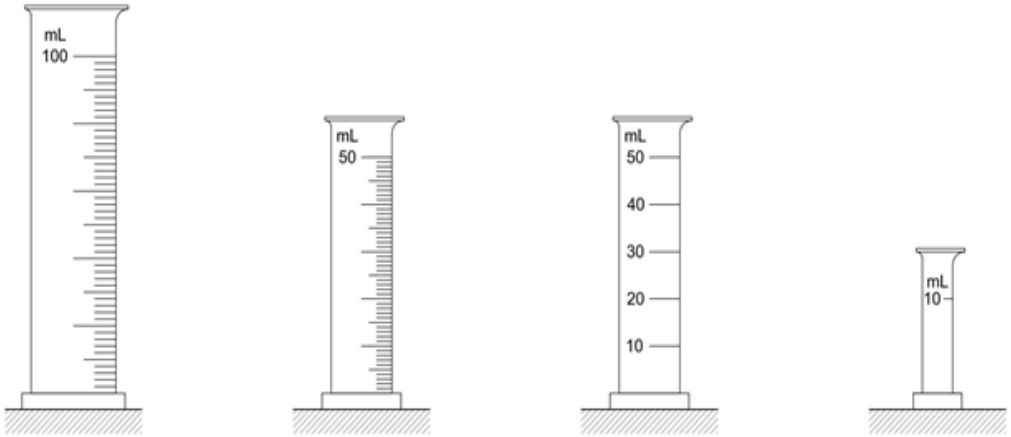
# Scientific reasoning of primary school children – Construct validation using Rasch modelling

**ERAS-APERA International Conference 2018**

**Diana Ng Yee Ping**

# What is scientific reasoning?

The diagram shows four pieces of measuring cylinders.




A B C D

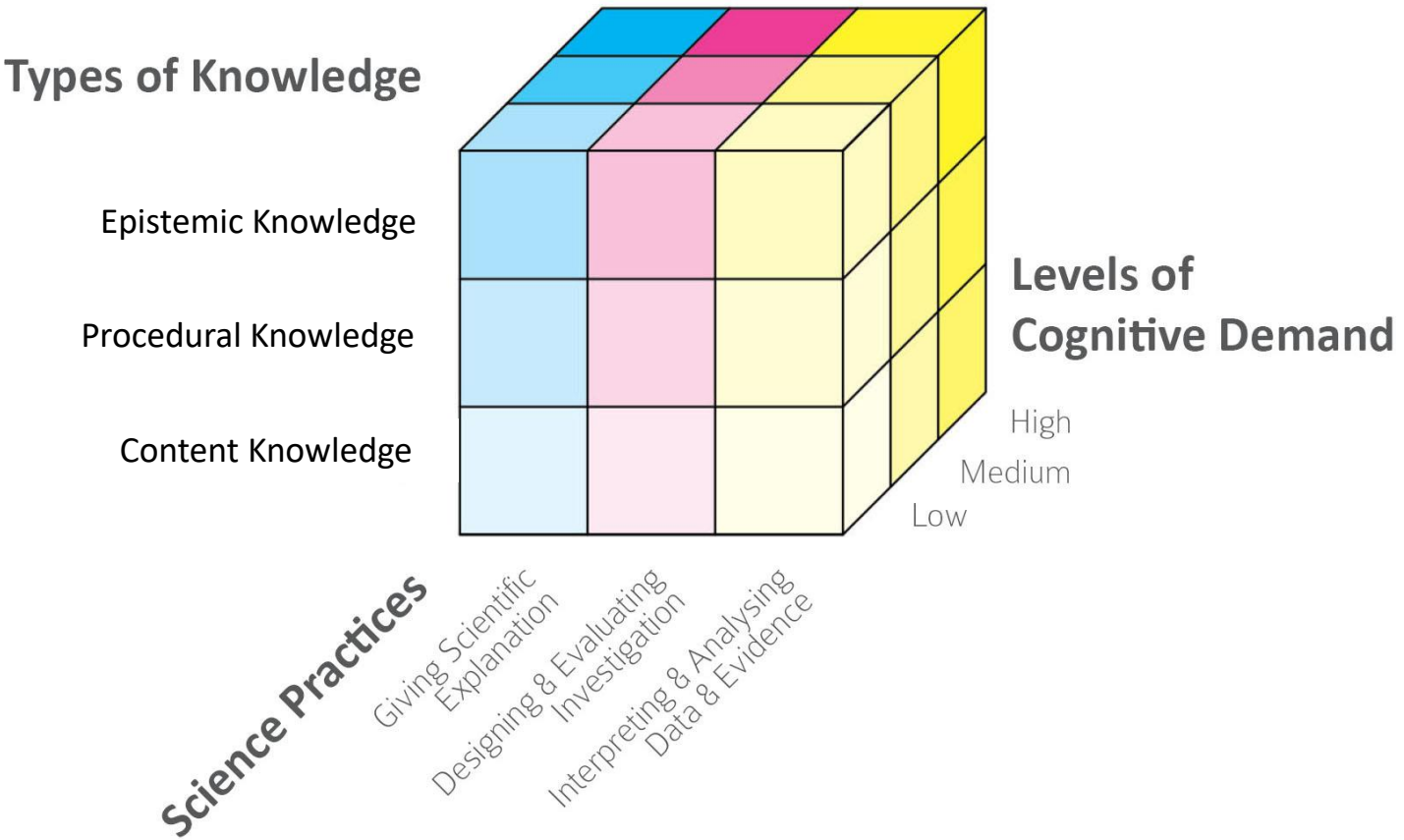
Which piece will measure 15 $cm^3$ of water most accurately? Circle your answer below.

A / B / C / D

# Presentation Outline

- Overview of Research
- *Primary Scientific Reasoning Test*
- Trait/Construct
- Item Characterisation Scheme
- Rasch Modelling
- Item Exemplars
- Discussion

# *Primary Scientific Reasoning Test (PSRT)*



## *Primary Scientific Reasoning Test (PSRT)*

- Paper and pencil test
- 30 questions in total, each of 1 to 7 items
- Total of 100 items in 5-booklet rotation design
- Pupils attempted only 1 booklet
- Testing duration of each booklet is 1 hour
- Variety of response formats
- Sample consisted of 431 primary six pupils in Singapore

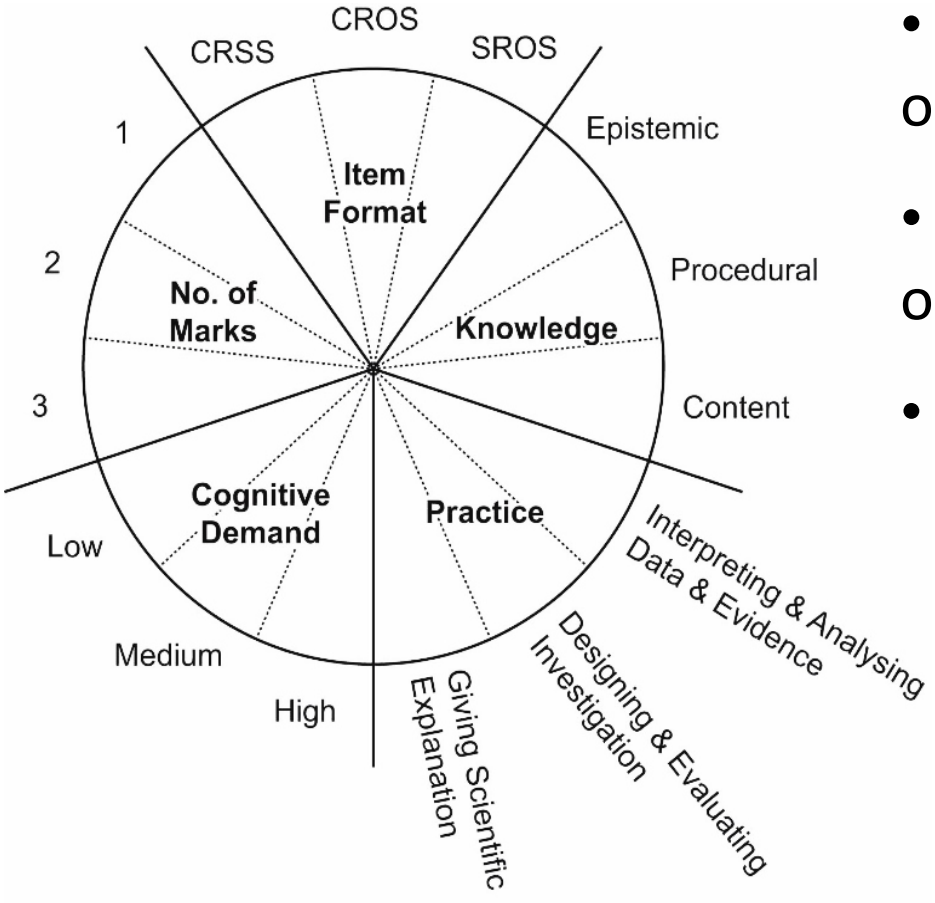
# *Primary Scientific Reasoning Test (PSRT)*

| Booklet | Anchor Block and Question Blocks | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Anchor | I | II | III | IV | | | | | | |
| 2 | Anchor | | | | IV | V | VI | VII | | | |
| 3 | Anchor | I | | | | V | | | VIII | IX | |
| 4 | Anchor | | II | | | | VI | | VIII | | X |
| 5 | Anchor | | | III | | | | VII | | IX | X |

# Validation Framework

| Design focus in phase | Validity evidence collected |
|---|---|
| **I - Initial Test**<br>**QUAL**<br>↓ | Test content for the development of the scientific reasoning construct and the identification of typical performance |
| **II - Expert Review**<br>**QUAN + qual**<br>↓ **III - Pilot Study**<br>**QUAN + QUAL**<br>↓ | Test content and response processes to initiate the development of the draft instrument |
| **IV - 2nd Expert Review**<br>**QUAL**<br>↓ | Test content to refine the draft instrument |
| **V - Main Study**<br>**QUAN + qual** | Test content, response processes, internal structure & relations to other variables to assess the construct validity of the finalised instrument |

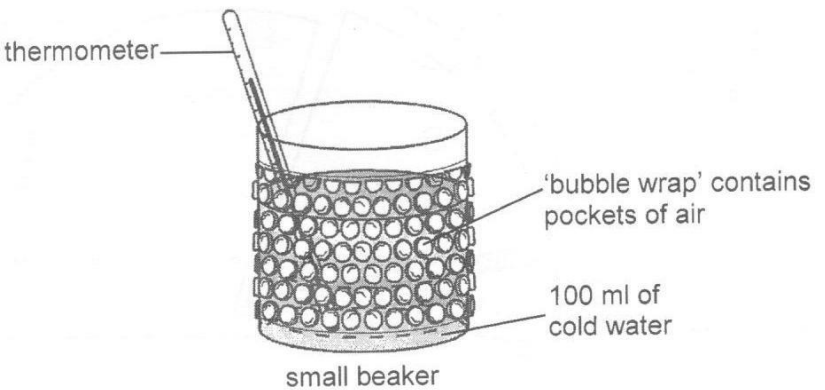
# *Primary Scientific Reasoning Test (PSRT)*




- Selected-response objective scoring (SROS)
- Constructed-response objective scoring (CROS)
- Constructive-response subjective scoring (CRSS)

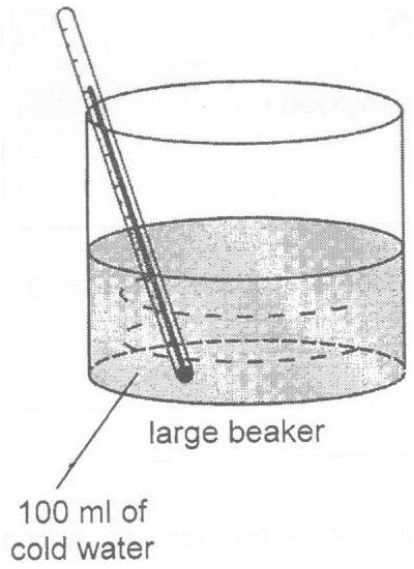
# Item Characterisation Scheme

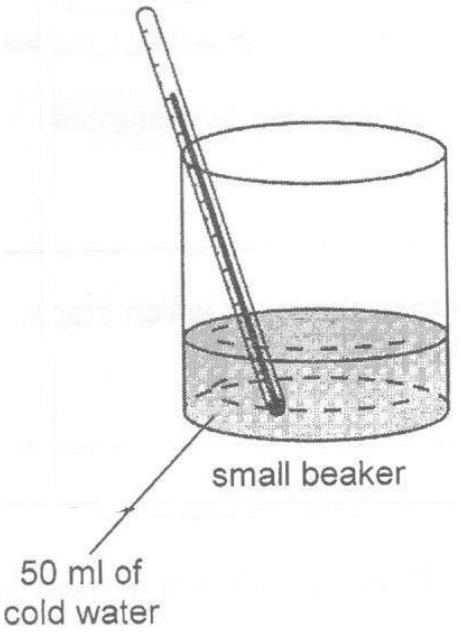
John wants to study whether the material, 'bubble wrap' affects how cold water gains heat.
The diagram below shows one set-up of his experiment.

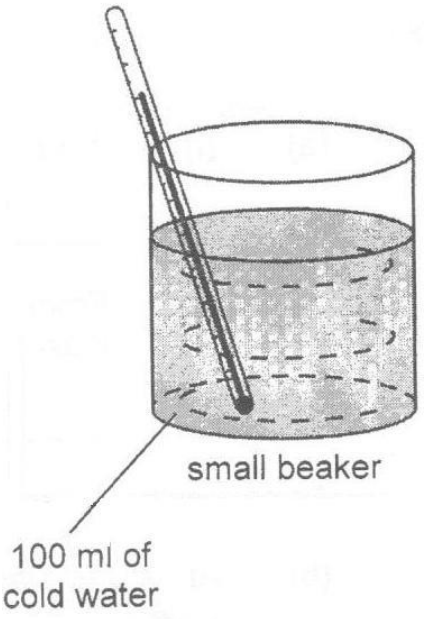


# Item Characterisation Scheme

Item 1: Which other set-up shown below must he use in his experiment?

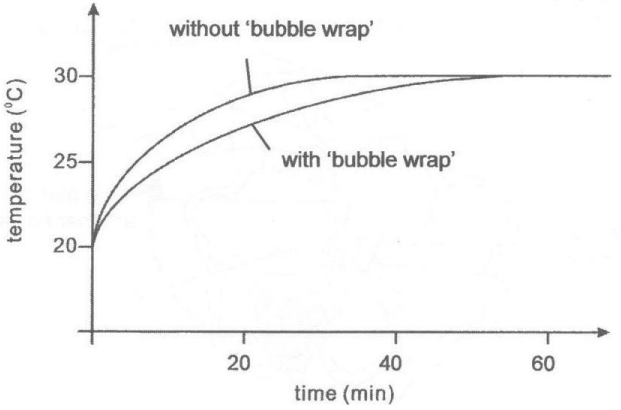


# Item Characterisation Scheme

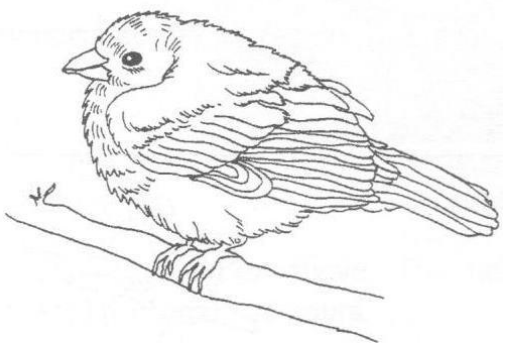
Item 2: John recorded the readings of the thermometer over time. His results are shown below.




Based on his results, John concluded that the beaker with the 'bubble wrap' gains heat more slowly. Is he correct?
Yes / No / Cannot Tell

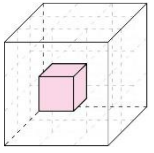
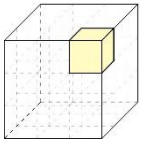
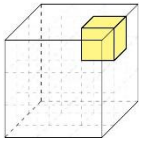
# Item Characterisation Scheme

Item 3: A bird with a thick layer of feathers is shown below. There are air pockets among the feathers.

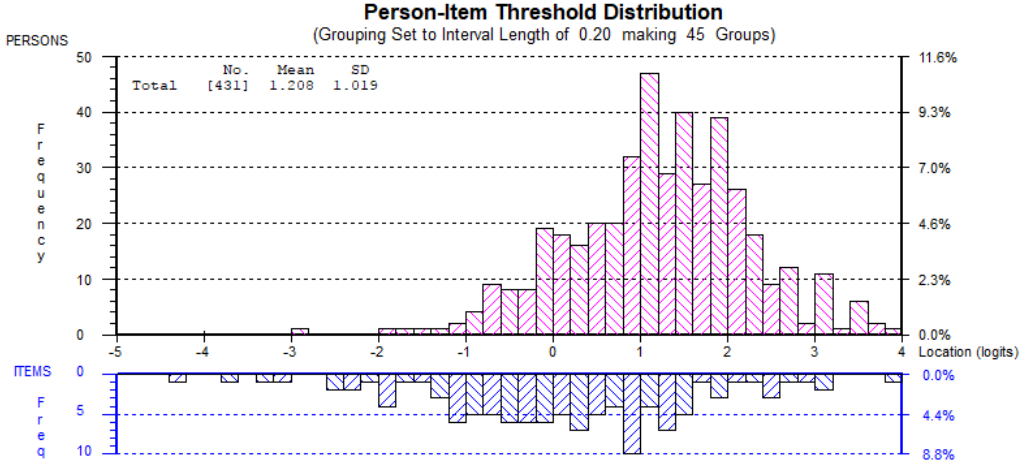


Based on the results of John's experiment, explain how the air pockets keep the bird warm in cold air.

# Item Characterisation Scheme

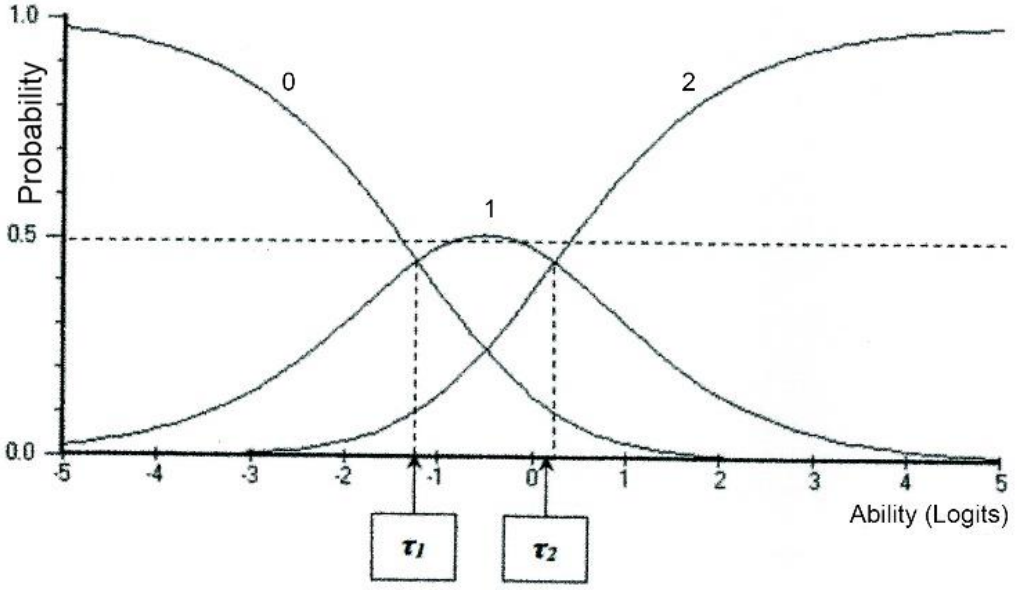
| | Item 1 | Item 2 | Item 3 |
|---|---|---|---|
| **Knowledge Type** | Procedural | Content | Content |
| **Practice** | Designing and evaluating investigation | Interpreting and analysing data and evidence | Interpreting and analysing data and evidence |
| **Cognitive Demand** | Low | Low | Medium |
| **Mark** | 1 | 1 | 1 |
| **Item format** | SROS | SROS | CRSS |
| **Grid model** |  |  |  |

# Initial Rasch Statistics of the PSRT



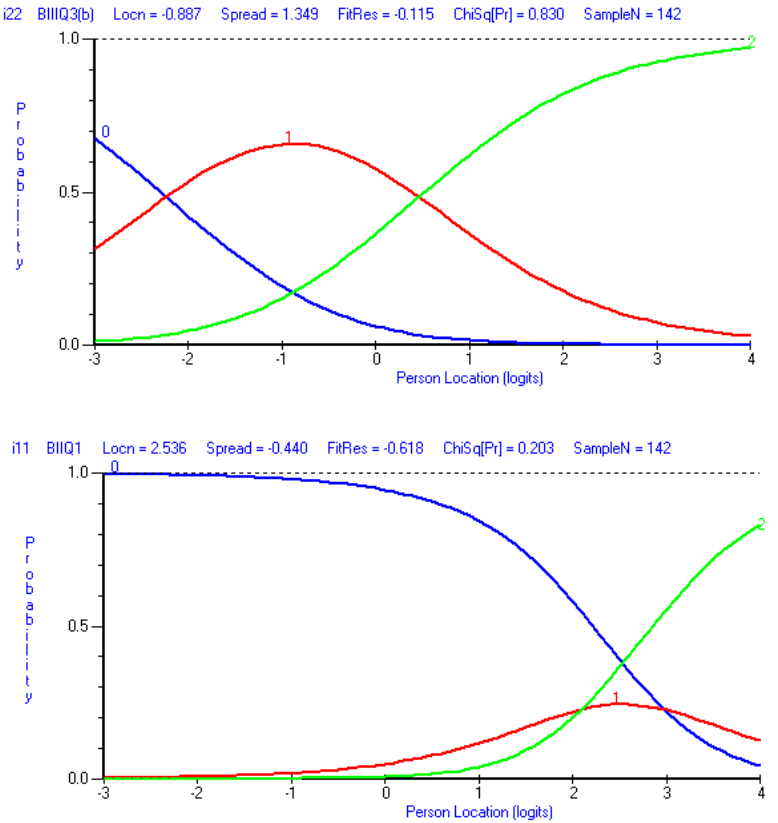

| Mean Item Location (SD) | Mean Fit Residual (SD) | Mean Person Location (SD) | Total Item Chi Square (Prob) | Person Separation Index | Power of Analysis of Fit |
|---|---|---|---|---|---|
| 0.000 (1.373) | -0.211 (1.350) | 1.208 (1.019) | 739.69 (0.000) | 0.864 | Excellent |

# Rationale for Ordered Response Categories



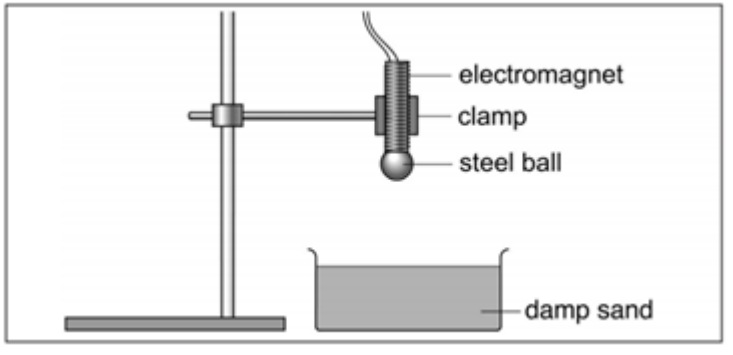

*van Wyke and Andrich (2006)*

# Rationale for Ordered Response Categories



# Item Exemplar 1 – 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.



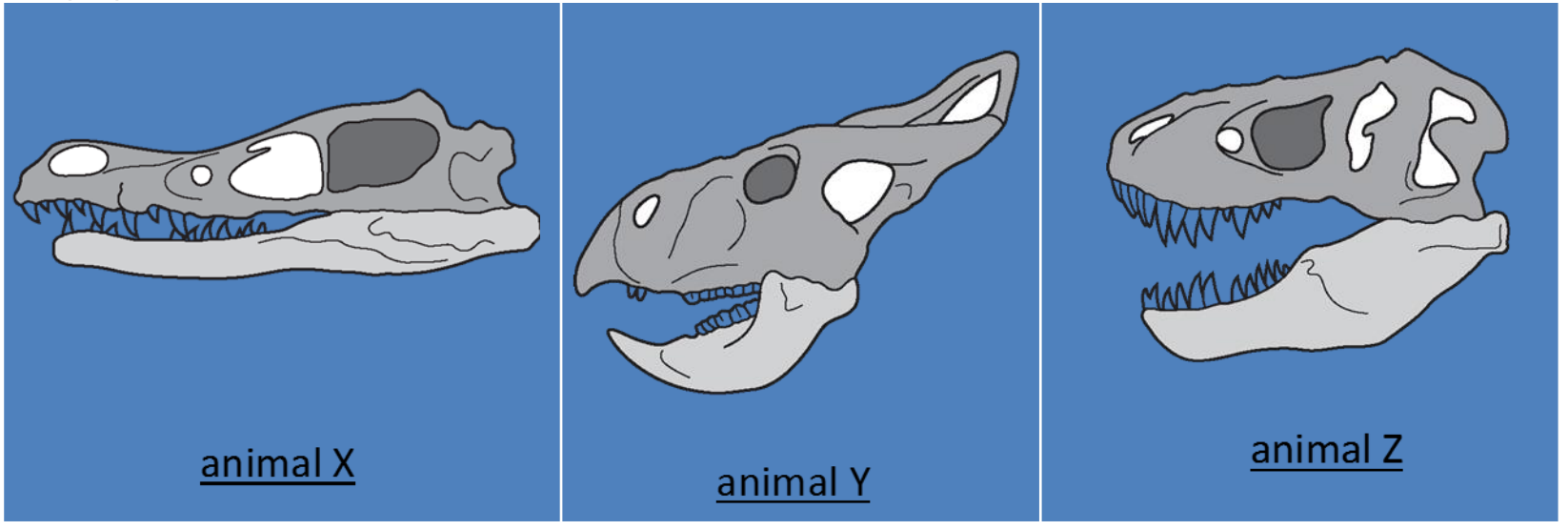

[2 marks]

| Framework category | |
|---|---|
| Knowledge Type | Procedural |
| Practice | Designing and evaluating investigation |
| Cognitive Demand | High |

| Item Characteristic | |
|---|---|
| Location, SE | 2.178, 0.148 |
| Fit Residual | 0.835 |
| Chi sq (prob) | 4.893 (0.558) |

# Item Exemplar – Disordered Thresholds

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




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]

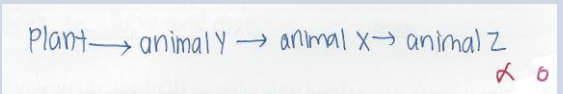
# Item Exemplar

| Framework category | |
|---|---|
| Knowledge Type | Content |
| Practice | Interpreting and analysing data and evidence |
| Cognitive Demand | High |

| | |
|---|---|
| **Two correct inter-related food chains in food web**<br>**2 marks**<br><br>**One correct inter-related food chains in food web**<br>**1 mark** | **Additional marking guidance:**<br>**Producer must be present**<br><br>**If three food chains given, of which one is incorrect, response will be penalised a mark** |

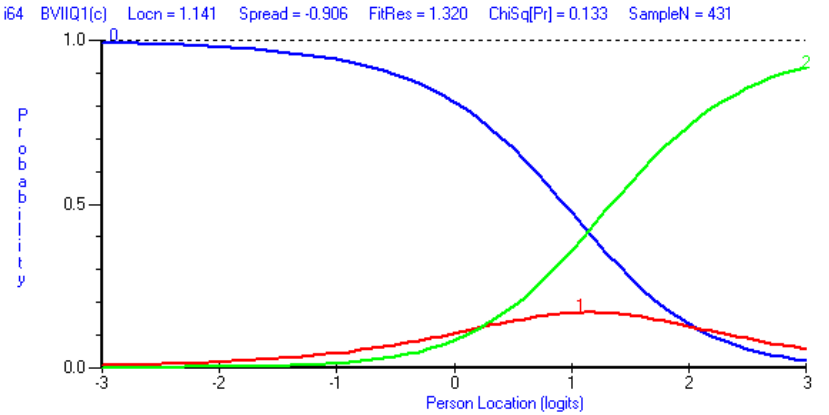
# Item Exemplar – Disordered Threshold

## Examples of errors




0 mark

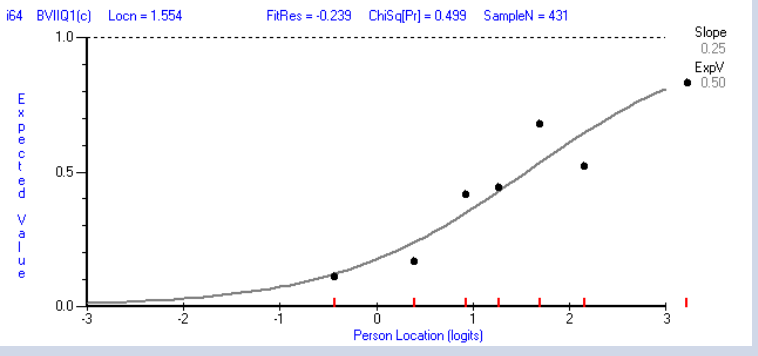

2 – 1 = 1 mark




| Item Characteristic | |
|---|---|
| Location, SE | 1.141, -0.906 |
| Fit Residual | 1.320 |
| Chi sq (prob) | 9.817 (0.133) |

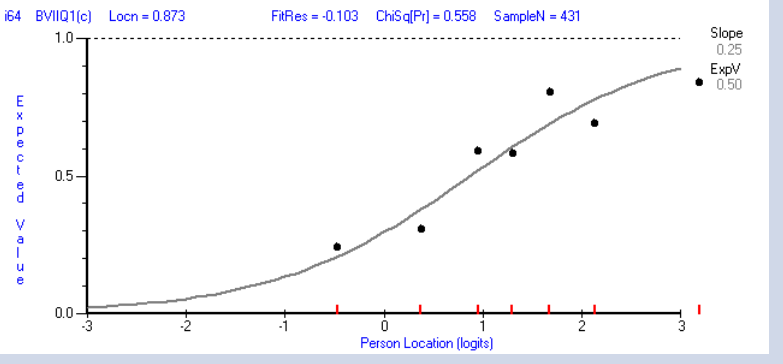
# Item Exemplar – After Re-scoring

## Rescored Mid-Category as Incorrect




| Item Characteristic | |
|---|---|
| Location, SE | 1.554, 0.170 |
| Fit Residual | -0.239 |
| Chi sq (prob) | 5.358 (0.499) |

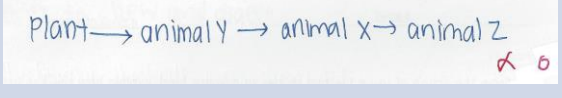
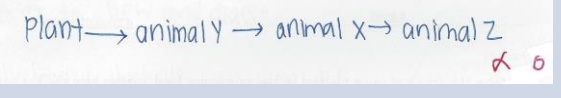
## Rescored Mid-Category as Correct



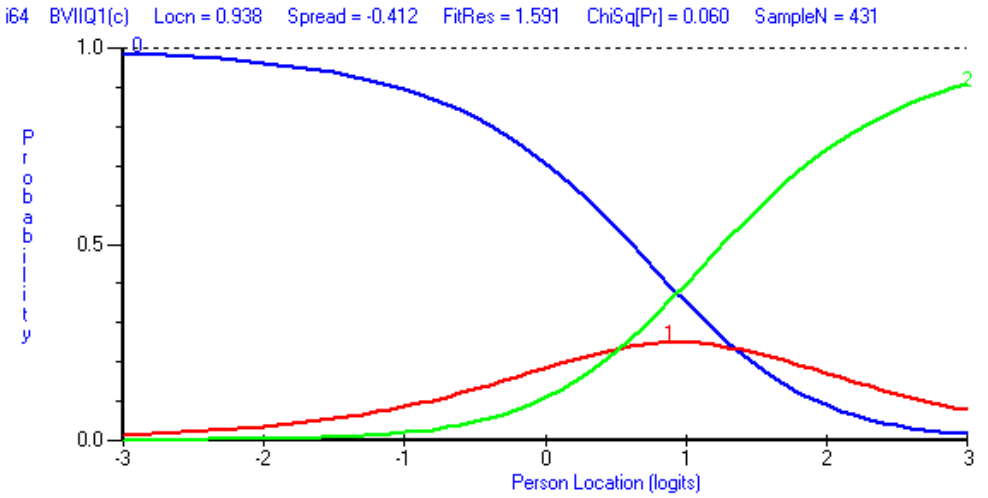

| Item Characteristic | |
|---|---|
| Location, SE | 0.873, 0.171 |
| Fit Residual | -0.103 |
| Chi sq (prob) | 4.891 (0.558) |

# Item Exemplar

| Old Mark Scheme | New Mark Scheme |
|---|---|
| Two correct inter-related food chains in food web - 2 marks<br><br>One correct inter-related food chain in food web - 1 mark | Two correct inter-related food chains in food web - 2 marks<br><br>One correct food chain - 1 mark |

| Application of MS | Application of MS |
|---|---|
| 0 mark<br>  | 1 mark<br>  |
| 2 – 1 = 1 mark<br>  | 2 marks<br>  |

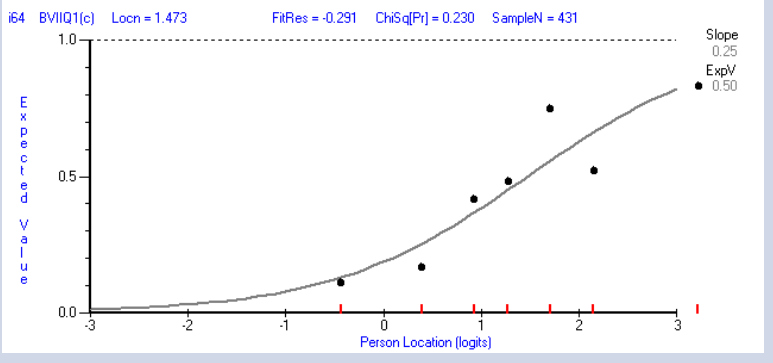
# After application of new Mark Scheme




| Item Characteristic | |
|---|---|
| Location, SE | 0.938, -0.412 |
| Fit Residual | 1.591 |
| Chi sq (prob) | 12.102 (0.060) |

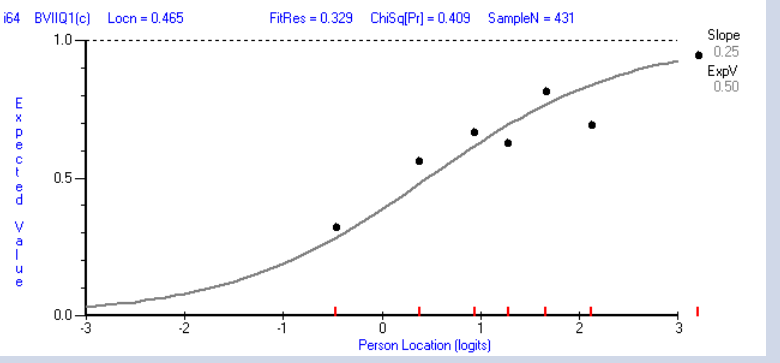
# After application of new MS - Re-scoring

## Rescored Mid-Category as Incorrect




| Item Characteristic | |
|---|---|
| Location, SE | 1.473, 0.170 |
| Fit Residual | -0.291 |
| Chi sq (prob) | 8.107 (0.230) |

## Rescored Mid-Category as Correct




| Item Characteristic | |
|---|---|
| Location, SE | 0.465, 0.177 |
| Fit Residual | 0.329 |
| Chi sq (prob) | 6.127 (0.409) |

# Discussion

- Improved test content validity
- Analysis of misconceptions and learning difficulties of pupils
- Learning gaps in curriculum
- Instructional opportunity