Rasch modeling - part I

Theory and principles

Intro

Today's goal:

Teach the general idea of Rasch modeling

Outline:

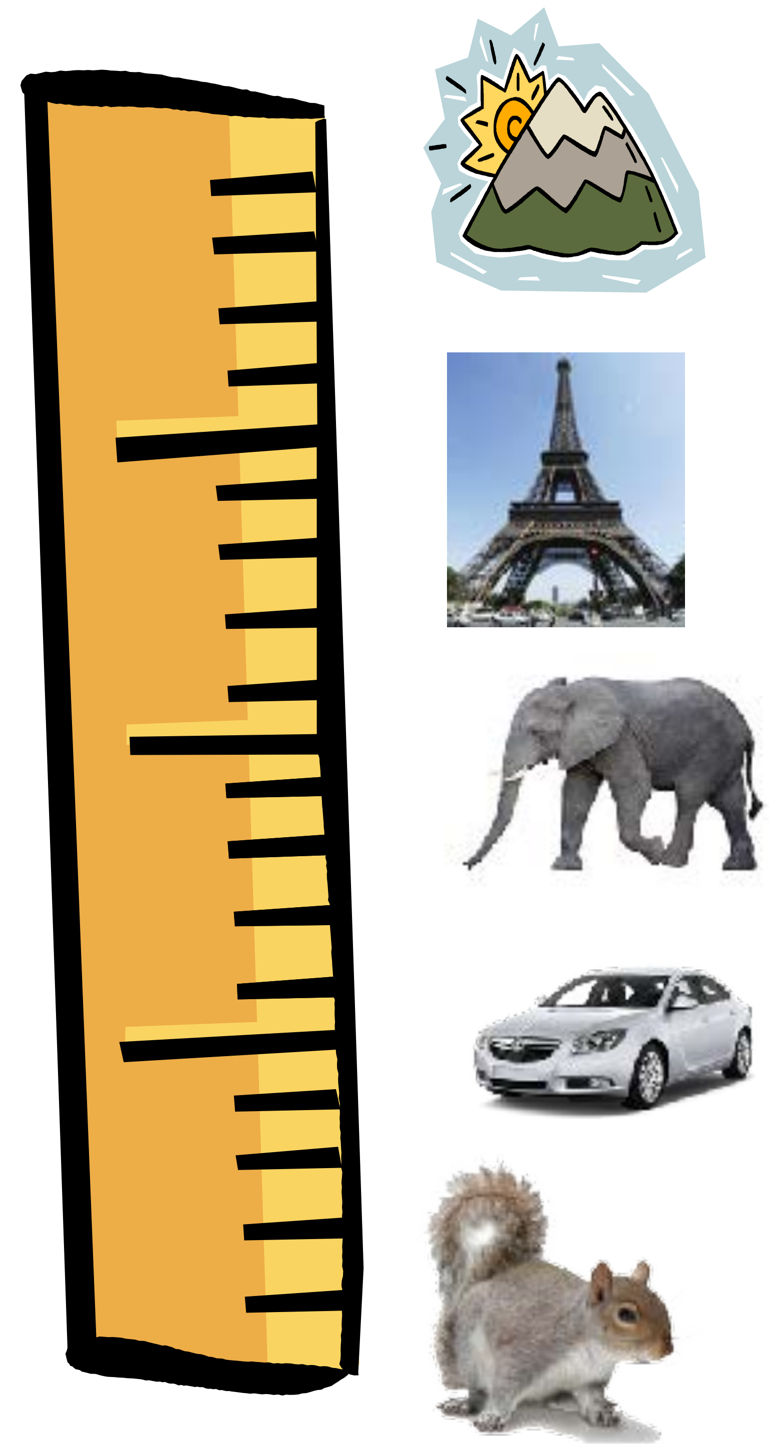
- Basic principles
- Interpreting Rasch model output
- Next lecture: running a Rasch model (in R?)

Rasch modeling

basic principles

Rasch modeling

Goal: making a scale



Psychometrics

Social science belief:

We can do the same thing with social and psychological concepts!"

For instance, concept: "peacefulness"

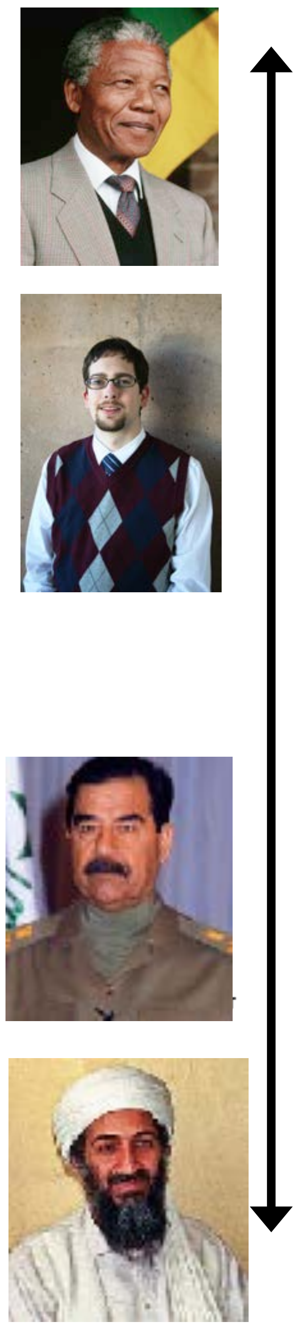
- Definition: "respect for the rights and lives of all living creatures"

We can put both persons and items on a scale!

- Rasch modeling: this scale will be unidimensional

Scale of persons

trait level



The higher the person, the more peaceful they are

Scale of items

item difficulty (to comply with the concept)

"If I discover a fly in my soup, I'll try CPR"

"War can never be justified"

"I never hurt someone on purpose"

"I respect people's feelings"

"War is needed to defend your country"

"human rights don't apply to criminals"

The higher the item, the more peaceful someone has to be to comply with it

Reversed items

Combining scales

“If I discover a fly in my soup, I’ll try CPR”

“War can never be justified”

“I never hurt someone on purpose”

“I respect people’s feelings”

“War is needed to defend your country”

“human rights don’t apply to criminals”

Rule:

An item I and a person P have the same level if there’s a 50% chance that someone with the same trait level as P complies with items with the same difficulty level as as I.

Why Rasch?

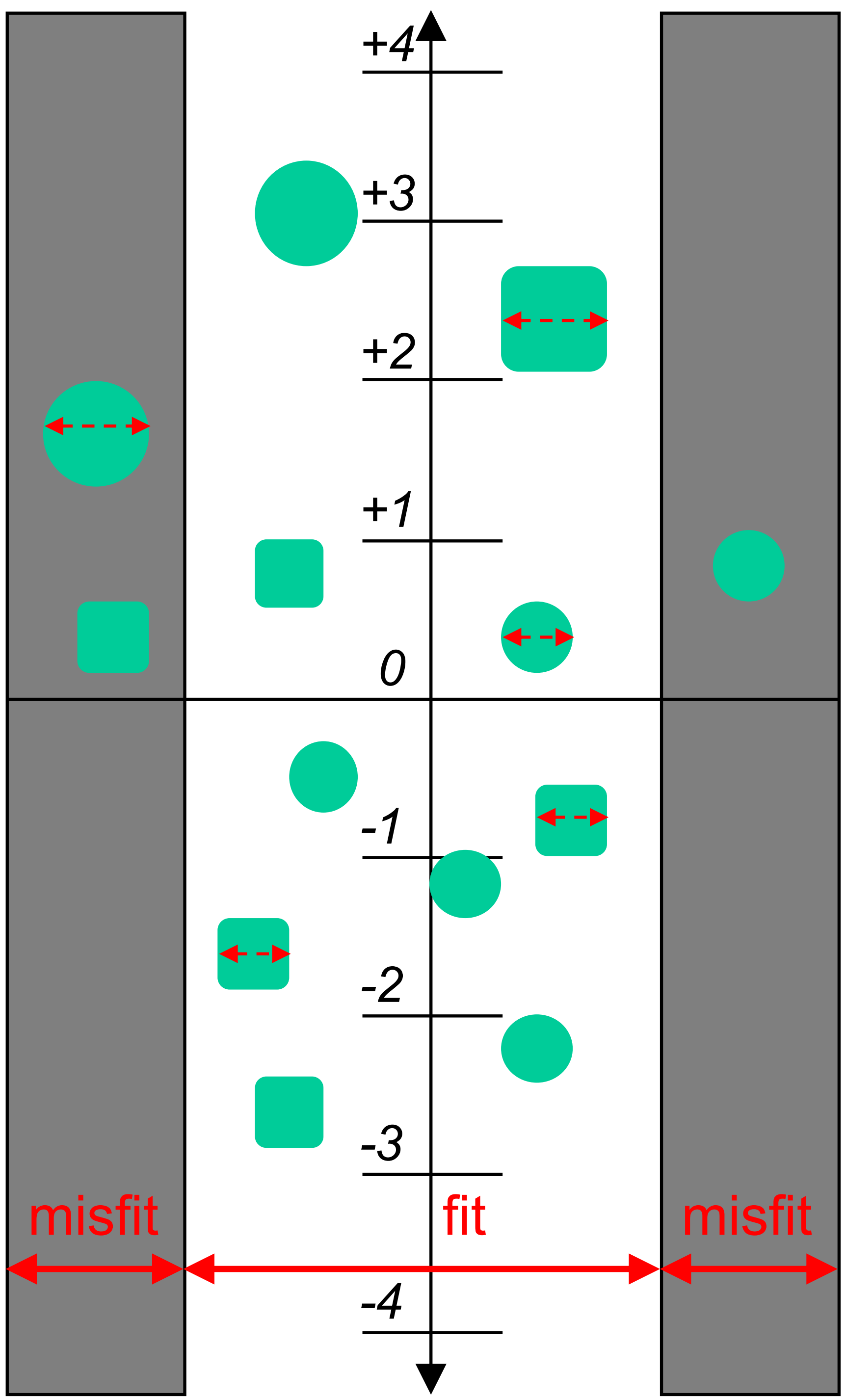
Use Rasch modeling if you want to know:

... **difficulty** of items and **ability** of persons. Both are unknown in advance!

... whether persons and items **fit** our unidimensional model

... the amount of **error**. Did we have enough persons and items to create an accurate measure?

Keywords



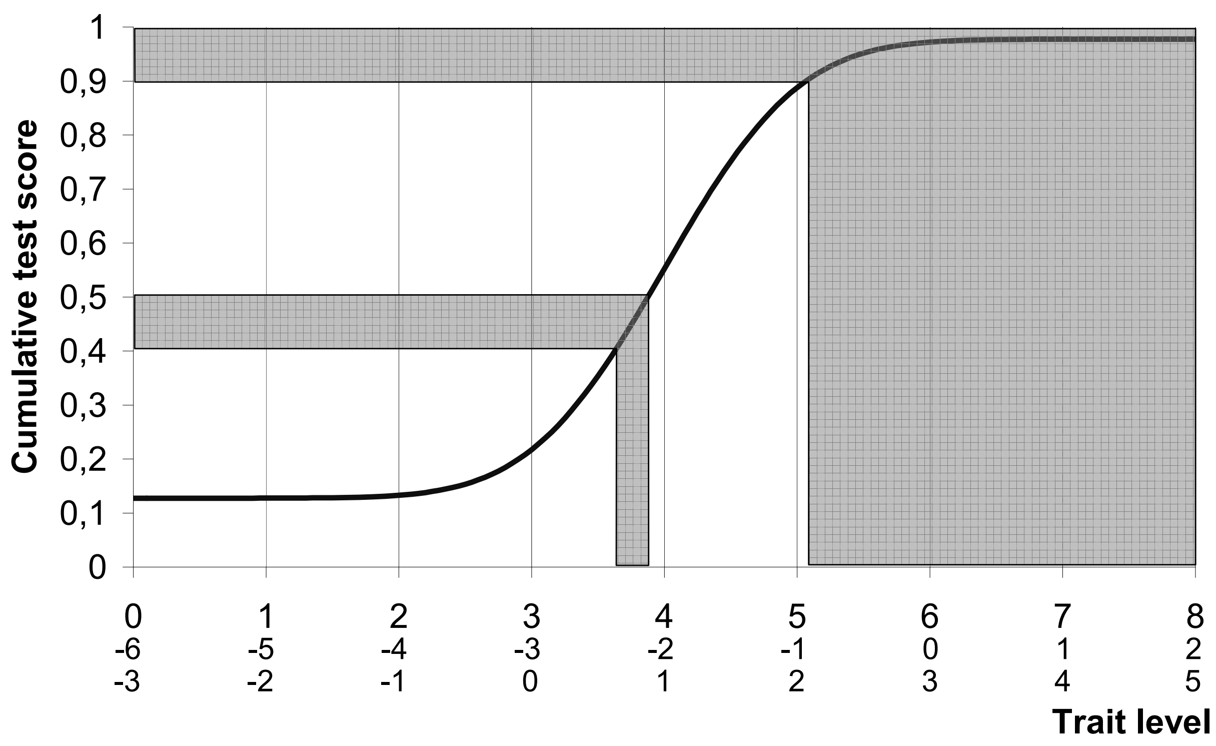
Persons (squares) and items (circles)

Person Trait level and Item Difficulty (vertical position)

Fit (horizontal position)

Error (square/circle size)

Logits

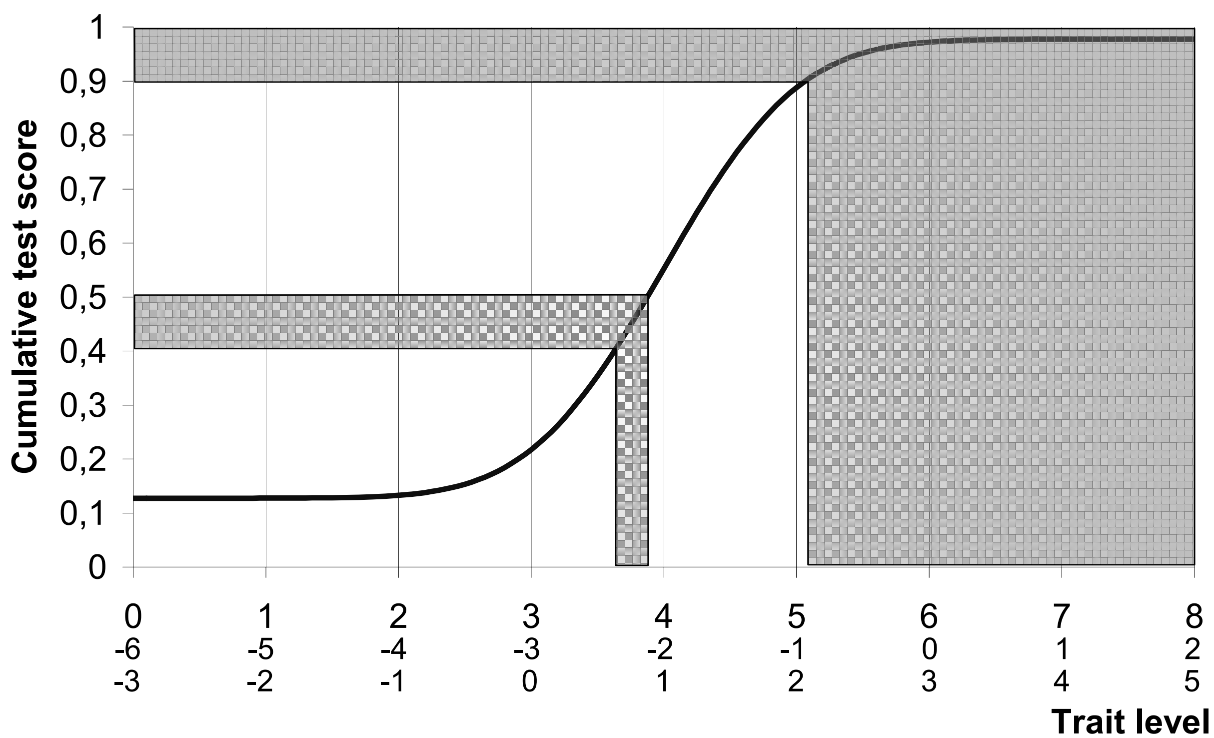


Rasch about persons:

There's a Log Odds relationship between test score and trait level

> The difference between getting 40% and 50% right is a smaller trait level difference than the difference between getting 90% and 100% right

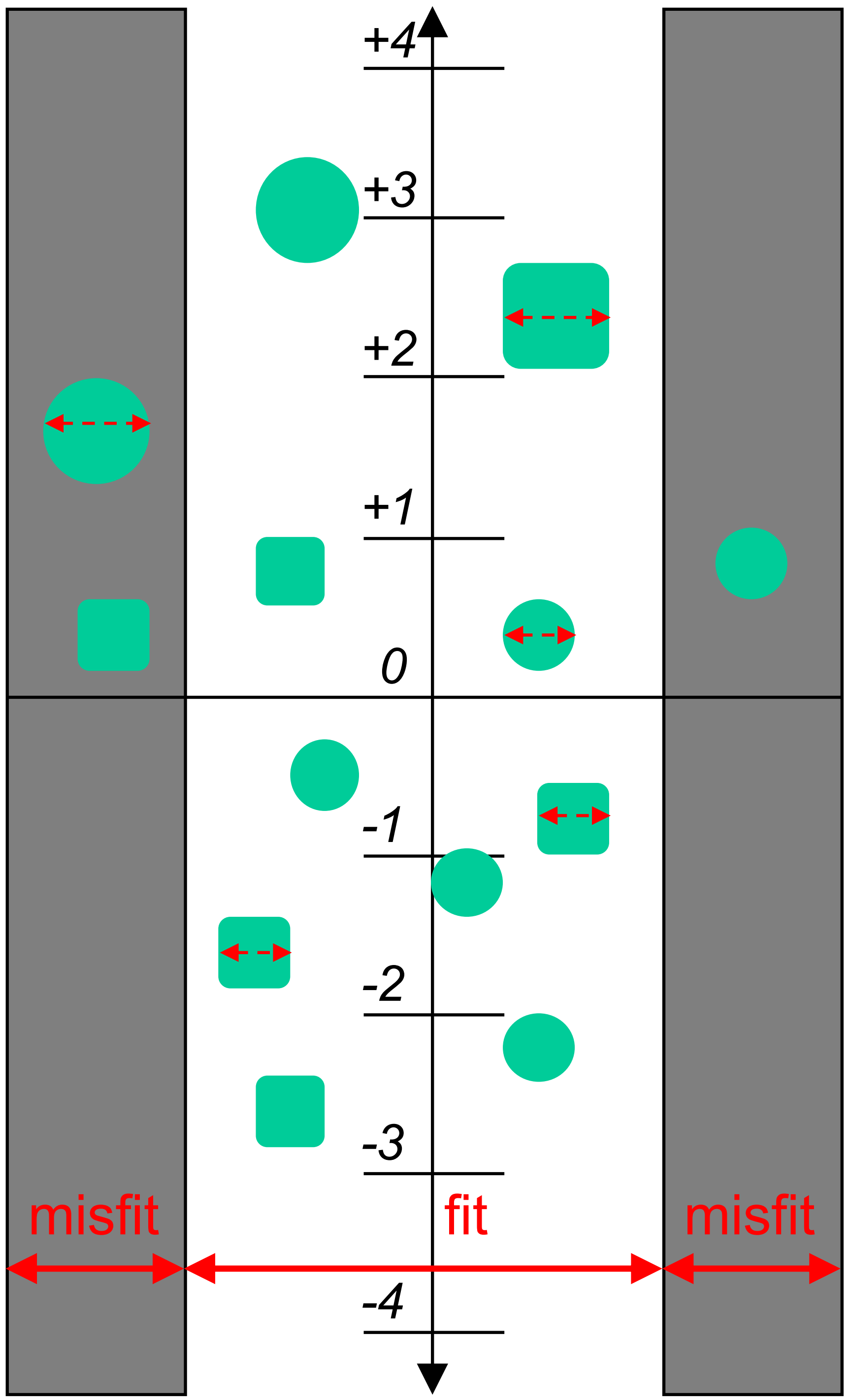
Logits



Rasch about items: There's a Log Odds relationship between correct responses and item difficulty

The difference between 40% and 50% correct responses is a smaller item difficulty difference than the difference between 90% and 100% correct responses

Logits



Item difficulty and person ability are expressed in a Log Odds Ratio (logits).

Log odds can be translated into probability scores and vice versa:

$$P\left(X_{is}=1\middle|\theta_s,\beta_i\right)=\frac{e^{\theta_s-\beta_i}}{1+e^{\theta_s-\beta_i}}$$

$$\ln\left[P_{is}/\left(1-P_{is}\right)\right]=\theta_s-\beta_i$$

Logits explained

P_{is} = chance person s answers item i correctly

θ_s = persona ability (trait level)

β_i = item difficulty

The log odds of a correct response equals the difference between person ability and item difficulty

$$\ln\left[P_{is}/\left(1-P_{is}\right)\right]=\theta_s-\beta_i$$

If we know the person ability and the item difficulty, we can calculate the probability that the person will answer this item correctly

$$P\left(X_{is}=1|\theta_s,\beta_i\right)=\frac{e^{\theta_s-\beta_i}}{1+e^{\theta_s-\beta_i}}$$

Example

Suppose we have 5 items, with known difficulties:

$$\beta = [-2, -1, 0, 1, 2]$$

Take person s, trait level unknown, who answered all items correct except the last one:

$$X = [1, 1, 1, 1, 0]$$

The likelihood of having this response pattern is:

$$L(X) = P_{1s}P_{2s}P_{3s}P_{4s}(1-P_{5s})$$

Example

Since we have the difficulties β, we can calculate the $L(X)$ for different trait levels:

$\theta_s = -1$:

$e^{-1+2}/(1+e^{-1+2})*e^{-1+1}/(1+e^{-1+1})*e^{-1}/(1+e^{-1})*e^{-1-1}/(1+e^{-1-1})*(1-(e^{-1-2}/(1+e^{-1-2})) = 0.0112$

$\theta_s = 0$:

$e^{2}/(1+e^{2})*e^{1}/(1+e^{1})*e^{0}/(1+e^{0})*e^{-1}/(1+e^{-1})*(1-(e^{-2}/(1+e^{-2})) = 0.0763$

$\theta_s = 1$:

$e^{1+2}/(1+e^{1+2})*e^{1+1}/(1+e^{1+1})*e^{1}/(1+e^{1})*e^{1-1}/(1+e^{1-1})*(1-(e^{1-2}/(1+e^{1-2})) = 0.2242$

Example

The trait level of person *s* is the trait level with the highest likelihood value

Person *t* with response pattern [1,1,1,0,1] has generally lower likelihood scores than person *s* (why?)

But person *t* has the same trait level as person *s* (why?)

Likelihood is a measure of fit! (why?)

Example

One can calculate item difficulty in a similar fashion, using known trait levels

Wait a minute...

We can determine person trait levels using item difficulties

We can determine item difficulties using person trait levels

But we start without knowing any of them!

Example

Solution:

The Rasch model runs an iterative process to determine both trait levels and difficulties

Target: optimal overall model fit

Rasch model output

how to interpret Rasch model results

TABLE 4.1
BLOT Item Difficulty Estimates With Associated Error Estimates for Each Item

Item	*Difficulty Estimate*	*Error Estimate*	*Infit Mean Square*	*Outfit Mean Square*	*Infit t*	*Outfit t*
1	−0.77	0.26	0.98	0.69	0.0	−0.8
2	−0.70	0.26	1.01	0.75	0.1	−0.6
3	0.74	0.2	0.98	0.9	−0.2	−0.5
4	0.00	0.22	1.00	0.88	0.0	−0.4
5	−0.98	0.28	0.98	0.76	−0.1	−0.5
6	−2.42	0.47	1.06	0.83	0.3	0.1
7	−0.64	0.25	0.97	0.65	−0.1	−1.0
8	0.85	0.19	0.91	1.00	−1.1	0.1
9	0.18	0.21	1.07	0.97	0.7	0.0
10	−0.19	0.23	0.92	0.68	−0.7	−1.1
11	0.18	0.21	1.02	0.96	0.2	−0.1
12	−1.76	0.36	0.69	0.24	−1.1	−1.5
13	1.00	0.19	1.16	1.32	2.0	1.8
14	−0.70	0.26	1.15	1.32	1.0	0.9
15	1.00	0.19	0.96	0.84	−0.4	−0.9
16	−0.30	0.23	1.13	1.03	1.0	0.2
17	0.39	0.2	0.87	0.75	−1.4	−1.2
18	−0.05	0.22	0.9	0.74	−0.9	−1.0
19	0.47	0.2	1.01	1.05	0.1	0.3
20	−0.84	0.27	0.91	0.81	−0.5	−0.4
21	2.33	0.2	1.27	1.75	2.6	3.4
22	−1.06	0.29	0.91	1.69	−0.4	1.4
23	0.35	0.21	1.06	0.92	0.7	−0.3
24	0.22	0.21	0.89	1.03	−1.1	0.2
25	0.51	0.2	1.07	1.26	0.8	1.2
26	0.78	0.2	0.89	0.75	−1.3	−1.4
27	−0.91	0.27	0.85	0.62	−0.8	−0.9
28	1.63	0.19	1.12	1.23	1.4	1.4
29	−0.46	0.24	0.94	0.71	−0.4	−0.8
30	1.07	0.19	1.19	1.15	2.3	0.9
31	0.18	0.21	1.07	1.55	0.7	2.0
32	1.14	0.19	0.96	0.85	−0.5	−0.9
33	−0.52	0.25	1.1	0.93	0.7	−0.1
34	−0.41	0.24	1	0.79	0.1	−0.6
35	−0.30	0.23	0.93	0.73	−0.5	−0.9

Note. Fit statistics are shown in their natural (mean square) and standardized forms (standardized as *t*).

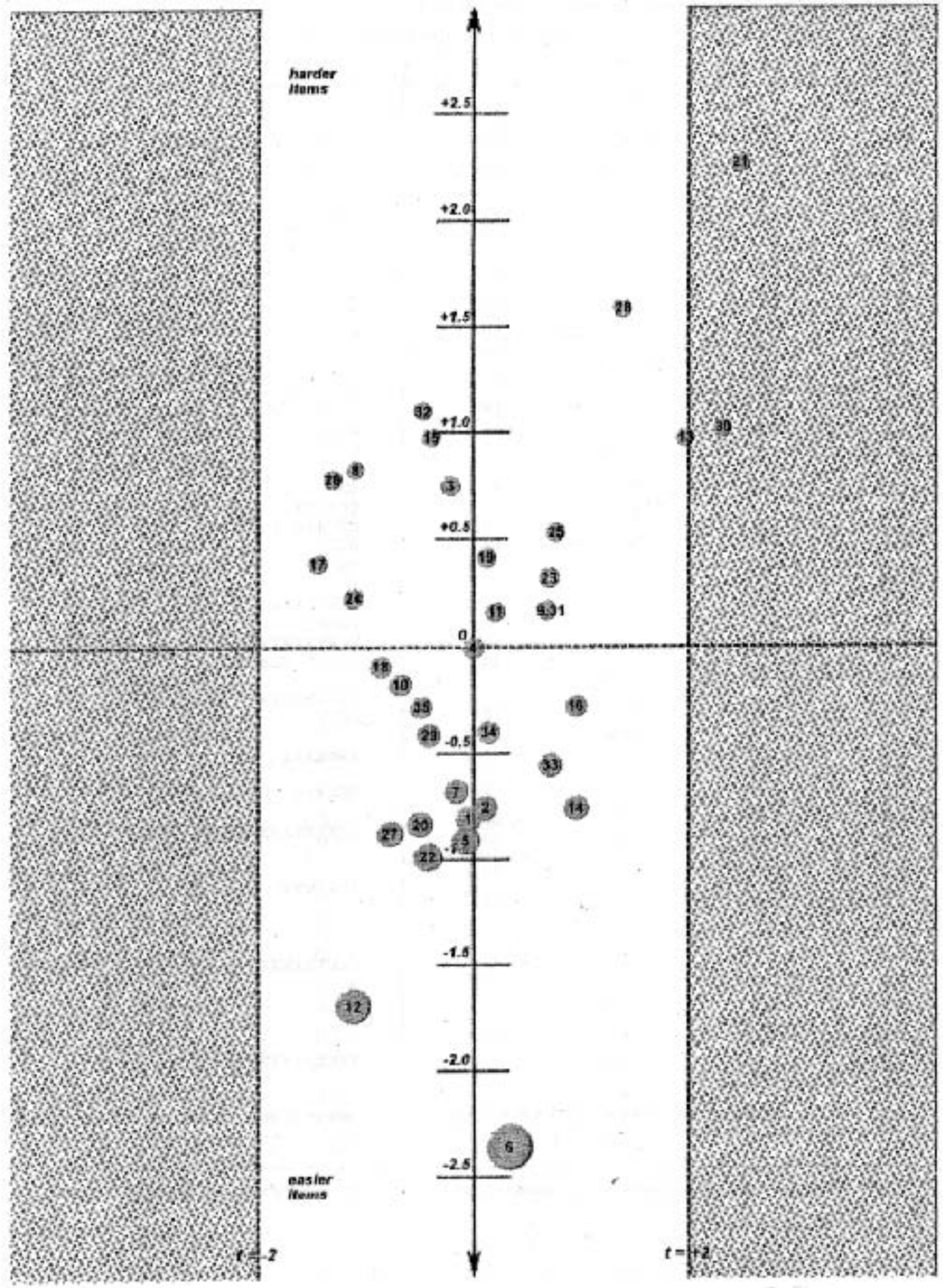


FIG. 4.1. BLOT item pathway.

Item estimates

Fit is not the same as error!

- Error is about precision (less data = more error)
- Fit is about expectations (unexpected behavior = lower fit)

Many items have a difficulty near 0

The logit scale is an interval scale: the difference in "difficulty" between items 15 and 4 is the same as between 4 and 22.

More on item fit

Item	*Difficulty Estimate*	*Error Estimate*	*Infit Mean Square*	*Outfit Mean Square*	*Infit t*	*Outfit t*
1	–0.77	0.26	0.98	0.69	0.0	–0.8
2	–0.70	0.26	1.01	0.75	0.1	–0.6
3	0.74	0.2	0.98	0.9	–0.2	–0.5
4	0.00	0.22	1.00	0.88	0.0	–0.4
5	–0.98	0.28	0.98	0.76	–0.1	–0.5
6	–2.42	0.47	1.06	0.83	0.3	0.1
7	–0.64	0.25	0.97	0.65	–0.1	–1.0
8	0.85	0.19	0.91	1.00	–1.1	0.1
9	0.18	0.21	1.07	0.97	0.7	0.0
10	–0.19	0.23	0.92	0.68	–0.7	–1.1
11	0.18	0.21	1.02	0.96	0.2	–0.1
12	–1.76	0.36	0.69	0.24	–1.1	–1.5
13	1.00	0.19	1.16	1.32	2.0	1.8
14	–0.70	0.26	1.15	1.32	1.0	0.9
15	1.00	0.19	0.96	0.84	–0.4	–0.9
16	–0.30	0.23	1.13	1.03	1.0	0.2
17	0.39	0.2	0.87	0.75	–1.4	–1.2

More on item fit

Expected fit mean square: 1.0

Fit mean square < 1.0: less variation than expected —> overfit

Ex.: 11111111110000000000

< 0.6 is problematic

Fit mean square > 1.0: more variation than expected —> underfit

Ex.: 11001001010101101001

> 1.2 is problematic

More on item fit

Underfit is worse than overfit

t-values show significance of misfit

- Can be too sensitive, especially for item fit!
- Otherwise, use +/–1.96 as a cutoff value

More on item fit

Outfit: all scores count equally

Infit: scores of persons with an ability near the item difficult are weighted more in determining the fit

Example:

11111110001110000000 has a higher infit, lower outfit

11111110100000000101 has a lower infit, higher outfit

Note: all of this also holds for persons!

Item summary

Summary of Item Estimates

Mean	0.00
SD	0.95
SD (adjusted)	0.92
Reliability	0.94

Fit Statistics

Infit Mean Square		*Outfit Mean Square*	
Mean	1.00	Mean	0.95
SD	0.11	SD	0.31

Infit t		Outfit t	
Mean	0.09	Mean	-0.05
SD	0.98	SD	1.10

fixed to zero

standard deviation of item scores

scale reliability (how certain we are of item order)

Means and SDs of fit mean squares and t-values

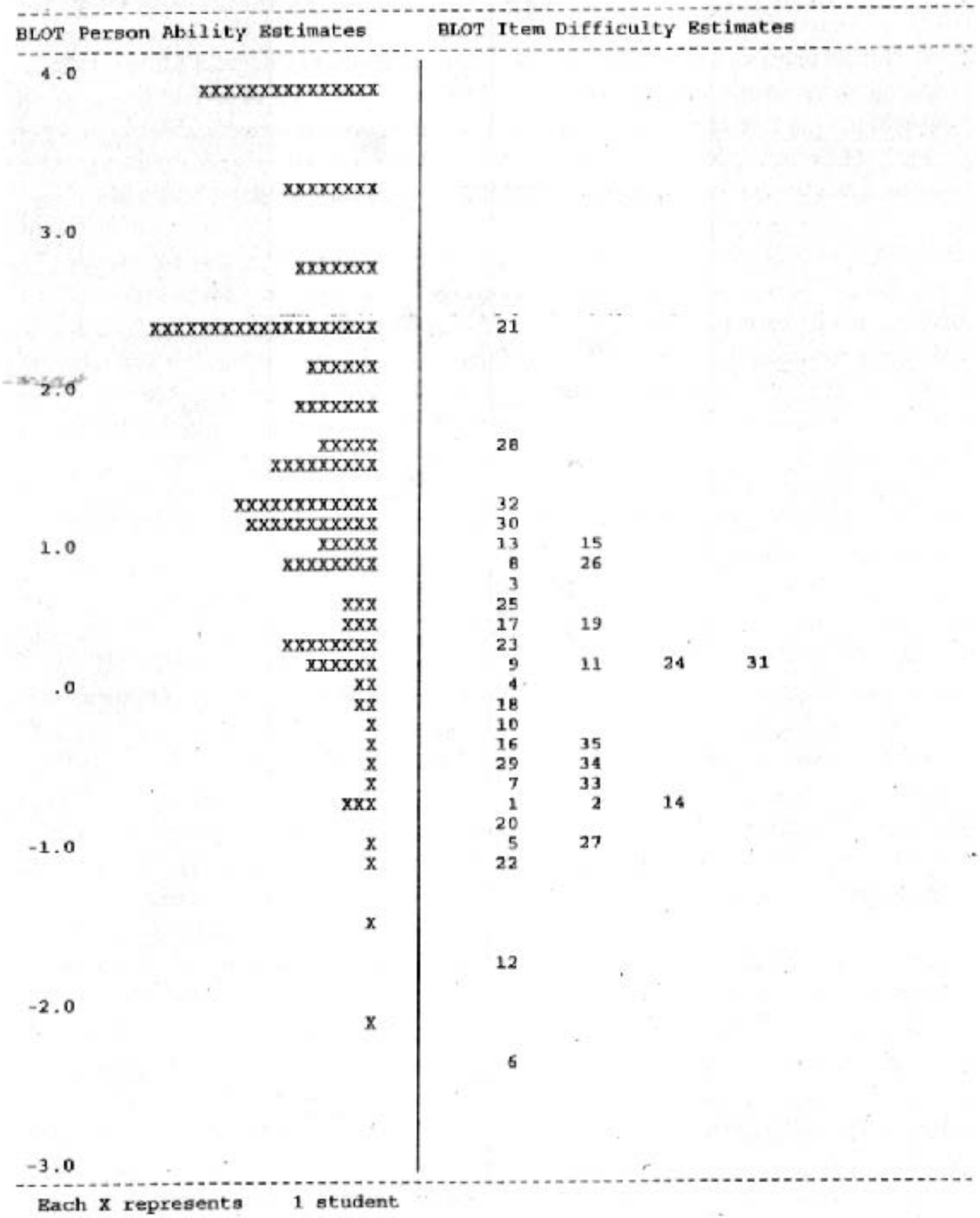


FIG. 4.2. Item–person map for the BLOT analysis (QUEST).

+ 3 persons with a perfect score

Item-person map

Many persons with a high trait level

- The scale is “too easy” for them

Few persons with a low trait level

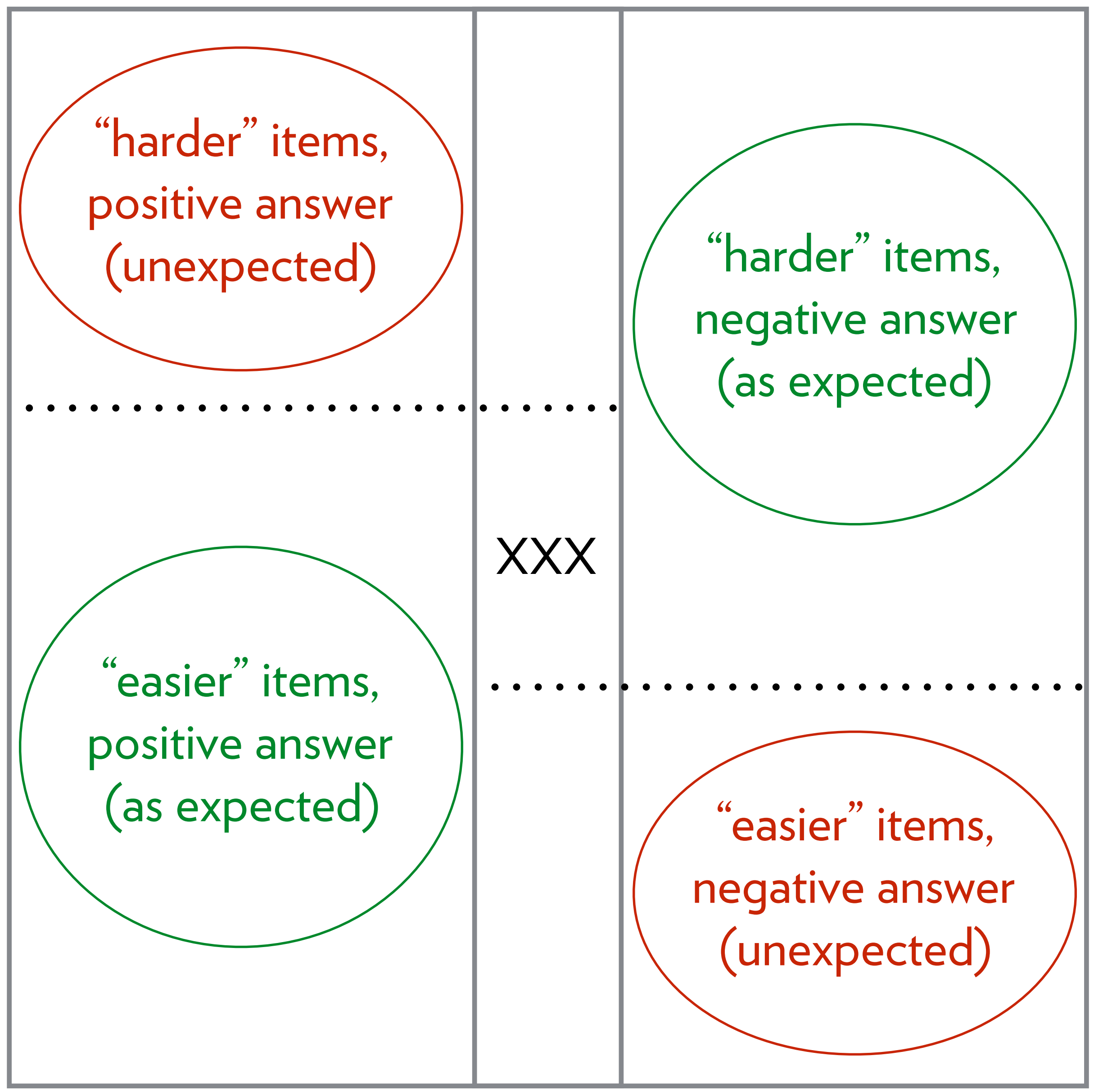
- Causes error in easiest items, e.g. item 6

Several perfect scores

- We don’t know anything about their trait level (other than that it’s high)

More persons than items

Kidmap



But... we expect at least some answers in the unexpected areas (otherwise we'd have overfit!)

"It is the mark of a truly intelligent person to be moved by statistics."

T H A N K S !

George Bernard Shaw