Assessing Clinical Reasoning: The Script Concordance Test

Dr. Linda O’Connor
Course Chair – 2nd Year Clinical Skills
Objectives

• Review why the SCT has been developed
• Review guidelines for construction of the SCT
• Lessons learned administering the SCT to medical students
Evaluating well-defined problems

• Current evaluation methods assess:
  – Factual knowledge (MCQ)
  – Behaviour (OSCE)

Assessing factual knowledge is great for assessing well-defined problems (all data is present, goals are clear, known solutions)
MCC Objectives - Scholar

• 2.0 Apply principles of research and information management to learning and practice
• 2.1 Describe the principles of evidence-based medicine
• 2.2 Retrieve information from appropriate sources
• 2.3 Assess the quality of information, using principles of critical appraisal:
  – 2.3.1 its relevance and importance
  – 2.3.2 the appropriateness of its methodology
  – 2.3.3 its conformity to ethical standards
• 2.4 Integrate retrieved information into clinical practice
• 2.5 Accept complexity, uncertainty, and ambiguity as part of medical practice
Evaluating ill-defined problems

How do we assess reasoning ability in clinical situations where uncertainty exists?

• Can use oral exams, but difficult to standardize and organize with large group of examinees
• Must take into account the variability in the clinical reasoning process of professionals
Script theory

• According to script theory, clinicians have networks of organized knowledge which have been developed over years of clinical encounters.

• Scripts are made up of links between illnesses, clinical features and management options.

• Clinicians retrieve these scripts to make judgments on the effect that each new piece of information has on a hypothesis or management option in order to progress toward solutions to clinical problems.
Script Concordance Test

Script Concordance testing is based on the principle that the steps in the clinical reasoning process can be assessed and compared to those of a panel of experts.
Who can use it?

• Concept originated by Feltovich in 1984, since then research has developed and validated the SCT

• Series of studies looking at fields in family medicine, surgery, gynecology, radiology, neurology and Emergency medicine have confirmed its reliability and construct validity of test scores (explores ability to interpret data in the making of clinical decisions, rather than simple recall of factual data)

• Can be used for assessment in undergraduate or postgraduate students or for CME/ small group learning sessions
Why is the SCT useful?

- Relatively easy to construct and score
- Reasonable time/ number of questions for validity
- Authentic/ face validity – clinicians find the test appealing because its cognitive tasks are the same as those they face in their daily practice; doesn’t assess trivial knowledge but rather explores real world clinical knowledge
- Most significant is the lack of “intermediate effect” – it has been repeatedly shown that novices score lower on SCT than expert clinicians
What is it?

• The test is case based. It starts with a clinical vignette which has uncertainty.
• This is followed by a series of questions in 3 parts:
  – First part (‘if you were thinking of’) contains a relevant diagnostic or management option
  – Second part (‘and then you were to find’) presents a new clinical finding eg a physical sign, pre-existing condition, imaging study or lab result
  – Third part (‘this option would become’) is a five point Likert scale that captures the examinees decisions
The examinee must decide what affect the new clinical finding in part 2 has on the information or decision given in part 1 and must note the direction (positive, negative or neutral) and intensity in part 3
Example

A 4 year old female presents to your family medical clinic with a diffuse erythematous maculo-papular exanthem of 24 hours duration. Her parents mention she attends daycare facilities fulltime with 8 other children.

If you suspect... And then... Your hypothesis is...

1. Viral exanthem
   You learn another child in her daycare has been diagnosed with strept throat. -2 -1 0 1 2

2. Measles
   The patient’s parents indicate she did not receive an MMR vaccination due to conscientious objection. -2 -1 0 1 2

3. Scarlet fever
   On physical exam you note peri-oral pallor and a strawberry red tongue. -2 -1 0 1 2

Circle one: -2 = greatly weakened, -1 = slightly weakened, 0 = no effect on hypothesis, 1 = strengthened, 2 = greatly strengthened
Test Construction

• Ask: what is the purpose of the evaluation?
  – Initiate a learning discussion
  – assess learning achievement in students or residents

• Consider content validity – the extent to which the test samples or covers the area of clinical reasoning under consideration
  – Relevant clinical scenarios to that specialty
  – Can measure diagnostic hypothesis, investigation (utility or risk-benefit) or management options (utility or risk-benefit)
Test Construction

• Team of 2 clinicians should write the questions

• Not necessarily complex or unusual, but should contain uncertainty in the case scenario and in the answer options (the situation must call for 2 diagnoses, investigation or management options)

• Even experts should not be able to provide one single solution to the problem
• The questions attached to a case should all address the same issue (ie all address diagnosis hypotheses or all address treatment issues)
  – Likert scale descriptors will be the same
  – So students don’t assume the information is cumulative

• Cases should not ‘build’ on each other ie probe diagnosis then investigation then treatment
How many cases, how many questions?

• Fewer cases with 2-3 questions per case is optimal (Norman, 2006)

• 20 cases with 50-60 questions for one hour of testing, reach Cronbach alpha values of 0.80 or higher

• Need to consider case specificity (successfully solving one case is not reflective of the ability to solve others) and have enough cases, but keep in mind that one question for each case is too cognitively demanding for examinees
Likert scale

• How many anchors?
  – 5 point anchor scales used most commonly
  – Research suggests a 3 point scale with single best answer may be just as effective (Bland, 2006)

• Structure of the scale should be the same for the whole test

• Use of wording is important
  – ‘Contraindicated’ or ‘indicated’ in investigation or treatment questions reinforces uncertainty by introducing a legal or risk-benefit issue
Examples of Anchor Descriptors

• If you were thinking of the following diagnosis
  -2 very unlikely  -1 unlikely  0 neither likely or unlikely
  +1 more likely  +2 very likely

• If you were considering the usefulness of the following investigation/treatment
  -2 useless  -1 less useful  0 neither more or less useful
  +1 useful  +2 very useful

• If you were considering the risk-benefit of the following investigation/treatment
  -2 strongly contraindicated  -1 contraindicated
  0 neither more or less indicated
  +1 indicated  +2 strongly indicated
Developing the Scoring Key

• Expert clinicians are asked to complete the test individually and their answers are used to develop the scoring key. (Charlin, 2002)

• The answer which most of the experts pick is given a value of 1. Answers not picked by the experts are given a value of 0. Other answers are given a proportionate value (equal to the number of experts who picked the answer divided by the modal value for the question).
• Questions where the experts all agree are discarded (no ambiguity, tests factual knowledge rather than reasoning skills)

• Questions where the experts all disagree are discarded (noise) (Charlin, 2006)
• Example
  • If 8/10 experts pick +2, then +2 would be worth 1 (8/8)
  • If 1 expert picks +1 then +1 would be worth 0.13 (1/8)
  • If 1 expert picks 0, then 0 would be worth 0.13 (1/8)
  • -1 and -2 would be worth 0

• Examinees then take the test and their score for each question is added to give a total score for the test (each question will have a maximum of 1 and a minimum of 0). That number is divided by the number of questions and multiplied by 100 to get a percentage score.
Expert Panel

How many experts? (Gagnon, 2005)
• For high stake examinations, panels should comprise 15 members (no increase of Cronbach’s alpha coefficient above 20 members)
• For low stake exams or formative assessment 10 experts

Who should the experts be?
Physicians with overall clinical experience that represent the goals of the students or residents who are being tested.
What we did

• Questions were developed by Brandon Girardi (Meds 2011) and myself
• Made up 16 cases and 50 questions and ended up discarding 9 questions.
• Panel of 10 expert Family Physicians completed the test online
• 34 students participated, but only 14 completed the online test
What we learned

• Convenient for the expert panel to complete the test online
• Students using an online format exited the test and looked up “answers”
• Students wanted incentive to do the test
• Expert panel should have an odd number of participants (won’t have to discard questions if 2 answers are split 50-50)
• StudentVoice.com is an online community to which Queen’s belongs and they will set the test up online. Contact person is Jennifer Massey at jennifer.massey@queensu.ca
Other interesting research with the SCT

• Study showed that assessing clinical reasoning skills by SCT (and not OSCE results) in clerks was predictive of reasoning skills 2 years later after residency training even if the tested domains were different (Brailovsky 2001)
• Used to assess ethics (Llorca 2003)
• Can be used to address specific educational objectives. Participants in a session answer the test individually, then have small group discussions to reach a consensus. It can be compared to answers made by an expert panel. If there are discrepancies, it can stimulate interesting discussion as the experts explain their reasoning (Labelle 2001).


