



# Measuring scientific reasoning: Construct Validation of the Primary Scientific Reasoning Test using Rasch modelling

Rasch User Day  
31 March 2017

Diana Ng  
Supervisors: Prof Jo-Anne Baird and Dr Josh McGrane

# Outline of Presentation

- Brief outline of overall research
- Interest in polytomously scored two-mark structured questions
- Understanding threshold order
- Item exemplars from PSRT
- Results of rescored items with disordered thresholds
- Discussion

# Research Study

- Scientific reasoning and its importance
- Current assessments of scientific reasoning
- Over-arching research question :  
**What is the construct validity of the Primary Scientific Reasoning Test?**

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed for proposed uses of tests.

*(Standards for Educational and Psychological Testing , 2014, p. 11)*

- Multi-method, sequential mixed research design
- 431 mixed-ability 12 year-old primary pupils from 6 Singapore schools



# Primary Scientific Reasoning Test (PSRT)

- Items test ability to apply understanding of *three knowledge types*, using *three science practices*
- Descriptors from past and current curricular initiatives and literature in science education
- Individually administered, paper and pencil test
- 5 booklets in rotation design
- 30 questions in total, each of 1 to 4 items
- Items in multi-item question could share same stem or stimulus, but each item has its own testing objective
- Testing duration of each booklet is 1 hour
- Each pupil is tested with 1 booklet
- Variety of response formats

# Primary Scientific Reasoning Test (PSRT)

## Types of Knowledge

Content Knowledge

Procedural Knowledge

Epistemic Knowledge

## Levels of Cognitive Demand

High

Medium

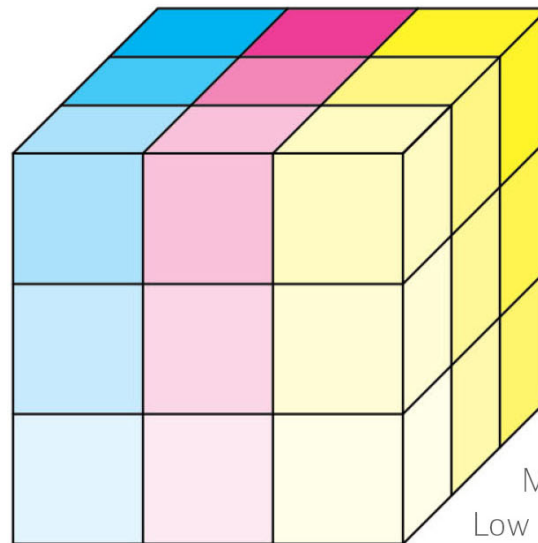
Low

## Science Practices

Giving Scientific  
Explanation

Designing & Evaluating  
Investigation

Interpreting & Analysing  
Data & Evidence



# PSRT Summary Statistics

## ITEM - PERSON INTERACTION

### ITEMS

	Location		Fit Residual
Mean	0.0000	Mean	-0.1910
Std Dev	1.4557	Std Dev	1.3093
Skewness	-0.2515	Skewness	0.8408
Kurtosis	-0.1304	Kurtosis	0.7271
	Correlation		0.1531
	[location/stdResidual]		

### PERSONS

	Location		Fit Residual
Mean	1.2298	Mean	-0.1980
Std Dev	1.0770	Std Dev	0.7898
Skewness	-0.2293	Skewness	0.4760
Kurtosis	0.2182	Kurtosis	0.8918
	Correlation		-0.1680
	[location/stdResidual]		

☒ Include  
Extremes

N = 431

## ITEM - TRAIT INTERACTION

Total - Item Chi Square	749.3218
Degrees of Freedom	600
Chi Square Probability	0.000030

## PERSON RELIABILITY INDICES

PerSepIdx:	Item98b
* with extms	0.86098
* NO extms	0.86098
CoeffAlpha	N/A
* with extms	N/A
* NO extms	N/A

[Coefficient Alpha not applicable with missing data]

## LIKELIHOOD RATIO TEST

Analysis	Likelihood	ChiSq
anaName1		DegF
anaName2		Prob

## SEPARATION INDICES

	Item	Person/Item
Index	0.97224	0.89320
Variance	1.94060	703.73580
Error	0.05387	75.16145

## POWER OF ANALYSIS OF FIT

This display is intended as a guide  
ONLY and should be used in  
conjunction with other analysis  
indicators

Details

Excellent
Good
Reasonable
Low
Too Low

< Display Control

File Text Format

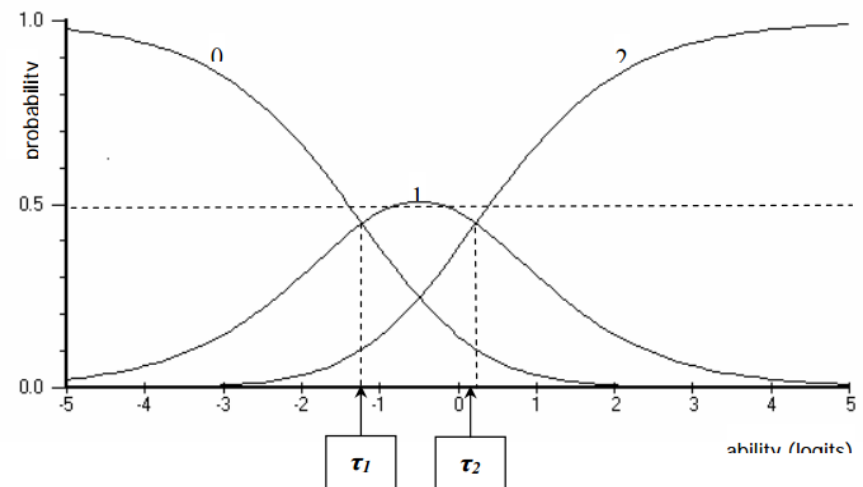
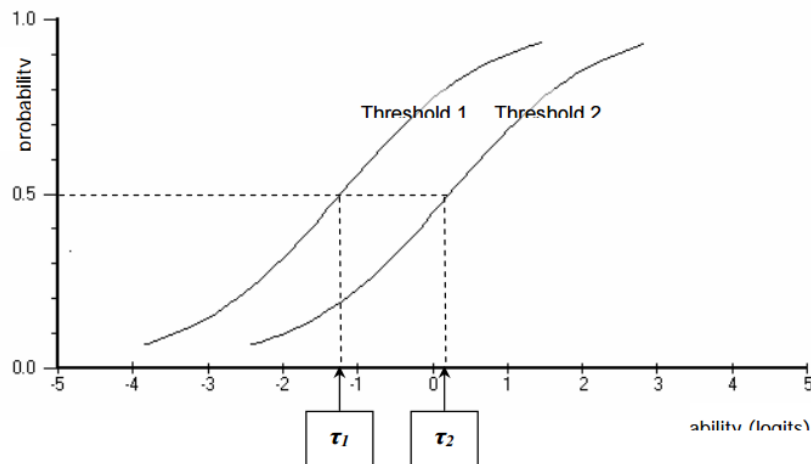
☒ Fixed ☐ Tab Delimit

Save

## Rationale for Ordered Response Categories

*In a design in which the objects of measurement are judged experimentally independently as successful or unsuccessful in meeting the requirements of successive, contiguous categories on a continuum, the categories will be said to be empirically ordered correctly if the relative difficulties of achieving a success increases with the intended ordering of the categories.*

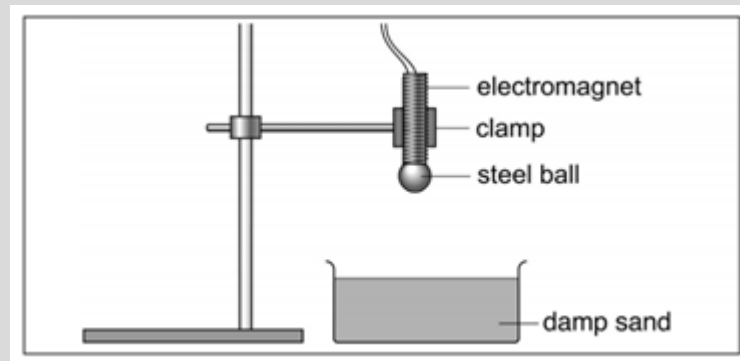
(Andrich, 2009)



van Wyke and Andrich (2006)

## Item Exemplar – Ordered Threshold

Jack said that the ball could be dropped using an electromagnet instead of dropping it by hand. Explain why this would make the results more accurate.



[2 marks]

Framework category	
Knowledge Type	Procedural
Practice	Evaluation of the planning and implementation of the scientific process
Cognitive Demand	High

Item Characteristic	
Location, SE	2.166, 0.148
Fit Residual	0.880
Chi sq (prob)	5.805 (0.445)

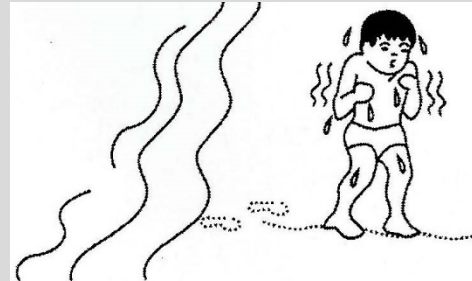


## Item Exemplar – Ordered Threshold

Eric went for a swim. When he came out of the water, he felt cold

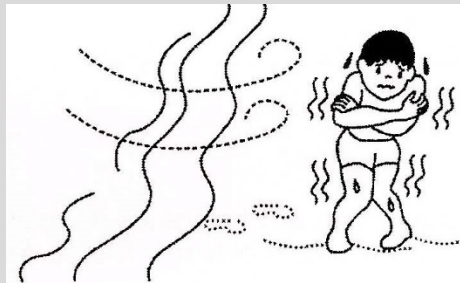


his body was dry before swimming



his body was wet after swimming

At that moment, a strong wind blew, and he felt even colder.



a strong wind blew over his wet body

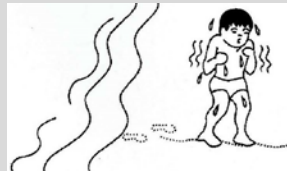
Explain why he felt even colder when a strong wind blew. [2 mark]

## Item Exemplar – Ordered Threshold

Eric went for a swim. When he came out of the water, he felt cold

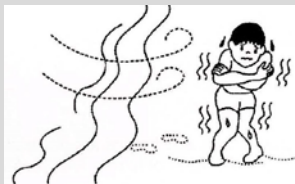


his body was dry before swimming



his body was wet after swimming

At that moment, a strong wind blew, and he felt even colder.



a strong wind blew over his wet body

Explain why he felt even colder when a strong wind blew. [2 mark]

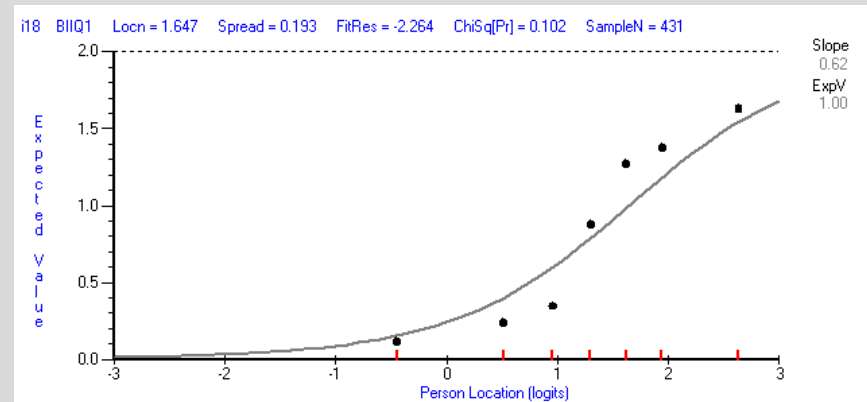
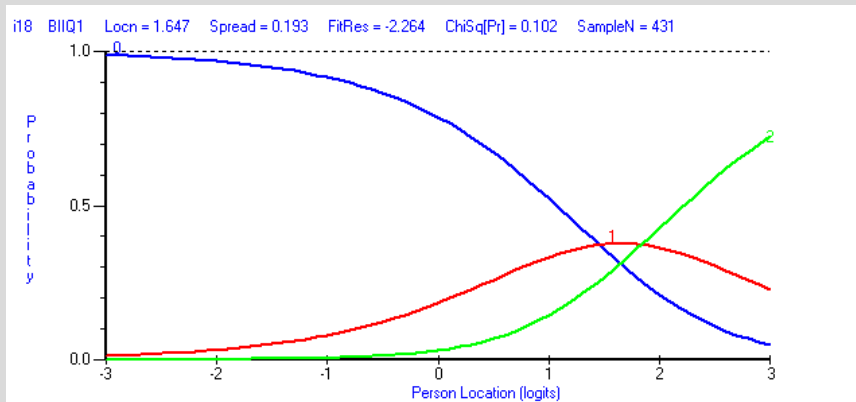
First mark point:

Water gained heat from his body  
(and evaporated)

Second mark point:

The presence of strong wind  
speeds up/increases the rate of  
evaporation

# Item Exemplar – Ordered Threshold

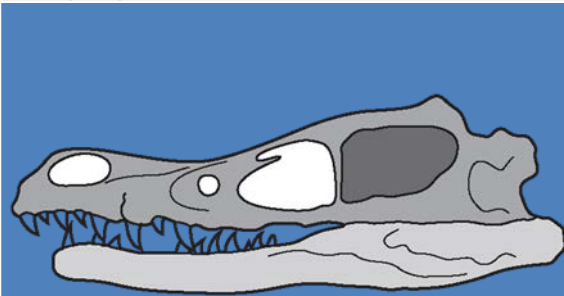


Framework category	
Knowledge Type	Content
Practice	Recall and apply appropriate scientific knowledge
Cognitive Demand	Medium

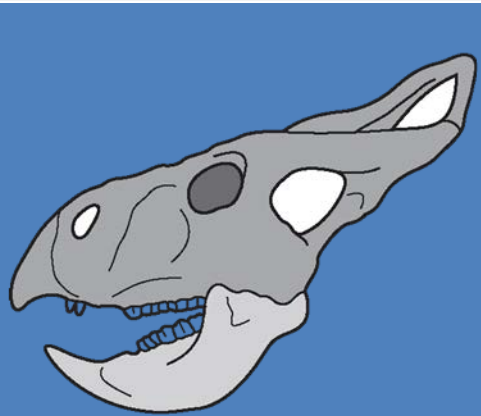
Item Characteristic	
Location, SE	1.647, 0.112
Fit Residual	-2.264
Chi sq (prob)	10.596 (0.1017)

## Item Exemplar – Disordered Thresholds

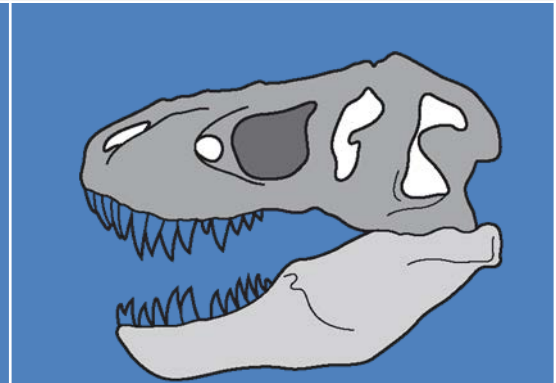
The skulls of three animals which lived on Earth millions of years ago are shown



animal X



animal Y



animal Z

Based on the shapes of the teeth found in the lion and the giraffe, draw a possible **food web** involving the following four organisms: plant, animal X, animal Y and animal Z in the space below.

[2 marks]

# Item Exemplar

Framework category	
Knowledge Type	Content
Practice	Recall and apply appropriate scientific knowledge
Cognitive Demand	High

**Two correct inter-related food chains in food web**  
**2 marks**

**One correct inter-related food chains in food web**  
**1 mark**

**Additional marking guidance:**  
**Producer must be present**

**If three food chains given, of which one is incorrect, response will be penalised a mark**

# Item Exemplar – Disordered Threshold

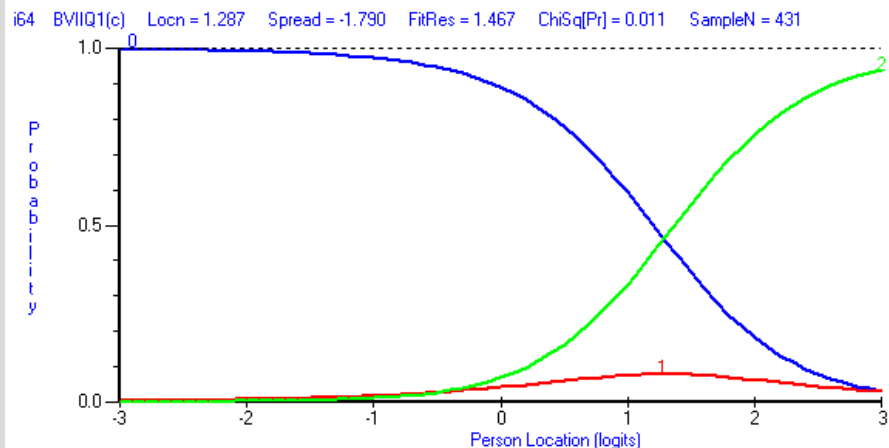
## Examples of errors

Plant  $\rightarrow$  animal Y, animal X  $\rightarrow$  animal Z

0 mark



2 - 1 = 1 mark



## Item Characteristic

Location, SE

1.287, 0.100

Fit Residual

1.467

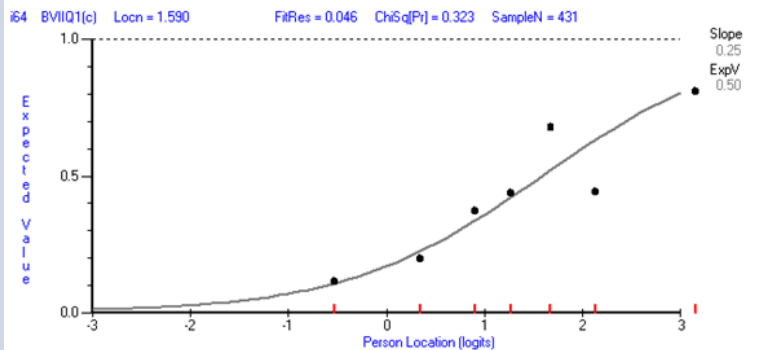
Chi sq (prob)

16.488 (0.011)

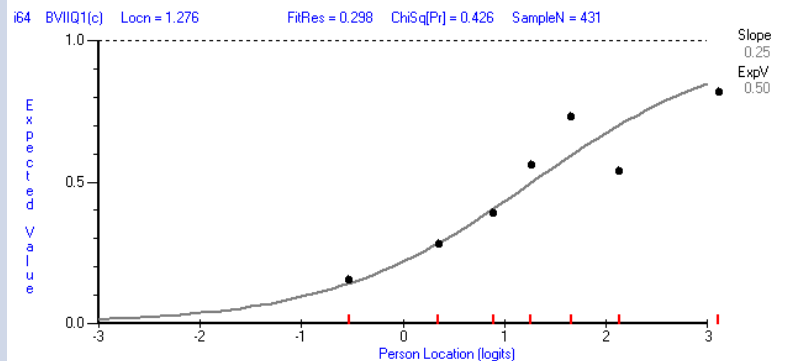


# Item Exemplar – After Re-scoring

## Rescored Mid-Category as Incorrect



## Rescored Mid-Category as Correct



Item Characteristic	
Location, SE	1.590, 0.172
Fit Residual	0.046
Chi sq (prob)	6.977 (0.323)

Item Characteristic	
Location, SE	1.276, 0.170
Fit Residual	0.298
Chi sq (prob)	5.976 (0.426)

## Discussion

- Usefulness of threshold order to understand reasoning ability
- Implications of threshold disorder
- Change marking key at rescoring item stage
- Analysis of errors and misconceptions that pupils demonstrate

[Diana.Ng@education.ox.ac.uk](mailto:Diana.Ng@education.ox.ac.uk)

Oxford University Centre for Educational Assessment