of possibilities can suit the assessment of clinical reasoning better than standard four-option MC questions.

O. ten Cate (✉)
Center for Research and Development of Education, University Medical Center Utrecht,
Utrecht, The Netherlands
e-mail: T.J.tenCate@umcutrecht.nl

© The Author(s) 2018
O. ten Cate et al. (eds.), *Principles and Practice of Case-based Clinical Reasoning Education*, Innovation and Change in Professional Education 15,
https://doi.org/10.1007/978-3-319-64828-6_7

The CBCR Test as Developed at the University Medical Center Utrecht

In 2010 a specific written assessment format was developed for the CBCR course at the University Center Utrecht, which has been in place ever since that time. The aim of the design was to align the test as closely as possible with both the CBCR course as delivered and the desired future skill of patient-oriented clinical reasoning.

The case-based questions all follow the course of the clinical encounter, starting with a case title that reflects the initial information the physician normally would have (age, gender, and main complaint) and a short clinical presentation vignette. Then, a series of questions about the case unfolds.

Features from different established item types as discussed in Chap. 5 have been incorporated:

- Relatively long lists of options, such as used in *extended matching* questions and long menu items
- A differential diagnostic approach from the *comprehensive integrated puzzle* items through the “alternative scenarios” approach
- The addition of new information to alter a hypothesis as used in the *script concordance test*.

The first and foremost feature of the Utrecht CBCR test is the integrated, patient-oriented nature of the test. The unit of focus is a patient with several option lists (diagnostic hypotheses, history findings, physical examination findings, diagnostic findings, management options, etc.) that all pertain to focused questions following from the initial patient presentation or from follow-up information about this patient. The test has a limited number of cases but a broad enough set to reduce the threat of case specificity and has a series of questions within each case. The CBCR test has the following characteristics.

Alignment with Actual Cases Discussed

An important purpose of the CBCR course for junior students is to mentally “install” a basic framework of a limited number of illness scripts. The Utrecht CBCR test therefore closely follows the cases discussed during the CBCR sessions. There is no expectation of substantial *transfer of learning*, i.e., of a benefit of studying cases for the ability to handle other cases and no aim to test that transfer ability early in the medical curriculum. Indeed, an additional aim of closely following the cases studied is to reinforce the learned scripts both by rehearsing for the test and during the test itself. For senior medical students, who have acquired a basic mental framework of illness scripts in their long-term memory, it may well be recommended to deviate from cases discussed, to gradually test a more general clinical reasoning ability, but this is not the aim of the end-of-course CBCR test for the preclinical students.

Cases, Options Lists, and Scenarios

The case is the unit of focus for test items. The question starts with a short case vignette that usually resembles an initial presentation of the patient at the primary care doctor's office, the emergency department, or elsewhere, similar to the start of a CBCR education case. Several option lists that may vary in length from 5 to 25 or more options accompany the case presentation. These lists include the following nine optional categories: (a) diagnoses, (b) history questions, (c) history findings, (d) physical examination procedures, (e) physical examination findings, (f) diagnostic test options, (g) diagnostic test findings, (h) management options, and (i) prognosis. Usually, a limited number of lists (4 or 5) will be used for one case. The distinction between history questions and history findings is that either the question can be "What are the [two] most relevant next questions to ask?" or "Which [three] history findings would you expect if hypothesis X were true?" (or equivalent questions). This also holds for physical examination and diagnostic tests. All questions may ask to check as many options from a list as the item writer finds suitable, e.g., "Which four lab findings would you expect to find given what you know about this patient?" However, the number of correct options should not be more than one third of the total list, and preferably much less, to avoid successful guessing.

Questions may be about a differential diagnosis following from the initial case vignette, but cases generally include one or more sequential *scenarios*. A scenario is defined as a deviation from the initial course of happenings or findings, but still relates to the same patient (age, sex, main complaint, and initial vignette). A scenario usually starts like this. "Scenario B. Presume, the radiology of this patient's thorax shows a lump in the left lower lung, which two hypotheses from the list (a) Diagnoses would now be most likely?" Or "Scenario C. Presume, the radiology of this patient's thorax would show no pathology, which two hypotheses from the list (a) Diagnoses would now be most likely?" Case title, initial vignette, and all options lists are identical throughout a case, but correct options (and the number of requested options) vary.

In practice, most cases include one to three scenarios. When there is only one scenario, the word *scenario* is not used.

CBCR Test Quality Findings Since 2010

Between December 2010 and April 2017, the test has been administered 14 times, with on average about 12 cases and 50 items (about four items per case), for cohorts of about 300 students. The test reliability averaged 0.73 (Cronbach's alpha) for tests with an average duration of not much more than 1 h. In this course, spread across 8 months, a test is administered twice (in December and April), and the final score

combines both sub-scores. This combined score would compare with a test with 24 cases, 2–2.5 h of testing and an estimated average reliability of 0.84 (estimated with the Spearman-Brown formula). This is more than satisfactory and more efficient than key feature tests (Page and Bordage 1995).

The test questions were derived from small group CBCR cases, supplemented with cases presented in a lecture hall. We will not expand on the latter cases, which constituted a minority of the items, but they were suitable for a similar approach to clinical reasoning testing, given the large group education format (Borleffs et al. 2003). Given the richness of the cases discussed in the CBCR sessions, it was possible to devise new questions for each test, hardly without using any previous questions.

Electronic and Paper Versions

All CBCR tests except one were administered electronically. The first version of the test was not much more than a protected interface. After logging in with a general password, candidates could access all questions per case on one screen: The left side of the screen showed a series of questions and open fill-in-the-blank-slots; on the right side of the screen, all relevant options lists were displayed. Candidates were then asked to enter a three-digit number into the relevant slots at the left side of the screen. The resulting file was exported as an Excel® file, available after the test administration and suitable for analysis with an elaborate Excel® analysis application and statistical software.

A next version of the application showed a professional interface with check boxes instead of fill-in-the-blank slots. Currently, a commercial test administration firm, TestVision® (www.testvision.nl), has incorporated the CBCR item format requirements into a professional application.

What the Utrecht CBCR Test Does Not Provide

The Utrecht CBCR test approach has limitations. In clinical reasoning, a logical question is “what is the most likely diagnosis?”. All question types that are not constructed response suffer from some cueing, as the answer can be chosen from a list. Script concordance testing and the comprehensive integrative puzzle approaches have simply given up on this requirement as diagnostic hypotheses are given and not asked (Charlin et al. 2000; Ber 2003; Groothoff et al. 2008). CBCR test questions provide a list to choose from that can be as long as the item writer wishes, thereby somewhat limiting this cueing, similar to extended matching questions and long menu questions (Case and Swanson 1998; Schuwirth et al. 1996). The recommended length is 20 options. In practice, they are often shorter and sometimes longer.

Another feature that is not supported is the possibility to evaluate a chain of interdependent reasoning questions and answers of a candidate, requiring conditional links between consecutive answers (*if the student chooses X on item 1, then item 2*

will be adapted). That possibility is not provided. It would be possible to value an answer differently if it follows upon a previous *wrong* answer, as the two answers may be correctly related to each other. If the chain of reasoning becomes longer, however, the potential branches to be evaluated would quickly become too many to manage. The use of parallel scenarios however compensates this by the possibility of branching options from the same patient. A question of a parallel scenario always starts with “Presume,..” e.g., “Presume, you have received result X from diagnostic test Y, what then would be the most likely diagnosis?”.

Finally, the current use of CBCR test methodology does not allow for weighing of items within answers (“Provide a differential diagnosis of three in a correct order of likelihood”); that now requires multiple questions (“What is the most likely diagnosis?” and “Name two other diagnostic hypotheses”). These are technical limitations that in the future may be solved with sophisticated software.

Rules and Regulations Around the Utrecht CBCR Test

In practice, the Utrecht CBCR test is administered twice a year, each for half of the final score. Students pass the course requirements if their final test score, combined with proof of active participation in the course as a student and as a peer teacher, is satisfactory. As participation also yields a score, test and participation scores are combined to cover 88% and 12% of the final score, respectively, which was found to be a useful ratio. We will not expand on how the participation rate per student is calculated, but details can be found in the model study guide in Chap. 10. Students who do not pass the requirements can opt for a retake of the examination.

Issues of Validity of the CBCR Test

The Utrecht CBCR tests as applied since 2010 have an undisputed content validity, as they are built upon cases that are used in the course, and they cover all cases.

The construct validity of the Utrecht CBCR test approach remains to be investigated.

Example

Table 7.1 shows an example of a CBCR test question, derived from a case that is used in education. The representation of the question may have different forms when presented as an actual test. The initial case vignette should remain visible while students proceed with scenarios through the case. Supplementary visual information, such as a photo of the patient, an X-ray image, or others may appear when indicated.

Table 7.1 Example of a CBCR case translated to test questions

<p>Case 1. A 36-year old woman with headache A 36-year old female patient presents at her family doctor with a severe headache.</p>	
<p>Case 1, Scenario A The patient has a diffuse, clamping headache, occurring frequently in the last month. Last year she had similar complaints when she was very busy at work. On physical examination you find no abnormalities.</p> <p>1. Which diagnosis is most likely? Choose 1 option from 'diagnoses' 2. Which diagnostic test option is now indicated? Choose 1 option from 'diagnostic test options'. 3. Which management options are now suitable? Choose 3 options from 'management options'.</p>	
<p>Case 1, Scenario B Presume, that the patient had experienced several episodes of severe headache during the past two months. In the past she has had similar complaints. There is no fever.</p> <p>4. One of the diagnoses you consider is migraine. Which history features would you expect to find if this diagnosis is true? Choose 5 options from 'history features'</p> <p>Presume, you are thinking of a cluster headache. 5. Which findings during physical examination make this diagnosis more likely? Choose 2 options from 'physical examination'.</p>	
<p>Case 1, Scenario C Based on the history and physical examination you consider temporal arteritis and cluster headache. 6. Which diagnostic test options can help to distinguish between them? Choose 2 options from 'diagnostic test options'.</p>	
<p>Diagnoses Acute glaucoma Alcohol hangover Analgesics headache Cerebral tumor Cluster headache Dentogenic headache Encephalitis Hypoglycemia Influenza Meningitis Migraine Posttraumatic headache Pre-eclampsia Refraction abnormalities Sinusitis Subarachnoid hemorrhage Subdural or epidural hematoma Temporal arteritis Tension headache Trigl neuralgiaemia</p>	<p>History findings Alcohol provokes the complaints Duration: 1 to 2 days Duration: several minutes Duration: several weeks First epileptic attack 1 week ago Menstruation provokes the complaints Morning vomiting Pain depends on position Pain is bilateral Pain is clamping Pain is throbbing Pain is unilateral Pain started acute Patient has dental problems Patient is 8 months pregnant Patient is 8 weeks pregnant Patient is hypersensitive to light and noise Patient is nauseous and vomits Patient sees flashes of light Patient uses painkillers in high doses Recent personality change</p>
<p>Physical examination findings Decreased visual acuity Fever Increased blood pressure Increased tear production Miosis (pupillary constriction) Neck stiffness Neurological impairment Pain on palpation of temporal artery Reduced level of consciousness</p>	<p>Diagnostic test options Biopsy temporal artery Cell count and differentiation CT-angiography CT-scan or MRI-scan EEG ESR No diagnostic tests Rontgen photo neck/skull</p>
<p>Management options Discuss measures to achieve stress reduction Headache diary Refer to specialist Send for emergency treatment Strong painkillers (opioids) Watchful waiting, unless complaints worsen</p>	

Scoring of Items

Once students have taken the test, a file results with all answers. A common format is an Excel® file with rows per student and columns of answers per option. See Fig. 7.1

for an example. The unit of scoring is the option. If Question 1 from Case 1, Scenario A, asks for a differential diagnosis of four hypotheses that need to be considered with this patient, all students will have an answer for a, b, c, d. Scores are counted for each question into a sum score of 0 to 4 or 5, depending on the number of options requested in that question. For psychometric purposes (calculation of Cronbach’s alpha reliability and item analysis) the unit of analysis is the question (but there are arguments to use scenarios or cases as units of analysis). As with any test, we recommend to conduct an item analysis to determine whether any items need to be removed before final scores for students are disclosed. Those final scores can be calculated in different ways. The Utrecht procedure is to first calculate the mean guessing rate per item (e.g. 20 or 25%) and take that percentage of the maximum score as a bottom score. E.g., a test of 50 questions and 120 options, with an average guessing score of 21% yields a bottom score of 25. Subtracting that from the maximum score of 120 means that all students receive a score between 0 and 95 points. As we want to end the CBCR course with a final score that combines two test scores (88%) and a score for participation (12%), and because we use a 100-point scale, each test score must be recalculated to a 0-44 points range.

Case	Case 1												Case 2											
Scenario	Case 1 – Scenario A						Case 1 – Scen. B						Case 2 – Scen. A				Case 2 – Scen. B							
Question	1		2		3		4			5			6		7		8		9		10			
Option	a	b	c	d	a	b	a	a	b	c	d	a	b	c	a	a	a	b	a	b	a	b	c	d
Student 1																								
Student 2																								
Student 3																								
Et cetera																								

Fig. 7.1 Possible data format for analysis

Checklist for Item Writers

We end with a checklist for writers of test questions for CBCR-tests. Box 7.1 includes a number of pitfalls to avoid and recommendations to follow.

Box 7.1 Checklist for Item Writers of the Utrecht CBCR Test

1. Always include *age, sex* and *main complaint*, and *sign or symptom* in the case title.
2. Always relate to *this* individual patient. Instead of “Which two physical examination findings do patients with complaint X always have,” ask: “Which two physical examination features do you expect to find when examining this patient, based on her complaint X?”. Stimulate students to think from a patient-oriented perspective, also during the test. “...do you expect to find...” is the aimed typical hypothesis-driven thinking mode and is regularly used in CBCR test items. In many cases, it is sensible to add “...if this hypothesis is correct.” Check whether an answer requires a preceding case vignette; if not, it is probably not a question about the individual patient.
3. When writing items, stay close to how the information arrived at the physician, e.g., the history as the patient presents its physical examination as what the physicians sees or hears), rather than interpreting and summarizing with semantic qualifiers (avoid “The patient provides a family history of cardiovascular disease,” and rather “...has yellow sclerae” than “...is icteric”), unless the presentation would be a discharge letter from a hospital.
4. Be specific about the number of requested options from the list (and avoid “at least” or “maximum number” of options to be checked).
5. Make sure that the list contains options that do not overlap or include each other. Also, do not include two items in the list that evidently exclude each other, while the list contains other options (“age younger than 30, age 30 or older, and age 50” should not all be used in one listing).
6. Formulate finding options specific rather than general (*Age 46* rather than *older than 40*; “Pain since three weeks” rather than “Pain since quite some time”); again formulate how this information is presented by the patient.
7. Intersperse questions with small follow-up information text. *The following lab results are reported: [...]*; then continue with a next question.
8. Start a new *scenario* if there is a deviation from earlier information or questions. The first question of a new scenario typically starts with “Scenario B. Presume, the patient had said/shown/....” If there is no branching or deviation from earlier information, “scenario” terminology is not needed. Students must understand the significance of the “scenario” terminology in their instruction about this test.

(continued)

Box 7.1 (continued)

9. Avoid linking a question to a previous question (e.g., “What are the two best management options for this diagnosis” after the question “What is the most likely diagnosis”) as this requires sophisticated analysis technology, unless the questions are scored by hand. The solution is to start with “Presume, the diagnosis X has been confirmed, what would then be...” Make sure that this diagnosis does not disclose the previous answer, e.g., by choosing as X something that is *not* the most likely diagnosis. Likewise, do not use the option lists *history questions*, *physical examination procedures*, and *diagnostic test options* in such way that a follow-up text or scenario reveals the answer to a previous question.
10. Try to find an adequate representation of questions across cases and an adequate balance of questions about history, physical examination, diagnostic tests, and management.
11. Make sure a CBCR test question is always reviewed by a colleague *without the model answer* before it is accepted. One reason is that option lists may include more items that must be considered correct than initially anticipated. A recommendation is to ask a few senior students to take the draft test. That should reveal most major flaws.

References

- Ber, R. (2003). The CIP (comprehensive integrative puzzle) assessment method. *Medical Teacher*, 25(2), 171–176.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347–364.
- Borleffs, J. C. C., et al. (2003). “Clinical reasoning theater”: A new approach to clinical reasoning education. *Academic Medicine: Journal of the Association of American Medical Colleges*, 78(3), 322–325.
- Case, S. M., & Swanson, D. B. (1998). *Constructing written test questions for the basic and clinical sciences* (2nd ed.). Philadelphia: National Board of Medical Examiners.
- Charlin, B., et al. (2000). The script concordance test: A tool to assess the reflective clinician. *Teaching and learning in medicine*, 12(4), 189–195.
- Groothoff, J. W., et al. (2008). Growth of analytical thinking skills over time as measured with the MATCH test. *Medical Education*, 42(10), 1037–1043.

- Page, G., & Bordage, G. (1995). The Medical Council of Canada's key features project: A more valid written examination of clinical decision-making skills. *Academic Medicine*, 70(2), 104–110.
- Schuwirth, L. W., et al. (1996). Computerized long-menu questions as an alternative to open-ended questions in computerized assessment. *Medical Education*, 30(1), 50–55.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

