Item Analysis (I)

C B Hazlett

Chinese University of Hong Kong

[gratefully acknowledge contributions by E Skakun, J Ware and Dept of Paed, CUHK]

Overview of Presentation

Part I

Introduction: What is Item Analysis (IA)?

Part II

IA for items marked in a binary fashion (right=1, wrong=0)

Part II

IA for items marked in a continuous manner
(scores from 0 to any value such 99.9)

Objectives

After this presentation you will

understand how some statistical indices can detect if an item is performing as desired

use item analysis (IA) regularly & competently in your review of any student assessments

Introduction: What is Item Analysis?

Part I

Item Analysis (IA)

Process by which assessment items are critically reviewed

- determine if items function according to expectation
- identify structural flaws
- improve item quality

Use both judgment & empirical data

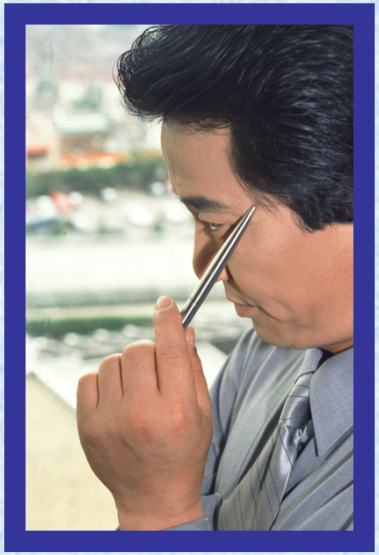
Judgmental Item Analysis

The Judgmental Criteria

Are content, processes & constructs being assessed by the item relevant?

Is the item properly structured?

Is the item free of bias?



Empirical Item Analysis

Relevant statistical properties of an ***overall*** assessment

Mean of total test (average score) $\bar{X}$

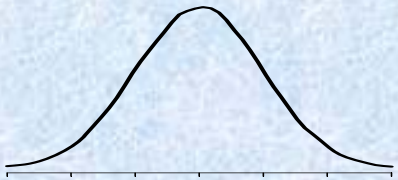
- if mean is near or below cut point for passing, ask
  - is test mismatched to course objectives?
  - was teaching adequate?
  - if this is formative assessment, maybe this is ok as it is good incentive for undertaking remediation

Empirical Item Analysis: Overall Assessment

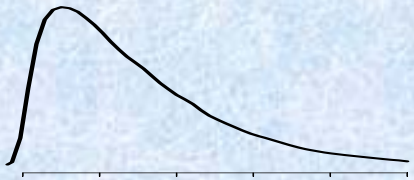
Relevant statistical properties of an *overall assessment*

Histogram & table of candidates' scores on test

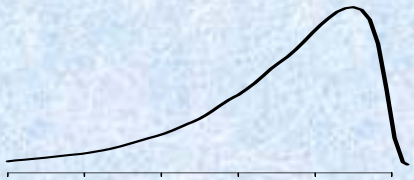
- examine shape of distribution: likely normal or negatively skewed; possible concern if positively skewed & mean is lower than desired



√



X



√

Empirical Item Analysis: Overall Assessment

Relevant statistical properties of an *overall assessment*

Reliability of assessment (KR_{20} or Cronbach's alpha: α)

- measures the level of commonality across all items

- important to use enough items/stations (i.e., reduces error and improves sampling of relevant content)

  reliability increases (from 0 to 1) as no. of items/stations increases

Empirical Item Analysis: Overall Assessment

Relevant statistical properties of an *overall assessment*

Standard error of measurement (SEM): $\sigma_{\bar{x}}$

- want this to be small in order to make more confident judgments

- candidate's true score is likely within ± SEM

- e.g. if a candidate's score is 65 & SEM is 10, then there is a 67% likelihood his/her true score is between 55 & 75 (i.e., ± 1 SEM)

  there is a 95% likelihood his/her true score is between 45 & 85 (i.e., ± 2 SEM)

Empirical Item Analysis: Properties of Items

Relevant properties of an *item* within an assessment

Item difficulty

Item's correlation with total mark on the assessment

(or correlation between item & a reference/gold standard)

Item discrimination

Empirical IA Facilitates Benchmarking

Difficulty levels

enable a school to benchmark its students' performances against those in other schools within IDEAL Consortium

Item Analysis: Items Scored as Right (1) or Wrong (1)

Part II

Student Responses Marked in a Binary Manner (0/1)

Involves Most Selected Response Formats

X - type (T / F)[1] or Multiple X-type (Multiple T/F)[1]

A - type (best one of n options)

R - type (extended matching)

---

[1] Consortium no longer accepts X-type items as most members regard them unsuitable

Example: An Item & Its Statistical Performance Properties

Among the common study designs used in clinical research, a study of rare harmful effects of an intervention requires use of which design in order to establish the most valid but also ethically obtained evidence?

A. randomized control trial (RCT)

B. case-control (retrospective) study *

C. case study

D. cohort (prospective) study

E. case series

* Originally keyed answer

Statistical Report for Item 1

ITEM 1: DIF=0.837, RPB= 0.179, CRPB= 0.049 (95% CON= -0.125, 0.220)

GROUP	N	INV	NF	OMIT	A	B*	C	D	E
TOTAL	129	0	0	0	.12	.84	.00	.01	.04
HIGH	39	0			.05	.95	.00	.00	.00
MID	58	0			.12	.81	.00	.02	.05
LOW	32	0			.19	.75	.00	.00	.06
DISCRIMINATING POWER					- 0.14	0.20	0.00	0.00	- 0.06

Definition & Interpretation of Statistical Indices

ITEM 1:DIF=0.837 RPB= 0.179 CRPB = 0.049 95% CON= [- 0.125, 0.220]

DIF: Difficulty (0.837) ~84%

proportion of examinees answering item correctly

range 0.0 – 1.00 (0 – 100%); DIF → 100, easier the item

RPB: Point-biserial correlation (0.179)

correlation between item and total test score

range -1.00 – 1.00

Definition & Interpretation of Statistical Indices (cont'd)

CRPB: Corrected point-biserial correlation (0.049)

correlation between item & test's total score not including item in this total score

95% CON: 95% Confidence Interval for the CRPB

.95 Confidence interval [- 0.125 --- 0.220]

Other Components in Item Analysis Report

GROUP N INV NF OMIT A B* C D E

N	Number of examinees in the group (129)
INV	No. of examinees NOT providing a valid response to this item (0)
NF	No. of examinees NOT finishing the test from this item onwards (0)
OMIT	Number of examinees omitting this item
A – E	Alternatives, best answer is * (B*)

Other Components in Item Analysis Report (Cont'd)

HIGH – approximately 27% of candidates scoring the highest on the test

MID – approximately 46% of candidates whose performance is 'midway'

LOW - approximately 27% of scoring scoring the lowest on the test

Other Components in an Item Analysis Report (cont'd)

DISCRIMINATING POWER

difference between the proportions of the HIGH and LOW groups that select the each option; want the value to be positive for the correct answer & negative for all wrong options

What Statistical Properties are Wanted?

Answer: It depends on purpose of the assessment item

- **is this a criterion or norm referenced assessment?**
  - *if criterion referenced*, difficulty levels, bi-serials & discrimination indices approaching zero can be just what you want
- **is this a formative or summative assessment?**
  - *if formative* & difficulty level is low, use this as a diagnostic tool for remediation
  - *if summative*, evaluate the difficulty level in reference to cut score

What Statistical Properties are Wanted ?

DIF (Difficulty Level)

- statistically items that are in the mid range (.30 to .70 difficulty) have more likelihood of demonstrating adequate discrimination & correlation with the total test
- educationally, often want item answered correctly by a majority of the class; if criterion reference assessment, maybe even 100%
- if difficulty drops below .2, either question or teaching is poor, or assessment is otherwise mismatched to targeted audience & course

What Statistical Properties are Wanted ?

CRPB (correlation between item and rest of the assessment)

- corrected point bi-serial will be smaller than point bi-serial
- want value to be positive, but don't expect many items to > 0.25
- as difficulty nears 0 or 1, correlation can't be much above 0.0
    - thus, if using a criterion referenced item & difficulty is 1.0, don't be concerned that CRPB is also zero

What Statistical Properties are Wanted?

Discrimination Power

- value should be positive for correct option (altho' close to 0 okay if item is criterion referenced & all or almost all candidates are answering correctly, i.e., Dif is near 1.0)
- if norm referenced item, a discrimination less than 0.10 is indication item has some room for improvement; values > 0.35 are not common
- values should be negative for incorrect options

Good Results for Item Designed to be Discriminating

GROUP	N	INV	NF	OMIT	A	B*	C	D	E
TOTAL	129	0	0	0	.10	.75	.08	.07	.10
HIGH	39	0			.00	1.00	.00	.00	.00
MID	58	0			.08	.74	.06	.05	.07
LOW	32	0			.15	.48	.11	.10	.16
DISCRIMINATING POWER					- .15	.52	- .11	- .10	- .16

Learning to Screen for Possible Problems

With a little practice, can look at item analysis output & quickly spot items with possible problems

Without items in hand what can you deduce from the following item performance indicators?

Identify Possible Problems

ITEM 150: DIF = 0.223 RPB= 0.061

GROUP	N	INV	NF	OMIT	A*	B	C	D	E
TOTAL	129	0	0	0	.22	.18	.20	.19	.21

Identify Possible Problems (Cont'd)

Keyed as correct answer

ITEM 2: DIF= 0.109, RPB= 0.021

GROUP	N	INV	NF	OMIT	A	B	C	D*	E
TOTAL	129	0	0	0	.01	.71	.04	.11	.13

Identify Possible Problems (Cont'd)

ITEM 3: DIF= 0.503, RPB= 0.291 Dis = .33

GROUP	N	INV	NF	OMIT	A	B	C*	D	E
TOTAL	229	0	0	0	.44	.00	.50	.00	.06

GROUP	N	INV	NF	OMIT	A	B	C*	D	E
TOTAL	229	0	0	0	.44	.00	.50	.00	.06
HIGH			0		.26	.00	.71	.00	.03
MID			0		.51	.00	.43	.00	.06
LOW			0		.53	.00	.38	.00	.09

Identify Possible Problems (Cont'd)

ITEM 4: DIF= 0.821, RPB= -0.181

GROUP	N	INV	NF	OMIT	A	B	C	D	E*
TOTAL	113	0	0	0	.00	.00	.18	.00	.82
HIGH		0			.00	.00	.25	.00	.75
MID		0			.00	.00	.24	.00	.76
LOW		0			.00	.00	.06	.00	.94

Item Analyses for Items Scored as Right (1) or Wrong (0)

Small Group Exercise

Use the item analysis to flag possible problems for the

9 Paediatric A-type items provided at end of this presentation

IA for Items Marked as Continuous Data

Part III

Items Marked as Continuous Data: 0 to Some Value

Most Constructed Response Formats

Modified Essay Questions (MEQ)

Short Answer Questions (SAQ)

OSCE (Objective Structured Clinical Examinations)

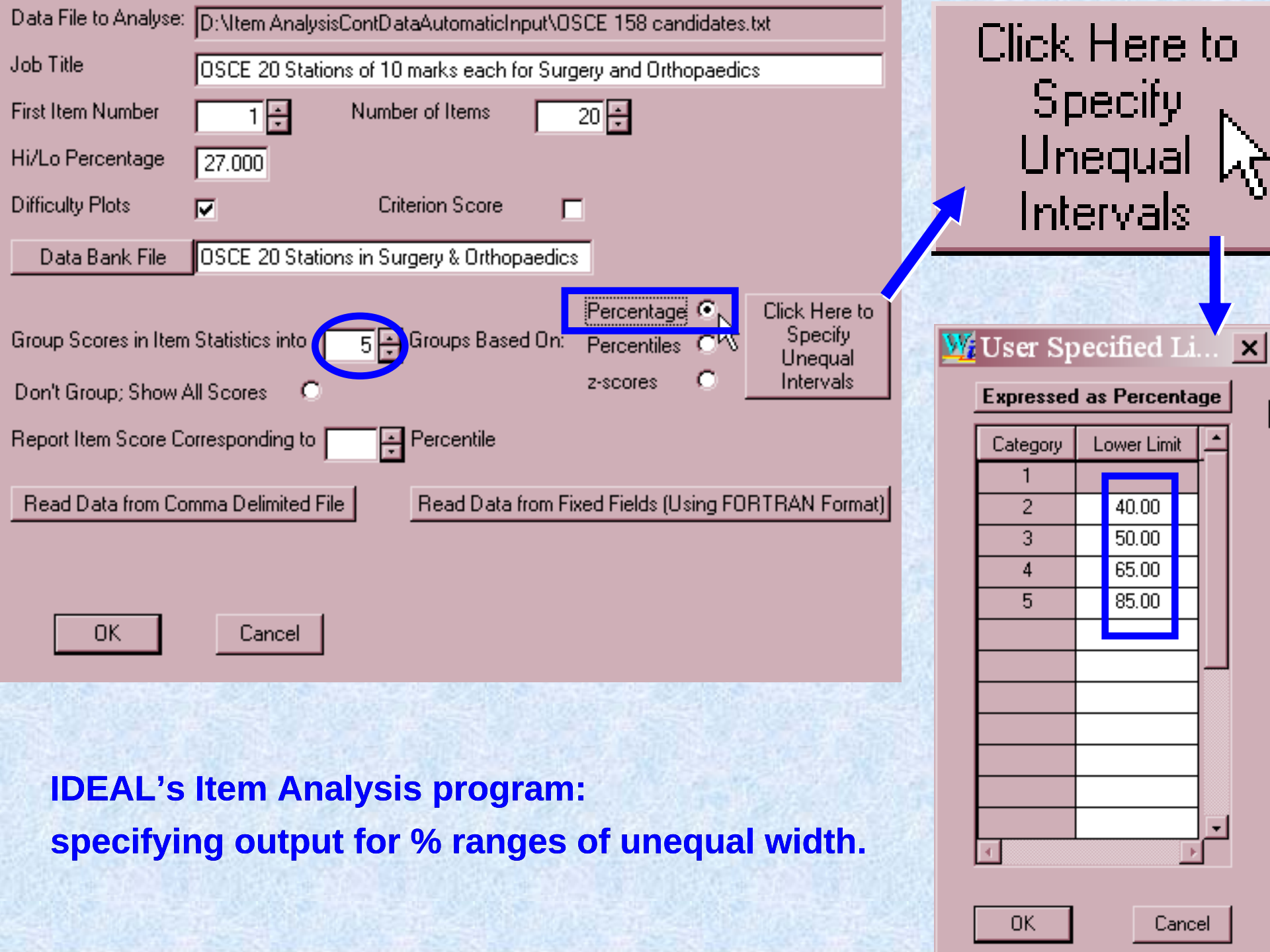
OSPE (Objective Structured Practical Examinations)

Some Selected Response Formats

Pick n of N (using either A-type or R-type format)

Multiple X - type (Multiple T / F)

- if total question is marked as 0 up to number of choices



IDEAL's Item Analysis program:

specifying output for % ranges of unequal width.

Example: Station 4 in a 20 Station Surgery OSCE in 5th Yr MBChB

Counseling patient management: skin lesion needing excision

AT THIS STATION: PHOTOGRAPH OF A PATIENT'S CHEEK WITH A LESION & A FAX SENT BY THE PATIENT'S RELATIVE. READ THE PATIENT'S DETAILS BELOW & MAKE PHONE CALL TO SON-IN-LAW ACCORDING TO INSTRUCTIONS.

History:

Mrs Wong, 70 yr-old, has come to Outpatient Clinic because of a growth on her cheek. She lives alone, is rather forgetful, & so has asked you to ring her son-in-law who lives in Singapore. His fax has details of what he feels he needs to know so that the family can advise Mrs Wong appropriately.

Telephone conversation [10 marks]

Station 4: Scoring Instructions

1. **Introduction:** 1 mark [details not on this slide]

   Candidate introduces him/herself & clarifies s/he is looking after Mrs Fung.

2. **Is this a cancer?** 2 marks [scoring details below]

   Mrs Wong has a typical seborrheic keratotic lesion which is benign; common among old people. Morphologically are neoplasms with variable melanin pigmentation.

   Score 2: correct diagnosis & conclusion all expressed in lay language
   Score 1: reasonable alternative explanation but conveys same message
   Score .5: misleading answer given inaccuracies and poor explanation
   Score 0: meaningless information & poor communication

3. **Does it have to be removed?** 1 mark [scoring details . . .]
4. **What would happen if not removed?** 1 mark [scoring details . . .]
5. **Would she need to be hospitalized?** 1 mark [scoring details . . .]
6. **Would there be a scar?** 1 mark [scoring details . . .]
7. **Inquire about the patient's use of aspirin?** 1 mark [scoring details . . .]

IA Report for Station 4: Unequal % Ranges

% Range: unequal ranges of performances base on school's cut points

ITEM 4:	DIF=0.798		CORR= 0.408			CR_R= 0.213	(95% CON= 0.058, 0.358)	
GROUP	N	INV	range:	1	2	3	4	5	Overall
TOTAL	156	0		0.01	0.03	0.16	0.26	0.53	0.80
HIGH	45	0		0.00	0.00	0.05	0.24	0.71	0.86
MID	69	0		0.01	0.01	0.12	0.30	0.55	0.82
LOW	42	0		0.02	0.10	0.34	0.22	0.32	0.69
DISCRIMINATING POWER:				-0.02	-0.10	-0.29	0.02	0.39	0.17

Station 4: Counseling patient management: skin lesion needing excision

Comparing Reports: Quintiles & Unequal % Ranges

Example: A Choice of Unequal Percentage Ranges

% Range	< 40%	40-49%	50-65%	66-85%	> 85%	
	1	2	3	4	5	Overall
TOT	0.01	0.03	0.16	0.26	0.53	0.80
HI	0.00	0.00	0.05	0.24	0.71	0.86
MID	0.01	0.01	0.12	0.30	0.55	0.82
LOW	0.02	0.10	0.34	0.22	0.32	0.69
DIS:	- 0.02	- 0.10	- 0.29	0.02	0.39	0.17

These values

less these

equal these

Interpreting IA Output for Station 4

DIF=0.798:

average score out of maximum value possible on the item (in this OSCE, each station had a maximum value of 10 so typical performance was 8)

CORR= 0.404, CR_R= 0.214 (95% CON= 0.058, 0.359)

correlation coefficient is between scores on this station & the whole OSCE; (corrected correlation is between this station & rest of the 19 stations);

size of correlation is comparatively good, indicating station tends to assess same construct that the overall OSCE measures (presumably clinical skill)

Station 4: Counseling patient management: skin lesion needing excision

Interpreting IA for Item 4

(cont'd)

- **Station discriminates between hi & low performance groups**
  - High Grp
    - increasing proportions of good candidates are represented as one moves from the cut point to high score categories [desired]
  - Low Grp
    - at & below cut point discrimination indices are negative [desired]
  - Discrimination Power
    - usually increases from low to high performance categories [desired]

Given helpful scoring scheme, markers able to discriminate more easily & probably more accurately

Interpreting IA for Station 4
Based on **Percentage Range** Performances

- Thus, Mean, Correlation & Discrimination Power all indicate station was okay
- Educational conclusion: skill has been adequately taught and/or learned
  - Only 1% are clear failures (2/3's of these were in low group)
  - Another 3% are borderline failures (10/11's were in low group)
  - All high (100%) and almost all (97+%) in middle groups were above the cut point
- This station measures what the overall OSCE measures

Station 5: Breaking Bad News: family of terminal cancer patient

YOUR PATIENT'S RELATIVE IS AT THIS STATION. PATIENT UNDERWENT AN EXPLORATORY LAPAROTOMY AT WHICH AN INOPERABLE CANCER OF THE STOMACH WAS CONFIRMED BY BIOPSIES TAKEN FROM LIVER & STOMACH.

RELATIVE IS PATIENT'S SON WHO WISHES TO KNOW NOW WHAT TO EXPECT. RELATIVE WILL OPEN THE CONVERSATION.

Patient is 58 yr-old male referred by his doctor to Surgical Outpatient Clinic with epigastric pain, anorexia, considerable weight loss & general weakness; admitted soon after for investigation & eventual laparotomy. As a result of these investigations family already has been told that tumour was advanced & there would be nothing gained from an operation; but family insisted patient be given a chance. Unfortunately, results of the past investigations were confirmed at operation.

Scoring Instructions:

8/10 – Excellent; 6/10 – Pass; 4/10 – Inadequate or weak

IA Report for Station 5: Unequal % Ranges

Unequal Percentage Ranges

ITEM 5: DIF=0.658, CORR= 0.234, CR_R= 0.078 (95% CON= - 0.08, 0.232)

GROUP	N	INV	range:	1	2	3	4	5	Overall
TOTAL	156	0		0.00	0.12	0.47	0.41	0.00	0.66
HIGH	45	0		0.00	0.10	0.34	0.56	0.00	0.69
MID	69	0		0.00	0.11	0.49	0.41	0.00	0.66
LOW	42	0		0.00	0.17	0.56	0.27	0.00	0.62
DISCRIMINATING POWER:				0.00	-0.07	-0.22	0.29	0.00	0.07

Station 5: Breaking bad news: terminal cancer

IA Report for Station 5: Unequal % Ranges

% Range	< 40%	40-49%	50-65%	66-85%	> 85%	
	1	2	3	4	5	Overall
TOT	0.00	0.12	0.47	0.41	0.00	0.66
HI	0.00	0.10	0.34	0.56	0.00	0.69
MID	0.00	0.11	0.49	0.41	0.00	0.66
LOW	0.00	0.17	0.56	0.27	0.00	0.62
Discrim =	- .00	- .07	- .21	.30	.00	.07

Diff = .658, CR_R = .078 [CI = - 0.08 - + 0.23]

Station 5: Breaking bad news: terminal cancer

Interpreting Station 5: Breaking Bad News: Cancer

- Mean (66%) is acceptable and well above cut point
- Correlation (.08, CI = -.08 to +.23) is not different from zero, indicating this
- communication station measures something independent to that of the overall OSCE
- Discrimination is not very good (.07 overall)
  - Overall 12% fail the station; 45% of the top performers are in bottom 3 performance categories
  - No one is excellent & none are gross failures (at least as marked)

  Why is this communication station so different than previous one?
    - Is station mismatched to skill level of students and/or the teaching that had been provided?
    - Does this station fail to simulate communicating bad news?
    - Is there an inherent language problem?
    - *Was the scoring guideline or marker inadequate?*

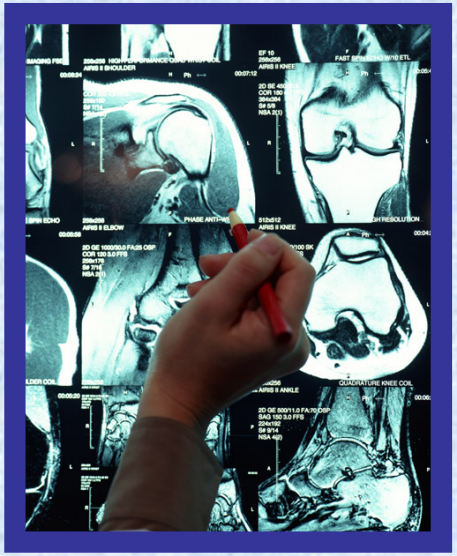
Final Comment on Item Analyses

Begin with the judgmental method & ensure responses to following questions are positive:

Is content and process contained in the item relevant?

Is item properly structured?

Is item free of bias?



If response to any question is negative, take corrective measures.

Then Consider the Item Statistics

For Items Marked in Binary fashion (most selected response)

Is item of appropriate difficulty?

Is item's & total test scores positively correlated?

Is discrimination index for the best answer positive?

Are the discriminating indices for the distracters negative?

Are each of the distracters functional (attractive to at least some candidates)?

Are the performance characteristics consistent with the purpose of your assessment (i.e., formative/summative or normative/criterion referenced)?

Then Consider the Item Statistics

For Items Marked in Non binary fashion
(most constructed response & OSCE/OSPE formats)

Is item of appropriate difficulty?

Is item/station's score & total test scores positively related?

Are discrimination levels for categories above the cut point positive?

Are discrimination levels below the cut point negative?

Are these performance characteristics consistent with the purpose of the assessment (criterion/norm referenced or summative/formative)?

References:

Case S.M. & Swanson D.B. (2001).

Constructing written test questions for the basic and clinical sciences. **Philadelphia: National Board of Medical Examiners.**

Osterlind S.J. (1998).

Constructing test items. **Boston: Kluwer Academic Publishers.**

Precht, D., Hazlett, C., Yip, S. & Nicholls, J. (2005)

IDEAL – HK™ Item Analysis Users' Guide: Selected and Constructed Item Formats. **Hong Kong: Candor Production Ltd**

Item Analyses for Items Scored as Right (1) or Wrong (0)

Small Group Exercises

Use the item analysis to flag possible problems

with the following 9 Paediatric A-type items

Notes for Facilitators

These items were administered to 4th year medical students as part of the summative assessments that were administered after candidates had rotated through a 11 week module in Paediatrics at the Chinese University of Hong Kong. In any given academic year, CUHK runs four rotations; some of the nine items were administered in module 1, others in modules 2 to 4, with some used in two modules. (This information is relevant if any in your small group ask why the sample sizes are not the same throughout the nine A-type items.)

The items are not necessarily representative of the entire assessment, but were selected to illustrate how item analysis can help an assessment committee to review the adequacy of items used in an examination. In an effort to be helpful I have provided some commentary for each item and its corresponding item analysis. Obviously, regard my comments with a jaundice eye as Ibrahim, who is the content expert in this discipline, may wish to negate or add to the commentary in terms of diagnosing why certain items are inadequate and/or how to properly revise the item in order to address its inadequacy.

As a teaching principle, however, the registrants should be able to (1) flag poor items using the psychometric data, (2) recognize the data identify if there is a problem but do not diagnosis the exact nature of the problem (& thus content expertise is req'd) and (3) should utilize item writing criteria (that John and Issy will cover in teaching A and R type items) in order to revise a poorly performing item.

Question 1; Keyed Answer: B

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 1: DIF=0.683 RPB= 0.045 CRPB= -0.022 (95% CON= -0.328, 0.287)

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
DISCRIMINATING POWER					0.10	-0.07	-0.03

Question 2; Keyed Answer: 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 2: DIF=0.31 RPB= -0.266 CRPB= -0.334 (95% CON= - 0.582,-0.029)

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
DISCRIMINATING POWER					-0.32	0.39	-0.08	0.00	0.01

Examples of a revised stem:

What developmental factor is most likely to contribute to cardiopulmonary arrest in paediatric patients? *[For the options, generate a list of developmental factors, one of which is epidemiologically the most likely to lead to an arrest.]*

What is the most likely outcome of a cardiopulmonary arrest in paediatric patients? *[For the options, generate a list of outcomes which occur in paediatrics who arrest, one of which occurs more often as supported by epidemiological evidence].*

However, both of the above revisions, while addressing the shortcomings of the original stem, probably have indeterminate answers as the correct answer would vary by confounding factors that are not provided in the stem's information; and similar to the original item, both revisions continue to encourage candidates to simply memorize clinical / epidemiological facts.

The more useful assessment approach, from a learning perspective, is to compose a scenario that describes a paediatric patient with developmental problems that are predictive of a probable forthcoming arrest, or conversely, one who is suffering from the consequences of an cardiopulmonary arrest. Then ask candidate to make a diagnosis, or, suggest the most appropriate management strategy, or, propose the most likely prognostic outcome. Candidate will have to synthesise information in the scenario in order to deduce an answer. If candidate has only memorized some related data s/he probably will not be able to answer this latter type of question.

Question 3; Keyed Answer: 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 3: DIF=0.500 RPB= -0.191 CRPB= -0.276 (95% CON= -0.547, 0.048)

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
DISCRIMINATING POWER					-0.30	0.00	0.09	0.20

Question 4: Keyed Answer: 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 4: DIF=0.421 RPB= -0.014 CRPB= -0.098 (95% CON= -0.405, 0.229)

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
DISCRIMINATING POWER					0.30	-0.03	-0.26

Queston 5; Keyed Answer: 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 5: DIF=0.675 RPB= 0.585 CRPB= 0.531 (95% CON= 0.262, 0.723)

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
DISCRIMINATING POWER					-0.25	-0.25	-0.17	0.67	0.00

Question 6; Keyed Answer: D

A 4-month-old infant has recurrent episodes of pneumonia. He is irritable during feeds with frequent large volume vomiting. Blood picture reveals hypochromic microcytic anaemia.

What is the most likely diagnosis?

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

ITEM 6: DIF=0.421 RPB= 0.473 CRPB= 0.390 (95% CON= 0.080, 0.631)

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
DISCRIMINATING POWER					0.09	-0.57	-0.16	0.64	0.00

Question 7; Keyed Answer: B

A two year old child presents with generalized tonic clonic seizures at the Emergency department.

What is the most important on-site investigation?

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

ITEM 7: DIF=0.789 RPB= 0.588 CRPB= 0.519 (95% CON= 0.239, 0.719)

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
DISCRIMINATING POWER					-0.50	0.75	-0.25

Question 8; Keyed Answer: D

What is the most frequent complication of congenital rubella?

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

ITEM 8: DIF=0.780 RPB= 0.411 CRPB= 0.352 (95% CON= 0.049, 0.595)

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
DISCRIMINATING POWER					-0.33	-0.08	0.00	0.42

Question 9; Keyed Answer: 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 good in between the coughing spells. 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.

What is the most likely diagnosis?

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

ITEM 9: DIF=0.732 RPB= 0.490 CRPB= 0.426 (95% CON= 0.136, 0.648)

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
DISCRIMINATING POWER					0.58	-0.42	-0.17

Item Analyses for Items Scored as Continuous Data (e.g., 0 to 100 including decimals such 7.5 or 20.75)

Small Group Exercises

Use the relevant item analysis to flag possible problems with the following 2 Surgical OSCE Stations administered in 5th (final) Yr

Assume the cut point (passing mark) is 50% as used in running the respective item analyses for unequal % performance groups (i.e., less than 40%, 40 to 50%, 50 to 65%, 65 to 85%, >85%)

(clear fails) (debatable fails) (debatable passes) (clear passes) (honors)

Station 17: Breast Examination: Lumps

THERE ARE FIVE BREAST SIMULATIONS (A – E). PALPATE & MAKE A DIAGNOSIS FOR EACH. STATE IF BREAST IS NORMAL OR IF ABNORMAL, INDICATE MOST LIKELY DIAGNOSIS BASED ON HISTORY & WHAT YOU HAVE PALPATED.

Each Breast Exam [2 marks]

The site shall be described correctly using either clock face method or 4 quadrant placement always with distance from areolar/nipple estimate in cm. Measurement shall be $\pm$ 2 cm accurate. Size & consistency shall be reasonably accurate.

Judgment should be exercised if complete (correct) answer isn't given & instead alternative or inaccurate / incorrect answer is given.

☐ FAIL	☐ BORDERLINE	☐ PASS

Item Analysis Report for Station 17

Unequal % Ranges

ITEM 17: DIF=0.700 CORR= 0.359 CR_R= 0.175 (95% CON= 0.018, 0.323)

GROUP	N	INV	range:	1	2	3	4	5	Overall
TOTAL	156	0		0.04	0.06	0.18	0.48	0.24	0.70
HIGH	45	0		0.00	0.02	0.12	0.44	0.41	0.77
MID	69	0		0.04	0.05	0.18	0.51	0.22	0.70
LOW	42	0		0.07	0.10	0.24	0.46	0.12	0.63
DISCRIMINATING POWER:				-0.07	-0.07	-0.12	-0.02	0.29	0.14

Station 17: Breast Examination: Lumps

Summarized IA Report for Station 17: Unequal % Ranges

% Range	< 40%	40-49%	50-65%	66-85%	> 85%	
	1	2	3	4	5	Overall
TOT	0.04	0.06	0.18	0.48	0.24	0.70
HI	0.00	0.02	0.12	0.44	0.41	0.77
MID	0.04	0.05	0.18	0.51	0.22	0.70
LOW	0.07	0.10	0.24	0.46	0.12	0.63
Discrim =	- .07	- .07	- .12	- .02	.29	.14

Diff = .70, CR_R = .175 [CI = .02 - .32]

Station 17: Breast Examination: Lumps

Station 11: CVP Setup and Reading

AT THIS STATION, AN OBSERVER - EXAMINER & VARIOUS EQUIPMENT, INCLUDING A MANIKIN FOR USE IN DEMONSTRATING SETTING UP OF A CVP LINE & TAKING OF AN ACTUAL READING. PROVIDE COMMENTARY ABOUT WHAT & WHY YOU ARE DOING. EXAMINER WILL ASK QUESTIONS

Assemble items for taking CVP [2 marks]

Measure CVP of the manikin [5 marks]

Equipment on table:

1. Drip (IV fluid) stand
2. Empty 500 ml containers of (a) saline, (b) colloid solution & (c) water
3. IV giving sets (a) JMS (simple) (b) pump set & (c) micro-drip set
4. CVP manometer set
5. 14 G IV cannula
6. 3 way tap set

CVP set up used with manikin:

1.Second drip (IV fluid) stand 2.Manikin on couch

3.Bag of normal saline (500 ml) connect via JMS IV set to manometer set

4.Plastic measuring cylinder of water with connecting tubing to simulate manikins CVP

Scoring for Station 11

1. Assembly of manometry set	2 marks
2. Positioning of manikin	1 mark
3. Demonstrate the zero position	1 mark
4. Manometre in vertical position (along drip stand)	1 mark
5. Prime and read correctly	2 marks
6. QUESTION 1: What is normal CVP?	1 mark
7. QUESTION 2: What are 4 *causes* of abnormally high CVP reading?	2 marks

Place tick in the box given your overall impression.

☐ FAIL	☐ BORDERLINE	☐ PASS

IA Report for Station 11: Unequal % Ranges

% Range	< 40%	40-49%	50-65%	66-85%	> 85%	
	1	2	3	4	5	Overall
TOT	0.01	0.04	0.14	0.41	0.39	0.77
HI	0.00	0.00	0.00	0.39	0.61	0.87
MID	0.00	0.03	0.19	0.41	0.38	0.77
LOW	0.05	0.12	0.20	0.44	0.20	0.66
Discrim =	- .05	- .12	- .20	- .05	.41	.20

Diff = .77, CR_R = .275 [CI = .12 - .41]

Station 11: CVP Setup & Reading

Station 13: Examination of Gait and Trendelenberg Sign

AT THIS STATION THERE IS A PATIENT & OBSERVER EXAMINER.

ASSESS GAIT & PERFORM A TRENDELENBURG'S TEST. GIVE A RUNNING COMMENTARY ABOUT WHAT & WHY YOU ARE DOING. SUMMARISE YOUR FINDINGS.

Patient is middle aged with intermittent right hip pain for last 2 yrs. Walking tolerance is reduced to 1 hr & there is limping from time to time, especially after prolonged walking. Patient is otherwise well.

Clinical examination [10 marks]

Place a tick in the box with your overall impression.

☐ FAIL	☐ BORDERLINE	☐ PASS

IA Report Station 13: Unequal % Ranges

% Range	< 40%	40-49%	50-65%	66-85%	> 85%	
	1	2	3	4	5	Overall
TOT	0.24	0.22	0.29	0.19	0.04	0.50
HI	0.07	0.17	0.37	0.29	0.10	0.58
MID	0.26	0.27	0.27	0.18	0.03	0.48
LOW	0.39	0.20	0.27	0.12	0.02	0.45
Discrim =	- .32	- .02	- .10	.17	.07	.13

Diff = .50, CR_R = .082 [CI = - 0.08 - + 0.24]

Station 13: Examination of Gait & Trendelenberg Sign