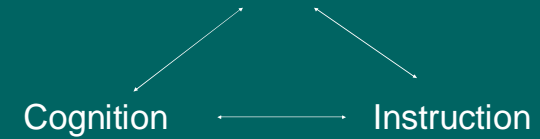


Diagnostic Item Analysis:
The Dense Item Perspective

William M. Bart
University of Minnesota
bartx001@umn.edu
Invited Lecture
The University of Tokyo
Tokyo, Japan
July 9, 2007

A Triangular Model of Education and Psychology

Testing and Measurement



A Central Aspect in the Model

The Test Item

The Purposes of the Test Item

- The test item should inform the user as to how the participant thinks when responding to the item (cognition).
- The test item should inform the user as to what the participant should do to improve the performance of the subject (instruction).

The Present Case

- Test items tend neither to inform the user as to how the participant thinks when responding to the item nor to inform the user as to what should be done to improve the performance of the participant.

Typical Test Item

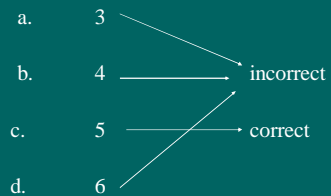
Select the letter of the response with the correct answer

$$2 + 3 = \underline{\quad? \quad}$$

- a. 3
- b. 4
- c. 5
- d. 6

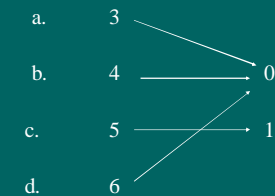
Qualitative Scoring Convention for The Test Item

$$2 + 3 = \underline{\quad? \quad}$$

- a. 3
 - b. 4
 - c. 5
 - d. 6
- incorrect
- correct
- 

Quantitative Scoring Convention for The Test Item

$$2 + 3 = \underline{\quad? \quad}$$

- a. 3
 - b. 4
 - c. 5
 - d. 6
- 0
- 1
- 

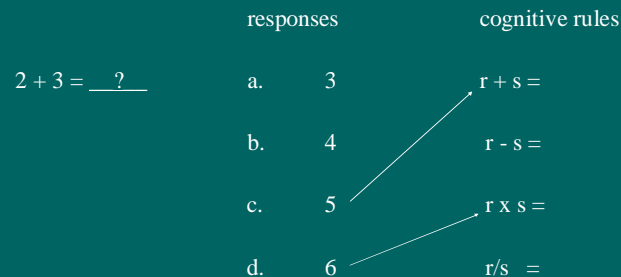
An Item Mapping

- Every scoring convention for an item may be viewed as a mapping between the set of responses to an item and some set of elements related to the item responses (e.g., item scores, cognitive rules, instructional sequences).
- In the example of the arithmetic item, $2 + 3 = ?$, the set of item responses is $\{3, 4, 5, 6\}$ and the set of elements related to the item responses is $\{0, 1\}$.

Two Qualifications

- Instead of Item Mapping, we will use the alternative term “Item Digraph”.
- We will limit our subsequent examples of Item Digraphs to those with diagnostic utility.

Item Digraph of Typical Test Item



Item Digraph

1. A set of item response options.
2. A cognitive rule set for the item consisting of a set of cognitive rules.
3. A representation of the mapping between the response options of an item and the cognitive rules for the item.
4. Each item digraph engenders a 0/1 item response \times rule matrix (A matrix) with cell $a(i,j) = 1$ if the j -th rule interprets the i -th response option and $= 0$ otherwise.

The A Matrix for the Arithmetic Test Item, $2 + 3 = ?$

Item	Cognitive Rules				Total
	r + s	r - s	r x s	r / t	
3	0	0	0	0	0
4	0	0	0	0	0
5	1	0	0	0	1
6	0	0	1	0	1

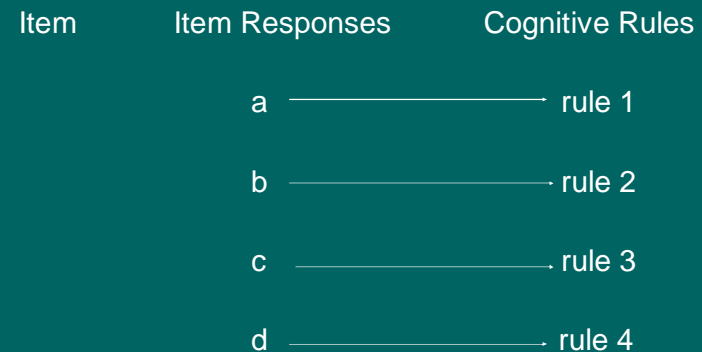
Semi-Dense item

- Any item for which each response option is interpreted by one and only one rule in the cognitive rule set for the item and for which each rule in the cognitive rule set for the item interprets one and only one rule.

The Semi-Dense Item

- Is a highly diagnostic item
- Is rich (i.e., dense) with qualitative inferential potential regarding the cognitive rules used by participants.

Item Digraph for a Semi-Dense Item



Algebraic Interpretation of a Semi-Dense Item

- An item is semi-dense if there is a bijection between the set of item responses for the item and the set of cognitive rules used to interpret the responses to the item.
- In other words, an item is semi-dense if each item response is related to only one rule and each rule is related to only one item response.

The A Matrix for a Semi-Dense Item

Item

Response Cognitive Rules

Options	1	2	3	4	Total
a	1	0	0	0	1
b	0	1	0	0	1
c	0	0	1	0	1
d	0	0	0	1	1

The Algebraic Interpretation of a Semi-Dense Item

- The A matrix for a semi-dense item is an identity matrix.

The Cognitive Rule Set for a Test Item

- The set of cognitive rules for a test item is derived from prior cognitive research on the test item.
- The cognitive rules may include misconceptions.
- Cognitive rules indicate how aspects of a test item stem are transformed into a response.

Cognitive Rules for a Test Item are not merely Attributes

- Some models of diagnostic testing such as those suggested by Tatsuoka and others focus on attributes.
- Listing some attributes for an item response provides no information as to how the attributes are to be combined and in what order.
- If one attribute is add 2 to the quantity (i.e., $a + 2 = ?$) and the second attribute is divide the quantity by 7 (i.e., $a/7 = ?$), then the first attribute followed by the second attribute results in the quantity $(a + 2)/7$ and the second attribute followed by the first attribute results in the quantity $(a/7) + 2$ and the $(a + 2)/7 < (a/7) + 2$.

The Diagnostic Inadequacy of Most Test Items

- Most test items are not semi-dense and thus most test items are not highly diagnostic.
- As a result, most tests lack diagnostic utility.
- In other words, few qualitative cognitive inferences tend to be drawn from the item response patterns of participants.

Item Diagnostic Properties

- From a dense item perspective, test items have five diagnostic properties.
- Those properties are related to algebraic properties of item digraphs.

Five Item Diagnostic Properties

1. Response Interpretability.
 - extent to which all response options for an item are interpreted by rules from the cognitive rule set for the item.
 - extent to which each item response option is mapped onto at least one rule for the item rule set.
 - extent to which there is a mapping from the set of item responses for an item to the rules in the rule set for the item.
 - this is the most basic item diagnostic property.

Five Item Diagnostic Properties

2. Response Discrimination
 - extent to which each response option for an item is interpreted by one and only one rule in the cognitive rule set for the item.
 - extent to which there is a function from the set of distinct responses to the item to the rules in the rule set for the item.

Five Item Diagnostic Properties

3. Rule Discrimination 1
 - extent to which the item has response discrimination and the extent to which distinct rules in the rule set for the item interpret distinct response options.
 - extent to which there is an injection from the item response set for an item to the rule set for the item
 - I.e., extent to which there is a function between the item response set for an item and the rule set for the item and the extent to which each rule in the rule set for the item has a distinct response related to the item.

Five Item Diagnostic Properties

4. Rule Set Usage
 - extent to which the item has response discrimination and the extent to which each rule in the cognitive rule set for the item interprets an item response option.
 - extent to which there is a surjection between the item response set for an item and the cognitive rule set for the item.
 - e.g., extent to which there is a function between the item response set for an item and each rule in the rule set for the item interprets an item response.

Five Item Diagnostic Properties

5. Semi-Density
 - extent to which each item response option is interpreted by one and only one rule and each rule interprets one and only one response option.
 - in other words, an item is semi-dense to the extent to which there is a bijection between the item response set and the rule set for the item.

Item Diagnostic Property Indices

1. Each of the five item diagnostic properties relates to a psychometric index indicating the extent to which a test item has a specific item diagnostic property.
2. Each item diagnostic property index ranges from 0 to 1.

Item Diagnostic Property Indices

1. Indices that are associated with the five properties are the following:
 - (1) I1 - Response Interpretability;
 - (2) I2 - Response Discrimination;
 - (3) I3 - Rule Discrimination 1;
 - (4) I4 - Rule Set Usage; and
 - (5) I5 - Semi-Density.
3. $I1 \geq I2 \geq I3 > \text{or} = \text{or} < I4 \geq I5$.

Index for Item Response Interpretability

1. $I1 = \frac{\sum_i t(i)}{k}$
with $t(i) = 1$ if $\sum_j a(i,j) > 0$
and $t(i) = 0$ otherwise and
 $k =$ number of response options for the item.

Index for Item Response Discrimination

2. $I2 = \frac{\sum_i t(i)}{k}$
with $t(i) = 1 / \sum_j a(i,j)$
if $\sum_j a(i,j) > 0$
and $t(i) = 0$ otherwise and
 $k =$ number of response options for the item.

Other Indices

- The other three indices for item diagnostic properties are derived from the index of response interpretability and the index of response discrimination.

Item Digraph for a Semi-Dense Item

Item	Item Responses	Cognitive Rules
	a	rule 1
	b	rule 2
	c	rule 3
	d	rule 4

Item Diagnostic Property Index Values for a Semi-Dense Item

- $I1 = 1$
- $I2 = 1$
- $I3 = 1$
- $I4 = 1$
- $I5 = 1$

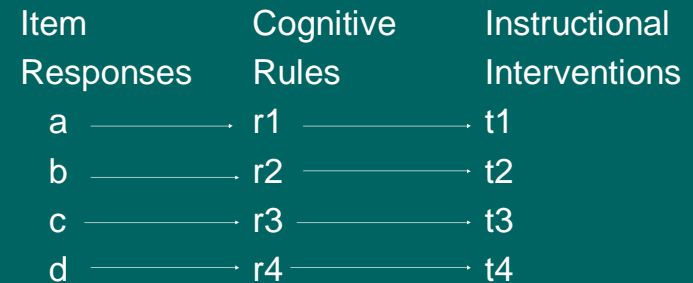
Item Digraph of Typical Test Item

	responses	cognitive rules
$2 + 3 = \underline{\quad?}$	a. 3	$r + s =$
	b. 4	$r - s =$
	c. 5	$r \times s =$
	d. 6	$r/s =$

Item Diagnostic Property Index Values for the Test Item, $2 + 3 = ?$

- I1 = .50
- I2 = .50
- I3 = .50
- I4 = .25
- I5 = .25

Item Digraph for a Dense Item



Item Diagnostic Analysis from the Dense Item Perspective

- One may engage in an *a priori* method of item analysis by computing the five item diagnostic property indices which indicate the extent to which an item has the five item diagnostic properties.
- This method of item diagnostic analysis acts as a filter to screen test items for diagnostic tests.

Recommendations

1. Research on how to assemble test items that are not highly diagnostic to maximize the diagnosticity of tests.
2. Research on how to increase the diagnosticity of test items and tests.

Recommendations

- 3. Research on how to integrate measurement and cognitive research to make informative and diagnostic test items and tests.
- 4. Research on how to use multiple scoring keys to score test response patterns to increase the diagnostic information extracted from item response patterns.