ERKNet

The European Rare Kidney Disease Reference Network

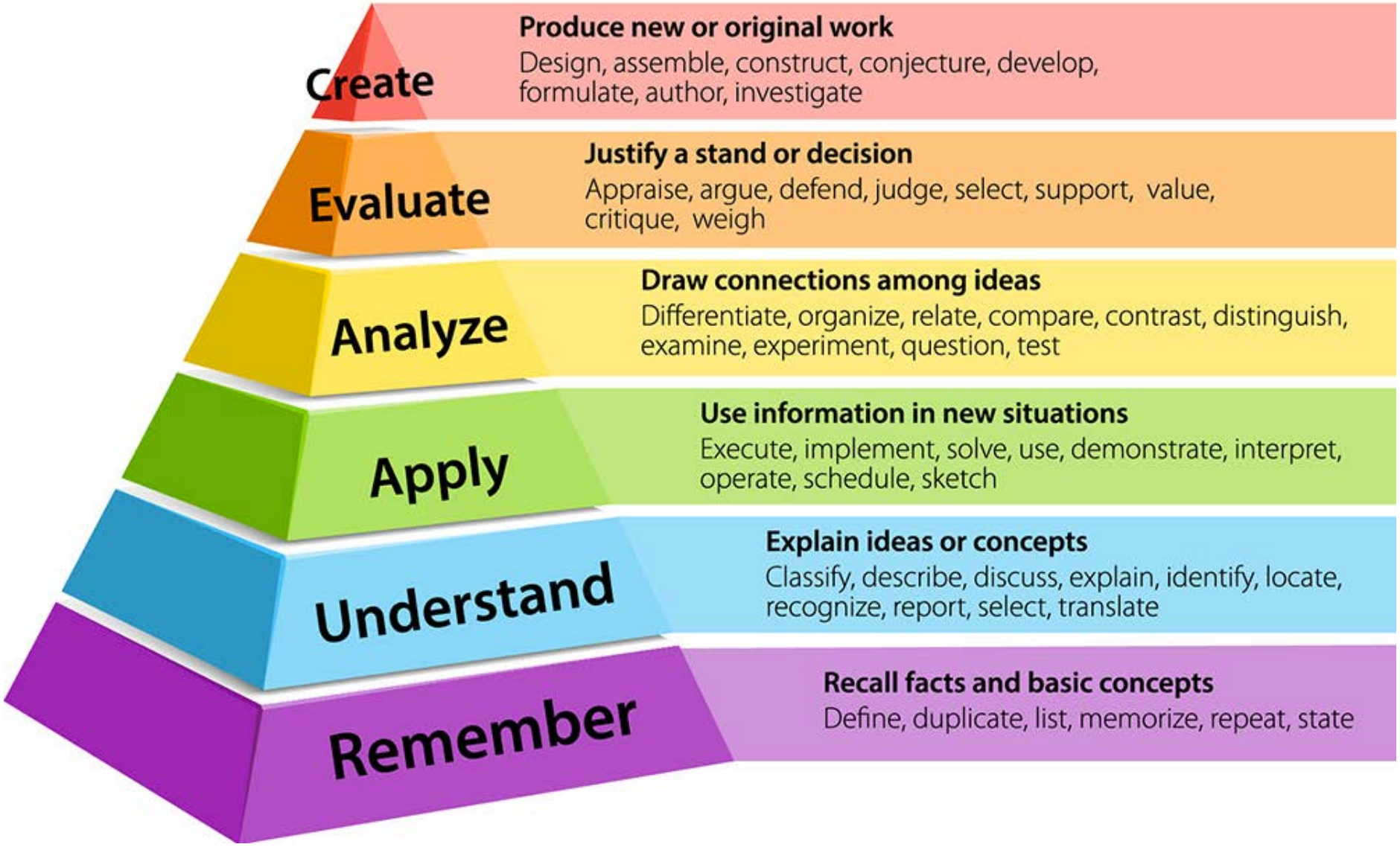
Writing MCQ for the ERKNet Case-based eLearning Module

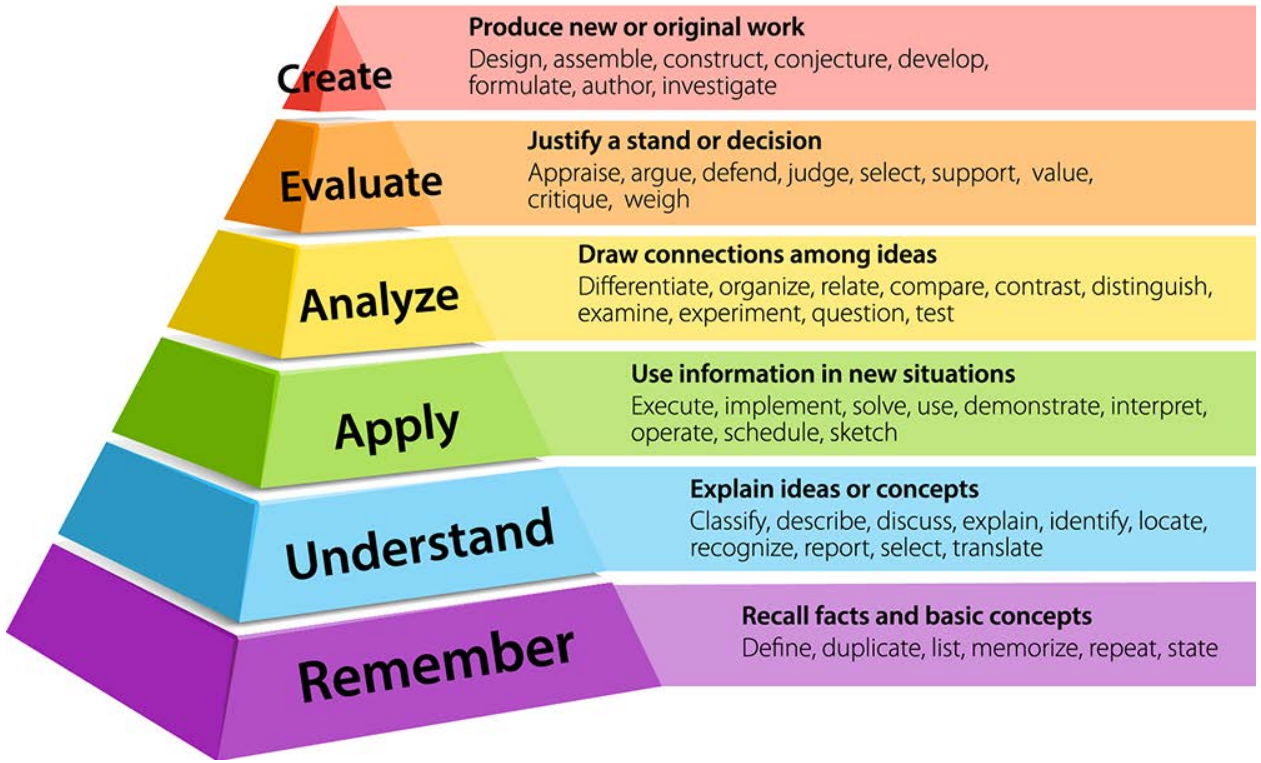
part 1: Overview

- **Life long learning**
- eLearning module
- Selecting a case

Life long learning

Lifelong learning





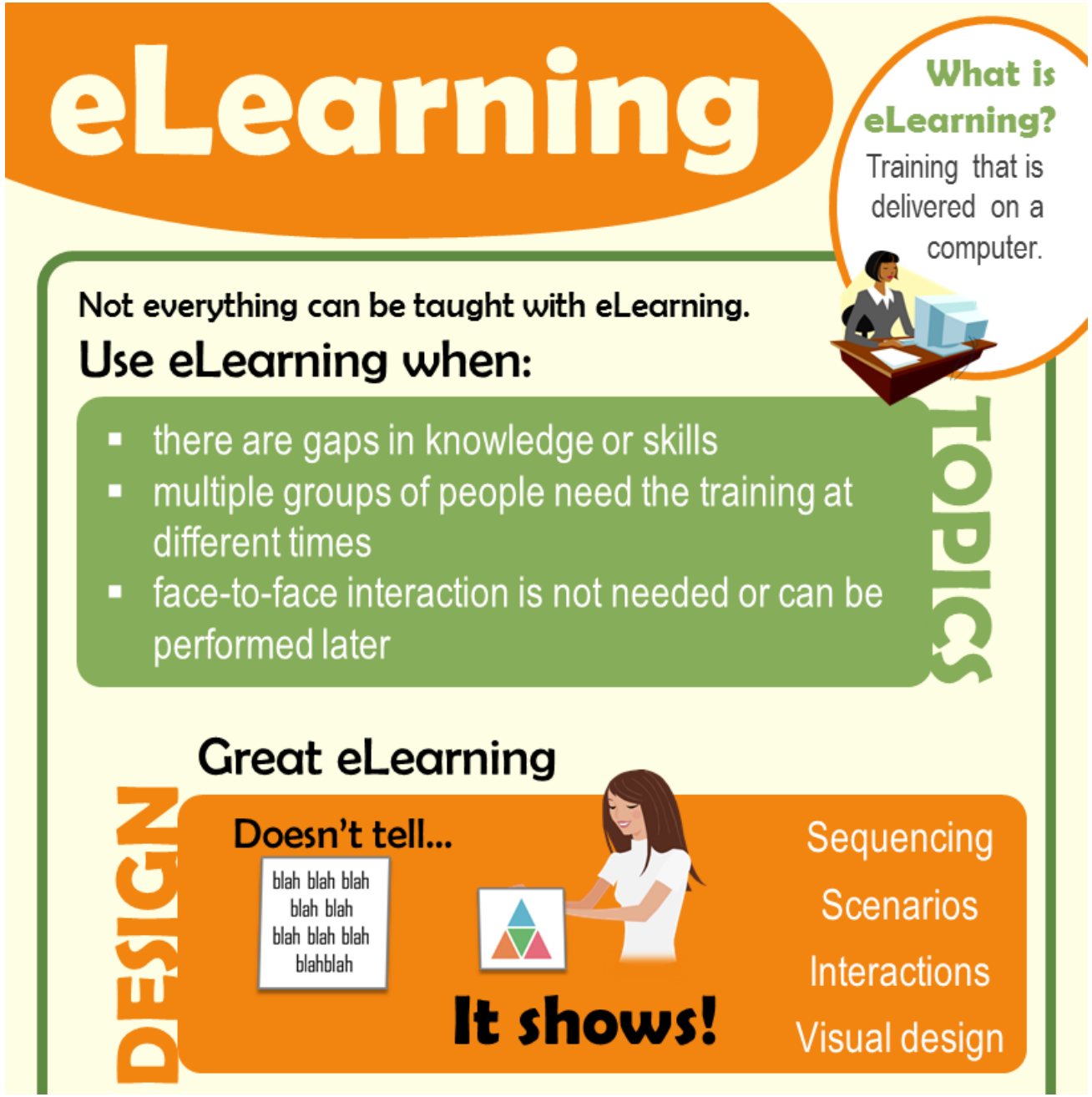
Script concordance testing

Case based MCQ/eLearning

True-False

part 1: Overview

- Life long learning
- **eLearning module**
- Selecting a case



- **Sequencing**:

- Case based
- Progression through the case

Sequencing

The NEW ENGLAND JOURNAL of MEDICINE

CASE RECORDS of the MASSACHUSETTS GENERAL HOSPITAL

Case 19-2018: A 15-Year-Old Girl with Acute Kidney Injury

Ann Y. Kao, M.D., M.P.H., Pallavi Sagar, M.D., Jean E. Klig, M.D.,
Amita Sharma, M.D., and Kristen J. Tomaszewski, M.D.

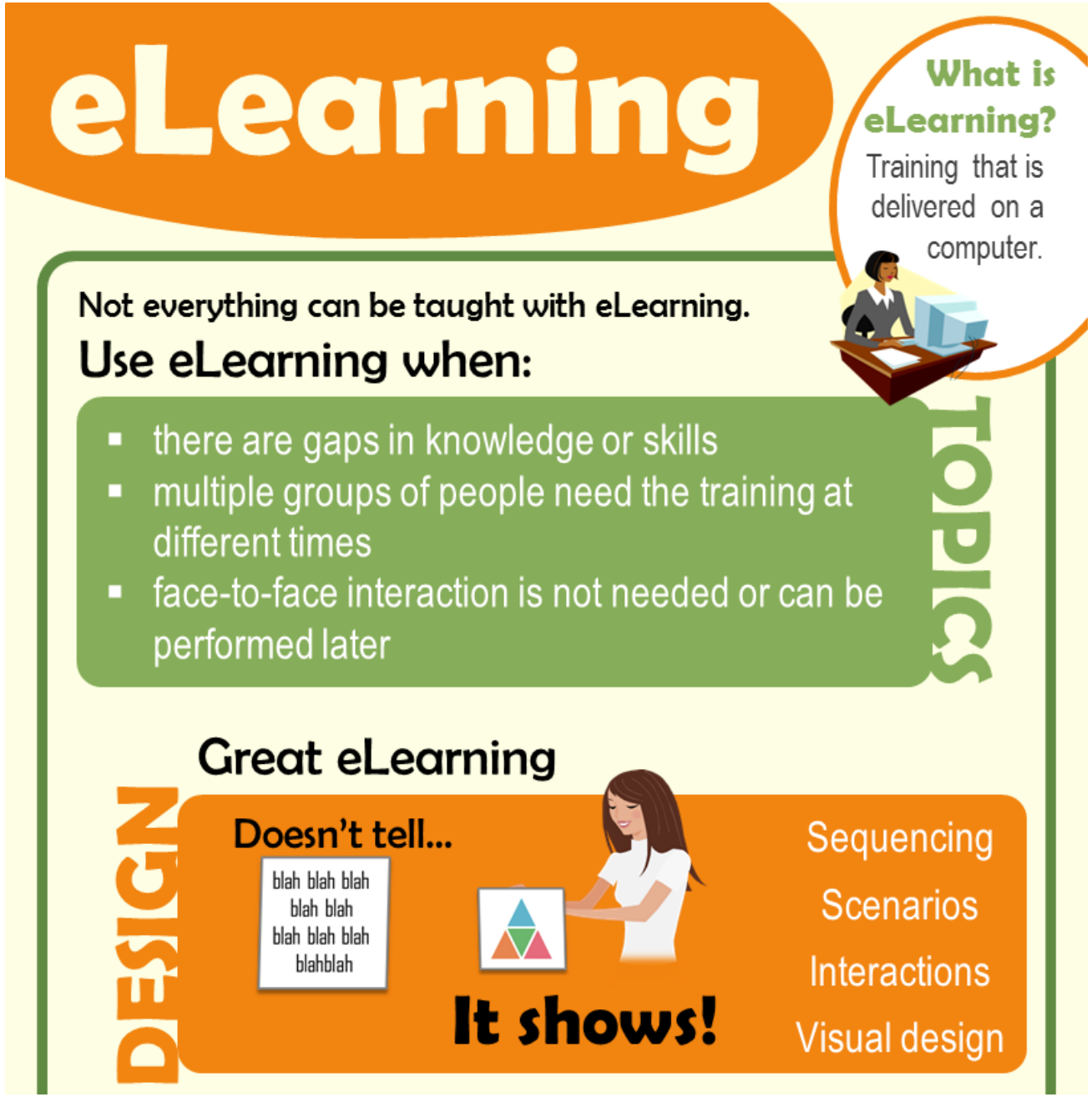
PRESENTATION OF CASE

Dr. Helen I. Healy (Pediatrics): A 15-year-old girl was admitted to this hospital during the summer because of acute kidney injury.

The patient had been well until 8 days before admission, when painful cramping in the lower abdomen and bloody diarrhea developed. Bowel movements occurred approximately every hour, and the patient was unable to sleep. She took ibuprofen but had no relief of the abdominal pain.

On the third day of illness, two episodes of nonbloody, nonbilious emesis occurred. The following day, the patient was seen by her primary care pediatrician.

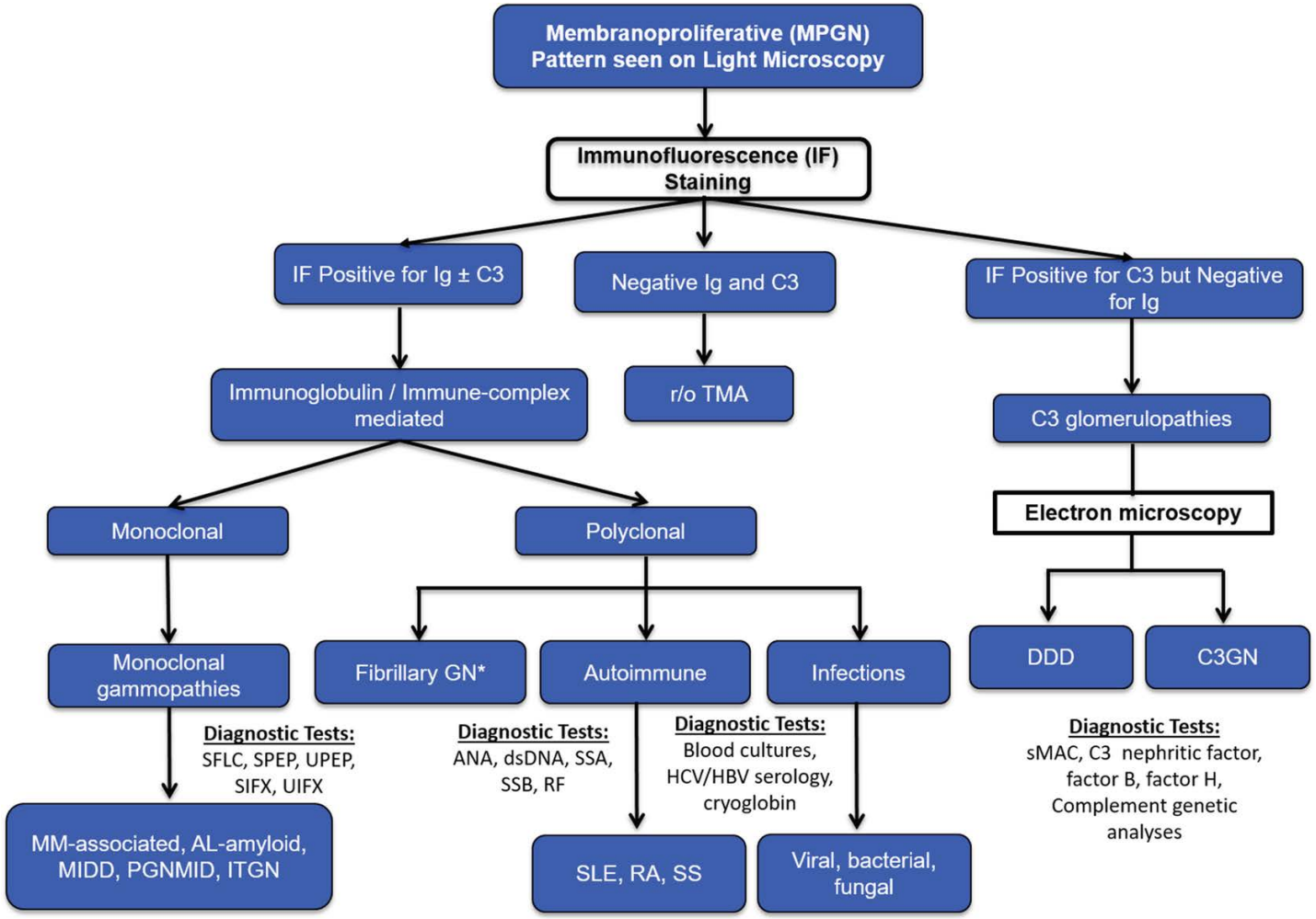
From the Departments of Pediatrics (A.Y.K., A.S.), Medicine (A.Y.K.), Radiology (P.S.), Emergency Medicine (J.E.K.), and Pathology (K.J.T.), Massachusetts General Hospital, and the Departments of Pediatrics (A.Y.K., A.S.), Medicine (A.Y.K.), Radiology (P.S.), Emergency Medicine (J.E.K.), and Pathology (K.J.T.), Harvard Medical School — both in Boston.

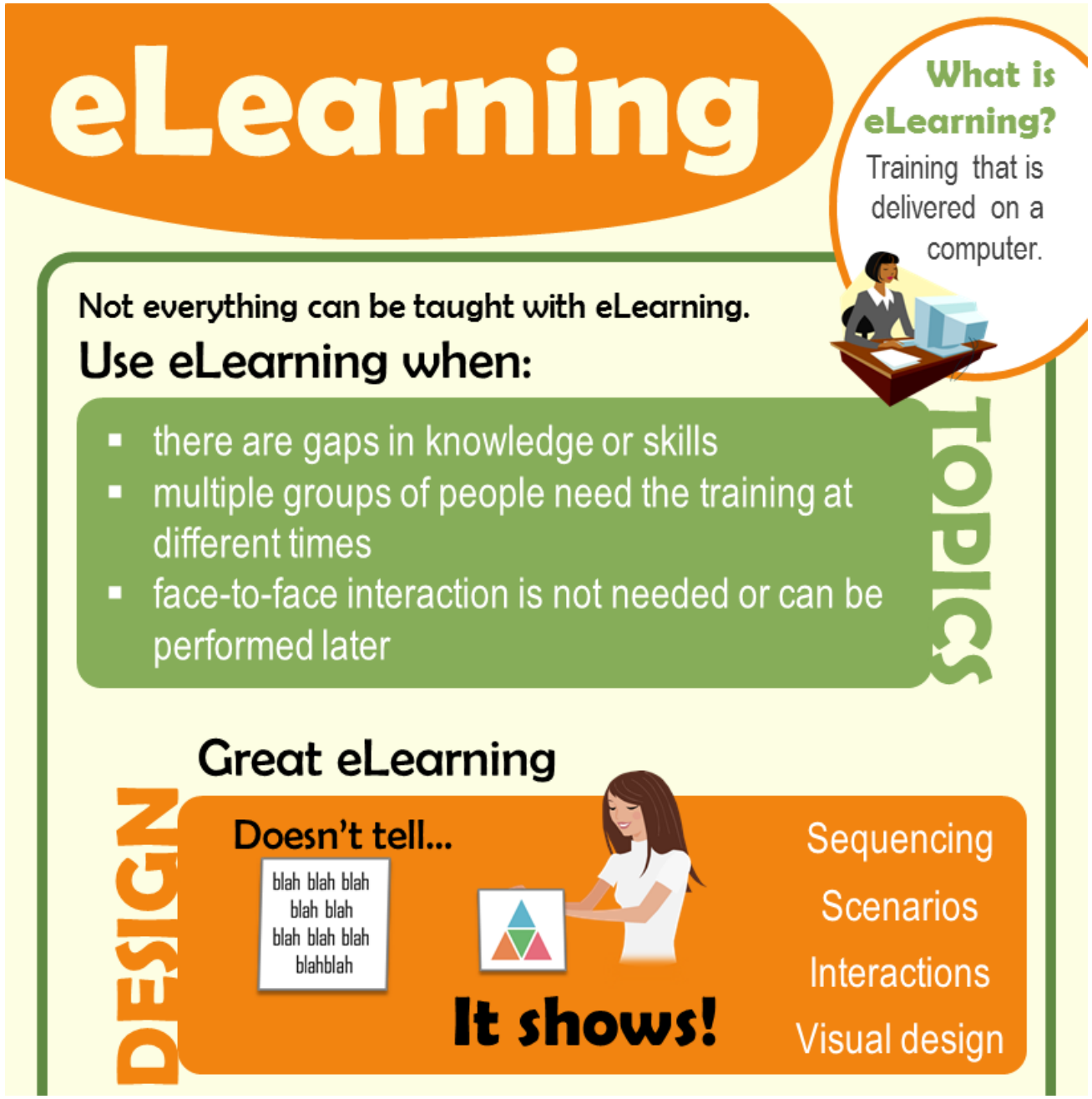


- **Scenarios**:

- Case based
- 'smoothened' real life cases (balance between typical versus 'messy')

Clinical algorithm versus ‘real-life’





- **Interactions**:
- Try to be as 'immersive' as possible
- Visual design:
- Images, test results, videos…

Not a ‘boring’ textbook…

Mind your audience (young > tablet or phone)

ERKNet
The European Rare Kidney Disease Reference Network

eLearning Instruction manual

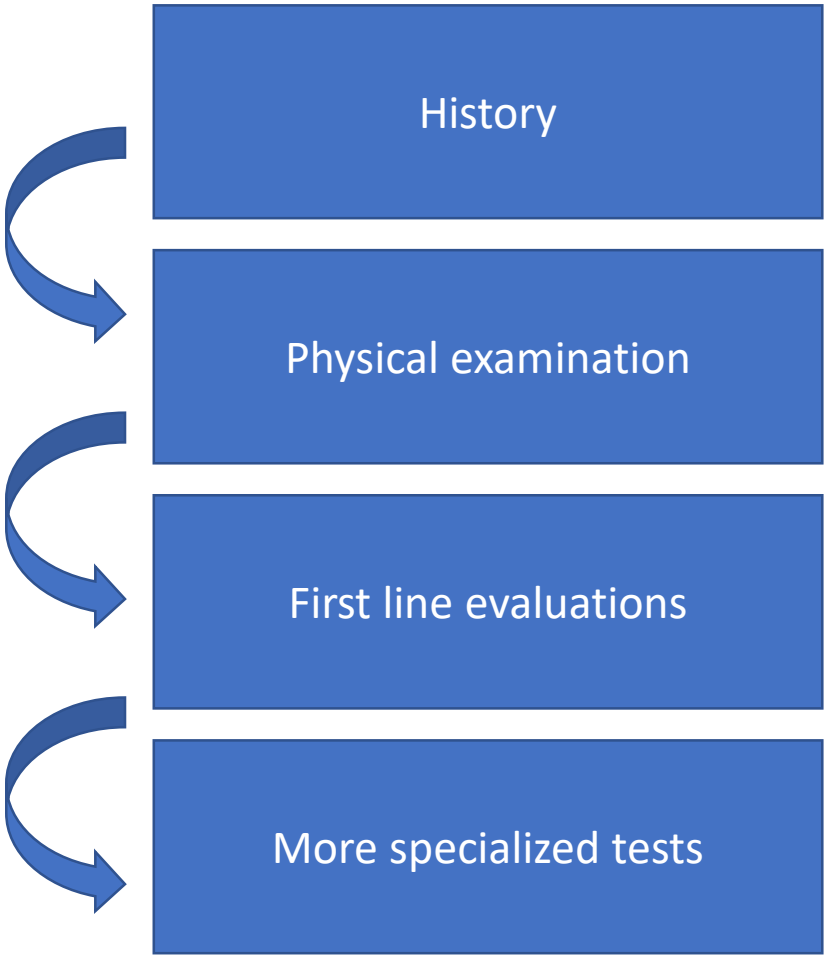
ERKNet Case-based eLearning Module

Instruction manual

Follow a ‘house style’…

- Long case
- Case description + question to be asked + possible solutions
- 6 types possible in platform:
  - MCQ
  - sorting
  - lab values
  - matrix
  - mapping
  - slider

The long case

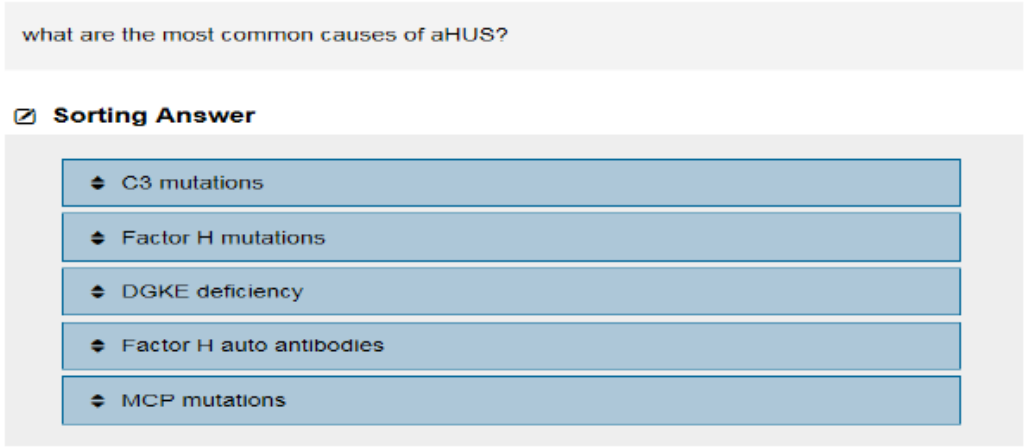


- Case is ‘divided’
- Use ‘logical’ moments
- Every question is a ‘learning experience’
- Think beforehand what message/skill you want to convey

The ‘question options’:

1) Multiple-choice question with one or several correct answers:
The learner receives a feedback as to which answer(s) is/are correct, with an explanation by the expert.

2) Sorting question:
The learner is requested to order a list of items according to a certain criterion.



3) Lab values question:

The Expert indicates a list of laboratory values and the learner rates whether each value is (a) relevant to the disease in question and (b) expected to be elevated, decreased or normal.

which laboratory abnormalities do you expect in aHUS?

Lab values answer

A	✔	Potassium	elevated	elevated
B	✔	Hemoglobin	decreased	decreased
C	✖	Leukocytes	decreased	elevated
D	✔	Thrombocytes	decreased	decreased
E	✖	Haptoglobin	elevated	decreased
F	✔	Creatinine	elevated	elevated
G	✔	Urea	elevated	elevated
H	✖	Homocysteine	decreased	normal
I	✔	LDH	elevated	elevated
J	✔	Stool culture	normal	normal
K	✔	Shigatoxin stool PCR	normal	normal

Expert answers: Potassium, Hemoglobin, Leukocytes, Thrombocytes, Haptoglobin, Creatinine, Urea, Homocysteine, LDH, Stool culture, Shigatoxin stool PCR

4) Network/Matrix question:

In this question type two sets of variables are organized in tabular format and the learner is asked to assign a value to each combination. There is no limitation to the number of table rows and columns.

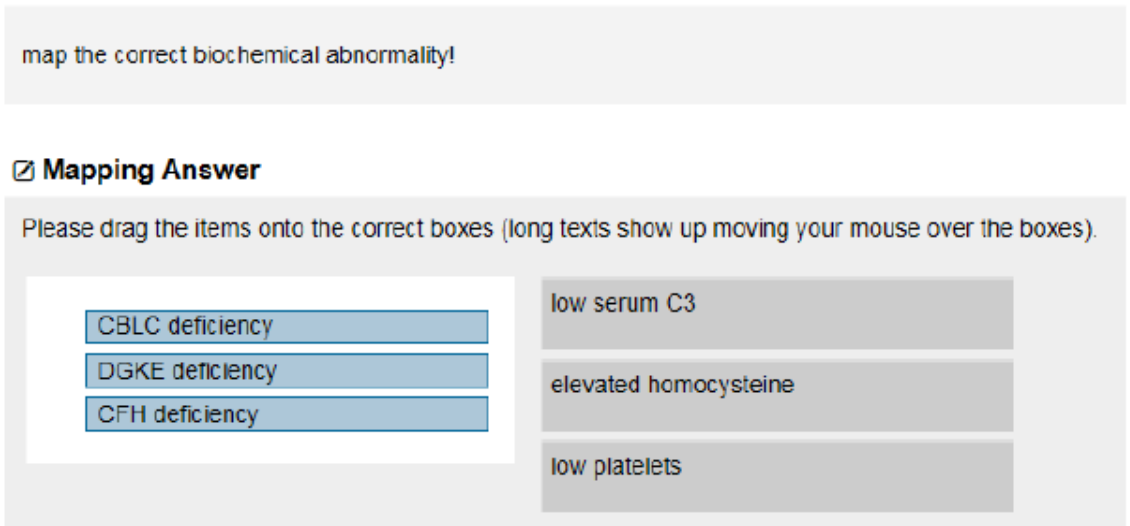
which etiologies differ in likelihood according to age?

Network/Matrix answer

Own Answer:	child	adult
genetic	+ ✗	+ + ✗
auto antibody	? ✗	+ + ✓

5) Mapping question:

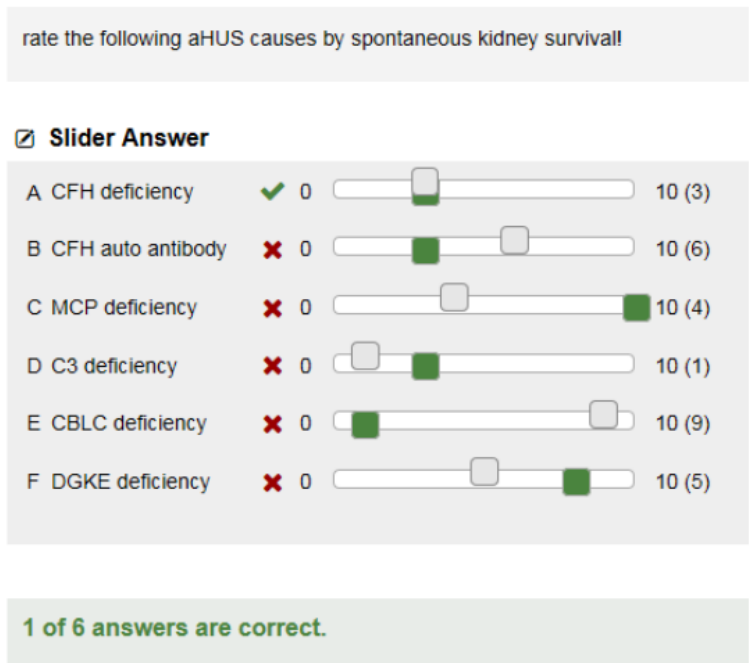
In this question type two sets of criteria are to be matched.



6) Slider question:

The learner can rate individual criteria according to a numeric Scale.

The Expert´s ranking will be indicated in green and the learner´s ranking in grey.



part 1: Overview

- Life long learning
- eLearning module
- **Selecting a case**

Selecting the perfect case

- ‘but it was an actual real case!’
- Reflect good and realistic clinical practice
- Peer review the case (no tunnelvision)

‘but it was an actual real case!’

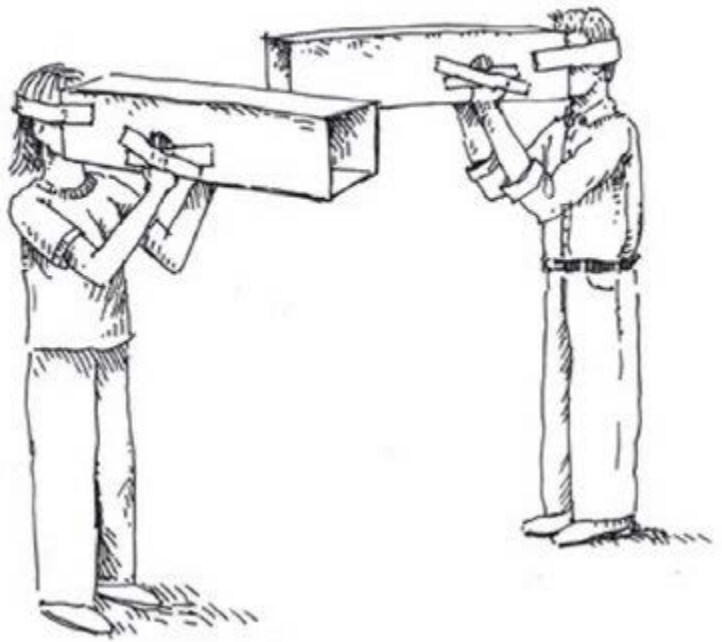
- Excellent ‘inspiration’ for cases
- Do not feel confined by the details of the ‘actual’ case
- Do not select an ‘unusual presentation’ as ‘the most likely’ scenario
- Priority: create good question, not tell a ‘great anecdote’

"I know this is an unusual sales presentation, but you're a very hard person to reach."

The case should reflect **good** and **realistic** clinical practice

- Sometimes a case can be an example of poor clinical practice or secondary to the structure of the health care system
- Only use this when it is ‘didactic’
- Otherwise try to adhere to good (international) guidelines

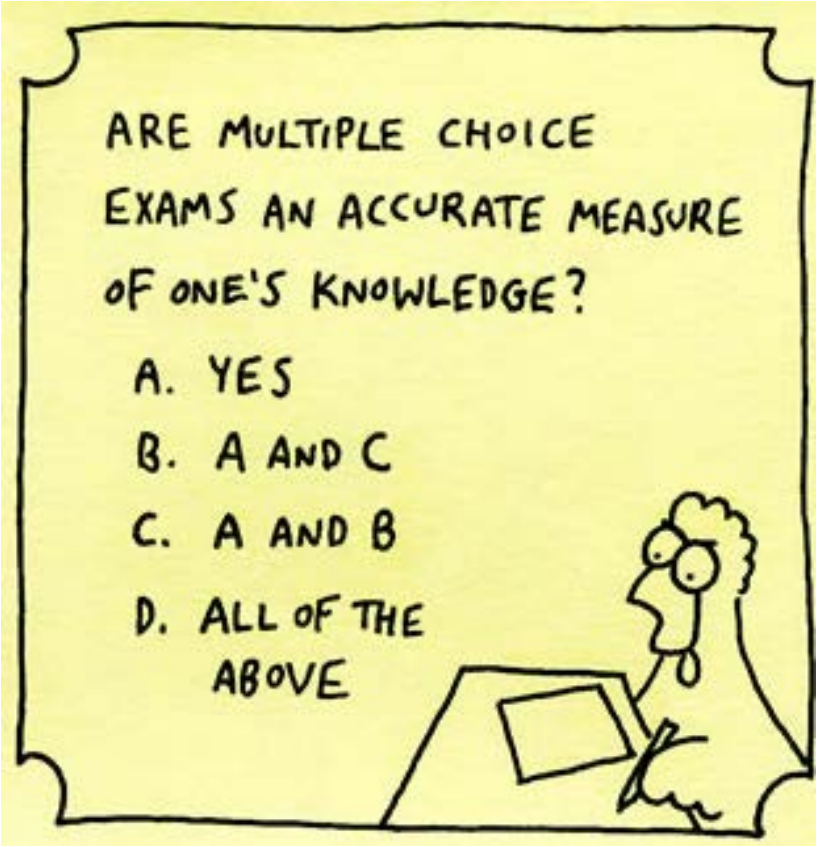
Peer review the case (no tunnelvision)



Part 2: MCQ

- An overview of the different methods of MCQ
- The concept of case based MCQ for teaching and evaluation
- How to formulate the ‘perfect question’

Multiple choice item formats



TRUE-FALSE VS. ONE-BEST-ANSWER ITEMS

- **True-false item formats that require test-takers select some set of options that are true:**
  - C-type (A/B/Both/Neither response items)
  - K-type (complex true-false items)
  - X-type (simple true-false items)

- **One-best-answer item formats that require test-takers select the single best response:**
  - A-type (4 or more options, single items or sets)
  - B-type (4- or 5-option matching items, in sets of 2 to 5 items)
  - F-type (items grouped into sets around specific content, where test-takers cannot return to previously seen items in the set)
  - G-type (items grouped into sets around specific content, where test-takers can return to previously seen items in the set)
  - R-type (extended-matching items, in sets of 5 to 20 items)

TRUE-FALSE VS. ONE-BEST-ANSWER ITEMS

- **True-false item formats that require test-takers select some set of options that are true:**
  - **C-type (A/B/Both/Neither response items)**
  - K-type (complex true-false items)
  - X-type (simple true-false items)

- **One-best-answer item formats that require test-takers select the single best response:**
  - A-type (4 or more options, single items or sets)
  - F-type (items grouped into sets around specific content, where test-takers cannot return to previously seen items in the set)
  - G-type (items grouped into sets around specific content, where test-takers can return to previously seen items in the set)
  - R-type (extended-matching items, in sets of 5 to 20 items)

Good example

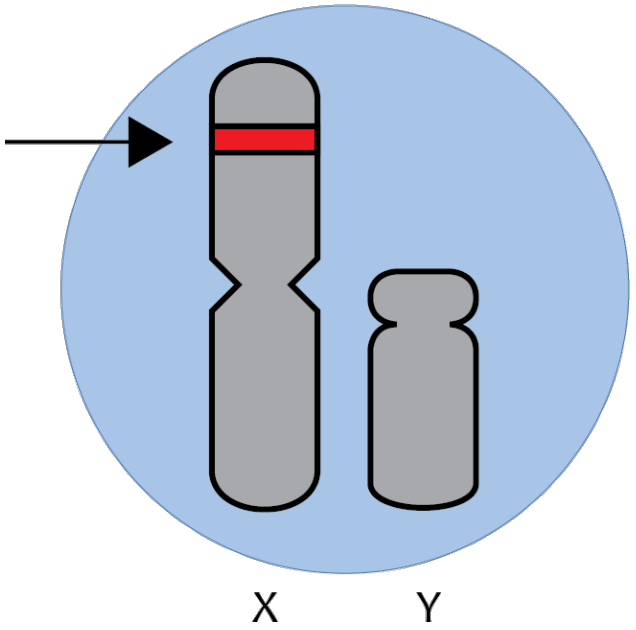
Which of the following are X-linked recessive conditions?

1 Cystic fibrosis

2 Duchenne muscular dystrophy

3 Hemophilia A (classic hemophilia)

4 Tay-Sachs disease



Good example

The stem is clear.

Which of the following are X-linked recessive conditions?

The question should be closed and focused.

1 Cystic fibrosis

2 Duchenne muscular dystrophy

3 Hemophilia A (classic hemophilia)

4 Tay-Sachs disease

The options are absolutely true or false with no ambiguity.

Options should be homogenous (all are conditions), clearly worded, and of similar length.

Bad example

True statements about cystic fibrosis (CF) include:

1 CF is an autosomal recessive disease

2 Children with CF usually die in their teens

3 Males with CF are sterile

4 The incidence of CF is 1:2000

Bad example

The question is not closed and focused.

True statements about cystic fibrosis (CF) include:

1 CF is an autosomal recessive disease

2 Children with CF usually die in their teens

3 Males with CF are sterile

4 The incidence of CF is 1:2000

A group of content experts would not necessarily agree on the answers.

Is this in Europe? Is this among all ethnic groups?

Options are non homogenous and vaguely worded, and of similar length.

GENERAL RULES FOR TRUE-FALSE ITEMS

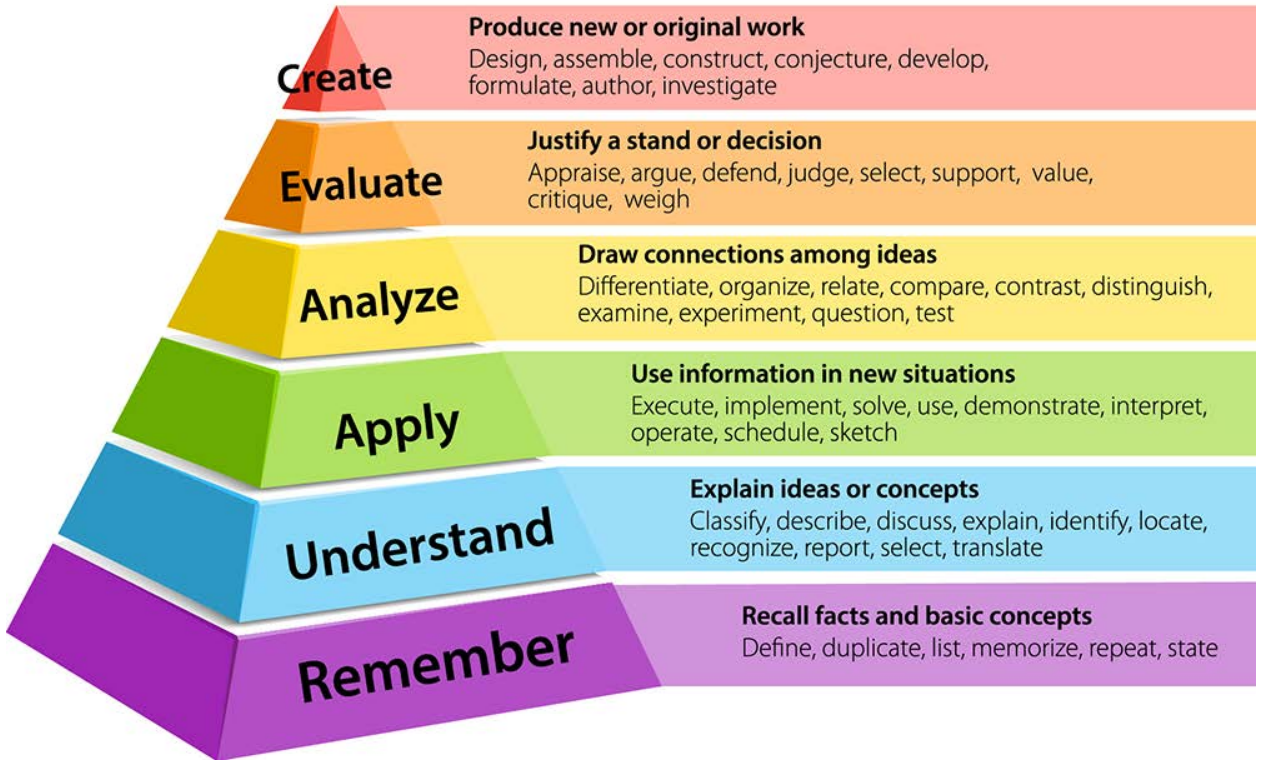
- Because test-takers are required to select all the options that are "true," true-false items must satisfy the following rules:

  - Item and option text must be **clear** and unambiguous. Avoid imprecise phrases such as "is associated with" or "is useful for" or "is important"; words that provide cueing such as "may" or "could be"; and vague terms such as "usually" or "frequently."
  - The lead-in question should be closed and **focused**.
  - Options must be **absolutely** true or false; no shades of gray are permissible.
  - Options should be **homogenous** so that they can be judged as entirely true or entirely false on a single dimension.

GENERAL RULES FOR TRUE-FALSE ITEMS

RECOMMENDATIONS FOR USING TRUE-FALSE ITEMS

- **Avoid** true-false questions if possible.
- The writer may have something **particular in mind** when writing the item, but careful review subsequently reveals subtle difficulties that were not apparent to the item author.
- Often the distinction between “true” and “false” is not clear, and it is not uncommon for **subsequent reviewers** to alter the answer key.
- In order to avoid ambiguity with this item type, item writers often lean toward assessing recall of an **isolated fact**.
- Application of knowledge, integration, synthesis, and judgment questions can be better assessed by **one-best-answer questions**.



Script concordance testing

Case based MCQ

True-False

TRUE-FALSE VS. ONE-BEST-ANSWER ITEMS

- **True-false item formats that require test-takers select some set of options that are true:**
  - C-type (A/B/Both/Neither response items)
  - K-type (complex true-false items)
  - X-type (simple true-false items)

- **One-best-answer item formats that require test-takers select the single best response:**
  - **A-type (4 or more options, single items or sets)**
  - F-type (items grouped into sets around specific content, where test-takers cannot return to previously seen items in the set)
  - G-type (items grouped into sets around specific content, where test-takers can return to previously seen items in the set)
  - R-type (extended-matching items, in sets of 5 to 20 items)

The One-Best-Answer family

- One-best-answer questions make explicit that **only one option** is to be selected.
- These items are the **most widely used** multiple-choice item format.
- They consist of a **stem** (e.g., a clinical case presen- tation) and a lead-in **question**, followed by a series of **choices**, with one correct answer and anywhere from three to five distractors.
- This question describes a **situation** (in this instance, a patient scenario) and asks the test-taker to indicate the most likely cause of the problem.

Stem:

A 32-year-old man has a 4-day history of progressive weakness in his extremities. He has been healthy except for an upper respiratory tract infection 10 days ago. His temperature is 37.8°C (100.0°F), pulse is 94/min, respirations are 42/min and shallow, and blood pressure is 130/80 mm Hg. He has symmetric weakness of both sides of the face and the proximal and distal muscles of the extremities. Sensation is intact. No deep tendon reflexes can be elicited. Babinski sign is present.

Lead-in: Which of the following is the most likely diagnosis?

Possible answers:

A. Acute disseminated encephalomyelitis

B. Guillain-Barré syndrome*

C. Myasthenia gravis

D. Poliomyelitis

E. Polymyositis

- Even though the incorrect answers are not completely wrong, they are less correct than the “keyed answer” indicated by the asterisk in the option set.
- Distractors in one-best-answer questions do not have to be totally wrong.

Note that the incorrect options are not wholly wrong. The options can be diagrammed as follows:

D C A E B

Least Likely Diagnosis — Most Likely Diagnosis

Recommendations:

- Use of a focused lead-in
- homogenous options that fall on a single dimension
- avoidance of vague terms

Bad example:

Which of the following is true about pseudogout?

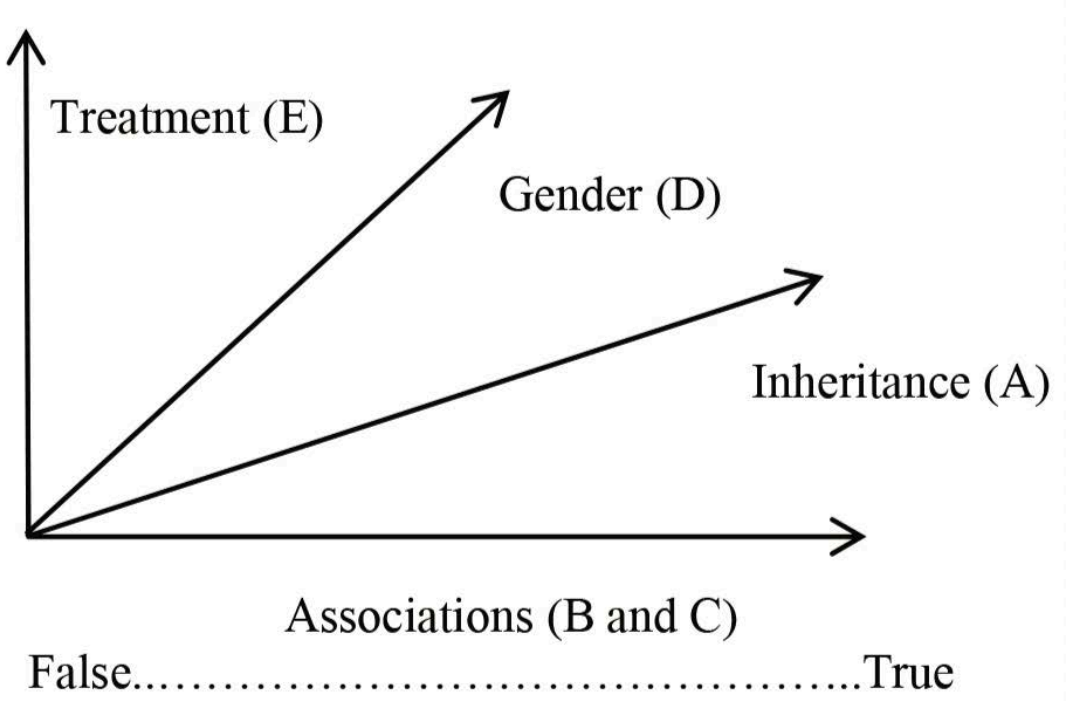
A. *It is clearly hereditary in most cases*
B. *It is seldom associated with acute pain in a joint*
C. *It may be associated with a finding of chondrocalcinosis*
D. *It occurs frequently in women*
E. *It responds well to treatment with allopurinol*

Bad example:

Which of the following is true about pseudogout?

A. *It is clearly hereditary in most cases*
B. *It is seldom associated with acute pain in a joint*
C. *It may be associated with a finding of chondrocalcinosis*
D. *It occurs frequently in women*
E. *It responds well to treatment with allopurinol*

- options are not listed on a single dimension.
- After reading the stem, the test-taker has only the vaguest idea what the question is about.
- In an attempt to determine the “best” answer, the test-takers have to decide whether “it occurs frequently in women” is more or less true than “it is seldom associated with acute pain in a joint.”



- The options are heterogeneous and deal with miscellaneous facts
- They cannot be rank-ordered from least to most true along a single dimension.
- Although this question appears to assess knowledge of several different points, its inherent flaws preclude this.
- The question by itself is not clear; the item cannot be answered without looking at the options.

“cover-the-options” rule

- If a lead-in is properly focused, a test-taker should be able to read the stem and lead-in, cover the options, and guess what the right answer is without seeing the option set.
- For example, in this next item, after reading the lead-in, you should be able to answer the question (ibuprofen) without seeing the options.
- When writing/assessing items, covering the options and attempting to answer the item is a good way to check whether this rule has been followed.

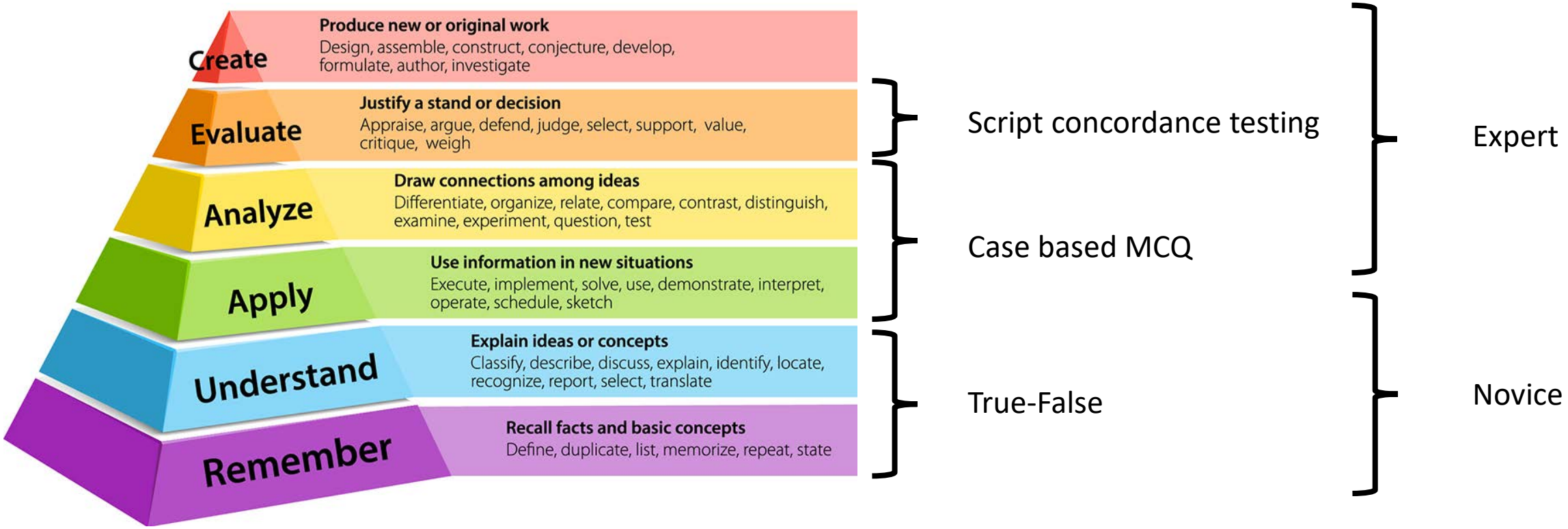
“cover-the-options” rule

A 58-year-old man comes to the office because of pain in the right knee for the past 3 days. He has a history of type 2 diabetes mellitus, hypertension, and hyperlipidemia controlled with an oral hypoglycemic drug and an ACE inhibitor. There is no family history of similar disorders. On physical examination, the knee is swollen, passive motion produces pain, and ballottement discloses an effusion. Synovial fluid is cloudy and contains positive birefringent crystals and no bacteria. X-ray shows chondrocalcinosis. **Which of the following is the most appropriate pharmacotherapy?**

A. *Allopurinol*
B. *Betamethasone*
C. *Ibuprofen**
D. *Infliximab*
E. *Levofloxacin*

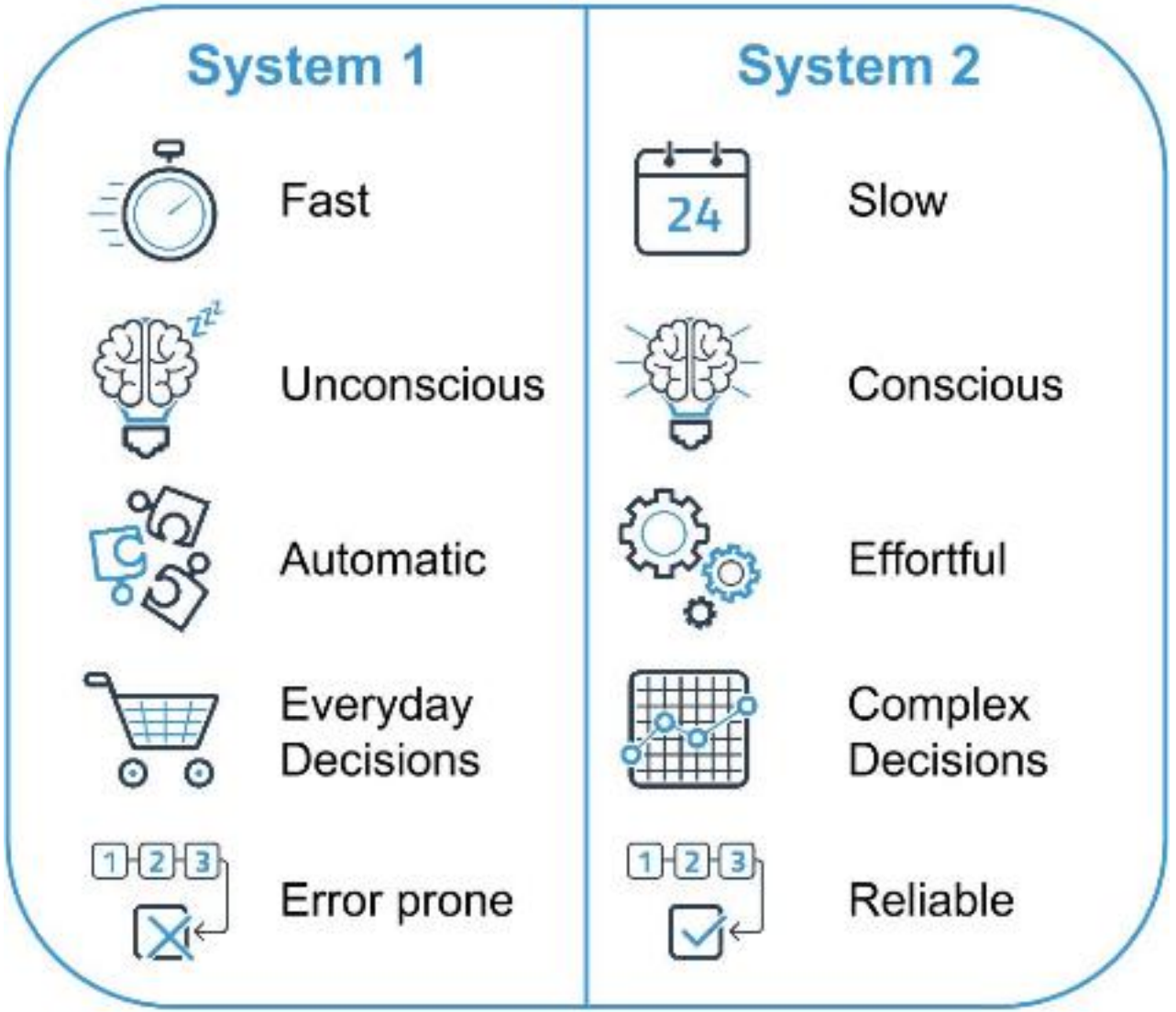
Part 2: MCQ

- An overview of the different methods of MCQ
- **The concept of case based MCQ for teaching and evaluation**
- How to formulate the ‘perfect question’



How does clinical reasoning work?

Dual process theory



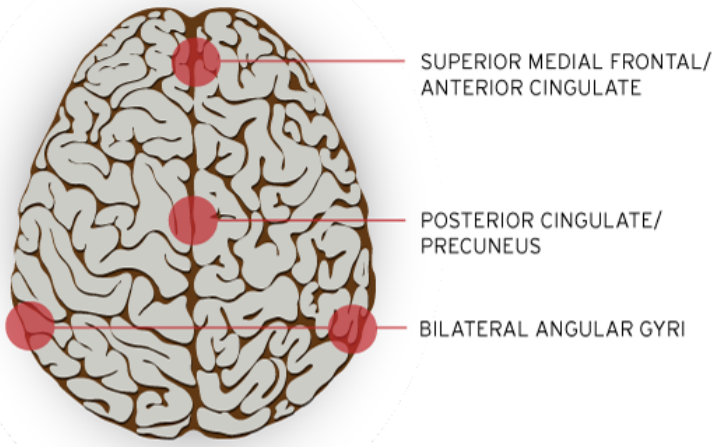
System 1 and System 2

Serum biochemistry and arterial blood gas

			Reference Range
FiO_2	0.50		
pH	7.05		7.35-7.45
pCO_2	66	mmHg	35-45
pO_2	247	mmHg	80-95
Bicarbonate	18	mmol/L	22-28
Base excess	-14		-3 - +3
O_2 saturation	99	%	> 95
Na^+	131	mmol/L	134-146
K^+	5.0	mmol/L	3.4-5.0
Cl^-	92	mmol/L	98-106
Urea	15	mmol/L	3-8
Creatinine	227	micromol/L	45-90
Glucose	50.9	mmol/L	3.5-5.5

Neurologic correlate

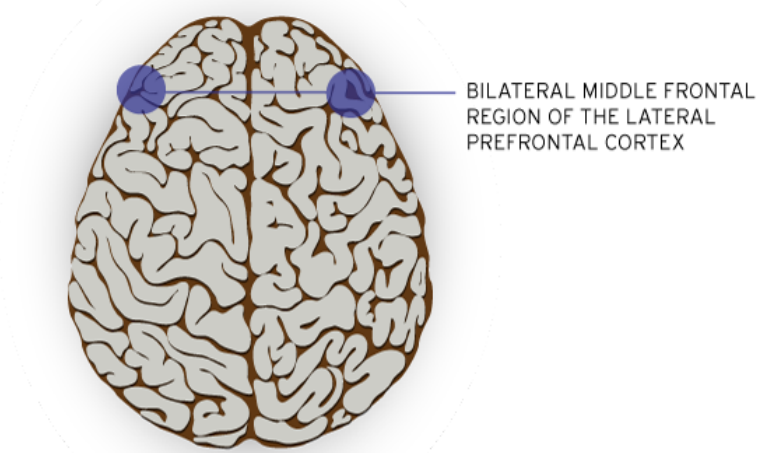
Areas of the Brain Affiliated with System 1 Processing



SYSTEM 1 CHARACTERISTICS

- Holistic
- Emotional; pleasure-pain oriented
- Behavior mediated by "vibes" from past experiences
- Encodes reality into concrete images, metaphors, and narratives
- More rapid processing; oriented toward immediate action
- Self-evidently valid; "experiencing is believing"

Areas of the Brain Affiliated with System 2 Processing

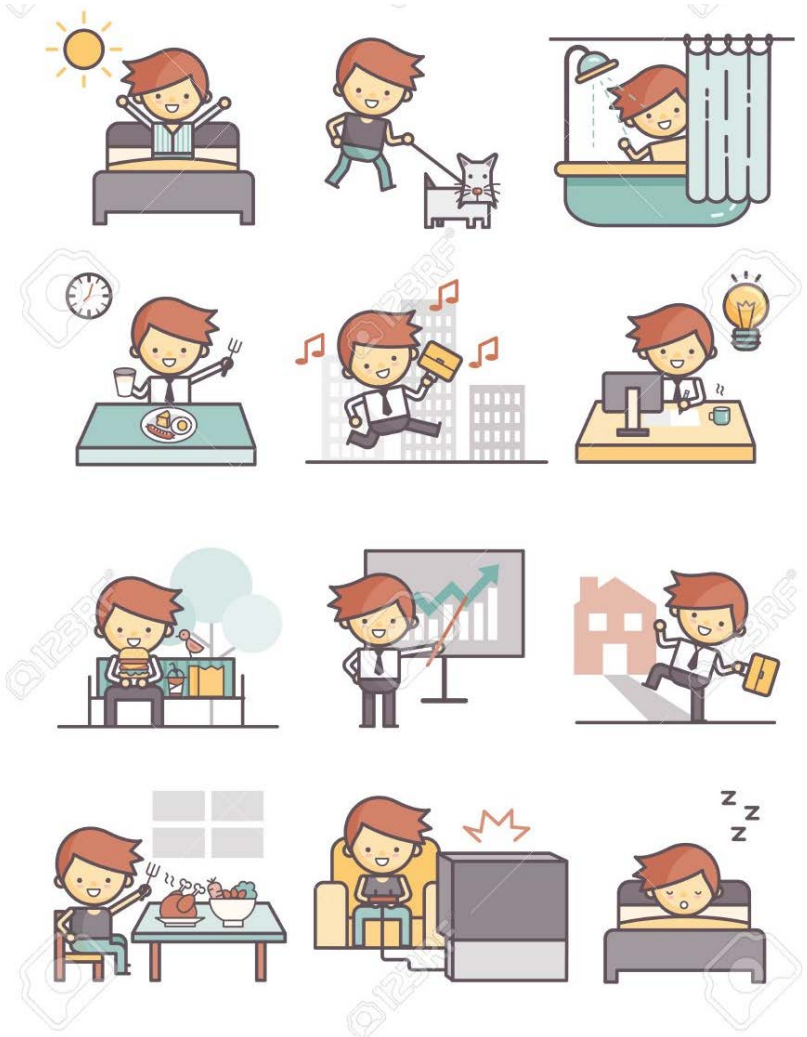


SYSTEM 2 CHARACTERISTICS

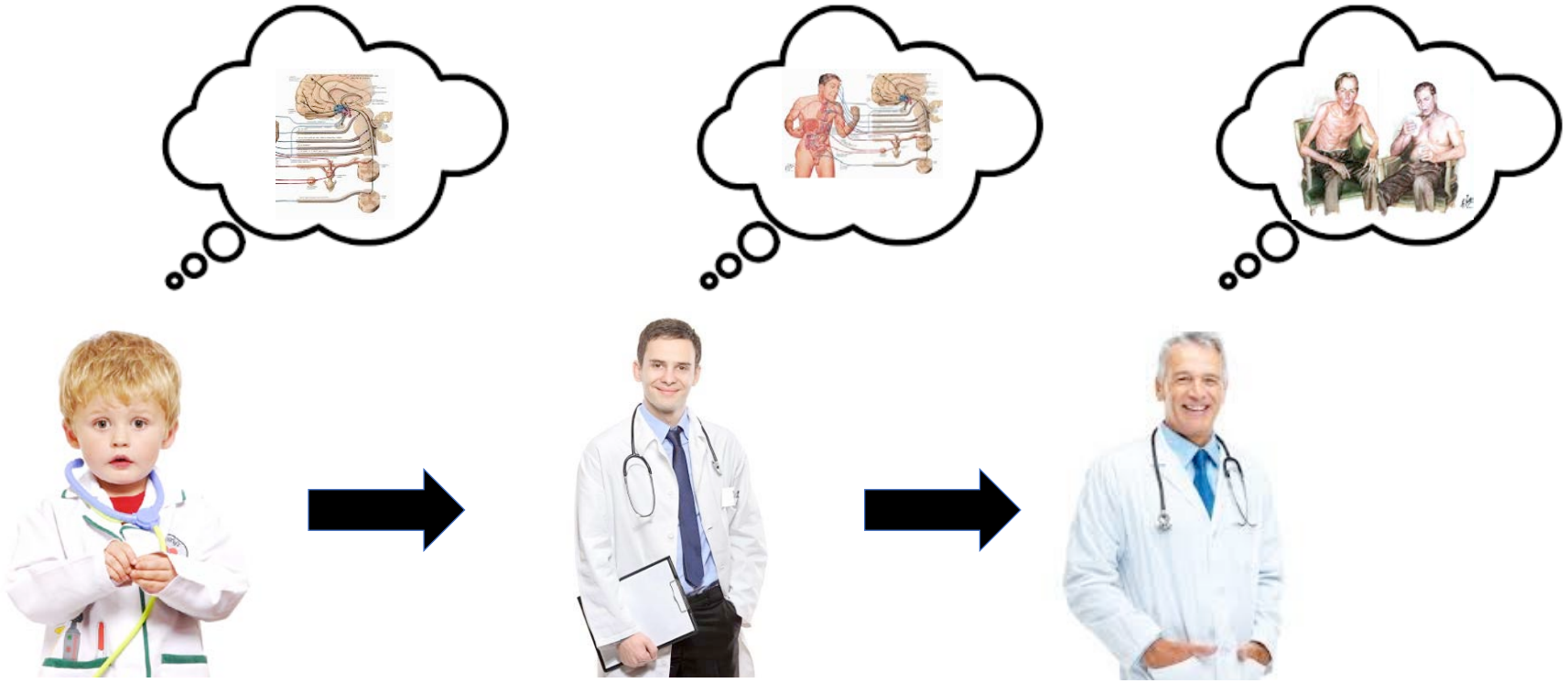
- Analytical
- Logical; reason oriented
- Behavior mediated by conscious appraisal of events
- Encodes reality in abstract symbols, words, and numbers
- Slower processing; oriented toward delayed action
- Requires justification via logic and evidence

From Farrell, Goh, and White (2014)

How often are we in system 1?



Clinical reasoning



Patient

Context:
Stress, stimuli
emotion
fatigue

Pattern recognized

Pattern processor

Pattern Not recognized

Intelligence
education – experience
Rationality
Critical thinking
Feedback

Type 1 thinking

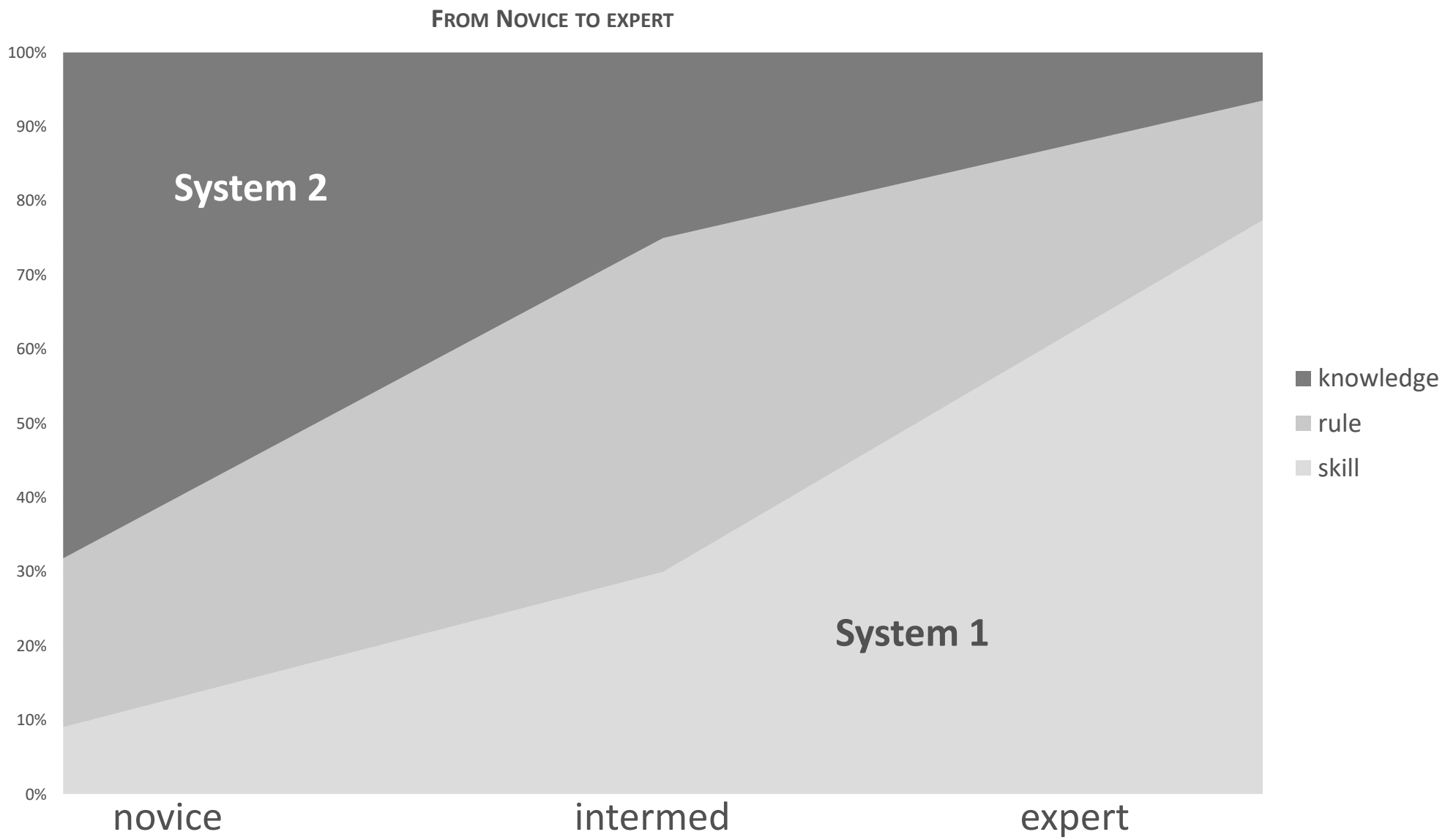
repetition

Type 2 thinking

Rational / dysrational 'overrule'

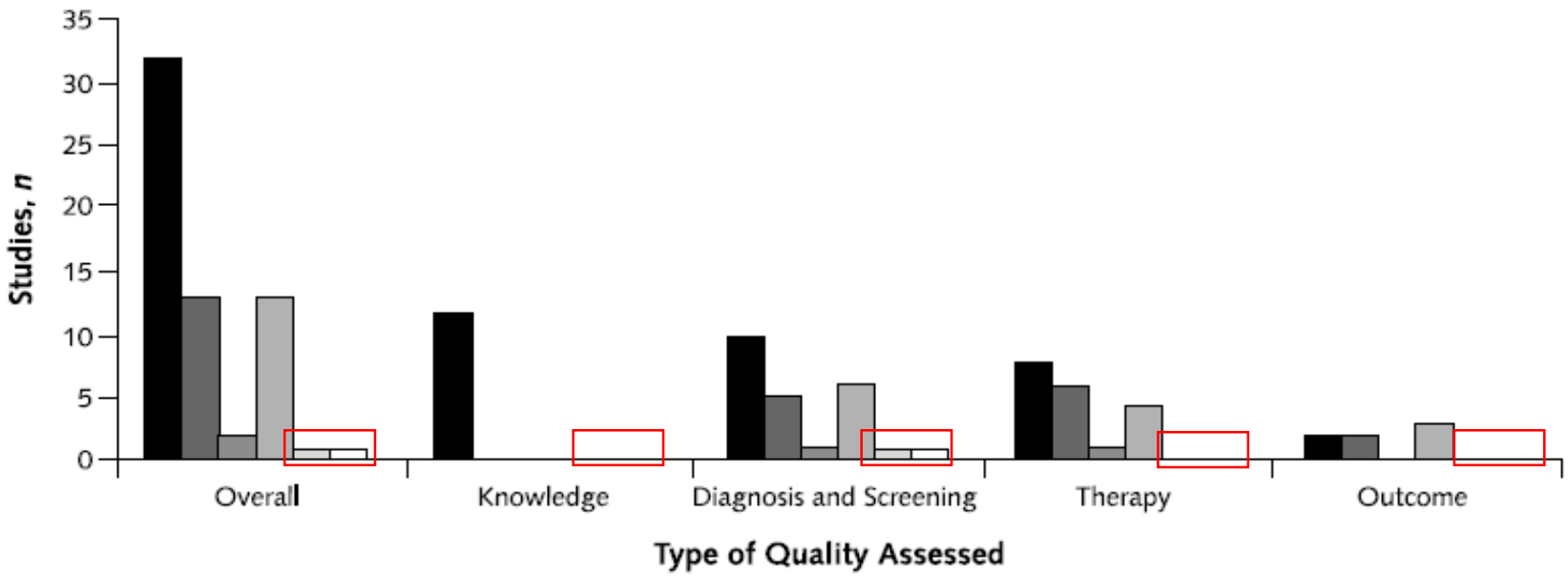
Calibration

Diagnosis



Dr. Pat Croskerry, 'How doctors think', EmergencyLondon, 16.11.2014

Experience is not a garantee!



- Studies in which length of time in practice or age was associated with lower performance for all outcomes.
- Studies in which length of time in practice or age was associated with lower performance for some outcomes; no effect was found for other outcomes.
- Studies in which there was a concave relationship between length of time in practice or age and performance.
- Studies in which no association was found between length of time in practice or age and performance.
- Studies in which length of time in practice or age was associated with higher performance for some outcomes; no effect was found for other outcomes.
- Studies in which length of time in practice or age was associated with higher performance for all outcomes.

Script concordance testing

Case description

A 72-year-old woman is referred from her nursing home to the emergency room for an acute shortness of breath. Her past medical history includes type II diabetes mellitus, congestive heart failure and heavy smoking. She was recently given morphine sulphate for intractable back pain. You are discussing the possibility of a pulmonary embolism.

If you were thinking of	And then you were to find	This hypothesis would become				
Q1- Ordering a computed tomography	A serum creatinine level of 110 µmol/L	-2	-1	0	+1	+2
Q2- Ordering a VA/Q lung scan	A history of chronic obstructive lung disease	-2	-1	0	+1	+2
Q3- Ordering a D-dimer test	A case of ovarian cancer in your patient's medical record	-2	-1	0	+1	+2

Investigative action

New information

5-anchor Likert-type scale

-2: strongly contraindicated; -1: contraindicated; 0: neither more or less indicated; +1: indicated; +2: strongly indicated

Part 2: MCQ

- An overview of the different methods of MCQ
- The concept of case based MCQ for teaching and evaluation
- **How to formulate the ‘perfect question’**

Basic Rules for Writing Clinical Vignettes and One-Best-Answer Items

Basic Rules for Writing ‘clinical vignettes’

WRITING CLINICAL VIGNETTES

“Recall Questions” are those that assess student knowledge of definitions or facts.

“Interpretation Questions” require students to review some information, and reach some conclusion, such as a diagnosis.

“Problem-Solving Questions” present a situation and require students to take some action (e.g., decide the next step in patient management).

For a high-stakes summative examination/ eLearning, the use of vignettes that require higher-order thinking skills and application of knowledge is preferable to simple recall items.

“Problem-Solving Questions”

A 28-year-old primigravid woman is at 11 weeks’ gestation. Medical history is unremarkable. Family history is unremarkable except that both of her brothers have intellectual developmental disorder, her mother died of breast cancer at age 55, and her father is estranged. No family health records are available.

Which of the following studies is appropriate?

- Amniocentesis for α-fetoprotein
- Blood test for fragile X carrier status
- Blood test for phenylketonuria carrier status
- Chorionic villus sampling for chromosome analysis
- Chorionic villus sampling for Duchenne’s muscular dystrophy

“Problem-Solving Questions”

A 28-year-old primigravid woman is at 11 weeks’ gestation. Medical history is unremarkable. Family history is unremarkable except that both of her brothers have intellectual developmental disorder, her mother died of breast cancer at age 55, and her father is estranged. No family health records are available.

Which of the following studies is appropriate?

- Amniocentesis for α-fetoprotein
- Blood test for fragile X carrier status
- Blood test for phenylketonuria carrier status
- Chorionic villus sampling for chromosome analysis
- Chorionic villus sampling for Duchenne’s muscular dystrophy

This item describes a patient and asks which study is most appropriate. The inclusion of the **vignette** leads to a more **realistic task**, because the test-taker would need to be able to both **recall** specific information and **synthesize** that information to know which studies should be ordered.

GUIDELINES FOR CLINICAL VIGNETTE CONTENT

- Focus items on **common** or potentially **catastrophic** problems; avoid "zebras" (maybe not easy for *rare diseases*…) and esoterica
- Pose **clinical decision-making tasks** that would be expected of a successful clinical practitioner
- Avoid clinical situations that would be handled by a **specialist/world expert** if not writing for a specialty exam
- Focus on areas in which **clinical reasoning mistakes are often made**

GUIDELINES FOR CLINICAL VIGNETTE CONTENT

Age, gender (e.g., 5-year-old girl)

Site of care (e.g., the emergency department)

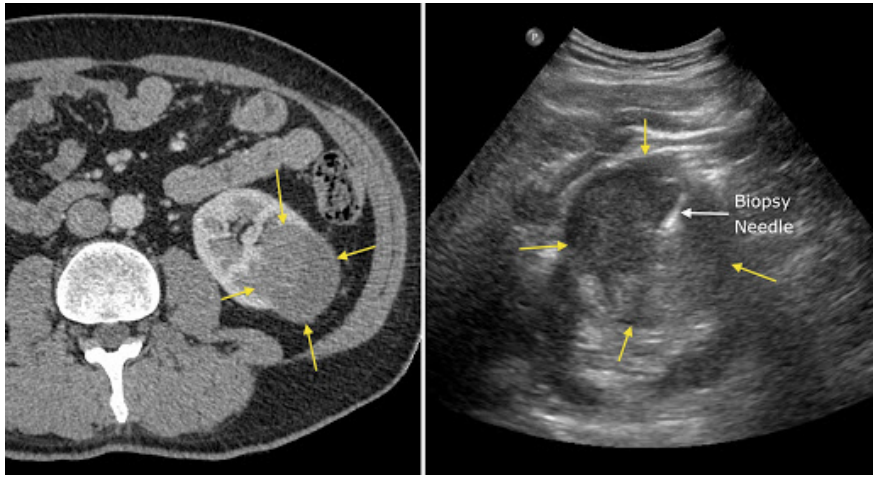
Presenting complaint (e.g., haematuria)
Duration of complaint (e.g., 2 days)

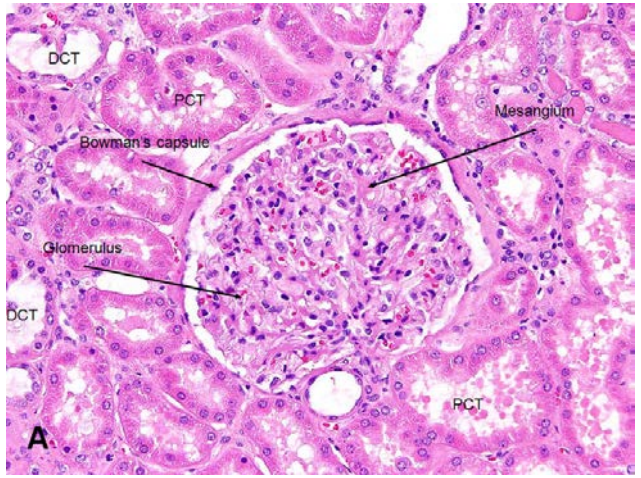
Patient history, including past medical history, family history, psychosocial history, and review of systems if important and plausible for the scenario
Physical findings
Results of diagnostic studies
Initial treatment, subsequent findings

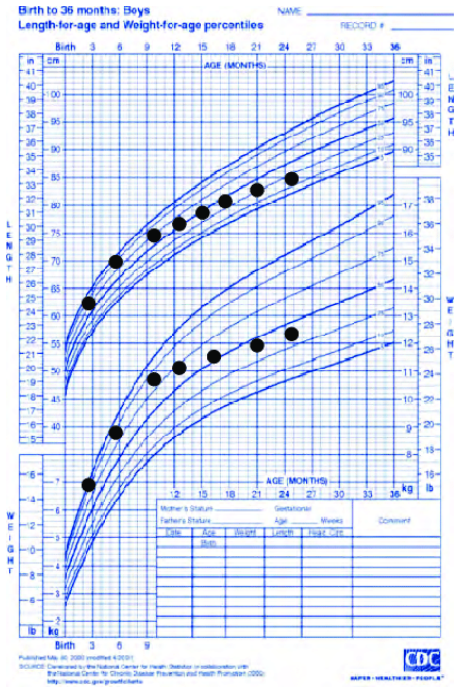
Make sure the item stems adhere to the following rules:

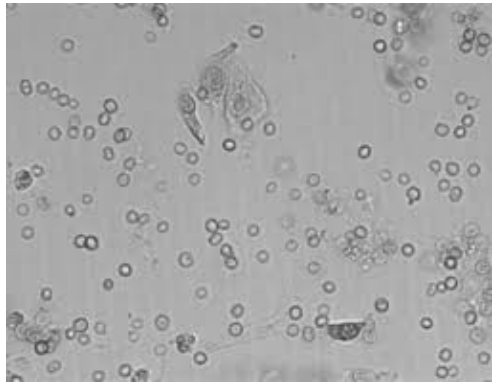
- Focuses on **important** concepts rather than **trivial** facts
- Can be answered **without looking** at the options
- Includes all relevant facts; **no additional data** should be provided in the options
- Is not “tricky” or **overly complex**
- Is **not negatively phrased** (e.g., avoid using *except* or *not* in the lead-in)

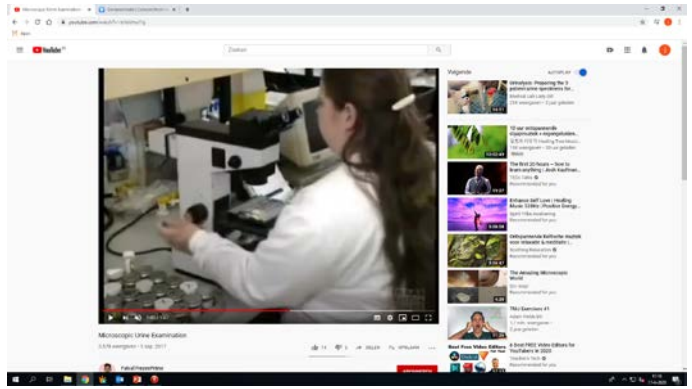
Add ‘immersive’ elements











ADDITIONAL POINTS ON VIGNETTES

Verbosity, Window Dressing, and Red Herrings

- Write clinical cases that are as short as possible
- avoid excessive verbosity, “window dressing”, and “red herrings” (information designed to mislead the learner or test-taker).

Patients Who Lie

- Ideally, patients in vignettes should tell the truth, or the physician’s interpretation of the patient’s story should be provided. Physicians use multiple cues to determine how truthful a patient is and many of these cues cannot be translated into written form.
  - Example: a patient’s alcohol consumption:
  - “The patient drinks 16 ounces of beer with dinner each night”
  - “The patient’s description of his alcohol consumption is contradictory.”
  - Do not write “The patient ‘claims’ to drink only one bottle of beer each night.”

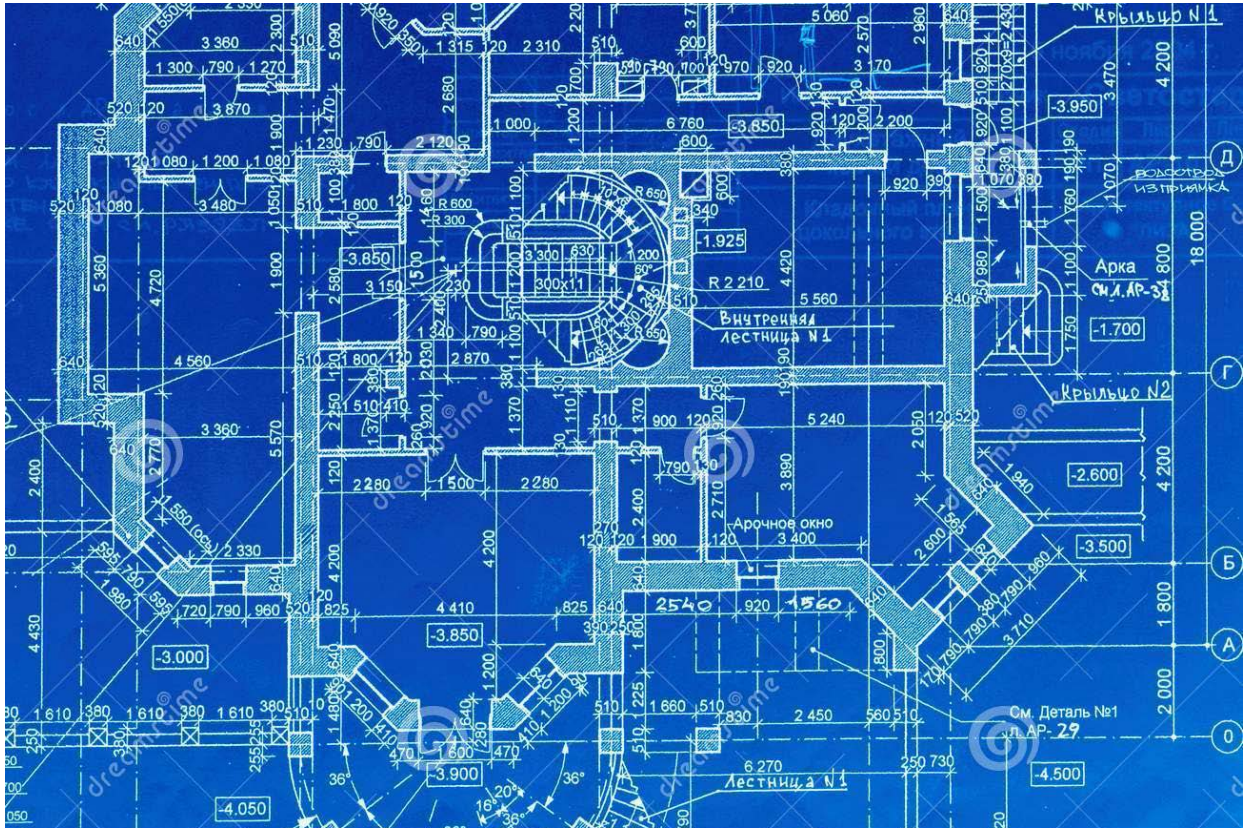
Basic Rules for Writing One-Best-Answer items/distractors

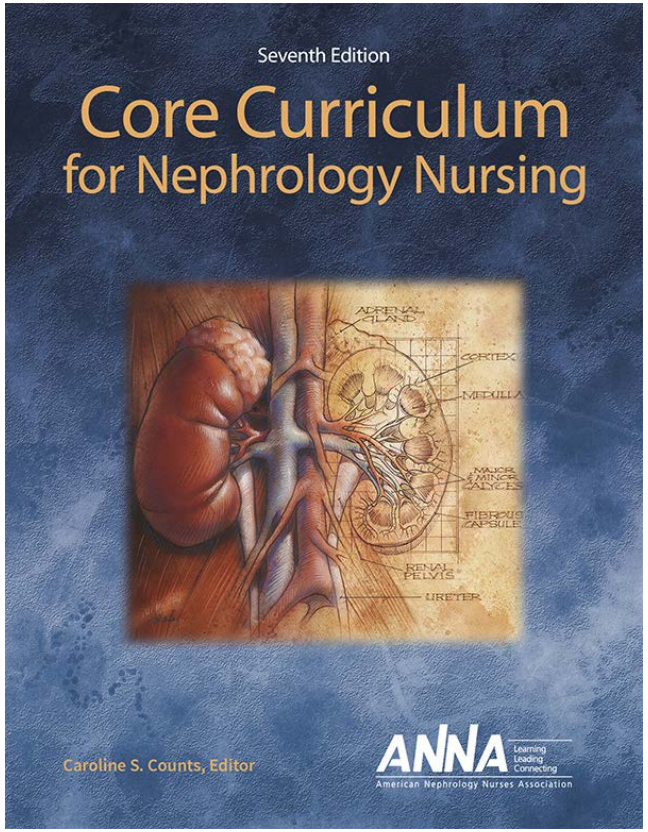
Rule 1: Each item should focus on an important concept or teaching point.

What do you want the student to learn, know or demonstrate?

- The topic of the item usually results from the blueprint, which is the outline of the major topics to be covered on the examination.
  - For instance, if an examination is developed to assess knowledge of the cardiovascular system, the blueprint might have two dimensions: 1) disease-based (e.g., hypertension, ischemic heart disease, systolic heart failure), and 2) task-based (e.g., assessment of foundational science principles, diagnosis, history, prognosis).
  - The blueprint would likely include items along both dimensions, and might call for six items on hypertension, four on systolic heart failure, two on diastolic heart failure, ten on ischemic heart disease, and so on.
- Along the task dimension there might be a similar distribution of topics.
- A clear and comprehensive blueprint or other set of test specifications should always be available so that item writers can stay focused on the important topics and write sufficient numbers of items for each topic.

Rule 1: get a blueprint or ‘core currciculum’





Rule 2: Each item should assess application of knowledge, not recall of an isolated fact.

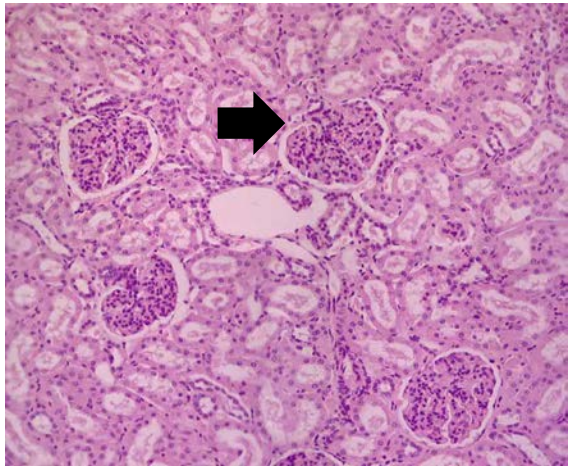
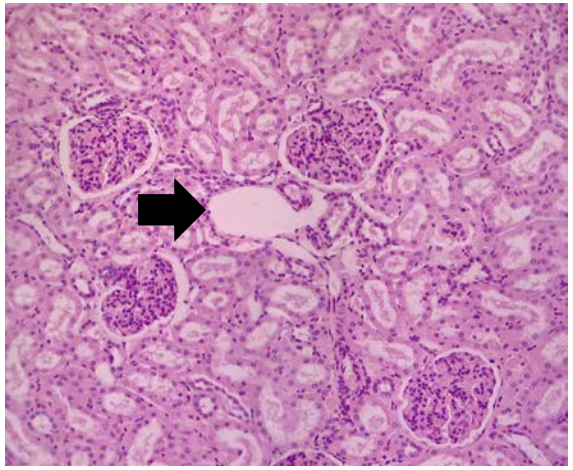
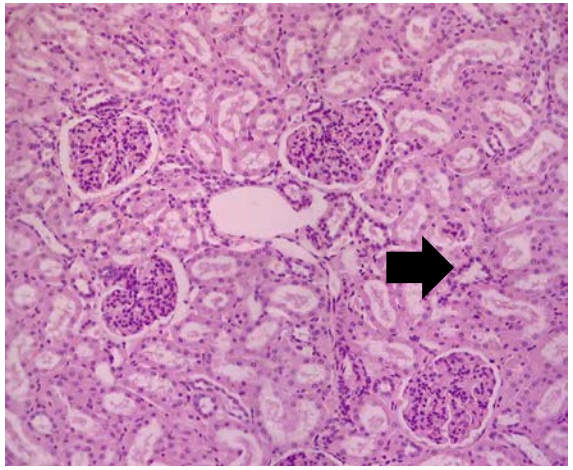
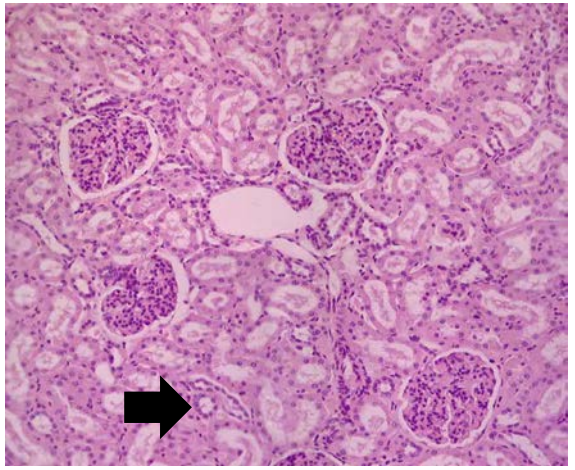
Rule 2

- Vignette
- Real patient as inspiration
- Beware of demographics / epidemiology
- At the level of the student/test-taker

- In some instances, such as the example with systolic heart failure, there will be an additional step that you must keep in mind: you should consider the underlying cause of the heart failure. **Patient demographics, past medical history, and other factors will differ depending on the etiology of the condition.**

    - Patients with systolic heart failure from a viral cardiomyopathy versus from ischemic heart disease may have different demographics and a different history; e.g., a younger patient with a viral illness preceding the onset of heart failure symptoms as compared to an older patient with risk factors for ischemic heart disease.

It does not have to be only text…

Rule 3: The item lead-in should be focused, closed, clear; and the cover-test should be possible.

- The next step is to ask the question with the use of a lead-in, and the accompanying vignette allows lead-ins to be focused on the patient.
- Such as:
  - “Which of the following is the next step in the management of this patient?”
  - “Which of the following is the most likely diagnosis?”
- An open-ended lead-in such as, “The diagnosis in the patient is:” should be avoided.
- The lead-in should be a single, closed, clear question. Ideally, after reading the vignette and the lead-in, a test-taker should be able to answer the item without seeing the options.
- Another reason to use a closed lead-in is because it helps to avoid certain item flaws, such as grammatical cueing.

Rule 4: All options should be homogeneous and plausible, to avoid cueing to the correct option.

- Next, generate the correct or keyed answer for the lead-in.
- Often, generating the correct answer is the easier step
- Generating plausible and parallel yet incorrect distractors is more challenging.

Note that the incorrect options are not wholly wrong. The options can be diagrammed as follows:

Rule 5: Always review items to identify and remove technical flaws that add irrelevant difficulty or benefit savvy test-takers.

Review

- Once you have written your item, you should take a step back and look closely at its structure. The bulk of the text (vignette or case information) should precede rather than follow the lead-in.
- The clinical or experimental vignette should make sense and follow a logical sequence:
  - first list patient demographics,
  - then history, physical examination, laboratory data, and so on.
- The lead-in should be closed, and the wording of the lead-in should logically generate a homogeneous option set.
- The use of a template to ensure all of these sections are in place and correctly structured is highly recommended.

Review

- As you review your item, ask yourself the following questions. If the options were removed, could a knowledgeable test-taker answer the question correctly?
- Is there anything in the phrasing or text that would confuse the knowledgeable test-taker?
- Are there any clues to help a testwise student guess the item correctly?
- Finally, you should ask a colleague to review the items you have written, in particular for content, clarity, and appropriateness for your particular test-taker population.

The lead-in:**Diagnosis**

- The diagnosis competency is subcategorized into more detailed concepts:
- Obtaining and Predicting History and Physical Examination
- Selecting and Interpreting Diagnostic Studies
- Formulating the Diagnosis
- Determining Prognosis/Outcome.

The lead-in:**Diagnosis**

Obtaining and Predicting History and Physical Examination

- *Which of the following factors in the patient's history most increased her risk for developing this condition?*
- *Which of the following additional information regarding this patient's history is most appropriate to obtain at this time?*

- *Which of the following is the most appropriate focus of the physical examination at this time?*

The lead-in:**Diagnosis**

Selecting and Interpreting Diagnostic Studies

- *Which of the following is the most appropriate diagnostic study to obtain at this time?*
- *Which of the following laboratory studies is most likely to confirm the diagnosis?*
- *Which of the following is the most likely explanation for these laboratory findings?*
- *Arterial blood gas analysis is most likely to show which of the following sets of findings?*

Formulating the Diagnosis

- *Which of the following is the most likely diagnosis?*
- *Which of the following is the most likely working diagnosis?*
- *Determining Prognosis/Outcome*
- *Based on these findings, this patient is most likely to develop which of the following?*
- *Which of the following is the most likely complication of this patient's current condition?*

The lead-in:**Management**

- The management competency contains a range of concepts:
- Health Maintenance and Disease Prevention
- Pharmacotherapy
- Clinical Interventions/Treatment.

- In most items that focus on the testing point of management, the patient's diagnosis is inferred so that the appropriate management can be determined.

The lead-in:**Management**

- *Which of the following immunizations should be administered at this time?*
  Which of the following is the most appropriate screening test?
  Which of the following tests would have predicted these findings?
  Which of the following is the most appropriate intervention?
- *For which of the following conditions is this patient at greatest risk?*
  Which of the following is most likely to have prevented this condition?
  Which of the following is the most appropriate next step in management to prevent [morbidity/mortality/disability]?
- *Which of the following should be recommended to prevent disability from this patient's injury/condition?*
- *Early treatment with which of the following is most likely to have prevented this patient's condition?*
- *Supplementation with which of the following is most likely to have prevented this patient's condition?*

The lead-in:**Management**

- Pharmacotherapy/Clinical Interventions and Treatments: These items assess principles of chronic and acute in inpa- tient and outpatient care.

  - *Which of the following is the most appropriate initial or next step in patient care?*
  - *Which of the following is the most effective management?*
  - *Which of the following is the most appropriate pharmacotherapy?*
  - *Which of the following is the first priority in caring for this patient?*

The lead in: **Mechanisms of Disease**

The items in this competency should evaluate test-takers' knowledge of pathophysiology in its broadest sense, including etiology, pathogenesis, natural history, clinical course, associated findings, complications, severity of illness, and intended or unintended effects of therapeutic interventions.

These items should be framed in a clinical context.

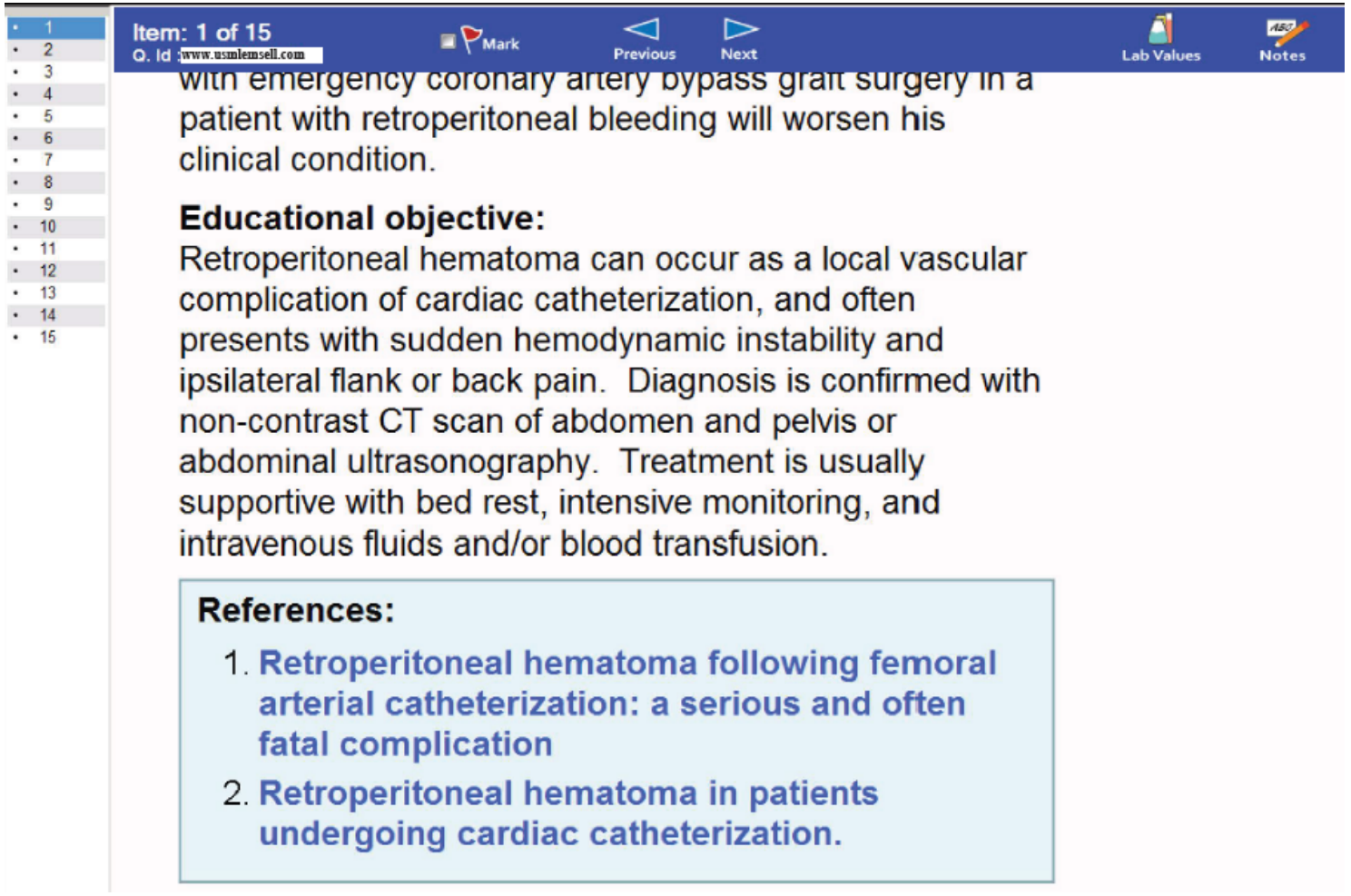
The lead in: **Mechanisms of Disease**

In general, the writer should open items on mechanisms of disease with a clinical vignette of a patient and his/her symptoms, signs, history, and lab study findings

Example lead-ins :

- *Which of the following is the most likely explanation for these findings?*
- *Which of the following is the most likely location of this patient's lesion?*
- *Which of the following is the most likely pathogen?*
- *Which of the following findings is most likely to be increased/decreased?*
- *A biopsy specimen is most likely to show which of the following?*

EBM supported questions



Common mistakes and how to avoid them

Technical item flaws

- **Flaws Related to Irrelevant Difficulty**
  - Options are overly long or complicated
  - Numeric data are not stated consistently
  - Terms in the options or the stem are vague
  - Language or structure of the options is not homogeneous
  - Options are not in a logical order
  - “None of the above” is used as an option
  - Stems are unnecessarily complicated
  - Stems contain negative phrasing

- **Flaws Related to Testwiseness**
  - Grammatical cues exist because one or more distractors don’t follow grammatically from the stem
  - Options are cued by being paired or exhaustive, where some options can be eliminated because other
  - options cover all possible outcomes
  - Absolute terms such as “always” or “never” are in some options
  - The correct answer is longer, more specific, or more complete than the other options
  - A word or phrase is included both in the stem and in the correct answer
  - Convergence (the correct answer includes the most elements in common with the other options) is present

Peer review committees in HMOs may move to take action against a physician's credentials to care for participants of the HMO. There is an associated requirement to assure that the physician receives due process in the course of these activities. **Due process must include which of the following?**

A. *Notice, an impartial forum, council, and a chance to hear and confront evidence*
B. *Proper notice, a tribunal empowered to make the decision, a chance to confront witnesses, and a chance to present evidence in defense*
C. *Reasonable and timely notice, an impartial panel empowered to make a decision, a chance to hear evidence and to confront witnesses, and the ability to present evidence in defense*

Options are overly long or complicated

- The stem contains **extraneous information**, and in fact, the stem is not needed to answer the question.
- The options themselves are **overly long** and complicated.
- The number of words in each option increases the reading load, which can shift the construct that is being measured from content knowledge to reading speed.
- Please note that this flaw relates only to options.

Options are overly long or complicated

- There are many well-constructed test questions that include a long stem, and decisions about stem length should be made in accordance with the testing point of the item.
- If the purpose of the item is to assess whether or not the student can interpret and synthesize information to determine, for example, the most likely diagnosis for a patient, then it is appropriate for the stem to include a fairly complete description of the situation.

After a second episode of infection, which of the following is the likelihood that a woman is infertile?

A. *Less than 20%*
B. *20 to 30%*
C. *Greater than 50%*
D. *75%*
E. *90%*

Numeric data are not presented consistently

- When numeric options are used, the options should be listed in numeric order and in a single format (i.e., as terms or ranges).
- Confusion can occur when formats are mixed or when options are listed in an illogical order.
- All options should be expressed as ranges or as specific percentages; mixing them is ill-advised.
- In addition C = D + E, which almost certainly rules out options D and E as correct answers for the test- wise examinee.

Severe obesity in early adolescence:

A. *has a 75% chance of clearing spontaneously*
B. *often is related to endocrine disorders*
C. *shows a poor prognosis*
D. *usually responds dramatically to dietary regimens*
E. *usually responds to pharmacotherapy and intensive psychotherapy*

Frequency terms that are vague and open to interpretation are used in the options

Vague frequency terms in the options (such as “often” or “usually”) are not consistently defined or interpreted by the readers, and sometimes not even by experts.

This can lead to multiple correct answers or a set of options that cannot be rank ordered in terms of correctness.

A 39-year-old man is brought to the hospital by his brother because he has become forgetful and confused and wanders at night because he cannot sleep. He has been drinking heavily and eating very little and has been slightly nauseated and tremulous for 4 weeks. On admission, 5% dextrose in water is initiated IV. Two hours later, the patient has ophthalmoplegia and is oriented to person only. ***Which of the following is the most appropriate next step in management?***

A. Administration of dabigatran
B. Administration of diazepam
*C. Administration of large doses of vitamin B1**
D. Administration of large doses of vitamin C
E. None of the above

“None of the above” is used as an option

- This is problematic in items where judgment is involved and the options are not absolutely true or false.
- If the correct response is intended to be one of the other listed options, knowledgeable students are faced with a dilemma because they have to decide between the option that the item writer has intended as correct and an option that encompasses everything not listed in the option set.

“None of the above” is used as an option

- Test-takers can often intuit an option that is more correct than the item writer intended to be correct, which would lead them to use the more expansive option.
- Use of “none of the above” essentially turns the item into a true-false item; each option has to be evaluated as more or less true than the universe of unlisted options.
- It is often possible to fix such items by replacing “none of the above” with an option that is more specific.
- “None of the above” could be replaced by “No management is indicated at this time,” to eliminate any ambiguity while still requiring the test-taker to commit to a management decision.

In a vaccine trial, 200 two-year-old boys were given a vaccine against a certain disease and then monitored for 5 years for occurrence of the disease. Of this group, 85% never contracted the disease. ***Which of the following statements concerning these results is correct?***

A. *The number of cases (i.e., 30 cases over 5 years) is too small for statistically meaningful conclusions*
B. *Vaccine efficacy (%) is calculated as 85-15/100*
C. *No conclusions can be drawn because the trial involved only boys*
D. *No conclusion can be drawn since no follow-up was done with nonvaccinated children*

Options are not homogeneous or parallel

- This item illustrates a common flaw where the options are not only too long but the structure of each option is different, both of which add to the reading time.
- Generally, this flaw can be corrected by careful editing to ensure that the options all have the same format and the same structure.
- In this particular item, the lead-in can be changed to “Which of the following is the most likely reason no conclusion can be drawn from these results?” Each option can then be edited to fit a logical and parallel answer to the lead-in.

Arrange the parents of the following children with Down syndrome in order of highest to lowest risk of recurrence. Assume that the maternal age in all cases is 22 years and that a subsequent pregnancy occurs within 5 years. **The karyotypes of the daughters are:**

I: 46, XX, -14, +T (14q21q) pat
II: 46, XX, -14, +T (14q21q) de novo
III: 46, XX, -14, +T (14q21q) mat
IV: 46, XX, -21, +T (14q21q) pat
V: 47, XX, -21, +T (21q21q) (parents not karyotyped)

A. III, IV, I, V, II
B. IV, III, V, I, II
C. III, I, IV, V, II
D. IV, III, I, V, II
E. III, IV, I, II, V

Stems are unnecessarily complicated

- This item requires that the student
- (a) understands the concepts of genetics that are represented
- (b) is able to rank order Roman numerals (the second of which is an irrelevant and unnecessarily difficult addition to the goal of the item).

- This item should be rewritten with the karyotypes arranged in the options themselves, so that the student who understands the order of risk of occurrence can more easily identify the correct answer.

Each of the following statements about cholesterol is true EXCEPT:

1. *Cholesterol contains numerous fatty acids*
2. *Cholesterol is not present in any foods of plant origin*
3. *Cholesterol is required in many complex bodily functions*
4. *Endogenous cholesterol is produced within the body*

Stems are negatively phrased

- A negative phrasing in the stem asks the test-taker to find the most false or least accurate option, with the rest being accurate, rather than to find the most accurate option.

- If most of the items on a test are positively phrased, the inclusion of a negatively phrased item stem carries the risk that the student will miss the word “**except**,” even when it is set in bold and/or capitalized.

FLAWS RELATED TO TESTWISENESS

A 60-year-old man is brought to the emergency department by the police, who found him lying unconscious on the sidewalk. ***After ascertaining that the airway is open, the first step in management should be intravenous administration of:***

A. CT scan of the head
B. diazepam
C. examination of cerebrospinal fluid
*D. glucose with vitamin B1 (thiamine)**
E. phenytoin

Presence of grammatical cues

- This flaw exists when an option does not follow grammatically from the stem or lead-in.
- In this example, testwise students can eliminate A and C as possible correct answers because they do not follow grammatically or logically from the lead-in. Testwise students then have to choose only between B, D, and E.
- This can happen when an item writer focuses more attention on writing the correct answer than on the distractors, leading to the potential for grammatical errors.
- To avoid this flaw, read each option immediately following the stem to ensure that the language is a good fit. Another way to avoid the flaw is to always use closed lead-ins, which helps the item writer avoid this problem.

Administration of furosemide results in:

A. *a decrease in urine potassium*
B. *an increase in urine potassium*
C. *improved glucose control in patients with type 2 diabetes*
D. *no change in urine potassium*
E. *requires decreasing the dose with renal failure*

Presence of grouped or collectively exhaustive options

- This flaw exists when a savvy student can identify a subset of options that cover all possible outcomes (are collectively exhaustive) and rule out the options not in that subset.
- In this item, options A, B, and D are exhaustive – urine potassium can only increase, decrease, or not change – and thus one of the three options must be the correct answer.
- A less testwise student might spend time considering C and E.
- Often, item writers add options like C and E only because they want to have a total of five options, but it is not an improvement of the item to add options that have no merit.
- The item writer should be able to rank order each option on the same dimension, and no subset of options should include all possible outcomes.

In patients with advanced dementia, Alzheimer type, the memory defect:

A. *can be treated adequately with phosphatidylcholine (lecithin)*
B. *could be a sequela of early parkinsonism*
C. *is never seen in patients with neurofibrillary tangles at autopsy*
D. *is never severe*
E. *possibly involves the cholinergic system*

Use of absolute terms

- In this item, options A, B, and E contain terms that are less absolute than those in options C and D.
- The testwise student will eliminate options C and D as possibilities, because they are less likely to be true than something stated less absolutely, and so this item is flawed with the inclusion of those terms.
- This flaw tends to arise when verbs are included in the options rather than in the lead-in. Focusing the stem, placing the verb in the stem, and shortening the options are possible ways to correct this flaw.

Secondary gain is:

A. *a complication of a variety of illnesses and tends to prolong many of them**
B. *a frequent problem in obsessive-compulsive disorder*
C. *never seen in organic brain damage*
D. *synonymous with malingering*

The correct option is longer, more specific, or more complete

- In this item, the correct answer, option A, is longer than the other options, and is the only “double” option. This flaw is another potential outcome when item writers pay more attention to constructing the correct answer than the distractors.
- One reason for this is that item writers likely create the correct answer first and then write the incorrect distractors.
- Another potential reason is that because item writers are often teachers, they will construct long correct answers that include additional instructional material, parenthetical information, caveats, and so on.
- This flaw can be avoided by reviewing the entire item set for length and removing language that is purely for instructional purposes only.

A 58-year-old man with a history of heavy alcohol use and previous psychiatric hospitalization is confused and agitated. He speaks of experiencing the world as unreal. ***This symptom is called:***

A. depersonalization
B. derailment
*C. derealization**
D. focal memory deficit
E. signal anxiety

Presence of word repetition (“clang clues”)

- This flaw arises when language used in the lead-in is repeated in the options. Here, the word “unreal” in the stem can clue test-takers to the fact that the correct answer, “derealization,” is the only option that also includes the word “real.”
- The same flaw can appear even if a word is repeated only in a metaphorical sense, such as when a stem mentions bone pain and the correct answer begins with the prefix “osteo-.”
- Item writers should scan the options and item stem to check for this word or phrase repetition.

Local anesthetics are most effective in the:

A. *anionic form, acting from inside the nerve membrane*
B. *cationic form, acting from inside the nerve membrane**
C. *cationic form, acting from outside the nerve membrane*
D. *uncharged form, acting from inside the nerve membrane*
E. *uncharged form, acting from outside the nerve membrane*

Presence of convergence

- The underlying flaw is that the correct answer is the option that has the most in common with the other options, and thus the test-wise student can converge on the right answer just by counting the number of times certain terms appear.
- In this example, the testwise student would eliminate “anionic form” as unlikely because “anionic form” appears only once; that student would also exclude “outside the nerve membrane” because “outside” appears less frequently than “inside.”
- The student would then have narrowed the options to B and D. Since three of the five options involve a charge, the testwise student would then select option B, which is in fact the correct answer.

Presence of convergence

- This flaw can also occur without being directly reflected in the language; for example, if an item is asking which pharmacotherapy is most effective, and three of the five options are in one class of drugs, the savvy student may rule out the other two as less likely.
- This flaw occurs when item writers start with the correct answer and write the distractors as permutations of the correct answer. The correct answer will then be more likely to have elements in common with the rest of the options, and the incorrect answers are more likely to be outliers.
- A useful check is to review all options and see if words or terms are repeated across options.

Take home messages

- Stick to the basic rules, follow the house style with a systematic approach
- Be inspired by the clinic, but write '*teaching*' cases
- Make the cases attractive (images, videos…)
- Have them reviewed by colleagues

Acknowledgements

- Large parts of this presentation are based on the handbook of the NBME. This can be found online: https://www.nbme.org/sites/default/files/2020-01/IWW_Gold_Book.pdf

- Interesting literature:
  - Thinking fast and slow - Daniel Kahneman
  - How Doctors Think - Jerome E. Groopman