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
- Investigating & Analysing
- Interpreting & Analysing Data & Evidence
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)
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.

![Diagram of electromagnet](image)

**Item Exemplar – Ordered Threshold**

<table>
<thead>
<tr>
<th>Framework category</th>
<th>Item Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Type</td>
<td>Location, SE 2.166, 0.148</td>
</tr>
<tr>
<td>Practice</td>
<td>Fit Residual 0.880</td>
</tr>
<tr>
<td>Cognitive Demand</td>
<td>Chi sq (prob) 5.805 (0.445)</td>
</tr>
<tr>
<td></td>
<td>Evaluation of the planning and implementation of the scientific process</td>
</tr>
</tbody>
</table>
Eric went for a swim. When he came out of the water, he felt cold.

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

Explain why he felt even colder when a strong wind blew. [2 mark]
Eric went for a swim. When he came out of the water, he felt cold. His body was dry before swimming, and his body was wet after swimming. At that moment, a strong wind blew, and he felt even colder. 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

<table>
<thead>
<tr>
<th>Knowledge Type</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>Recall and apply appropriate scientific knowledge</td>
</tr>
<tr>
<td>Cognitive Demand</td>
<td>Medium</td>
</tr>
</tbody>
</table>

### Item Characteristic

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Location, SE</td>
<td>1.647, 0.112</td>
</tr>
<tr>
<td>Fit Residual</td>
<td>-2.264</td>
</tr>
<tr>
<td>Chi sq (prob)</td>
<td>10.596 (0.1017)</td>
</tr>
</tbody>
</table>
The skulls of three animals which lived on Earth millions of years ago are shown below.

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

<table>
<thead>
<tr>
<th>Framework category</th>
<th>Knowledge Type</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>Recall and apply</td>
<td>appropriate scientific knowledge</td>
</tr>
<tr>
<td>Cognitive Demand</td>
<td></td>
<td>High</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Two correct inter-related food chains in food web</th>
<th>2 marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>One correct inter-related food chains in food web</td>
<td>1 mark</td>
</tr>
</tbody>
</table>

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

0 mark

2 – 1 = 1 mark

<table>
<thead>
<tr>
<th>Item Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location, SE</td>
<td>1.287, 0.100</td>
</tr>
<tr>
<td>Fit Residual</td>
<td>1.467</td>
</tr>
<tr>
<td>Chi sq (prob)</td>
<td>16.488 (0.011)</td>
</tr>
</tbody>
</table>
Item Exemplar – After Re-scoring

Rescored Mid-Category as Incorrect

<table>
<thead>
<tr>
<th>Item Characteristic</th>
<th>Location, SE</th>
<th>Fit Residual</th>
<th>Chi sq (prob)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.590, 0.172</td>
<td>0.046</td>
<td>6.977 (0.323)</td>
</tr>
</tbody>
</table>

Rescored Mid-Category as Correct

<table>
<thead>
<tr>
<th>Item Characteristic</th>
<th>Location, SE</th>
<th>Fit Residual</th>
<th>Chi sq (prob)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.276, 0.170</td>
<td>0.298</td>
<td>5.976 (0.426)</td>
</tr>
</tbody>
</table>
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