IDEAL

Assessing Medical Health Science Students Competencies

The Relevant Psychometric Theory

Reliability & Validity

Reliability:

theoretical definition

- ratio of ‘true’ variance to total variance

some lay definitions:

- repeatability (as in administered more once)
- dependability (as in same result over time)
- consistency (among items, forms, judges)

Validity

Theoretical Definition

- ratio of accurate variance to observed variance

Lay Definition

- did test instrument actually measure what was intended

What is 'True' & 'Accurate' & 'Observed' Variance?

A person's assessment score (on a question or a test) is composed of two components, i.e., some truth & some error & can be expressed as

$$X_i = T_i + E_i$$

where

X_i is the recorded observed score for individual i

T_i is the true score component of X_i

E_i is the error score component of X_i

What is 'True' & 'Accurate' & 'Observed' Variance?

The E_i component is a non systematic influence that does not correlate with the true score (randomness = error)

Sometimes E_i causes X_i to be inflated, sometimes deflated

- with a very large number of observed measures, E_i will sum to the value of zero (if it did not, then the E_i component would be systematic)

What is 'True' & 'Accurate' & 'Observed' Variance?

T_i (true score) component is the systematic, or consistent or repeatable aspect in the observed score.

It is not necessarily 'truth' in the sense of representing what was wanted or intended, simply that is predictable

- T_i may represent what is intended (e.g. clinical skill) but it may also include a systematic influence
  - such as a bias that is repeatedly present in the measure (as in an OSCE which only simulates a clinical situation)

Understanding the True Score

So the 'true' score component is dependable but may not be necessarily accurate (what was intended to be measured)

Degree of reliability is the amount of variance attributable to dependable measurement (i.e., the ratio of true to observed variance

- this is equivalent to the proportion of observed variance that is true variance).

Understanding the Accurate Score

Degree of validity of any measure depends on degree of reliability (i.e., reliability has a ceiling effect on validity):

- both reliability & validity have an underlying concept of **consistency**
- something cannot be accurate if it is not at least reliable (consistent)
- theoretically, validity cannot be more than the square root of the reliability
- reliability is a necessary condition in measurement but it is not a sufficient condition
- validity is a sufficient condition (i.e., if something is valid, it is necessarily reliable)

Estimating
'True' & 'Accurate' & 'Observed' Variance?

One can find the true score:

Take a very large number (n) of observed scores & sum them and divide each summed component by n

$$X_{i1} = T_{i1} + E_{i1}$$

$$X_{i2} = T_{i2} + E_{i2}$$

$$X_{i3} = T_{i3} + E_{i3}$$

. . . . . .

$$X_{in} = T_{in} + E_{in}$$

$$\Sigma X_{in} / n = \Sigma T_{in} / n + \Sigma E_{in} / n$$

What is 'True' & 'Accurate' & 'Observed' Variance?

What is each of these components in this form?

$$\Sigma X_{in}/n = \Sigma T_{in}/n + \Sigma E_{in}/n$$

Mean of the observed scores	Mean of the True scores	Mean of the error scores

What is mean of the error scores? _____

So the mean of the observed scores equals what? _____

What is 'True' & 'Accurate' & 'Observed' Variance?

One can find the true score:

X_{in} - E_{in} = T_{in}

observed score | True score | error score

One can also find the error score (observed minus the true score)T

Xin - Tin = Ein

One can Estimate True & Error Variance

Algebraically one can go on and also derive estimates of true and error variance

I will not go into the mathematical proof for this (as this is not a course in psychometrics)

- but if you have an keen interest in this area of investigation read Gulliksen's or Lord & Novick's texts on classical test theory

The concept to remember is that reliability is:

$$R_{(\text{reliability})} = \frac{\text{True Score Variance}}{\text{Observed Score Variance}}$$

In Parallel, Validity:

In parallel, validity can be thought of as:

$$V_{(validity)} = \frac{\text{Accurate Score Variance}}{\text{Observed Score Variance}}$$

Variety of Methods Used to Estimate Reliability

Administer a test twice & determine if results are same

- some correlate the marks between the 1st & 2nd administrations, others compare the means & variances,

- time between administrations is a factor to consider in interpreting the estimated reliability coefficient

Variety of Methods Used to Estimate Reliability

Administer two parallel tests

- some correlate the marks and use the correlation coefficient as an estimate of reliability

- some compare the means and variances between the tests

  - key factor to consider is that “parallel” tests, by definition, have the same mean & variance

    - but this may not be the case in reality

Variety of Methods Used to Estimate Reliability

One test is scored by two or more judges

- some correlate the marks between pairs of judges
- some compare the means & variances of the judges
  - if the marks are correlated, will the means & variances necessarily also be the same? _____ (yes/no)

Variety of Methods Used to Estimate Reliability

Split the test in halve & compare the halves

- some correlate the marks between the 1st & 2nd halve and use that as an estimate of reliability

- how the halves of the test are chosen can affect result

- if one takes every possible split half and calculates the correlation for each and then averages these, the reliability estimate is referred to as

  - Kuder Richardson Twenty (KR_{20}) coefficient if the data are binary (e.g., 1=right, 0=wrong)

  - Cronbach's alpha if the data are continuous

Many Types (forms) of Validity are Reported

From Weakest to Strongest

- Face Validity
- Content Validity
- Concurrent Validity
- Predictive Validity
- Construct Validity

Face Validity

Definition

Assessment appears to measure the intended attributes

- appearance to whom?

- the examinee in particular, but also anyone interested in the results of the examinee

How is it estimated?

- query people being assessed; if large % (75-80%) regard instrument as measuring the relevant attribute, then stated to have face validity

Content Validity

Definition

- each measure (item) in the assessment form comes from the domain (i.e., from the population of measures) that constitute the construct or attribute

- the collection of items in the assessment form samples the domain in a representative fashion

Content Validity

How is it estimated?

- query experts, and possibly only a widely or divergent group in terms of their orientation is needed to properly capture the concept of expert (e.g., specialist & GPs & nurses)

- if a large majority of each subset and the collective overall is in agreement that each item and the entire test comes from the domain, then content validity is established

  - large majority is often regarded as 75-80%

Concurrent Validity

Definition

Consistency between a gold standard (or best available measure) and another measure with both measures taken at the same point in time

How it is estimated

- Often the two sets are correlated
- If so, as the statistical model does not adhere to the same theoretically possible range of values, some caution is required

Predictive Validity

Definition

Consistency between a measure taken at an earlier time (i.e., T_1) with another measure taken at a later period (T_2)

How is it estimated

- prospectively: take the first measure and await for the predicted outcome to occur; compare the predicted outcome and the actual outcome

- retrospectively: measure a phenomenon & correlate that with some existing record from the past

Construct Validity

Definition

Consistency of the measure used in assessment is related to a non-operational construct or attribute (i.e., but one can't observe or measure the construct directly, only indirectly)

- the most powerful form of validity as it offers the most generalizablity

- widely useful as it applies to a multitude of situations

Construct Validity

How is it estimated

- a series of measures assumed to measure the same attribute cluster together (e.g., factor analysis)
- groups that should differ in the ability described by the construct, are compared to see if they do differ and in the direction that would support validation
    - this is weak evidence, as the groups might vary for other reasons
- pattern of convergence & divergence is consistent with validation (e.g., multi-method mult-trait matrix)


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Spearman-Brown prediction formula

$$r' = \frac{Nr}{1+(N-1)r}$$

$$N = \frac{r'(1-r)}{r(1-r')}$$

where **N =** $\dfrac{\text{\# of items in adjusted test}}{\text{\# of items in original test}}$

r = reliability of original test

r' = reliability of adjusted test

Spearman-Brown Prophecy Formula

Assume original reliability was .807 for test of 55 items

Assume you used only 25 items that loaded well on one common factor

- **when calculating reliability for the 25, the reliability would likely be _____ (higher, same, less)**
- **why?**

Spearman-Brown Prophecy Formula

See hand out


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

IDEAL

Assessing Medical Health Science Students Competencies

The Relevant Statistical Properties

for

Scored Responses: Binary & Continuous Data

Student Responses Score as Binary Data

Paediatrics: Item Analyses of A-type Items

Flagging the Quality of Questions

The most advanced language function a 14-month-old child usually can perform is to

A. babble
B. speak several recognizable words
C. combine two different words
D. speak in complete sentences
E. count from one to ten

```
ITEM  32: DIF=0.683, RPB=  0.045, CRPB= -0.022 (95% CON= -0.328, 0.287)
                     RBIS= 0.058, CRBIS=-0.029, IRI=0.021
    GROUP    N   INV   NF  OMIT     A        B*       C
    TOTAL   41     0    0    0    0.02     0.68     0.29
     HIGH   10     0              0.10     0.60     0.30
     MID    19     0              0.00     0.74     0.26
     LOW    12     0              0.00     0.67     0.33
    TEST SCORE MEAN %:              64       58       57
    DISCRIMINATING POWER          0.10    -0.07    -0.03
    STANDARD ERROR OF D.P.        0.09     0.21     0.20
```

Ques 30 [Module 3 (2003-04) EOM Exam] ItemID CP5-230 Ans A

Which of the following statements best describes events that contribute to the development of cardiopulmonary arrest in children?

A. arrest in most often end result of deterioration in respiratory and circulatory function
B. arrest in most often precipitated by ventricular arrhythmias
C. arrest in most likely the result of neurologic dysfunction and hypoventilation
D. arrest in most often an indicator of cardiovascular disease
E. arrest in most likely the result of electrolyte disturbance

ITEM 30: DIF=0.317, RPB= -0.266, CRPB= -0.334 (95% CON= -0.582,-0.029)
RBIS=-0.348, CRBIS=-0.436, IRI=-.124

GROUP	N	INV	NF	OMIT	A*	B	C	D	E
TOTAL	41	0	0	0	0.32	0.44	0.12	0.02	0.10
HIGH	11	0			0.18	0.64	0.09	0.00	0.09
MID	18	0			0.28	0.44	0.11	0.06	0.11
LOW	12	0			0.50	0.25	0.17	0.00	0.08
TEST SCORE MEAN %:					59	65	59	61	62
DISCRIMINATING POWER					-0.32	0.39	-0.08	0.00	0.01
STANDARD ERROR OF D.P.					0.20	0.21	0.14	0.00	0.12

Ques 44 [Module 4 (2003-04) EOM Exam] ItemID CP5-344 Ans A

In developing countries, the most important measure in the prevention of cholera in children would be:

A. Clean water supply
B. Cholera vaccination of all the toddlers
C. Parents washing their hands before preparing food for the children
D. Ample supply of oral rehydration fluid
E. Breast feeding up to 9 months of age

ITEM 44: DIF=0.500, RPB= -0.191, CRPB= -0.276 (95% CON= -0.547, 0.048)
RBIS=-0.240, CRBIS=-0.346, IRI=-.096

GROUP	N	INV	NF	OMIT	A*	C	D	E
TOTAL	38	0	0	0	0.50	0.03	0.03	0.45
HIGH	11	0			0.45	0.00	0.09	0.45
MID	19	0			0.42	0.05	0.00	0.53
LOW	8	0			0.75	0.00	0.00	0.25
TEST SCORE MEAN %:					61	64	69	63
DISCRIMINATING POWER					-0.30	0.00	0.09	0.20
STANDARD ERROR OF D.P.					0.23	0.00	0.10	0.22

Ques 77 [Module 4 (2003-04) EOM Exam] ItemID CP5-377 Ans C

In the management of febrile convulsion, which of the following is true:

A. An abnormal EEG is predictive of a higher risk of developing epilepsy
B. Regular oral antipyretic during recurrent febrile convulsions is effective in preventing epilepsy
C. There is no evidence that anti-pyretic treatment prevents the recurrence of febrile convulsions.
D. Regular rectal diazepam should be given to the child for at least 48 hours
E. Regular oral phenobarbitone should be given to the child for at least 48 hours

```
ITEM  77: DIF=0.421, RPB= -0.014, CRPB= -0.098 (95% CON= -0.405, 0.229)
                     RBIS=-0.018, CRBIS=-0.123, IRI=-.007
    GROUP    N   INV   NF  OMIT     A       B       C*
    TOTAL   38     0    0     0    0.47    0.11    0.42
     HIGH   11     0                0.55    0.09    0.36
     MID    19     0                0.53    0.11    0.37
     LOW     8     0                0.25    0.13    0.63
    TEST SCORE MEAN %:               62      61      62
    DISCRIMINATING POWER             0.30   -0.03   -0.26
    STANDARD ERROR OF D.P.           0.23    0.14    0.23
```

Ques 63 [Module 2 (2003-04) EOM Exam] ItemID CP5-163 Ans D

A term Chinese baby girl was delivered vaginally with birth weight of 3 kg. 24 hours after birth she was noticed to develop jaundice with serum bilirubin of 200 umol/l. What is the most likely cause of her jaundice?

A. Physiological jaundice
B. Glucose-6-phosphate dehydrogenase deficiency
C. Breast milk jaundice
D. Haemolytic disease of newborn due to ABO incompatibility
E. Biliary atresia

ITEM 63: DIF=0.675, RPB= 0.585, CRPB= 0.531 (95% CON= 0.262, 0.723)
RBIS= 0.762, CRBIS= 0.690, IRI=0.274

GROUP	N	INV	NF	OMIT	A	B	C	D*	E
TOTAL	40	0	0	0	0.15	0.10	0.05	0.68	0.03
HIGH	11	0			0.00	0.00	0.00	1.00	0.00
MID	17	0			0.18	0.06	0.00	0.71	0.06
LOW	12	0			0.25	0.25	0.17	0.33	0.00
TEST SCORE MEAN %:					58	57	48	67	65
DISCRIMINATING POWER					-0.25	-0.25	-0.17	0.67	0.00
STANDARD ERROR OF D.P.					0.14	0.14	0.12	0.20	0.00

A 4-month-old infant has recurrent episodes of pneumonia. He is irritable during feeds with frequent large volume vomits. Blood picture reveals hypochromic microcytic anaemia. The most likely diagnosis is:

A. Excessive milk intake
B. Pyloric stenosis
C. Immunodeficiency
D. Hiatus hernia
E. Choanal atresia

ITEM 32: DIF=0.421, RPB= 0.473, CRPB= 0.390 (95% CON= 0.080, 0.631)
RBIS= 0.597, CRBIS= 0.492, IRI=0.233

GROUP	N	INV	NF	OMIT	A	B	C	D*	E
TOTAL	38	1	0	1	0.03	0.37	0.13	0.42	0.03
HIGH	11	0			0.09	0.18	0.09	0.64	0.00
MID	19	1			0.00	0.32	0.11	0.47	0.05
LOW	8	0			0.00	0.75	0.25	0.00	0.00
TEST SCORE MEAN %:					68	59	59	65	62
DISCRIMINATING POWER					0.09	-0.57	-0.16	0.64	0.00
STANDARD ERROR OF D.P.					0.10	0.23	0.17	0.22	0.00

Ques 76 [Module 4 (2003-04) EOM Exam] ItemID CP5-376 Ans B

A two year old child presents with generalized tonic clonic seizures at the Emergency department. The most important on-site investigation:

A. Lumbar puncture
B. Blood glucose
C. Blood electrolytes
D. Skull X ray
E. White cell count

```
ITEM  76: DIF=0.789, RPB=  0.588, CRPB=  0.519 (95% CON=  0.239, 0.719)
                     RBIS= 0.830, CRBIS= 0.733, IRI=0.240
    GROUP    N   INV   NF  OMIT     A       B*      C
    TOTAL   38     0    0     0    0.13    0.79    0.08
     HIGH   11     0               0.00    1.00    0.00
     MID    19     0               0.05    0.89    0.05
     LOW     8     0               0.50    0.25    0.25
    TEST SCORE MEAN %:              56      64      55
    DISCRIMINATING POWER           -0.50    0.75   -0.25
    STANDARD ERROR OF D.P.          0.19    0.22    0.14
```

The following may be an indication for stopping breastfeeding

A. The infant becomes jaundiced before the 5th day of life
B. The infant develops diarrhoea with secondary lactose intolerance
C. The mother develops mastitis and is receiving antibiotic treatment
D. The mother is taking regular inhaled steroid therapy for asthma
E. The infant develops jaundice and vomiting and the urine is positive for reducing substances

```
ITEM  84: DIF=0.474, RPB=  0.370, CRPB=  0.286 (95% CON= -0.038, 0.555)
                     RBIS= 0.464, CRBIS= 0.358, IRI=0.185
    GROUP   N   INV   NF  OMIT     A       B       C       D       E*
    TOTAL   38    0    0    0    0.03    0.24    0.21    0.05    0.47
     HIGH   11    0              0.00    0.18    0.09    0.00    0.73
     MID    19    0              0.00    0.11    0.32    0.11    0.47
     LOW     8    0              0.13    0.63    0.13    0.00    0.13
    TEST SCORE MEAN %:            50      59      61      62      64
    DISCRIMINATING POWER        -0.13   -0.44   -0.03    0.00    0.60
    STANDARD ERROR OF D.P.       0.10    0.22    0.14    0.00    0.23
```

Ques 15 [Module 1 (2003-04) EOM Exam] ItemID CP5-15 Ans E

The following syndrome diseases are usually associated with obesity EXCEPT:

A. Prader Willi Syndrome
B. Turner Syndrome
C. Pseudohypoparathyroidism
D. Panhypopituitarism
E. Russell-Silver Syndrome

ITEM 15: DIF=0.488, RPB= 0.286, CRPB= 0.215 (95% CON= -0.100, 0.490)
RBIS= 0.359, CRBIS= 0.269, IRI=0.143

GROUP	N	INV	NF	OMIT	B	C	D	E*
TOTAL	41	0	0	0	0.37	0.10	0.05	0.49
HIGH	10	0			0.20	0.00	0.00	0.80
MID	19	0			0.42	0.00	0.11	0.47
LOW	12	0			0.42	0.33	0.00	0.25
TEST SCORE MEAN %:					57	50	59	60
DISCRIMINATING POWER					-0.22	-0.33	0.00	0.55

Select the correct statement regarding Duchenne muscular dystrophy

A. It is an autosomal recessive condition
B. Intellectual impairment is a feature
C. Ability to walk is lost at around 5 years of age
D. Alveolar hypoventilation does not occur
E. Lung function is abnormal at birth

ITEM 40: DIF=0.683, RPB= 0.505, CRPB= 0.438 (95% CON= 0.151, 0.657)
RBIS= 0.659, CRBIS= 0.572, IRI=0.235

GROUP	N	INV	NF	OMIT	A	B*	C	D	E
TOTAL	41	0	0	0	0.05	0.68	0.10	0.12	0.05
HIGH	10	0			0.00	1.00	0.00	0.00	0.00
MID	19	0			0.11	0.68	0.11	0.11	0.00
LOW	12	0			0.00	0.42	0.17	0.25	0.17
TEST SCORE MEAN %:					57	60	50	52	54
DISCRIMINATING POWER					0.00	0.58	-0.17	-0.25	-0.17
STANDARD ERROR OF D.P.					0.00	0.20	0.12	0.15	0.12

Ques 63 [Module 1 (2003-04) EOM Exam] ItemID CP5-63 Ans D

The most frequent complication of congenital rubella is:

A. cataracts
B. microcephaly
C. patent ductus arteriosus
D. deafness
E. thrombocytopenia

```
ITEM  63: DIF=0.780, RPB=  0.411, CRPB=  0.352 (95% CON=  0.049, 0.595)
                     RBIS= 0.575, CRBIS= 0.492, IRI=0.170
    GROUP   N   INV   NF  OMIT     A       B       C       D*
    TOTAL   41    0    0    0    0.17    0.02    0.02    0.78
     HIGH   10    0              0.00    0.00    0.00    1.00
     MID    19    0              0.16    0.00    0.05    0.79
     LOW    12    0              0.33    0.08    0.00    0.58
    TEST SCORE MEAN %:             51      53      58      59
    DISCRIMINATING POWER         -0.33   -0.08    0.00    0.42
    STANDARD ERROR OF D.P.        0.17    0.09    0.00    0.18
```

Ques 95 [Module 1 (2003-04) EOM Exam] ItemID CP5-95 Ans B

In a 2-year-old with persistence of the ductus arteriosus

A. The murmur is mainly systolic
B. There is a wide pulse pressure
C. The murmur quietens on lying the child flat
D. The ductus will be closed by administration of indomethacin
E. There will be cyanosis on exercise

ITEM 95: DIF=0.756, RPB= 0.442, CRPB= 0.380 (95% CON= 0.082, 0.616)
RBIS= 0.605, CRBIS= 0.521, IRI=0.190

GROUP	N	INV	NF	OMIT	A	B*	C	D	E
TOTAL	41	0	0	0	0.02	0.76	0.05	0.15	0.02
HIGH	10	0			0.00	0.90	0.00	0.00	0.10
MID	19	0			0.00	0.95	0.00	0.05	0.00
LOW	12	0			0.08	0.33	0.17	0.42	0.00
TEST SCORE MEAN %:					45	59	50	51	69
DISCRIMINATING POWER					-0.08	0.57	-0.17	-0.42	0.10
STANDARD ERROR OF D.P.					0.09	0.21	0.12	0.18	0.09

Ques 64 [Module 3 (2003-04) EOM Exam] ItemID CP5-264 Ans A

A 2-month old baby boy presents with cough for a few days. His mother describes the cough to be bad. No other noises are heard. His mother is worried as he seems to turn colour toward the end of cough. His feeding is well in between the coughing spell. His father has been coughing for the last few weeks and takes some over-the-counter medications with some improvement. He is not febrile and physical examination of the baby is normal. His chest radiograph is also normal.

The most likely diagnosis is:

A. Pertussis
B. Asthma
C. Bronchiolitis
D. Pneumonia
E. Laryngomalacia

Ques 64 [Module 3 (2003-04) EOM Exam] ItemID CP5-264 Ans A

ITEM 64: DIF=0.732, RPB= 0.490, CRPB= 0.426 (95% CON= 0.136, 0.648)
RBIS= 0.659, CRBIS= 0.572, IRI=0.217

GROUP	N	INV	NF	OMIT	A*	C	E
TOTAL	41	0	0	0	0.73	0.22	0.05
HIGH	11	0			1.00	0.00	0.00
MID	18	0			0.78	0.22	0.00
LOW	12	0			0.42	0.42	0.17
TEST SCORE MEAN %:					64	57	53
DISCRIMINATING POWER					0.58	-0.42	-0.17
STANDARD ERROR OF D.P.					0.19	0.17	0.12


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

IDEAL

Penalty Scoring in Selected Response Items

Does the adjustment improve validity?

C B Hazlett

Rationale & Evidence Supporting Use of Penalty Scoring

Rationale: guessing correctly is intellectual inaccuracy:

if there was no guessing, better information would be available re what each individual really does know

- thereby guiding teaching & decision marking re candidates

Evidence: reliability of a test with a correction formula applied for guessing is often more reliable with T/F items

- typically improvement is in 0.05 range for the test's internal consistency (i.e., Cronbach's alpha or KR_{20})

Evidence for Not Using Penalty Scoring

Evidence: medical students rarely make random guesses

- penalty formulae are designed to adjust for random guesses

Evidence for Not Using Penalty Scoring

Evidence: when penalty for guessing is used, it does not remove effect of guessing from test scores:

- students who have partial knowledge & are risk takers will 'guestimate', and on average do better
- students with partial knowledge & lacking confidence will omit answering & are thus penalized in comparison to risk takers

Thus, penalty scoring introduces unintended (error) variance

Evidence for Not Using Penalty Scoring

Doctors have to make most decisions under uncertainty & the use of penalty scoring can attenuate measuring that ability

Evidence for Not Using Penalty Scoring

Evidence:

- probability of a good test score by guessing alone is negligible

- likelihood of doing so in longer MCQ exams is <.000001 (i.e., less than 1 in a million)

Reasons for Not Using Penalty Scoring

Evidence: the correlation between corrected & uncorrected scores is very high

- score drops, but rank order of students is same
- with good questions, relative differences among the students are also the same

Reasons for Not Using Penalty Scoring

If concerned that without use of a penalty some students may inappropriately cross the bar of acceptability (i.e., pass instead of fail) then:

- the bar must be properly set (i.e., use standard setting methods)
- bar should not be an arbitrary, fixed % point (e.g., 50%)

Evidence: few tests are composed so that achieving this common, fixed % point correctly identifies the competent candidates

- or not achieving the % correctly identifies the incompetent ones

Evidence for Not Using Penalty Scoring

If a MCQ test:

- is reasonably long (minimum number of items is 40)
- is composed with quality, well-written items
- uses a sufficient number of options (unlike T/F formats)
- targets the level of teaching & expected learning by students

Then the evidence is: medical students do not guess because it is an ineffective strategy in getting good marks

Evidence for Not Using Penalty Scoring

If test is composed of

- poorly constructed, ambiguously worded items
- uses few options (such as all T/F items)
- uses implausible options
- misses level of coverage provided (i.e., is too difficult)

Then the evidence is: guessing is regarded by some students as an effective strategy, unless the penalty for guessing is high

Given the Evidence

Assumption there is a reasonable amount of random guessing is false

Assumption that penalty scoring corrects for guessing is false

Assumption that normal penalty scoring discourages guessing is false

Using a large penalty to stop students guessing on poorly constructed items is useless: i.e., results with or without a penalty are less valid

Does this Evidence apply Locally?

Compared penalty & non penalty scoring with A- & X-type items used in HK Licensing Examination 1997

- Quality of A type items regarded acceptable (internationally)
- Quality of X type typically of lower standard (but used in CUHK)

Licensing group: N=178

- Assessed B. Sc, Surg, OBG using 100 A-type items (best 1 of 5)
- Assessed Med, Ped, Psy, Orth, Ethics using 120 X-type items (i.e., Multiple T/F)

Does the Evidence Apply Locally

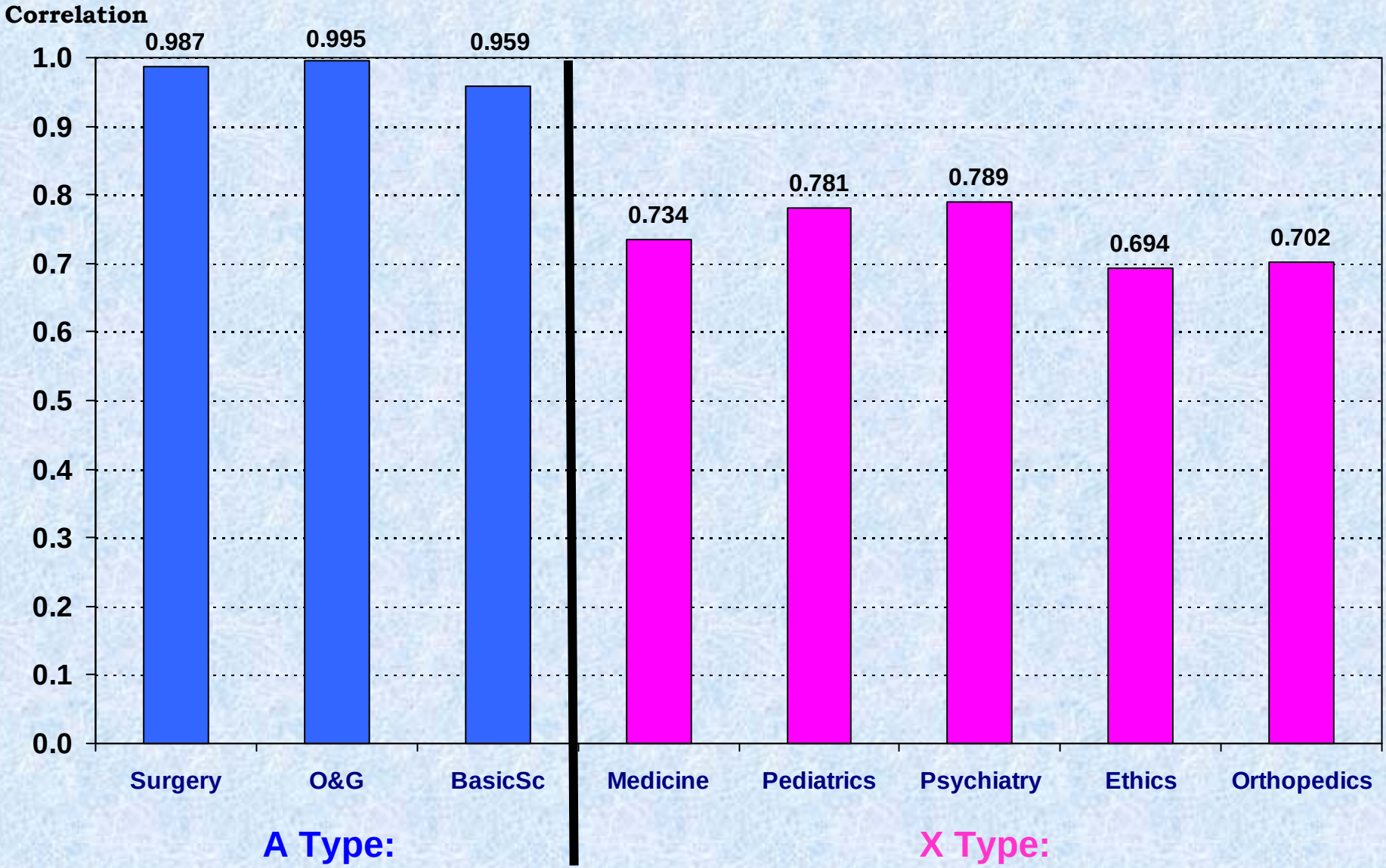
For 25 of the A-type items in Surgery, also used to assess 5th yr CUHK students

Findings:

- reliability for A-type items was 0.877
- reliability for X-type items was 0.87722
- correlation between candidates scores with and without penalty applied was high in X-type and very high with A-type

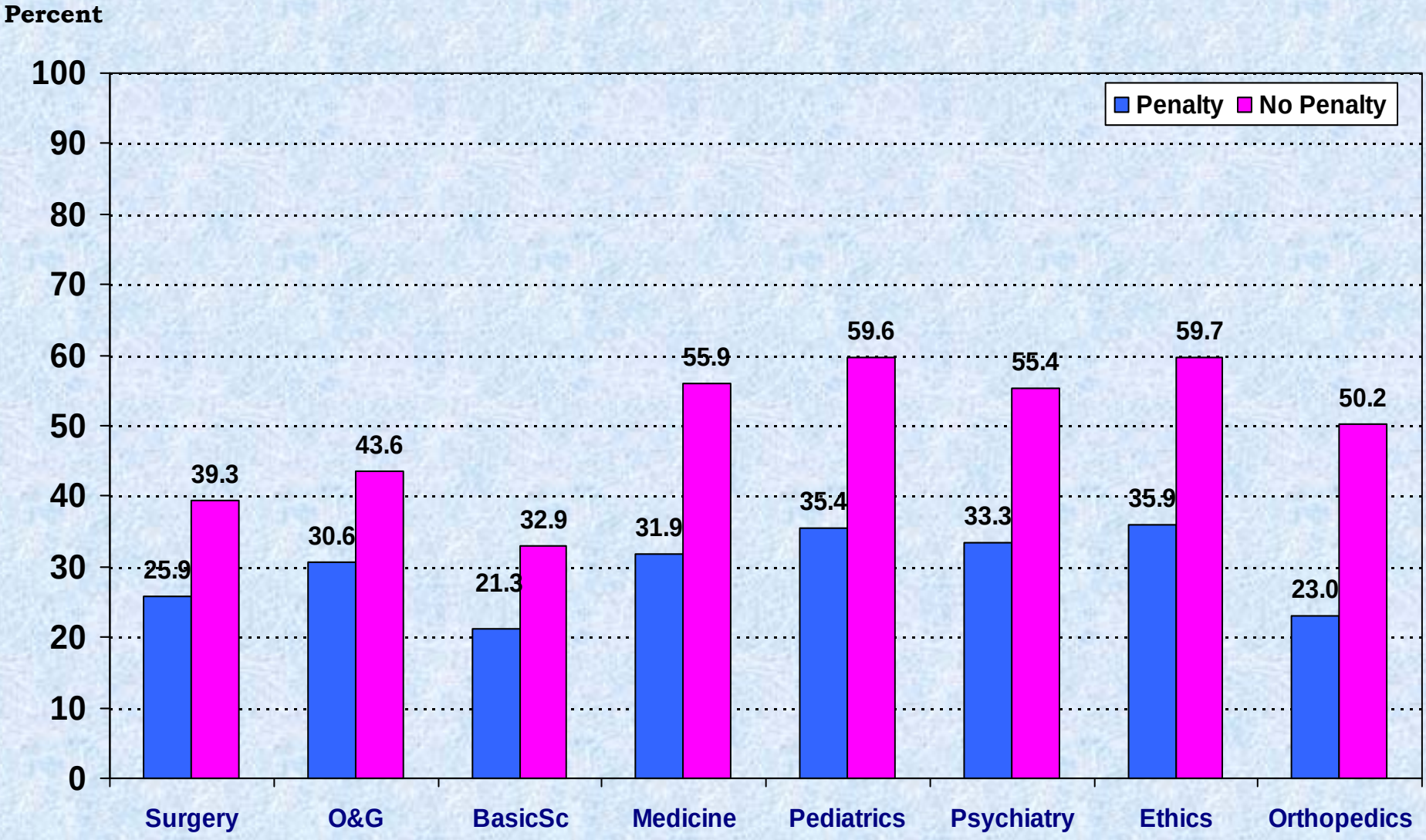
Licensing Examination: September 1997

Correlation Between Scores with and without Penalty for Guessing on A- & X-type Exams



Licensing Examination Results: September 1997

Average % Score on Each Test: With/Without Penalty

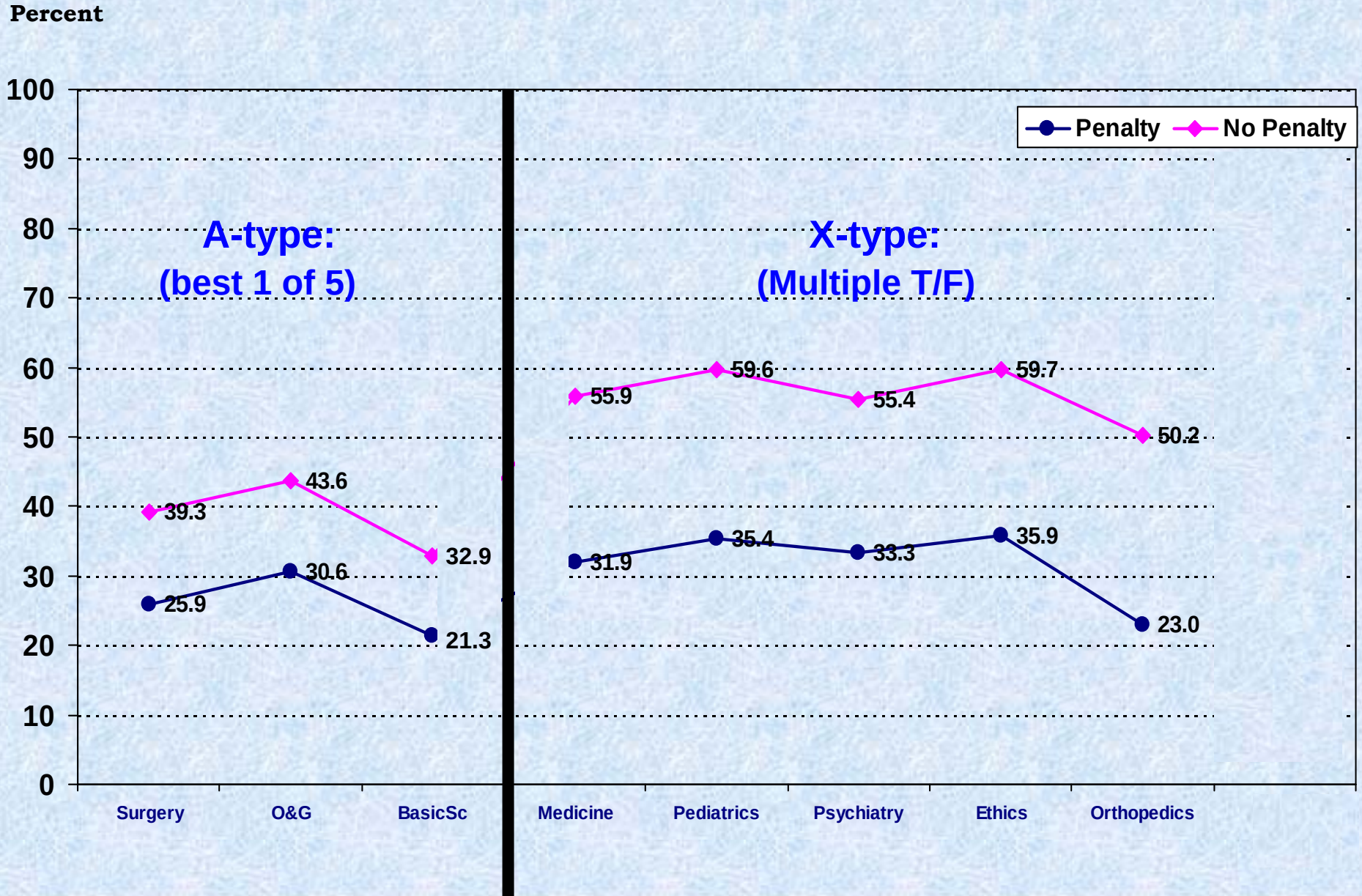


A Type: Average drop due to penalty is 12.9%

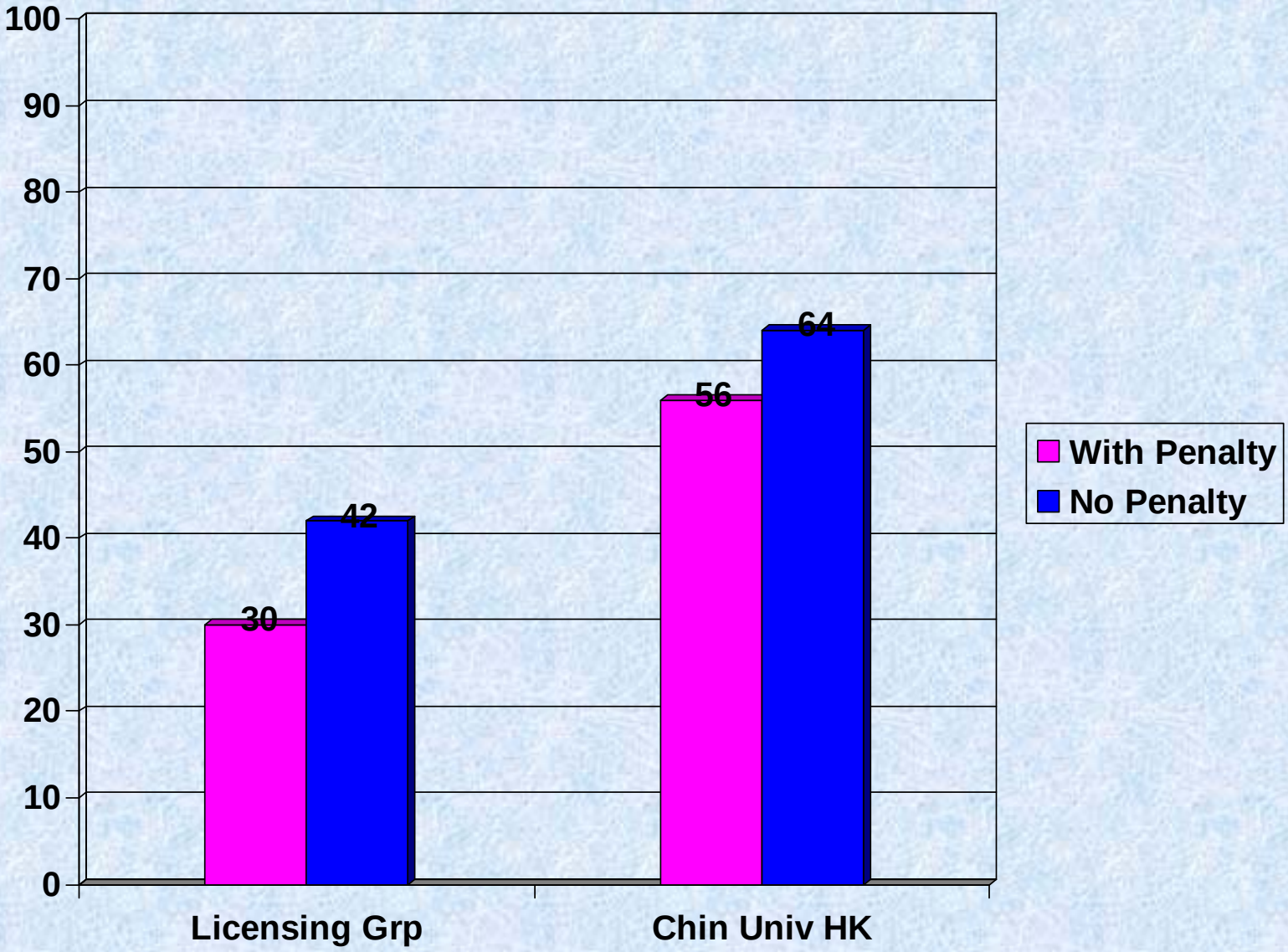
X Type: Average drop due to penalty 24.3%

Licensing Examination Results: September 1997

Average Performance With and Without Penalty



Comparative Effect of Penalty Scoring Using 25 A-type Items in Surgery



Summary & Conclusion

Given the evidence, few support use of penalty scoring

Some recent literature (hand out) shows that there are still some remnants --- but the arguments rely on issues of reliability, not validity, and unwarranted assumptions about what a test can achieve in terms of intellectual accuracy

"We must not expect more precision than the subject admits"

Aristotle: *Nicomachean Ethics*

IDEAL

Q & A

Faculty of Medicine, Curriculum Retreat 2005


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Assessment

Importance of assessment: legitimate concern to

1. **those who learn**
2. **those who teach**
3. **those who accredit**

Importance of Assessment in Learning

Brown & Knight (1994)

"Assessment is at the heart of the student experience

- **Defines what students see as important**
  - how students spend their time
  - how students come to see themselves as students and later as graduates"

Importance of Assessment

Students take cues from what is assessed rather than from lecturers assert as important

“If you want to want to change student learning then change the methods of assessment”

What is Assessment?

Taking a sample of what students do

Making inferences based on this sample

Estimating the worth of particular actions

Sample may include:

- use of computers
- writing essays
- completing tests/checklists
- solving problems and reporting solutions
- carrying out practical procedures
- recalling & reporting orally, in writing, by actions, with demonstrated feelings

Inferences are made about student's

- achievements
- potential
- intelligence
- aptitudes
- attitudes
- motivation
- personality

Estimating Worth

Estimating worth of particular actions include

- **worth in the form of grades**
- **worth in the form of marks**
- **worth in the form of recommendations**

and may be used for judgmental or for developmental purposes

Weaknesses

Each component of assessment may suffer weaknesses

Sampling may

- not be representative of the student's capabilities
- not match learning objectives of the course
- draw on too narrow a domain (e.g., only written examinations)
- over-weighted towards particular skills or methods

Weaknesses

Inferences drawn may vary among assessors (e.g., regarding a student assignment)

- **this is particularly likely if explicit criteria are not available**

Weaknesses

Estimates of worth in terms of marks or grades may also vary

The common weaknesses in assessment systems

Overload of students and staff.

Too many assignments with the same deadline set in the department/school

Insufficient time for students to complete assignments in the time available.

Insufficient time for staff to mark the assignments before the next semester.

Inadequate or superficial feedback provided to students.

Wide variations in assessment demands of different modules.

Wide variations in marking across modules.

p. 9, Brown, Bull and Pendlebury

The common weaknesses in assessment systems

Wide variations in marking within a module.

Wide variations in marking by demonstrators.

Fuzzy or non-existent criteria.

Undue precision and specificity of marking schemes or criteria.

Students do not know what is expected of them.

Students do not know what counts as a good or bad assignment/project.

Assessment viewed by some departments/schools as an extra rather than a recognized used of staff time.

Project supervision seen as an extra or the real time involved is not recognized.

p. 9, Brown, Bull and Pendlebury

The list of common weaknesses in assessment systems

Overload of students and staff.

Too many assignments with the same deadline set in the department/school.

Insufficient time for students to complete the assignments in the time available.

Insufficient time for staff to mark the assignments before the next semester.

Inadequate or superficial feedback provided to students.

Wide variations in assessment demands of different modules.

Wide variations in marking across modules.

Wide variations in marking within a module.

Wide variations in marking by demonstrators.

Fuzzy or non-existent criteria.

Undue precision and specificity of marking schemes or criteria.

Students do not know what is expected of them.

Students do not know what counts as a good or bad assignment/project.

Assessment viewed by some departments/schools as an extra rather than a recognized used of staff time.

Project supervision seen as an extra or the real time involved is not recognized.

p. 9, Brown, Bull and Pendlebury

Purposes of Assessment

Derivation of the term "assessment" is "*ad sedere*"

(i.e., 'to set down beside')

-implication of the word assessment's etymology is primary that of providing guidance and feedback to the learner

Why do we assess? What are the purposes?

I'll type in your ideas

Why do we assess? What are the purposes?

Why do we assess? What are the purposes?

Purposes of assessment

To provide feedback to students to improve their learning.

To motivate students.

To diagnose a student's strengths and weaknesses.

To help students to develop their skills of self-assessment.

To provide a profile of what a student has learnt.

To pass or fail a student.

To grade or rank a student.

To licence to proceed.

To select for future courses.

To licence to practice.

To predict success in future courses.

To predict success in employment.

To select for future employment.

To provide feedback to lecturers.

To improve teaching.

To evaluate a course's strengths and weaknesses.

To make the course appear 'respectable' and creditworthy to other institutions and employers.

p. 11, Brown, Bull and Pendlebury

The terms for the words assessment, appraisal, audit and evaluation do vary in meaning

- assessment is operationally defined here as any procedure used to estimate student learning for any purpose

Assessment

Types of Assessment

- Criterion–referenced
- Norm–referenced

Function of Assessment

- Summative
- Formative
- Diagnostic

Homo Australopithecus

The first assessment of learning

He commanded his son “Go out and kill your first bear”

Task: What would he say if he wanted to compare the hunting skills of all the off-springs in the tribe?

Hint: is the first command criterion referenced or norm reference?

Homo Australopithecus

The first assessment of learning

He commanded his son “Go out and kill your first bear”

- this is a criterion referenced learning tasks; if the son kills a bear, he achieves the learning required

He might have said “Go out and kill as many bears as you can” -- if he wanted to grade all the tribe’s off-springs

- this would be norm referenced; ‘do as well as possible in comparison to your peers’

Trends is Assessment

Examination in Ancient China

See Handout

Trends in assessment

From	Towards
Written examinations	Coursework
Tutor-led assessment	Student-led assessment
Implicit criteria	Explicit criteria
Competition	Collaboration
Product assessment	Process assessment
Objectives	Outcomes
Content	Competencies
Course assessment	Modular assessment
Advanced levels	Assessed prior learning

p. 13, Brown, Bull, Pendlebury


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Student Learning

Before one can assess a phenomenon,

need to know what it is

need to know what are its components

need to know how it is acquired

Learning Defined

(What it is) **(Its Components)**

Change in knowledge, understanding, skills & attitudes

brought by experience & reflection on that experience

(how it is acquired)

How Do Students Learn

reading

thinking

writing

listening

note-taking

observing

talking

doing

Key Processes in Learning

Forms of thinking

- searching for understanding
- problem solving
- creativity
- evaluating
- remembering (rote or identifying patterns or recalling through understanding)

Key Processes in Learning (cont'd)

Forms of reflection

- using feedback
- studying
- assessing

Forms of motivation

- ownership
- contextual relevance
- purposeful

Two Principle Types of Learners

Knowledge-seekers: reproductive style of learning

- **Search for facts, information**
- **Mechanical, surface learners**
- **Idea of deep thinking is viewed as “comprehensive information”**
- **Serialists building their overall view gradually and in parts**

Two Principle Types of Learners (cont'd)

Understanding-seekers: creative style of learning

- prefer global picture before exploring details (holistic)
- intrinsically motivated (less likely to be responders to a system)
- deep learning, creative, independent

Two Principle Types of Learners (cont'd)

Few students are equally comfortable with both styles of learning

Knowledge-seekers prefer highly structured forms of teaching and assessment (e.g., lectures and written examinations that are excessively MCQs)

Understanding-seekers prefer flexible approaches to teaching (including having a choice of what is studied) and a variety of forms of assessment (projects, open-ended assignments)

Types of Learners and Assessment

Do the examination questions require personal understanding or a reproduction of the lecturer's understanding?

Structure of Observed Learning Outcome (SOLO)

Biggs, 1996: SOLO (Structure of Observed Learning Outcome)

- **Five levels of learning tasks matched to five levels for assessment**

Prestructural

- **task is not attacked appropriately (student didn't understand the point)**

  [barely satisfactory]

Unstructured

- **a few learning tasks achieved/used but treated separately**

Structure of Observed Learning Outcome (SOLO) (cont'd)

Multi-structural

- several learning tasks achieved/used but treated separately

Relational

- learning tasks are integrated into a coherent whole with each component contributing to overall meaning

Extended Abstract

- the integrated whole set of learning tasks are reconceptualized enabling generalization to a new topic or area (meta-cognition)

[outstanding performance]

Structure of Observed Learning Outcome (SOLO) (cont'd)

In parallel Entwistle and Marton (1994) describe these five levels as

1. Absorbing facts, details and procedures related to exams without consideration of structure

   [barely satisfactory]

2. Accepting and using only the knowledge and logical structures provided in the lecture notes

3. Relying mainly on notes to develop summary structures solely for purposes to cope with answering examinations

Structure of Observed Learning Outcome (SOLO) (cont'd)

4. Developing structures from strategic reading to represent personal understanding and also to answer examinations
5. Developing structures from wide reading that relate personal understanding to the nature of the discipline

[outstanding performance]

Promoting deep thinking in surface learners

How students learn depends on

- **method of studying, and**
- **method of assessment, and**
- **organizational climate of department/school (its use of various approaches to teaching and assessment)**

Teaching Conducive to Promoting

Surface Learning	Deep Thinking
Ineffective lecturing	Effective lecturing
Heavy workload	Help with specific difficulties
Inappropriate assessment	
Lack of freedom to learn	Perceived freedom to learn
Promotion of reproductive	Promotion of personal understanding
Learning in the assessment	

Caution

Note: knowledge-seekers are necessarily superficial learners, and not all understanding-seekers are deep thinkers

Do Not Dismiss Knowledge Seeking

Knowing what & how is often as important as knowing why

- **e.g., Neurologist needs to know the specific neurological function of the optic nerve before s/he can make a diagnosis**

Key Point

Encourage students to develop both a knowledge orientation and an understanding orientation

- knowledge base is necessary for understanding
- knowledge and understanding are diametrically opposed
  - reproductive learning merges into understanding

Surface Learning can be Encouraged by Type of Assessment Method

Assessment methods that

- Emphasize recall or application of trivial procedural knowledge
- Create anxiety
- Omit formative assessment and relevant feedback
- Inconsistently or inaccurately marked (rewards)
- Follow from an excessive amount of material in the curriculum
- Lack contextual relevancy

Deep Learning can be Encouraged by Assessment Method

Assessment methods that

- **Foster active long-term engagement in learning tasks**
- **Relate to clearly stated academic expectations**
- **Stress meaning and relevance to student**
- **Focus on contextual relevancy**
- **Reward those who have exercised a responsible choice in learning method and content**

When Does Effective Learning Most Likely to Occur?

Structure & learning tasks are built on earlier learning

Students develop repertoire of skills from

- **direct teaching**
- **explicit modeling**
- **practice**
- **feedback**
- **reflection on their own approaches**

When Does Effective Learning Most Likely to Occur? (cont'd)

Students take responsibility for their own learning

- **Have some freedom of choice**
- **Opportunities to explore alternative approaches**
- **Encourage and trained in self-assessment**
- **Workload is moderate**
- **Variety of assessments is used**
- **Context of learning conducive to active engagement in learning**
- **Students perceive effort leads to success and recognition**

Learning Behavioral-based Skills

Skill : an organized and coordinated pattern of mental and/or physical activity informed by mental

: patterned sequences of actions in response to cues rather than isolated instances of behaviour

: goal-directed sequences of actions that maybe learned and routinized

: once learned, skills have built-in feedback mechanisms which enable the skilled to adjust to tasks in hand

Skills are not just Motor Skills

Skills have perceptual as well as cognitive components (as well as motor components) in varying proportions

- The more complex the skill, the greater is its cognitive component

When scope of skill includes cognitive skill, then the skills become not just the basis but centre piece of work in higher education

- Skill in speaking (vocational)
- Skill in negotiating/discussion (educational)

Skills are not just Motor Skills (cont'd)

Essential cognitive skills for practitioners in Health Sciences

- **in handling information**
- **evaluating evidence**
- **thinking critically**
- **solving problems**
- **arguing rationally**
- **learning to learn (meta-cognition)**

Skills are not just Motor Skills (cont'd)

Some essential social skills for leadership roles in Health Sciences

- **working with others**
- **leading others**
- **communicating with others**

Skills are not just Motor Skills (cont'd)

Some essential affective skills for success in Health Sciences

- **taking initiative**
- **exhibiting independence**
- **taking risks**
- **will to change**
- **valuing others**
- **valuing clarification**

Skills

To assess skills, this is relevant background research

- **Skills are identical across subjects**
- **Skills do not readily transfer across contexts**
- **Transfer is most likely when the student understands the underlying principles**
- **Knowledge and understanding of different contexts are necessary for transfer**
- **Training in transfer is promoted by providing a variety of assessment tasks**

Skills

Example:

Clinical competence

- **Competence in taking Hx of patient with acute abdominal pain does not correlate with competence in taking Hx of patient with chest pain**

Implication

- **Develop then assess understanding of the underlying mechanisms for both conditions**
- **Assess by pencil and paper and OSCE**
- **Repeat with other conditions**

Assessment Methods and Strategies

See Handout: Methods of Assessment

Methods of Assessment

Essays

A standard method. Has potential for measuring understanding, synthesis and evaluative skills. Relatively easy to set. Marking for grading using impressionistic marking is relatively fast. Marking for feedback can be time consuming. Variations between examiners can be high.

Problems

A standard method. Has potential for measuring application, analysis and problem solving strategies. Complex problems and their marking schemes can be difficult to design. Marking for grading of easy problems is fast. For complex problems marking can be slow. Marking for feedback can be slow. Variations between markers is usually low.

Reports on practicals

A standard method. Has potential for measuring knowledge of experimental procedures, analysis and interpretation of results. Measures know-how of practical skills but not the skills themselves. Marking for grading using impressions or simple structured forms is relatively fast. Marking for feedback with simple structured forms is faster than without them. Variations between markers, without structured forms, can be high. Method is often over-used. Different foci for different experiments recommended.

Methods of Assessment (cont'd)

Multiple choice questions (MCQs)

A standard method. Can sample a wide range of knowledge quickly. Has potential for measuring understanding, analysis, problem-solving skills and evaluative skills. Measures of knowledge are relatively easy to set. More complex questions require more time to set. Easy to mark and analyse results. Feedback to students is fast. Reliable. Danger of testing only trivial knowledge.

Short Answer questions

A standard method. Has potential for measuring analysis, application of knowledge, problem-solving and evaluative skills. Can be easier to design than complex MCQs but still relatively slow. Marking to model answers is relatively fast compared with problems but not compared with MCQs. Marking for feedback can be relatively fast.

Cases and open problems

Has potential for measuring application of knowledge, analysis, problem-solving and evaluative skills. Short cases are relatively easy to design and mark. Design of more complex cases and their marking schemes is more challenging. Marking for grading and feedback are about as fast as essay marking.

Mini-practicals

A series of mini-practicals undertaken under timed conditions. Potential for sampling wide range of practical, analytical and interpretative skills. Initial design is time-consuming. Marking is done on the spot so it is fast. Feedback to students is fast. Reliable but training of assessors is necessary.

Methods of Assessment (cont'd)

Projects, group projects and dissertations

Good all-round ability testing. Potential for sampling wide range of practical, analytical and interpretative skills. Wider application of knowledge, understanding and skills to real/simulated situations. Provides a measure of project and time management. Group projects can provide a measure of teamwork skills and leadership. Motivation and teamwork can be high. Marking for grading can be time consuming. Marking for feedback can be reduced through peer- and self-assessment and presentations. Feedback potential. Tests methods as well as end results. Can be variations between markers. Use of criteria reduces variability.

Orals

Test communication, understanding, capacity to think quickly under pressure and knowledge of procedures. Feedback potential. Marking for grading can be fast but some standardisation of interview procedure is needed.

Presentations

Test preparation, understanding, knowledge, capacity to structure information oral communication skills. Feedback potential: from tutor, self and peers. Marking for grading based on single criteria is fast and potentially reliable.

Poster sessions

Test capacity to present findings and interpretations succinctly and attractively. Danger of focusing unduly on presentation methods can be avoided by the use of simple criteria. Feedback potential: from tutor, self and peers. Marking for grading is fast. Use of criteria reduces variability.

Methods of Assessment (cont'd)

Reflective practice assignment

Measure capacity to analyse and evaluate experience in the light of theories and research evidence. Relatively easy to set. Feedback potential from peers, self and tutors. Marking for feedback can be slow. Marking for grading is about the same as for essays. Use of criteria reduces variability.

Single essay exams

Three hours on prepared topic. Relatively easy to set. Attention to criteria needed. Wider range of ability tested including capacity to draw on a wide range of knowledge, to synthesise and identify recurrent themes. Marking for feedback is relatively slow. Marking for grading is relatively fast providing the criteria are simple.

p. 46-7, Brown, Bull, Pendlebury

Terminology (Jargon)

Sources

- **The student, the lecturer, peer group, the employer**

Instruments

- **Marking scheme, the explicit or the implicit criteria**

Methods of assessment

- **Approach used to assess learning (MCQ, essay, etc)**

Differences Between Marking Schemes and Criteria

Criteria

- Assessed dimensions – e.g., organization, style, use of evidence

Schemes

- Linear and attached to specific operations or procedures, suitable for sequential tasks and simple design

Explicit Criteria: 4 functions

- Provide meaningful feedback to students

- Demonstrate links between assessment task and learning outcomes

- Ensure assessors are in broad agreement

- Pinpoint areas of disagreement between markers

Explicit Criteria (cont'd)

Caution

- **Undue specificity, detailed in criteria shown to yield low-level learning (Slavin, 1990)**
  - **useful for research, less so in learning**
  - **time consuming**

Designing and Marking Assessment (Course assignment or written paper)

Questions to ask in designing the assessment

1. What are the outcomes to be assessed?
2. What are the capabilities/skills in the outcomes?
3. Is the method of assessment constant with the outcome and skills?
4. Does the specific assessment task match the outcome and skills?

Designing and Marking Assessment (Course assignment or written paper) (cont'd)

5. Is the method relatively efficient in terms of student time?
6. Is the method relatively efficient in terms of assessor's time?
7. Are the marking schemes or criteria appropriate?
8. What alternatives are there and what are their relative advantages and disadvantages?

Items for Designing Questions

- Brainstorm with colleagues to generate a series of questions on a topic
- Keep a notebook of possible questions
- Use new types of questions in course rather than examinations
- Be sure working is clear
- Eliminate unwanted ambiguities
- Remember more open questions make it harder to devise a marking scheme

Ideas for Marking Questions

Marking for grading and feedback are two different purposes

Marking for Grading

- Faster than if marking for feedback
- Potentially less reliable (deadlines)
- Mark by question not entire set of questions (faster, more reliable)
- Mark intensively without break until marking scheme is fixed
- Return marks promptly

Marking for Feedback

- Return promptly
- Use few succinct, supportive statements
- Long, warm, woolly comments ineffective
- Short, sharp, critical comments ineffective
- Choose a few points that will produce greatest improvement (hold back on rest)
- State what is good and what could be improved
- Simply notify errors is not helpful – student needs to know how to correct them
- Balance of encouragement and some criticisms better then only litany of errors

If Marking Criterion – Reference Assessment

Skim read all answers, sort into piles of

Fail, Pass, Good, Excellent

Re-read in detail and adjust sort

Formative Feedback marking: Additional Tips

- Course should allow familiarization with assessment method
  - allow time for students to tackle assignments meaningfully
- evidence: very tough (grade deprivation) or very easy (grade in _) marking is ineffective
  - prompt return of assignments through course more effective
- don’t criticize if students reluctant to put forth best effort if work doesn’t count
- Assess only important skills, knowledge and understanding

Coping Strategies in Marking if Overloaded

- Set fewer but more challenging assignments
- Space assessments out more evenly
- Devolve marking or feedback to tutors and TAs
- Make scripts easier to mark (require word processing)
- Provide feedback to students most in need
- Feedback to groups not individuals

Use with caution


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Purposes and Goals of Essays

- Obtain an estimate of higher-order cognitive skills
- Encourage integrated thinking
- Develop communication skills in written form

Assessing Essays

But essays can be perceived by students as requiring

- **An argument, well supported by evidence**
- **A distinctive viewpoint on a problem/issue**
- **An ordered arrangement of facts and ideas**

Assessing Learning (cont'd)

These three are consonant with styles of learning

1. **Akin to deep active approach in search for and constructing meaning**
2. **Has elements of deep processing but largely concerned with reproducing knowledge**
3. **Entirely concerned with reproducing other's facts and ideas**

Quality of Essays

Quality of essays determined by

- **students' knowledge, perceptions, skills and understanding of topic**

- **the question/task required by the question posed**

Pros and Cons of Essays

Easy to set (deceptively so however)

Hard to mark (disturbingly difficult to mark objectively)

- **Must have at least an idea of what constitutes a good, indifferent and poor answer**
- **Must know one's own values and able to distinguish between views which are only different from one's own and those which are different and wrong**

Pros and Cons (cont'd)

Much argument if one should specify precisely what is required in a question

- **Those who prefer essays to pose questions which are obvious and transparent want specificity**
- **Those who want to develop analytical and problem solving skills for use in a wide variety of contexts want the students to de-code the question**

Pros and Cons (cont'd)

Question

Specificity

Design an experiment to test the hypothesis that short men are quick-tempered

De-code

"Write an essay on sex and epidemiologist"

Reducing Variability in Marking

See hand-out

Suggestions on marking essays

Know your values and prejudices about style, punctuation, grammar, handwriting and particular ideologies or perspectives.

Know when you are likely to feel tired or irritable.

Plan and pace your marking.

Re-mark the first few essays that you did to check your consistency.

Don’t be afraid to shift categories after marking.

If in doubt, give the furthest mark from the mid-point, but

Ask yourself, ‘Is this really a fail?’

Have a set of criteria. Keep them simple.

Consult them regularly.

Give the students the criteria.

Get a colleague to second mark a sample of your essays.

If there is a serious disagreement of more than one grade, get an independent marker to mark using the same criteria.

p. 66, Brown, Bull, Pendlebury

Marking

Marks	Notes on the criteria
	Interpretation and introduction
5	States clearly the key issues and provides a framework for answering the question
4	Weaker version of 5
3	Introduction is perfunctory or overlong. Framework is present but unclear.
2	Weaker version of 3
1	Little or no attempt at an introduction or description of the framework.
Quality of argument	
5	Develops logical argument and marshals clearly the relevant ideas and evidence and their strengths and limitations.
4	Weaker version of 5
3	Some aspects of the argument and use of evidence are weak.
2	Weaker version of 3.
1	Major weaknesses in argument and use of evidence or no argument is discernible.
Use of evidence	
5	Uses and evaluates evidence. Shows connections between evidence and the framework of the question.
4	Weaker version of 5.
3	Some inclusion of irrelevant or unlinked material.
2	Weaker version of 3.
1	Most of the material is irrelevant or repetitive.

Marking (cont'd)

Presentation

5 Correct and effective use of English. Referencing accurate. Legible.

4 Weaker version of 5

3 Some minor errors in grammar, syntax and referencing. Fairly legible.

2 Weaker version of 3.

1 Several errors. Illegible.

Conclusions

1 Draws together the strands of the argument. Creates a coherent perspective on the question and indicates tentative links to related deep issues.

4 Weaker version of 5.

3 Conclusions are clear and based on the argument and evidence presented.

2 Weaker version of 3.

1 Little or no conclusion or not based on the argument and evidence in the essay

Comments

p. 68, Brown, Bull, Pendlebury

Helping Students to Prepare Quality Essays

See Hand-out

Student Checklist

Covering information

My title/cover page shows clearly:

- **My Name**

Course title and number

Tutor's name

The question I have chosen

Date assignment handed in

Introduction

The introduction:

- **Sets the question topic against a wider background**
- **Clarifies my understanding of the question/topic**
- **Defines key or problematic terms**
- **Outlines the approach I will be taking to the question/type**

Main text

In the main body of the assignment:

- **My key points are clearly presented**

The points I make are systematically backed up by facts/evidence/examples/ arguments

Quotations and references to other works are accurately cited

Any diagrams, figures or tables are labeled properly

Conclusions

The conclusion:

- **Brings together the main points**
- **Links back to the question/topic**
- **States clearly my conclusion(s)**

Style and presentation

Overall, the assignment:

- **Reads clearly throughout**
- **Makes correct use of spelling, grammar and punctuation**
- **Accurately lists the background reading I have consulted**

Is within the word limits specified

From Hounsell (1995), p. 71, Brown, Bull, Pendlebury

Essays for Summative Purposes

Marking essays for summative purposes: fastest, most reliable

Fastest, most reliable method is "structured impressiontic" marking

a. **Establish a set of brief, precise criteria for each grade to be assigned**

b. **Read each essay quickly for overall impression rather than detail**

c. **Each essay is marked twice, independently**

d. **Marking is timetabled and undertaken within a limited period of time**

Modified Essays

Consists of a set of linked short answers, questions given in small sets, answers to first set can and be changed after subsequent information and next set of questions given

Highly suited to assessing competency in patient management

Easily match to PBh strategies used in tutorials

Issues of Reliability and Validity

Depends on

- **question posed**
- **whether model answer is agreed to by markers**

Evidence

- **students get better marks in MEQs as they progress**
- **experienced doctors are better in MEQs than trainees**


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Assessing Practical Work

Introduction

- Time honoured approach in health sciences
- Most expensive, most time-consuming part of any course
- Practical work and writing up report occupies significant student time, but always reflected in course marks

Assessing Practical Work (cont'd)

- Main principle underlying laboratory and practical work:
  - students learn more effectively when actively engaged on practical tasks
- Main limitations:
  1. task has to be perceived as meaningful and relevant by students
     - otherwise involvement is limited
  2. students need constructive guidance and feedback
     - otherwise learning is minimal

Goals of Practical Work

- Improve technical skills relevant to subject
- Improve understanding of methods of inquiry
- Reinforce theory with practice
- Develop problem-solving skills
- Nurture professional attitudes

If Goal is to Develop Scientific Enquiry

1. Design activities so they build on prior learning of basic concepts and lab skills
2. Explicit teaching of (i) what scientists do and (ii) nature of scientific enquiry
3. Must design special learning activities – laboratory cookbooks not effect

   Hegarty (1982)

Students' Experiences with Labs (Bliss, 1990)

- All good labs had inquiry-based projects
- All bad labs were cookbook approaches

Auditing Lab Work

Starting point : cross reference objectives, content and assessment of lab classes

: see following check list (also handout)

Objectives, content and assessment

Objectives of labs	Content & assessment for each lab A B C D
1.To instill confidence 2.To learn basic practical skills 3.To familiarize with important standard apparatus and measuring techniques 4.To illustrate material taught in lectures 5.To train in observation 6.To train in making deductions from measurements and interpretations of experimental data 7.To use experimental data to solve specific problems 8.To learn some theoretical material not taught in lectures 9.To foster a critical awareness (e.g., extraction of all information from the data; the avoidance of systematic errors) 10.To help bridge the gap between theory and practical 11.To improve report writing 12.To stimulate and maintain interest in scientific methods	

Assessing Laboratory Reports & Notebooks

Marking of reports/notebooks is often unreliable

To increase reliability

- **use explicit criteria**
- **double mark samples of work**
- **train demonstrators as markers**
- **use standard lab marking sheet (give to students before lab, and a marked one when returning their lab reports)**

Structured Mini-Practicals

1. Simple, effective, quick, wide sampling
2. OSPEs – next week
3. OSCEs – next week
4. Performance Evaluation Guides (PEG)
   - developed by Brown and Pendlebury (1996) for Dentistry
   - consists of a set of explicit criteria for each of 4 levels of competence (see handouts)

Designing a PEG

Define the task	What should be achieved? How does it relate to the objectives of the course? What skills are required?
Analyse the procedures	Which aspects of the procedure are needed to achieve the most important desired outcome?
Define four levels for the task for the target group of students (beginners/final year students)	Unsatisfactory level. Minimum level of competence. Good level of competence. High level of competence.

p. 106, Brown, Bull, Pendlebury

Example of a PEG in Dentistry

	Descriptions	Aesthetics	Structural and biological integrity
Grade D	Unacceptable outcomes as a result of treatment or lack of treatment which has already caused irreversible damage to the patient's oral environment, or will cause severe damage in the future.	Restoration looks obviously false. No characteristics of natural morphology. Width, length and position of restoration incorrect. Shade is many units wrong and does not blend with the surrounding dentition.	Restoration made on a tooth with undiagnosed non vitality. Restoration made on a tooth with inadequate root filling which would be easy to correct. Restored tooth has active apical pathology.

Example of a PEG in Dentistry (cont'd)

Grade C	Outcomes as a result of treatment or lack of treatment which have the potential for damage, or where reversible damage has already occurred.	Shade is incorrect by one unit. The core shows through body and neck zones. Black/green restoration margins.	Restoration made on a vital tooth with inadequate pulpal protection if the cavity is deep. Restoration made on a tooth with an inadequate root filling that would be difficult to correct. Restored tooth has non-active apical pathology which has not been treated.

Example of a PEG in Dentistry (cont'd)

Grade B	Outcomes which achieve the minimum acceptable standard below which there is a potential for damage to the patient.	Restoration can be seen by the professional observer but not by the patient or his or her friends. The patient is satisfied with the result.	Restoration made on a tooth with healthy pulp. Restoration made on tooth with healing apical area.

Example of a PEG in Dentistry (cont'd)

Grade A	Outcomes achieved by a standard of excellence where there are no clinical limitations imposed by affordability, time, or the patient's wishes.	The presence of the restoration cannot be detected by any observer 3 feet from the patient. The patient is extremely pleased with the result.	Restoration made on a tooth with healthy pulp. Restoration made on a tooth with technically proven root filling. Restoration is stable at try-in without cement.

p107, Brown, Bull, Pendlebury

One Way to Reduce Assessment Time

Instead of requiring a report for every practical, ask students to do:

a team presentation at a seminar

a poster session

Alternatives to Laboratory Work

See handout

Alternative approaches to laboratory work

Paper and pencil activities which require a student to solve an experimental problem or create and experimental design.

Provide the experimental data such as the output of a mass spectrometer, and ask groups of students to interpret it.

Present a video-recording of an experiment which shows readings on various instruments. Ask the students to note, calculate and interpret the results.

Set up a spreadsheet so that students can enter the results and obtain immediate graphic displays or calculations of data.

Set up the apparatus for the students so all they have to do is take readings.

Side-step part of the experiment. Ask students to perform the parts of an experiment which can be done quickly and provide pre-prepared materials for the slow part of the experiment (as is done in some Open University courses).

Ask the students to carry out a set of brief, simple experiments which exemplify some fundamental principles.

Design an MCQ to test understanding and knowledge of specific laboratory procedures. Go through the results with students so they are aware of their strengths and weaknesses.

Note: Each of these require detailed planning.

p. 112, Brown, Bull, Pendlebury

Alternatives to Laboratory Work (cont'd)

The alternatives:

Moore, I. and Exley, K. (eds) (1994) *Alternative Approaches to Teaching Engineering* Sheffield: CVCP UCoSDA.

Exley, K. and Dennick, R. (eds) (1996) *Innovations in Teaching in Medical Sciences* Birmingham: SEDA.

Tips for developing deep learning from Lab work

Brewer, I.M. (1985) *Learning More and Teaching Less* London: SRHE.

Providing Guidance and Feedback on Practical Work

To instill confidence in students one must

- **Provide guidance, feedback, reinforcement**
- **To be effective you or the tutor/lab instructor:**
  - be able to do the experiment, analyze the findings and interpret them
  - provide clear instructions
  - develop a set of hints, prompts, cues, questions to guide students without telling them what to do
  - provide encouragement
  - mark lab reports
    - fairly, consistently

Training the Demonstrator

Particularly on assessment and in giving feedback is important (see handout)

Demonstrator skills

Observe students at work.

Anticipate major difficulties of understanding.

Recognise major difficulties of understanding.

Give brief, clear explanations of processes and procedures.

Give directions.

Ask questions that clarify difficulties of understanding.

Ask questions that guide students.

Answer questions in a simple, direct and non-critical way.

Offer supportive and encouraging remarks.

Know when to help and not help a student.

Know how to mark fairly and consistently.

p. 115, Brown, Bull, Pendlebury

Assessing Projects

Stated advantages of projects

- **enable students to explore deeply a topic**
- **develop student initiative and resourcefulness**
- **enhance time and project-management skills**
- **provide ownership of the learning**
- **foster independence and creative-learning**

Disadvantages of Projects

- time-consuming to set up, monitor and provide feedback
- difficult to assign fairly
- variation in help sought by students
- opportunity for plagiarism

None of these are insurmountable, assessment no more than essay/written papers

Guides for Project Work: Science and Medicine

Exley, K. and Moore, I. (eds) (1992) *Innovations in Science Teaching* Birmingham: SEDA.

Moore, I. and Exley, K. (eds) (1994) *Alternative Approaches to Teaching Engineering* Sheffield: CVCP UCoSDA.

Exley, K. and Dennick, R. (eds) (1996) *Innovations in Teaching in Medical Sciences* Birmingham: SEDA.

Guides for Project Work: Helping Professions

Oriented for students

Herbert, M. (1990) *Planning a Research Project: A Guide for Practitioners and Trainees* London: Cassells.

The Research Evidence

Not a lot on assessment of projects

What is given here are more survey results than experimental findings

The Research Evidence (cont'd)

Wright, J. and Lodwick, R. (1989) 'The process of the PhD: a study of the first years of doctoral study' *Research Papers in Education* 4: 22-56.

Students value the following supervisory skills:

- provides feedback
- checks on progress
- provides support and encouragement
- provides academic guidance
- allows students to work independently some of the time

The Research Evidence (cont'd)

Wright, J. (1991) 'Studies of research supervision' Unpublished PhD thesis, University of Nottingham.

Supervisors who were successful as measured by their students successful completion rates

- had regular, frequent scheduled meetings with their students
- set tasks
- reviewed progress regularly
- commented on drafts
- taught their students time-management skills

The Research Evidence (cont'd)

Brown, G., Neerinck, D and Lapiere, D. (1991) *Supervision of Science Research: An International Study* Brussels: Ministry of Education and Science.

Survey research students in 4 countries wanted

- more help in planning, analyzing and interpreting results
- drafting theoretical models
- writing up and publications

The Research Evidence (cont'd)

Gabb, R. (1981) 'Playing the project game' *Assessment in Higher Education* 6:26-48.

Students who tried to get good marks on projects assumed these strategies worked well

- Pick the right supervisor
  - helpful, approachable, friendly
  - choice of topic was less important
- Find out about the assessment procedure
  - even if it is secret, friendly staff will reveal it

Gabb (1981) cont'd

Results

- assessors are interested in results more than any other aspect of projects (results are easy to mark so tell your supervisor any results obtained)

Length

- make report 30 pages, pad if necessary, but never exceed the limit

References

- make a long list (just look at the journal abstracts in order to get enough)

What We Do Know: Set Objectives for the Project

Examples: see handout

Examples of objectives

By the end of the module, it is expected that students will be better able to:

Select and use appropriate qualitative and quantitative methods of research in clinical education.

Read and evaluate research papers in clinical education.

Design investigations in clinical education.

Analyse data collected in investigations in clinical education.

Interpret data collected in investigations in clinical education.

Write a focused evaluative review of research literature in clinical education.

Present a well-constructed dissertation in clinical education.

p. 123, Brown, Bull, Pendlebury

Use a Structured Approach to Projects

These assist students to produce good projects (and theses/dissertations)

Split project into a set of assessed tasks:

- develop a design plan
- review of literature
- report of data collection
- report of data analysis

Mark each section; provide feedback

Assessing the Report

Major methods of assessing projects

- **Assessing artifacts**
- **(products – computer programme, a design, a working model, a port folio, an active learning programme)**
- **Assessing the report**
- **Vivas**
- **Poster sessions**
- **Presentations**
- **Log Books**
- **Written papers**

Assessing Project: An Experiment

See handout

Assessing an experimental project

Guidelines	Max Mark	Your mark	Comment
Introduction			
Sets out clearly the content and structure of the project.			
Identifies clearly the problem to be examined.			
Review			
Well structured. Relates closely to the problem/hypotheses of the dissertation. Is critical and probing of the literature.			
Shows an awareness of different types of evidence.			
Design			
Describes clearly so that the experiment could be replicated by a fellow student.			
Results			
Methods of data collected and analysis indicated clearly.			
Results			
Presented clearly. Some brief discussion of the results.			

Assessing an experimental Project (cont'd)

Results summarized

Interpretation and discussion. Results put into context of the literature. The limitations of the experimental study and data analysis indicated.

Conclusions drawn and related to hypotheses. Possible next steps of research and other implications indicated.

Layout, referencing, bibliography

Clear and accurate.

Global view of project and project management skills.

Response to comments and advice offered.

Supervisor's mark

Second marker's mark

External examiner's mark

Mark awarded

Note: A small percentage, say 5% could be reserved for overall impressions of the dissertation since the overall quality of a project may be greater than the sum of its parts. A modified checklist may be necessary for a project based on a series of linked experiments.

p. 128, Brown, Bull, Pendlebury


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Assessing Oral Communications

Students do better in oral presentations (such as seminars, discussions, vivas, consultations) if they are provided:

- **learning opportunities**
  - guidance
  - feedback

Assessing Oral Communications (cont'd)

Evidence shows key factors are:

- **providing clarity**
- **motivating audience**
- **maintaining interest**

Assessing Oral Communications (cont'd)

Most effective methods in improving oral communications are based on video feedback

- but feedback is not enough
- need also guidance on the analysis of presentations
- opportunity to use these criteria in order to develop self-assessment skills
  - e.g., micro-teaching (pre-post)
    - two day workshop

Guidelines for Clarity

Guidance for students pertaining to making things clear

See handout

Presentations and their assessment:

Making explanations clear

Know what you intend to explain.

Produce a structured summary before the talk.

Use the structuring tactics of:

Signposts – statements which indicate the content, e.g.: 'I want to deal briefly with lactation. First I want to outline the composition of mild; second, its synthesis; third, the normal lactation curves.'

'Most of you have heard the old wives' tale that eating carrots helps you to see in the dark. Is it true? Let's have a look at the basic biochemical processes involved.'

Frames – statements which indicate the beginning and ending of sub-topics, e.g.: 'So that ends my discussion of adrenaline. Let's now look at the role of glycogen.'

Presentations and their assessment: (cont'd)

Foci – statements (and gestures) which highlight the important points of the talk, e.g.: 'So the main point is…'

'Now this is very important…'

'But be careful. This interaction with penicillin occurs only whilst the cell walls are growing.'

Links* - statements which link the different parts of the talk together and the talk to the listeners' experience and knowledge, e.g.: 'So you can see that reduction in blood sugar levels is detected indirectly in the adrenaline glands and directly in the pancreas. This leads to the release of two different hormones…'

Use aids if appropriate.

Summarise during the presentation and at the end.

*Links are often conspicuous in their absence. Often an explainer misses out a chunk of an explanation because he/she assumes the listeners are familiar with the missing section. The more knowledgeable one is about a topic the more likely one is to commit this error. One way of avoiding it is to look at the topic from the standpoint of the explainees. This will help you to pitch the explanation at the appropriate level.

p. 155, Brown, Bull, Pendlebury

Guidelines for Motivating Audience

Tell audience why the talk will be of use to them

Tell an anecdote or story (humor adds to it)

Provide an overview, i.e., signposts (road map)

Guidelines for Maintaining Interest

Making the Explanations Interesting

See handout

Making explanations interesting

Show your own interest in and commitment to the topic.

Think of and use examples, analogies, metaphors and models that are apt for the audience and the topic.

If the material is unfamiliar, begin with several examples.

Use a mixture of modes of explaining but particularly the narrative mode (informal, personal and story-telling).

Play on the intellectual curiosity of the audience through the use of puzzles, problems and questions.

p. 156, Brown, Bull, Pendlebury

Assessment of Oral

Criteria can be grouped into the three themes of Aristole

Logos - persuasive use of argument

Pathos - emotional engagement with the audience

Ethos - perceived personality of the presenter and the position adapted to the topic and audience

Assessment of Oral

Cockroft, R. and Cockroft, S. (1992) *Persuading People: An Introduction to Rhetoric* London: Macmillan.

The authors operationalize these three domains

Logos (persuasive) – systematic, business like or unsystematic, slipshod, disorganized

Pathos (engaging) – enthusiastic and interesting or boring and dull

Ethos (personality) – warm, friendly, humble, helpful or cold, hostile, opinionated, avoids questions or answers aggressively

Assessment of Discussions

Assessment of Discussions (or the Discussion Stage in a Presentation

See handouts of example rating scales

Explanations evaluated

Opening	
Does your opening gain the group's attention?	yes/no
Does it establish rapport with the group?	yes/no
Does it indicate what you intend to explain?	yes/no
The Key Points Are your key points clearly expressed?	yes/no
Are your examples apt and interesting?	yes/no
Are your qualifications of the key points clearly stated?	yes/no
Is each key point summarized?	yes/no
Are the summaries clear?	yes/no
Are the beginnings and ends of the key points clearly indicated?	yes/no
The Summary	
Does your summary bring together the main points?	yes/no
Are your conclusions clearly stated?	yes/no
Do you come to an effective stop?	yes/no
Presentation	
Can the group hear and see you?	yes/no
Do you use eye contact to involve but not to threaten?	yes/no
Do you use audio-visual techniques effectively?	yes/no
Are you fluent verbally?	yes/no
Is your vocabulary appropriate to the group?	yes/no
Do you make use of pauses and silences?	yes/no
Do you vary your intonation?	yes/no
Is the organization of your material clear?	yes/no
Do you avoid vagueness and ambiguities?	yes/no
Is the presentation as interesting as you can make it?	yes/no

Any 'no' answer indicates that your explanations are not as explanatory as they might be. (But everybody gives quite a lot of 'no' answers if he or she is honest.)

This checklist may be adapted for self, peer or tutor feedback.

p. 158, Brown, Bull, Pendlebury

Seminar presentation and discussion

Twenty per cent of marks are awarded for your presentation at the seminar. The criteria are as follows:

Criteria	Marks	Comments
Content Opening, clarity of argument or explanation, summary and conclusion.	0 1 2 3 4 5	
Evidence Experimental design, literature review, interpretation.	0 1 2 3 4 5	
Presentation Fluency, use of audio-visual aids, handouts, body language.	0 1 2 3 4 5	
Discussion skills Listening, responding to questions, engaging others in the discussion, managing the group and individuals.	0 1 2 3 4 5	
Total mark	_____	

What's good?

What needs improving?

Tutor/student……

Date……..

Notes:

The form may be used for peer, self or tutor feedback or assessment.

Whatever method you use, do give the students the criteria before they prepare the presentation.

Some tutors remind their students that the tutor has the right of arbitration.

The form ma be adapted for use for individual or group presentations.

p. 159, Brown, Bull, Pendlebury

Assessing Discussions in Small Group Teaching

Criteria which were often failed (Brown, 1994)

- **Goals were unclear to students**
- **Tutors spoke too much**
- **Questions rarely rose above the level of recall**
- **Discussion was not focused**
- **Discussion was descriptive rather than analytical**

Assessing Students Contributions in Discussion

Video tape the discussion

Break group into small groups (n=3)

Ask each member to rate themselves in terms of these criteria

Assessing Students Contributions in Discussion (cont'd)

- Was the contribution from a class member in direct response to a question from the tutor, or was it an unsolicited comment or question?
- Did the contribution help the discussion (kept on track) or diverted the discussion?
- Did the contribution link to what anyone else had said?
- Did the contribution change the level of discussion?
  - was the level improved or reduced?

Quality of Assessing Discussion

What level of thought processing characterized the discussion?

See handout

Levels of discussion

Classify the predominant type of discussion in each one- (or two-) minute segment.

Type	Description
5 Explanatory	Searching for connections, causes, reasons, underlying assumptions and perspectives.
4 Interpretative	Exploring meanings, providing different interpretations of the evidence, formalizing definitions of terms.
3 Procedural	Outlining what was done in a task or situation rather than why. Listing what an author said rather than what the author was driving at.
2 Descriptive	Relating and exchanging experiences or opinions and assertions without supporting evidence.
1 Social	Friendly conversations of a personal kind that are marginal to the task.
0 Social	Unfriendly conversations of a personal kind that are marginal to the task.

p. 164, Brown, Bull, Pendlebury

Assessing Vivas

Vivas often used to decide

- borderline cases in medicine
- degree classification
- assessment of project/thesis

Always establish rapport and make candidate feel comfortable

- candidate should be able to do best possible, not a measure of coping with stress

Always make clear purpose of viva

Try to use a set of guidelines – see handout

Assessing structures and vocabularies

80+	Outstanding	As close to native use of structures, idioms and vocabulary for the purposes of discussion as can be detected. (Stating, describing, reasoning, persuading, reacting.)
70-79	Excellent	Confident and appropriate use of a wide range of structures, idioms and vocabulary for the purposes of discussion.
60-69	Very good	Some appropriate use of more complex structures and or good idioms. Good range of vocabulary for purposes of discussion.
50-59	Satisfactory	Some attempt at complex structures and idioms with some success. Suitable vocabulary for purposes of discussion.
45-49	Weak	Inconsistent. Some attempt at a wider range of structures. Beginning to use vocabulary for the purposes of discussion.
40-44		Barely adequate Simple sentence structure. Little vocabulary beyond the basics.
39-	Inadequate (fail)	Struggles to create sentences or use authentic words.

p. 165, Brown, Bull, Pendlebury

Assessing Consultations

Consultations: conversations with a purpose between a professional and a client

: characterized everyday life of most health professionals

Consultations: The Evidence

Brown, G. and Atkins, M. (1996) 'Explaining' in 0. Hargie (ed.) *Handbook of Communication Skills* London: Routledge.

1. The doctor is perceived to be competent trustworthy and friendly
2. The doctor has a clearly structured consultation
3. The doctor allows the patient to tell his own story

Consultations: The Evidence (cont'd)

These characteristics are good predictors of

- **patient recall**
- **patient understanding**
- **patient satisfaction**
- **patient's willingness to follow treatment**
- **improved health status of patients**

Common Errors Made in Consultations

See handout

Common errors made by young doctors in consultations

Use of jargon.

Lack of precision.

Avoidance of personal issues.

Failure to pick up verbal leads.

Undue repetition of questions.

Inappropriate questions.

Lack of clarification.

Lack of control.

Non-facilitation.

Assumption there is only one problem.

Time management of consultation.

p. 166, Brown, Bull, Pendlebury

Assessing the Consultation Structure

See handout

The well-ordered consultation

4 = Very good, 3 = Good, 2 = Satisfactory, 1 = Inadequate, 0 = Not present

Note: Very good does not mean perfect

OPENING

Welcoming into the surgery

Mutual introduction

4 3 2 1 0

RAPPORT

Forming a social link between patient and dentist, enabling fruitful communication

4 3 2 1 0

DIAGNOSIS

Questioning, listening, watching and leading the patient where necessary to disclose the patient's real problem or worry, medical and dental including the unstated ones (hidden agenda)

4 3 2 1 0

EXPLANATION

Discussing the problem and implications with the patient, using a suitable vocabulary that is understood by the patient

4 3 2 1 0

The well-ordered consultation (cont'd)

RESPONSIVENESS

The recognition and follow-up of verbal and non-verbal clues given by the patient when proposing, negotiating and carrying out treatment

4 3 2 1 0

EDUCATION

Inserting into the consultation a preventive and dental health message with encouragement of self care

4 3 2 1 0

DISMISSAL

A clear and mutually acceptable termination to the consultation containing a definite indication of continuing care

4 3 2 1 0

GLOBAL IMPRESSION

4 3 2 1 0 **Total Score /32**

Overall impression is more than the sum of the parts!

p. 167, Brown, Bull, Pendlebury

Assessing a Dissertation

See handout

Reading a project report or dissertation

Reflect: Fundamental questions

Is the standard of literary presentation adequate?

Is the dissertation (substantially) the candidate's own work?

Does the candidate have a general understanding of the relevant field and how this dissertation relates to it?

Has the candidate thought through the implications of the findings?

Is there evidence of originality?

Does the study add to existing knowledge of the subject?

Is there evidence that the candidate has developed skills in research at this level?

Is it worth publishing, albeit in modified form?

p. 131, Brown, Bull, Pendlebury

Reading a report or dissertation

Review of the literature

To what extent is the review relevant to the research study?

Has the candidate slipped into ‘here is all I know about x’?

Is there evidence of critical appraisal of other work, or is the review just descriptive?

How well has the candidate mastered the technical or theoretical literature?

Does the candidate make the links between the review and his or her design of the study explicit?

Is there a summary of the essential features of other work as it relates to this study?

Design of the study

What precautions were taken against likely sources of bias?

What are the limitations of the design? Is the candidate aware of them?

Is the methodology for data collection appropriate?

Are the techniques used for analysis appropriate?

In the circumstances, has the best design been chosen?

Has the candidate given an adequate justification for the design used?

Reading a report or dissertation (cont'd)

Presentation of results

Does the design/apparatus appear to have worked satisfactorily?

Have the hypotheses in fact been tested?

Do the solutions obtained relate to the questions posed?

Is the level and form of analysis appropriate for the data?

Could the presentation of the results have been made clearer?

Are patterns and trends in the results accurately identified and summarized?

Is a picture built up?

Discussion and conclusions

Is the candidate aware of possible limits to confidence/reliability/validity in the study?

Have the main points to emerge from the results been picked up for discussion?

Are links made to the literature?

Is there evidence of attempts at theory-building or re-conceptualisation of problems?

Are there speculations? Are they well grounded in the results?

p. 132, Brown, Bull, Pendlebury

Criteria for Assessing a Dissertation

See handout


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Assessing Poster Sessions & Exhibition

Efficient way to assess group work

Issues to consider (see handout)

Some issues when considering poster sessions

Who should assess (tutors, audience, peers and self)?

Is the poster the sole means of assessment?

- one sheet of A4 as a back-up?
- Oral explanation as a back-up?

Normal use is for assessing a group product.

Presentation content.

- be clear about criteria for presentation and content
- Assign some marks for design and visual presentation.

Efficiency and resources.

- limited space for presentation e.g., one flip-chart sheet
- reduced marking time
- suitable for peer assessment e.g., by ranking
- large display area needed (especially for peer assessment purposes)
- need to provide production materials.

p. 133, Brown, Bull, Pendlebury

Assessing Posters

Prepare an assessment sheet

See handout

Examples of poster assessment sheets

Visual impact

Immediate, striking impact, attractive	5	4	3	Ugly, messy, poor visual impact 2 1
Clarity of design				
Clear, self-explanatory	5	4	3	Confusing: needs explanation 2 1
Creativity of presentation				
Novel, surprising, different	5	4	3	Dull, obvious 2 1
Rationality of argument				
Clear, rational argument	5	4	3	Lacks justification or rationale 2 1
Practicality				
Workable, realistic	5	4	3	Unworkable, unrealistic 2 1
Overall evaluation of poster				
Excellent content	5	4	3	Poor content 2 1
Excellent presentation	5	4	3	Poor presentation 2 1
Date:_____				Total score /35

p. 134, Brown, Bull, Pendlebury

Poster assessment

MBChB Poster Assessment Year Two

Name of assessing student:

Name of assessed student:

Poster display assessment

Please include a mark and comment for each of the criteria indicated below:

Criteria	Comment	Mark 1-4 1 = weak 4 = good
Impact		
Was the display:		
Visually interesting and attractive?		
Effective in its use of colour?		
Balanced in its use of a variety of images and text?		

MBChB Poster Assessment Year Two (cont'd)

Content

Did it:

Convey a clear message?

Interest the reader/viewer?

Contain accurate technical information?

Structure

Did it:

Have a clear and logical structure?

Show evidence of planning and preparation?

Further comments

Total mark

p. 135, Brown, Bull, Pendlebury

Assessing Group Projects

Types:

- **A review**
- **A field study**
- **A simulation**
- **A task based on local problem needing a solution**

Purpose:

- **Encourage students to work together**
- **To study, think, reflect and present findings**

Assessing Group Projects

See objectives for group projects

See handouts

Assessing group projects: Some objectives of group projects

Your objectives?	Students know	Objective assessed?
Explore more deeply a topic		
Apply knowledge to a problem		
Provide training in research		
Develop initiative		
Foster independent learning		
Enhance time-management skills		
Develop project management skills		
Improve written communication of findings		
Improve oral presentations of findings		
Develop teamwork		
Develop leadership skills		
Others (please specify)		

p. 136, Brown, Bull, Pendlebury

Time for assessing projects:

Suggestions for marking group projects

Everybody gets the same mark.

The group decides the marks of individuals at the end of the project. (Group project 62: 5 students x 62 = 310 marks to be allocated.)

Group decides criteria at beginning of project and allocates marks at end of project.

Group allocates roles at beginning of project and criteria for each role. It allocates marks at end of project.

Group agrees that everybody will contribute equally to each task of the project. At the end of the project those that did little get marks below the mean, those that did more get marks above the mean. The decisions are taken by the group.

Tutor plus group use any one of 2-5. (For convenience in a discussion you might want to label these 6.1, 6.2, 6.3, 6.4, 6.5.)

Tutor only does any one of 2-5 (7.1, 7.2, 7.3, 7.4, 7.5).

Individual viva.

Individual project exam mark plus group project mark.

Yellow card and Red card. Group gets the same mark. If group reports a malingerer, he/she gets a Yellow card and his/her mark reduced by 10%. If no improvement, Red card issued at the end of project and student gets zero.

Everybody gets the same mark for the project but additional marks are given for contributions to the project.

Variations on the above approaches are possible.

p. 137, Brown, Bull, Pendlebury

Some issues for discussion in a department

How do we assess projects at the moment? Are there any additional steps that we can take to ensure that the assessment of projects is fair and assesses the skills that we want to assess?

Should we ask ex-students for their views on the value, organization and marking of projects?

Could we try a pilot including some self-assessment of some aspects of project work? Should it be based upon the students' comments in the final chapter of the project report? Should it be based upon a separate report?

Are there any other aspects of project work such as project presentation and teamwork that we wish to assess? Could peer-assessment be used? How? Would it be helpful?

What can we do about students' marks if negligent project supervision occurs?

What can we do about a student's marks if his or her work is part of a larger research project which is foundering?

What can we do about a student's marks if the apparatus, test rig or library resources promised were not available?

What can we do to minimize the likelihood of plagiarism? Is it a problem?

p. 138, Brown, Bull, Pendlebury


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

incorporate

Research on peer-assessment

The research evidence on peer-assessment indicates that it can promote critical thinking, the skills of task management, increases in self confidence, responsibility and awareness of group dynamics. These findings are subject to the proviso that:

Adequate training in small-group work is given.

Adequate training in self- and peer-assessment are given.

The learning task is clearly defined.

There is a learning contract or agreement between the members of the group and between the group and the tutor.

The assessment procedure is clear and known to the students.

The assessment marks reflect the time and effort invested by students.

p. 173, Brown, Bull, Pendlebury

incorporate

Criteria for self- and peer-assessment of group processes

Skill	Criteria for self/peer-assessment	Weighting
Ability to work In a group	Ability to arrive at consensus/overcome difficulties	
	Giving and accepting support/facilitating group effort	
Contribution To group	Attendance and time-keeping	
	Application; taking fair share of work	
Initiative	Ability to generate good ideas/solve problems	
	Gathering/researching additional/new information	

p. 175, Brown, Bull, Pendlebury

incorporate

Peer-group assessment: a global approach

Confidential

Please complete you assessment of everyone's contribution to the project, including your own. Provide a mark out of 10.

10-9 = Outstanding contribution

8-7 = Very good contribution

6-5 = Good contribution

4-3 = Fairly satisfactory contribution

2-1 = Unsatisfactory contribution

0 = No or virtually no contribution

Student **Mark** **Comments**

(Name in alphabetical order)

p. 176, Brown, Bull, Pendlebury

incorporate

How our group worked

Most of our meetings were confused	1 2 3 4 5	Most of our meetings were well organized
We often got side-tracked	1 2 3 4 5	We stuck to the task most of the time
We didn't listen to each other	1 2 3 4 5	We did listen to each other
Some talked too much and some did not talk enough	1 2 3 4 5	We all contributed to the discussion
We did not think through our ideas sufficiently	1 2 3 4 5	We thought through our ideas well
Some got aggressive and some got upset	1 2 3 4 5	We were able to argue and discuss without rancour
Most of us seemed to be bored by the discussion	1 2 3 4 5	Most of us seemed to enjoy the discussion
Most of us did not improve our discussion skills	1 2 3 4 5	Most of us did improve our discussion skills
Most of us did not learn much	1 2 3 4 5	Most of us did learn through the group work

Note: the discussion is more important than the rating.
How could the group have worked better?
Name _____Group_____
Thank you for your views.
There will be an opportunity to discuss the overall reactions of the group at our next meeting.
p. 177, Brown, Bull, Pendlebury

IDEAL

incorporate

Uses and usage of self-assessment

The long-term uses of self-assessment are summarized by Boud et al. (1995) as:

For individual self-monitoring: students monitor their own work for their own purposes. They may be introduced to ways of doing this but are left to their own devices.

To promote good learning practices and learning to learn.

For diagnosis and remediation.

As a learning activity to improve professional or academic practice.

To consolidate learning over a range of contexts: this use is of increasing importance for integrating learning from different modules. Guidance and support from tutors is required.

To review achievements as a prelude to recognition of prior learning.

To promote self-knowledge and understanding. Here the predominant interests are self reports, analysis and reflection that move beyond the present task towards long-term development.

As a substitute for other forms of assessment. Here the predominant interest is the generation of marks to be used in formal assessment.

p. 179, Brown, Bull, Pendlebury

incorporate

Meta-cognition: learning to learn

For learners to develop the capability of transferring learning strategies from one context to another, they need to:

- **Be able to articulate their strategies.**
- **Have opportunities to practice them in different contexts.**
- **Have opportunities to assess their strategies and outcomes.**
- **Receive feedback on their efforts to assess.**

incorporate

Meta-cognition: learning to learn

A lecturer can facilitate a student's capacity to learn in seven ways:

As a direct teacher of skills and strategies.

As a model who makes explicit the processes going on in his or her own mind as a problem or task is tackled.

As a provider and assessor of learning tasks.

As a guide during learning tasks.

As a provider and facilitator of peer-group assessment.

As a provider of opportunities for self-assessment.

As an assessor of self-assessment tasks.

p. 181, Brown, Bull, Pendlebury

incorporate

Getting started with self- and peer-assessment

Start:	**Don’t prevaricate, give it a try.**
Start small:	**Don’t be over-ambitious at first, festina lente.**
Convince:	**Ensure the students know that what they are doing is useful.**
Clarify:	**Let students know what is expected of them and be aware that criteria can be problematic.**
Be relevant:	**Make sure that what you are doing relates to course aims and philosophy.**
Be flexible:	**There are no hard and fast rules, adapt as you go along.**
Reflect:	**Think the processes through carefully so that you are aware of the implications of what you are doing.**
Innovate:	**Treat each event as new but recycle the useful elements.**
Rehearse:	**Let students have an opportunity to try things out before the assessment that counts.**
Liaise:	**Share good practices with friends and colleagues and support each other, don’t try to invent the wheel, persist, don’t let failure discourage you.**

p. 182, Brown, Bull, Pendlebury

incorporate

Skills expected of graduates

Knowledge skills

Graduates should

- have a body of knowledge in the field(s) studied;
- be able to apply theory to practice in familiar and unfamiliar situations;
- be able to identify, access, organize and communicate knowledge in both written and oral English; and
- have an appreciation of the requirements and characteristics of scholarship and research.

Thinking skills

Graduates should

- be able to exercise critical judgment;
- be capable of rigorous and independent thinking;
- be able to account for their decisions;
- be realistic self-evaluators;
- adopt a problem solving approach; and
- be creative and imaginative thinkers.

Personal skills

Graduates should have

- the capacity and desire to continue to learn;
- the ability to plan and achieve goals in both the personal and the professional sphere; and
- the ability to work with others.

Personal attributes

Graduates should

- strive for tolerance and integrity; and
- acknowledge their personal responsibility for (i) their own values judgments; and (ii) ethical behaviour towards others.

Practical skills (where appropriate)

Graduates should be able to

- collect, correlate, display, analyse and report observations;
- apply experimentally obtained results to new situations; and
- test hypotheses experimentally.

Based on Candy et al (1994)

p. 37, Brown, Bull and Pendlebury


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Modifications to Bloom's Taxonomy

1. See hand outs and references pertaining to these modifications
2. An overview: see following slides

Remember:

Retrieve relevant knowledge from long-term memory

- RECOGNIZING (Identifying)
- RECALLING (retrieving)

Understand:

Construct meaning from instructional messages, including oral, written & graphic communication

INTERPRETING	**(clarifying, paraphrasing, representing, translating)**
EXEMPLIFYING	**(illustrating, instantiating)**
CLASSIFYING	**(categorizing, subsuming)**
SUMMARIZING	**(abstracting, generalizing)**
INFERRING	**(concluding, extrapolating, interpolating, predicting)**
COMPARING	**(contrasting, mapping, matching)**
EXPLAINING	**(constructing models)**

Apply:

Carry out or use a procedure in a given situation

- EXECUTING (carrying out)
- IMPLEMENTING (using)

Analyze:

Break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose

DIFFERENTIATING (discriminating, distinguishing, focusing, selecting)

ORGANIZING (finding coherence, integrating, outlining, structuring)

ATTRIBUTING (deconstructing)

Evaluate:

Make judgments based on criteria and standards

CHECKING **(coordinating, detecting, monitoring, testing)**

CRITIQUING **(judging)**

Create:

Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure

GENERATING	**(hypothesizing)**
PLANNING	**(designing)**
PRODUCING	**(constructing)**


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

Short Answer and Modified Essay Questions

Why ?

How ?

Tim Neild

Flinders University School of Medicine

Adelaide, Australia

Why ?

The free form allows the possibility of testing almost anything:

- knowledge,
- reasoning,
- synthesis….

Once upon a time, short essay questions were the only sort of written test used.

Why has this changed?

What are we trying to do?

Advantages

They allow the student to show large quantities of information, and to demonstrate that:

- the information is known
- the relationship between different pieces of information is known (eg. as in control of blood pressure, where the relationships are usually taught)
- the relationship between different pieces of information is understood (eg. as in interpretation of an unseen scenario)

These questions are easy to set (compared to multi-choice).

Disadvantages

- They take a long time to mark
- Their intention has not always been clear to students

The practical aspects of two types

Short answer questions

The traditional exam question
A single entity
May be subdivided, but all parts are presented together.

“Mini cases” – “modified essay questions”

Multiple entities presented sequentially.

General principles that apply to all question types

Define the problem. What are we trying to find out about the students?

We want to know:

- whether the students ***know enough*** to move to the next stage
- whether the students know what ***we expect them to know***
- whether the students know what ***we have informed*** them they should know (Have we been clear about this?)
- whether the students know what …….(as above)
  and ***have the skills to use the information*** in the way we expect (have informed them?)

General principles……..

Once we have decided on the subject area (blueprinting?),
are we aiming to asses:

- Factual knowledge? (several Q types useful here)
- Understanding of a biological process (either normal or abnormal)?
  short answer questions are good for this
- Ability to reason or investigate efficiently?
  mini cases are designed for this

Short answer questions

Decide the “scope” and “value”, and find a way to indicate this to the students.

Scope

Questions requiring students to show extensive knowledge of a broad area will

- be expressed in language that makes this intention clear
- have sufficient time allocated to them
- be worth a proportionately large number of marks (high value)

Value

Sometimes important things can be expressed in few words, but are worth high marks. This is fine as long as it is clear to the students.

The extreme is the “killer question” – one that must be passed to pass the exam.

Decide on the difficulty of the question

Easy for students who are novices in the area, hard as their experience increases.

Difficult, expert

"Discuss the possible causes of unilateral limb weakness"

30 marks

- Completely free form, requires ability to express a synthesis of a complex area for top marks.
- Potential to distinguish excellent from good students.
- Requires an expert marker.

It may be desirable to indicate something about what is required:
eg.

"Discuss the possible causes of unilateral limb weakness. Indicate which causes are the most commonly encountered in general practice."

Still expert but lower "value"

"List the possible causes of unilateral limb weakness"

8 marks

- Form is specified (list), and does not imply detailed analysis.
- Could be marked by a non-expert using a checklist

Novice

"List the possible causes of unilateral limb weakness, using to the categories:

upper motor neurone lesions

lower motor neurone lesions"

8 marks

Form is specified (list), and does not imply detailed analysis.

- Q contains a reminder of the type of thing the answer should contain, which may be appropriate for novices in the field.
- Could be marked by a non-expert using a checklist

These distinctions between value and difficulty, and the ways of adjusting questions for them, are useful when assembling large tests from blueprints.

Marks allocated and/or time allowed are clear indicators of value that are usually understood
by students.

Wording is always a problem (is Cantonese better than English in this respect?).

Schools or National Tests tend to develop their own wording conventions that students learn.
This is one of the main reasons that students get better at assessment as a course progresses – they learn the course language and culture.

Therefore, for reliable summative assessment there must be prior formative assessment in the same form.

Who's fault is it if the student didn't understand the question……?

At the same time as setting the question, set the "answer".

The "answer" comes in two forms:

A model answer that will be released to students as part of a formative feedback process.

- A model answer is one that would get full marks, so make sure that it is ***not more*** than was required.
- Try and write it yourself in the time allocated. Adjust the question or the time if you find this impossible!

Model answers are useful to others who are reviewing your questions.

It is quite common (at Flinders anyway) for a review panel to suggest a change to a question because they could not see how it was asking for the "model answer", and only that answer.

The model answer is almost always a clearer indication of the examiner's intention than the question.

Closely related to the model answer is the "**marking scheme"**

The marking scheme is usually the model answer annotated to show the features that should be awarded marks.

A clear marking scheme is essential when

- there are multiple markers for the same question
- there is the possibility that the mark will be disputed and the mark will have to be explained or a re-mark obtained.

However, there will always be a place for expert judgement in marking.

[Standard setting would probably come next – I'm not going to talk about it]

Marking

Allocate marks according to the marking scheme, supplemented by expert judgement and experience……..

This is the biggest problem with short answer questions – the time and effort needed to mark them.

If an accurate grade or ranking for the student is required, there may be no alternative to hard work by the examiners.

But if the decision is simply “progress / do not progress”, there may be a quicker way, as shown by a recent experiment at Flinders….

Experienced examiners often say that they can read an answer and quickly say whether it is good enough or not.

In a recent test at Flinders, an experienced examiner quickly graded a series of answers on the scale:

• excellent	pass
• good enough to pass	pass
• borderline	
• not good enough	fail
• very poor	fail

The borderline answers (about 20% of the class) were then examined in detail with the help of another experienced examiner, and assigned to either the “good enough” or “not good enough”

There were 10 questions in the test, and students were required to pass 7. Results correlated closely with past performance.

The entire process took about 2/3 the time of conventional marking.

Finally, evaluate the question

Was the Q too hard or too easy?

- There is no absolute standard, it depends on the course.
- Was the Q easier or harder than others in the test?
- Was it expected to be that way?

Was the question “Reliable”?

Did the students who got high marks overall also get high marks for this Q?

The usual assumption is that all questions are testing the same student qualities. “Reliability” statistics can reveal questions that seem to be testing something else, or were testing nothing because they were too easy or too hard.

Some examples…..

These are two real examples from Flinders this year.

The test as a whole (14 questions) was highly “reliable”, but answers to one question did not correlate. Which one was it, and why?

Question 2.
Explain the role of the autonomic nervous system in keeping arterial blood pressure constant in a person who goes from a recumbent (lying down flat) position to standing.

(15 marks)

Question 3.
A drug is observed to produce the following effects:

- dry mouth
- reduced sweating
- hypotension
- increased blood flow to skin
- dilation of the pupils
- inability to focus on near objects

How is this drug acting? Explain your reasoning.

(5 marks)

Question 4.
Ingestion of a typical meal consisting of a mixture of carbohydrates, proteins and lipids results in a moderate increase in blood glucose. The glucose concentration then returns to fasting values within the following two hours.
Describe the processes leading to rapid removal of glucose from the circulation following a meal. Discuss the major fates of this glucose in **liver** and **muscle**. (You should outline any relevant metabolic pathways but do not need to provide details of most of the intermediate metabolites or enzymes.)

(11 marks)

Another test (10 questions in total) that was statistically “reliable” overall, but had one rogue question.

Which one, and why?

Millie Valasquez, aged 8, is accidentally pushed through a plate glass door ……

7. If the median nerve does not grow, what functional deficits will Millie have? Which muscles will be involved?

(10 marks)

Alessandra DiCesare, aged 68 years, reports persistent pain in her lower back……..

8. Discuss the factors that can contribute to abnormal regulation of bone metabolism and the development of osteoporosis in the elderly.

(10 marks)

9. Muscles of the back and abdomen contribute to both moving and supporting the vertebral column. Compare the actions of the muscle groups involved in each of these functions.

(10 marks)

……, Bob McDonaugh, a 47-year-old farmer, develops the signs of tetanus infection…..

10 How do muscles normally move the jaws, cheeks and mouth during eating? Tetanic infection of which nerves will lead to spasm of the masseter and muscles around Bob’s lips?

(10 marks)

Testing Reasoning with Mini-cases

Information → hypotheses
Need to distinguish between hypotheses → need for further information

Eventually, after cycles of information gathering,
Further information → conclusion

For *Scientific Reasoning*, the conclusion is understanding

For *Clinical Reasoning*, the conclusion is a diagnosis
and perhaps a management plan.

The mini-case format is often tailored to mimic the PBL format in a PBL course, but the basic idea can be applied to any course which values reasoning in the style outlined above.

Mini-case example

First, a clinical scenario

Next, the generic questions, which ask for
- hypotheses (number should be specified)
- information required to distinguish between them

Students need to be familiar with these questions and know how to answer them. *The answers differ for scientific and clinical reasoning.*

A 53-year-old man was found at home by his relatives, unconscious and unresponsive but with continuous epileptiform movements. He has a long history of alcoholism and had been drinking heavily for several days prior to this seizure.
On arrival in the Emergency Department he appeared to be aware of questions but did not respond verbally. His temperature was 38.5ºC and his neck was stiff.

Q.
Give 3 possible explanations for this clinical picture and explain your reasoning.
Q
What tests used in common practice would help you differentiate between your hypotheses? Give reasons for your choices.

Writing the model answer is as important as ever.

Q

What tests used in common practice would help you differentiate between your hypotheses? Give reasons for your choices.

For instance, what does “used in common practice” mean? Are the students sufficiently experienced to know this?

It needs to be agreed before marking starts.

An important feature of a mini case is progressive disclosure.

Students answer the initial questions, then get some more information.

They are not allowed to change their answers to the initial questions when they get the subsequent information.

There are various way to achieve this in practice:

- collect the answer papers
- put the answer papers on the floor
- have an answer book and tell students that they cannot turn back

After the initial stage, mini case questions usually focus on the specific areas that the examiners want to test.

Further examination

heart rate 130 bpm
respiratory rate 30/min
crepitations heard over the right chest wall

Chest -X-ray shows consolidation in the right middle and upper lobes of the lung
A CT scan of the head was normal
A full blood examination showed

Hb	*155 g/l*	*(normal 130 – 175)*
WBC	*18.7 x 10^9 /l*	*(normal 4 – 11)*

93% neutrophils, which show toxic granulation and a shift to the left

Q
What information is gained from the blood test results and what is its significance to the clinical presentation? Explain the mechanisms leading to these results.

Model answer was about findings typical of infection, and mechanisms responsible for neutrophilia. (This was a first year exam)

Mini cases may have only 2 parts, but this one had a third part.

Blood cultures grew gram positive diplococci from the aerobic bottle.
A lumbar puncture showed large numbers of polymorhs and gram positive diplococci.

Q
What diagnosis do these results suggest? 3 marks
Q
What is the most likely organism causing his problems? 2 marks

He was started on IV Benzylpenicillin (Penicillin G).

Q
Discuss the mechanism of action of beta-lactams, and how resistance to these antibiotics develops.
15 marks

The final question is a conventional short answer question.

This is OK; exams can have a mixture of question types.

Conclusion. What really matters.

It's best practice to get everything right, but two points stand out:

1. *Decide what you are trying to test*
   Check that it is fair to expect the students to perform to this level.
   Choose the right question format.

2. *Write the answer when you write the question*
   Check that you can do it well in the time allowed.
   Check with someone else that your answer follows logically from your question.


Howe Knowledge International Consortium for Sharing Medical Student Assessment Banks

How Clinicians Learn

Chinese University of Hong Kong

December 15 2000

Gordon Page

Professor, Faculty of Medicine
Director, Health Sciences Division of Educational Support and Development

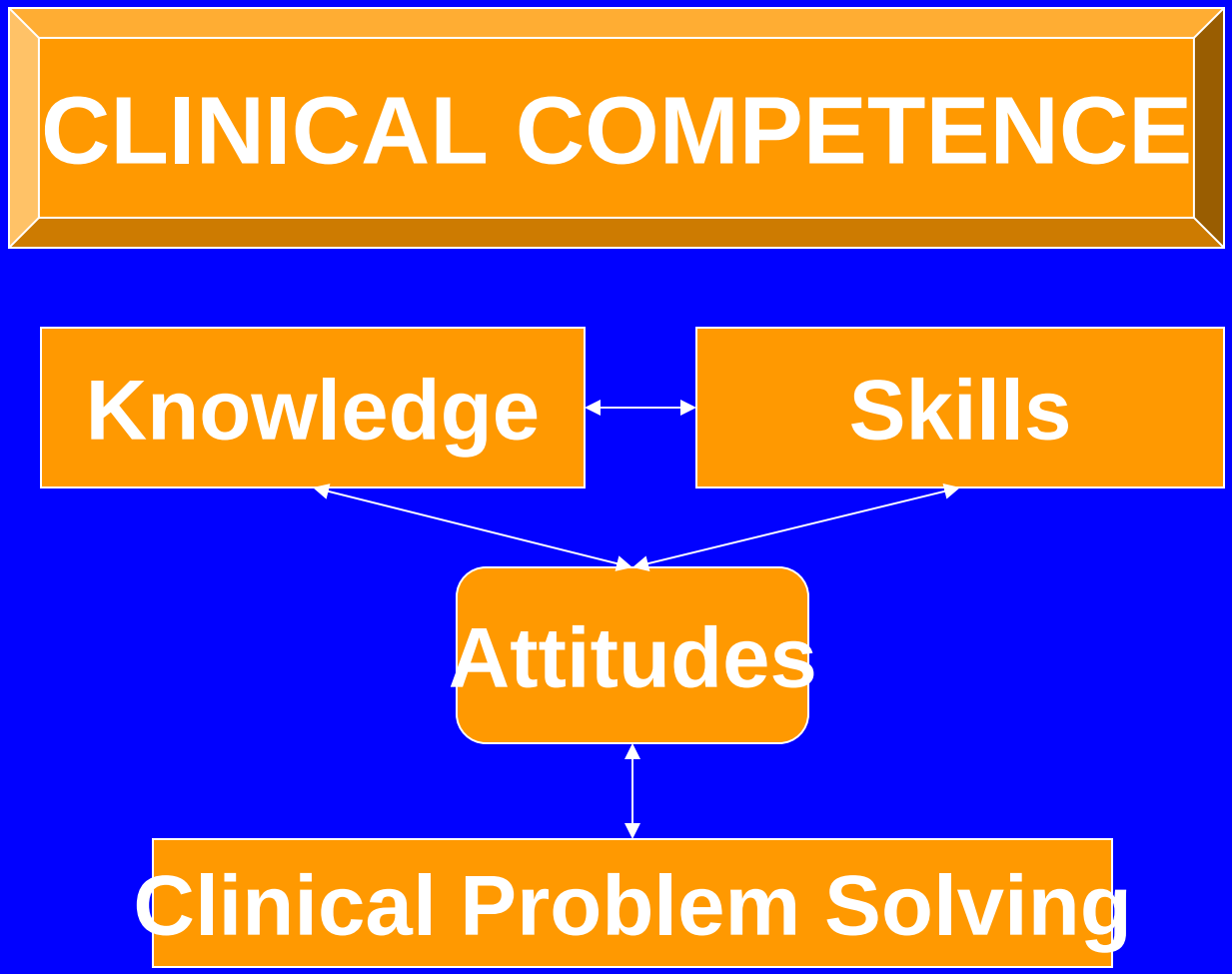
University of British Columbia
Vancouver Canada

Download Presentation

- www.health-sciences.ubc.ca/
- desd/presentations/

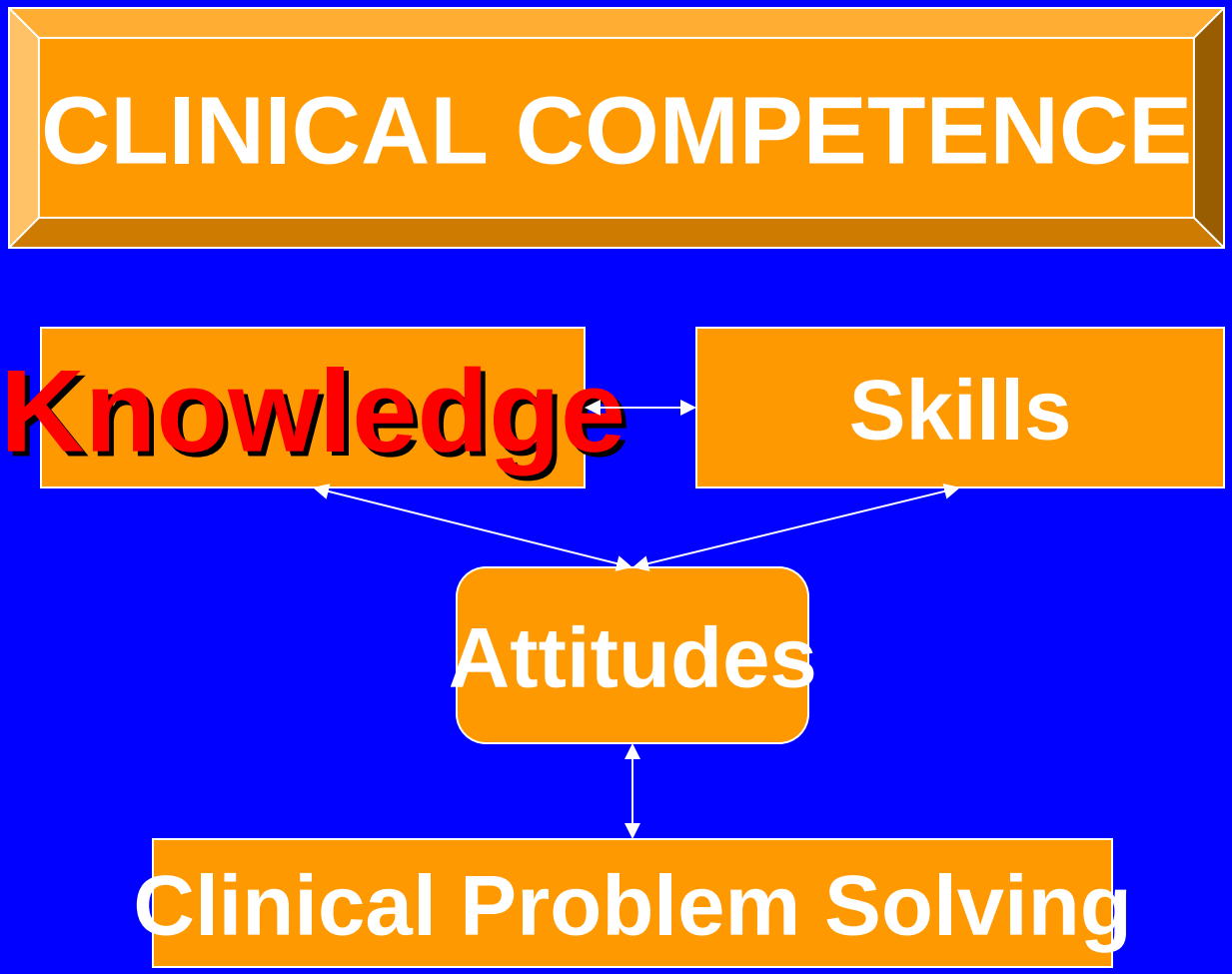
- *(after Dec 20)*

How do Clinicians Learn?



Newble, 2000

How do Clinicians Learn?



Newble, 2000

How do Clinicians Learn? Overview of Presentation

- *Clinical expertise -- How do competent clinicians problem solve?*
- *Psychological views of learning*
- *The role of "context" in learning*

How do Clinicians Learn? Overview of Presentation

- ***Clinical expertise -- How do competent clinicians problem solve?***
- ***Psychological views of learning***
- ***The role of "context" in learning***

Clinical Expertise

- ***How do competent clinicians arrive at diagnostic and management decisions (i.e., "think through") a clinical problem?***

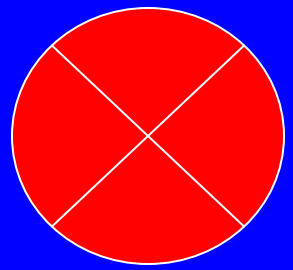
Clinical Expertise

(Bordage, G., Academic Medicine, 1994, 1999)

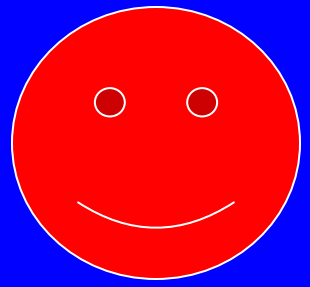
- ***Reduced*** **(“empty mind”)**
- ***Dispersed*** **(“cluttered mind”)**
- ***Elaborated*** **(“deductive thinking”)**
- ***Compiled*** **(“recall/recognition”)**
-

Clinical Expertise

- *Reduced*
- *Dispersed*



- *Elaborated*
- *Compiled*



-

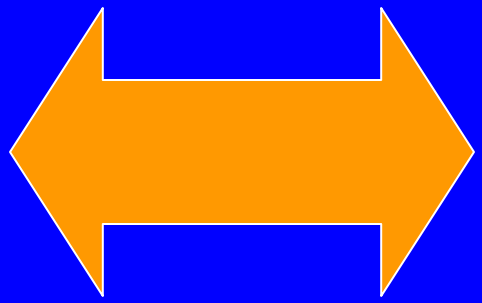
Clinical Expertise

- Competent clinicians use deductive reasoning (elaborated thinking) or recall/recognition (compiled thinking) processes
  - **********
- Expertise is “case/problem specific” (central role of knowledge)
- Clinician’s knowledge is organized around cases
- Thoroughness of data gathering is unrelated to diagnostic accuracy

How do Clinicians Learn? Overview of Presentation

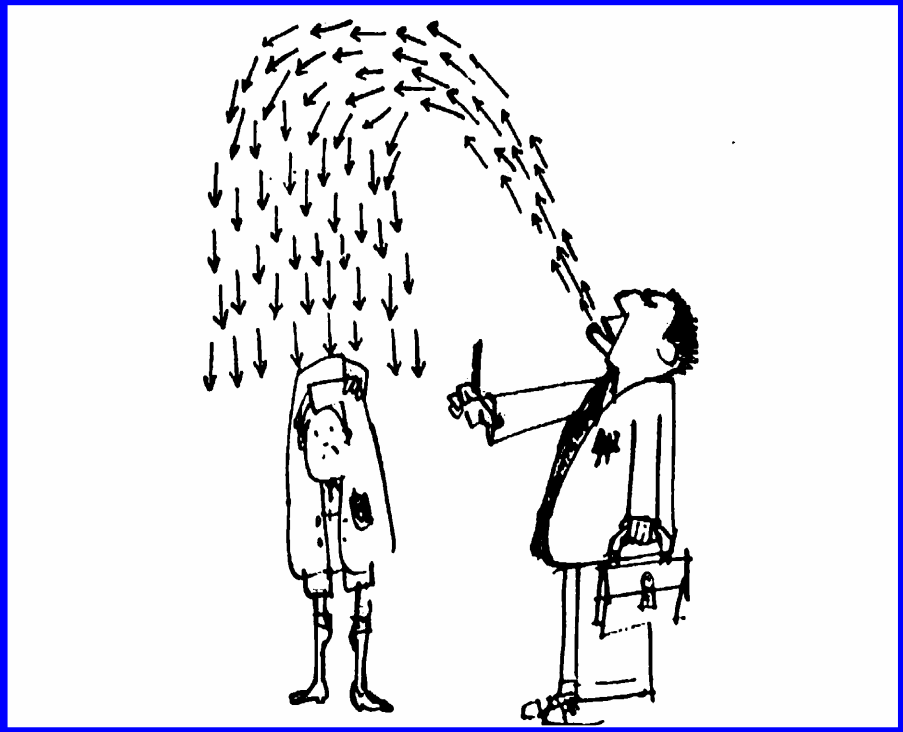
- *Clinical expertise -- How do competent clinicians problem solve?*
- *Psychological views of learning*
- *The role of "context" in learning*

Contrasting views of Teaching and Learning

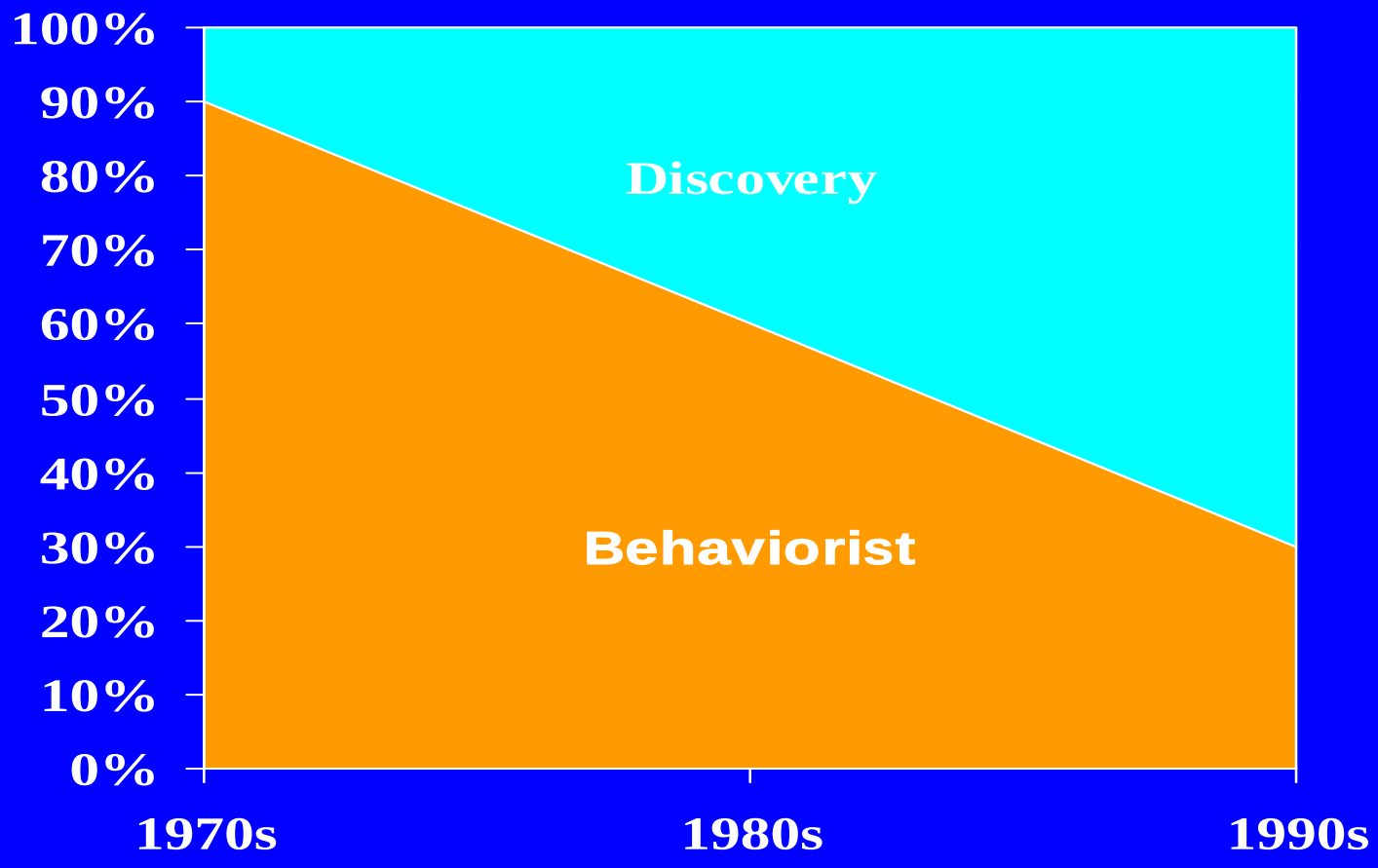
Behaviorist  *Discovery*

Behaviorist View of Learning

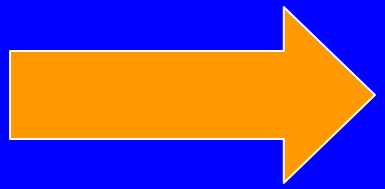
- Behavioral objectives
- Didactic instruction
- Acquisition of facts



Contrasting views of Teaching and Learning



Contrasting Views of Teaching and Learning

Behaviorist  ***Discovery/ Adult Learning/ Constructivist***

1970s *1980s* *1990s*

The teaching/learning process tends to be more effective if:

- ✓ Individuals are clear about what to learn
- ✓ Individuals are intellectually active in the learning process
- ✓ Individuals have opportunities for practice with feedback, the goal of the latter to help them identify the success of their learning

The teaching/learning process tends to be more effective if:

- ✓Individuals are clear about what to learn
- ✓Individuals are intellectually active in the learning process
- ✓Individuals have opportunities for practice with feedback, the goal of the latter to help them identify the success of their learning

Sample “1970s” Behavioral Objectives

- Define and describe the clinical manifestations, pathophysiology and principal methods of diagnosis and treatment for pyloric stenosis and cystic fibrosis
- Describe the pathophysiology, clinical manifestations and treatment of congenital hypothyroidism

Sample “1970s” Behavioral Objectives

- Define and describe the clinical manifestations, pathophysiology and principal methods of diagnosis and treatment for pyloric stenosis and cystic fibrosis
- Describe the pathophysiology, clinical manifestations and treatment of congenital hypothyroidism

Sample “1990s” Objective

- *“Develop diagnostic and management plans for children with the following clinical problems, and explain the rationale for these plans: cough, fever, vomiting, heart murmur, seizures, …”*
-

Sample “1990s” Objective

- *“Develop diagnostic and management plans for children with the following clinical problems, and explain the rationale for these plans: cough, fever, vomiting, heart murmur, seizures, …”*
-

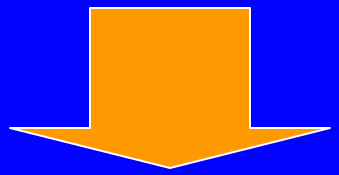
Sample “1990s” Objective

- *“Develop diagnostic and management plans for children with the following clinical problems, and explain the rationale for these plans:* ***cough, fever, vomiting, heart murmur, seizures****, …”*

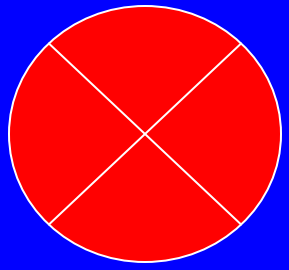
- “bottom-up” vs “top-

Top-Down

Diagnosis

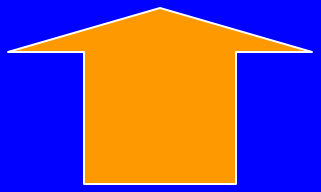


Clinical Features

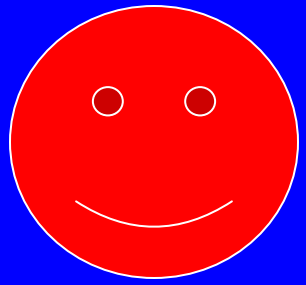


Bottom-Up

Diagnosis

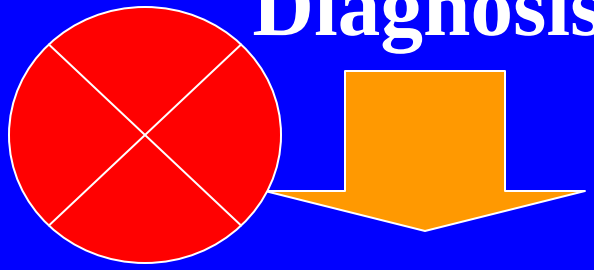


Clinical Features



Top-Down

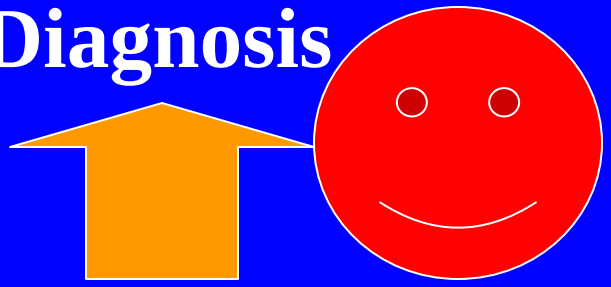
Diagnosis



Clinical Features

Bottom-Up

Diagnosis



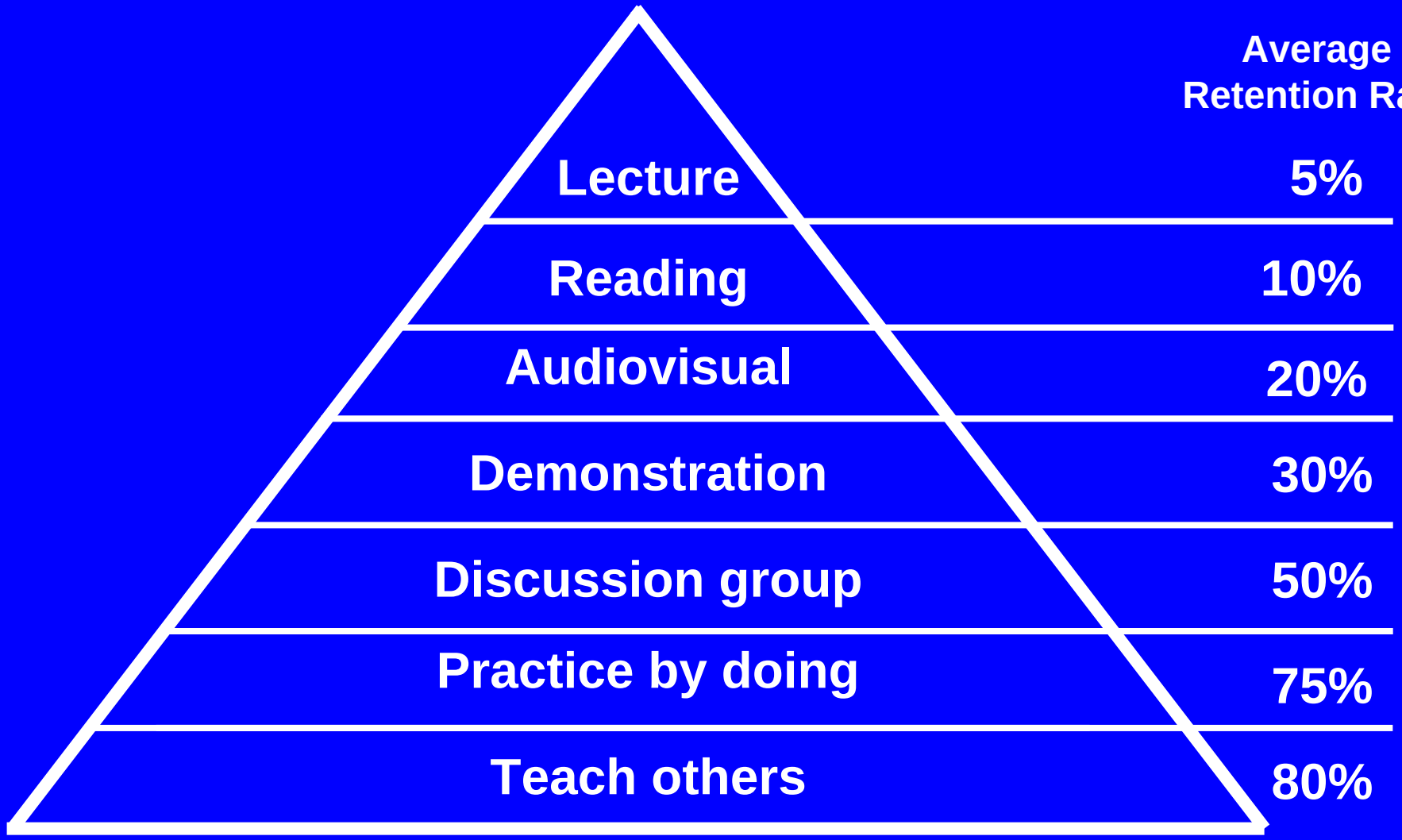
Clinical Features

Objectives ⟷ Instruction ⟷ Assessment

The teaching/learning process tends to be more effective if:

- ✓ Individuals are clear about what to learn
- ✓ Individuals are intellectually active in the learning process
- ✓ Individuals have opportunities for practice with feedback, the goal of the latter to help them identify the success of their learning

The learning pyramid



National Training Laboratories, Bethel, Maine, USA

How do Clinicians Learn? Overview of Presentation

- *Clinical expertise -- How do competent clinicians problem solve?*
- *Psychological views of learning*
- *The role of "context" in learning*

Context

- Context serves as an important memory aid/information retrieval cue

- *The context in which clinicians learn something should match as closely as possible the context in they must ultimately use that learning.*

Summary

The learning process for clinicians tends to be more effective if …

- ✓ Learning occurs in the context of cases -- the situation in which the clinician must ultimately use their learning
- ✓ The learning process is "bottom-up"
- ✓ Clinicians are intellectually active in the learning process
- ✓ Opportunities for practice with feedback ("experience") are provided

Summary

The learning process for clinicians tends to be more effective if …

- ✓ Learning occurs in the context of cases -- the situation in which the clinician must ultimately use their learning
- ✓ The learning process is "bottom-up"
- ✓ Clinicians are intellectually active in the learning process
- ✓ Opportunities for practice with feedback ("experience") are provided

Changes in Education
Medical School

- **1960s - 1970's**

- Normal --> abnormal --> clinical (basic sciences --> clinical exposure)
- Instruction organized by discipline
- Didactic "teacher-centred" instruction

- **1990's- 2000**

- Integration of basic sciences and clinical learning
- Instruction organized by organ system or around cases
- Interactive "student-centred" instruction
- Training in out-of-hospital settings

Changes in Education for *Medical Practitioners*

1960's - 2000

- **College requirements for "continuing professional development"**
- **Formal CME is less didactic, more interactive and "clinical practice or case-based"**
- **Increasing opportunities for self-assessment and office-based peer review**

Sir William Osler – “the natural method of teaching” (1932)

- ***“The student begins with the patient, continues with the patient, and ends his studies with the patient, using books and lectures as tools, as a means to an end”.***

Download Presentation

- www.health-sciences.ubc.ca/
- desd/presentations/

- *(after Dec 20)*