Objective Assessment and Item Analysis

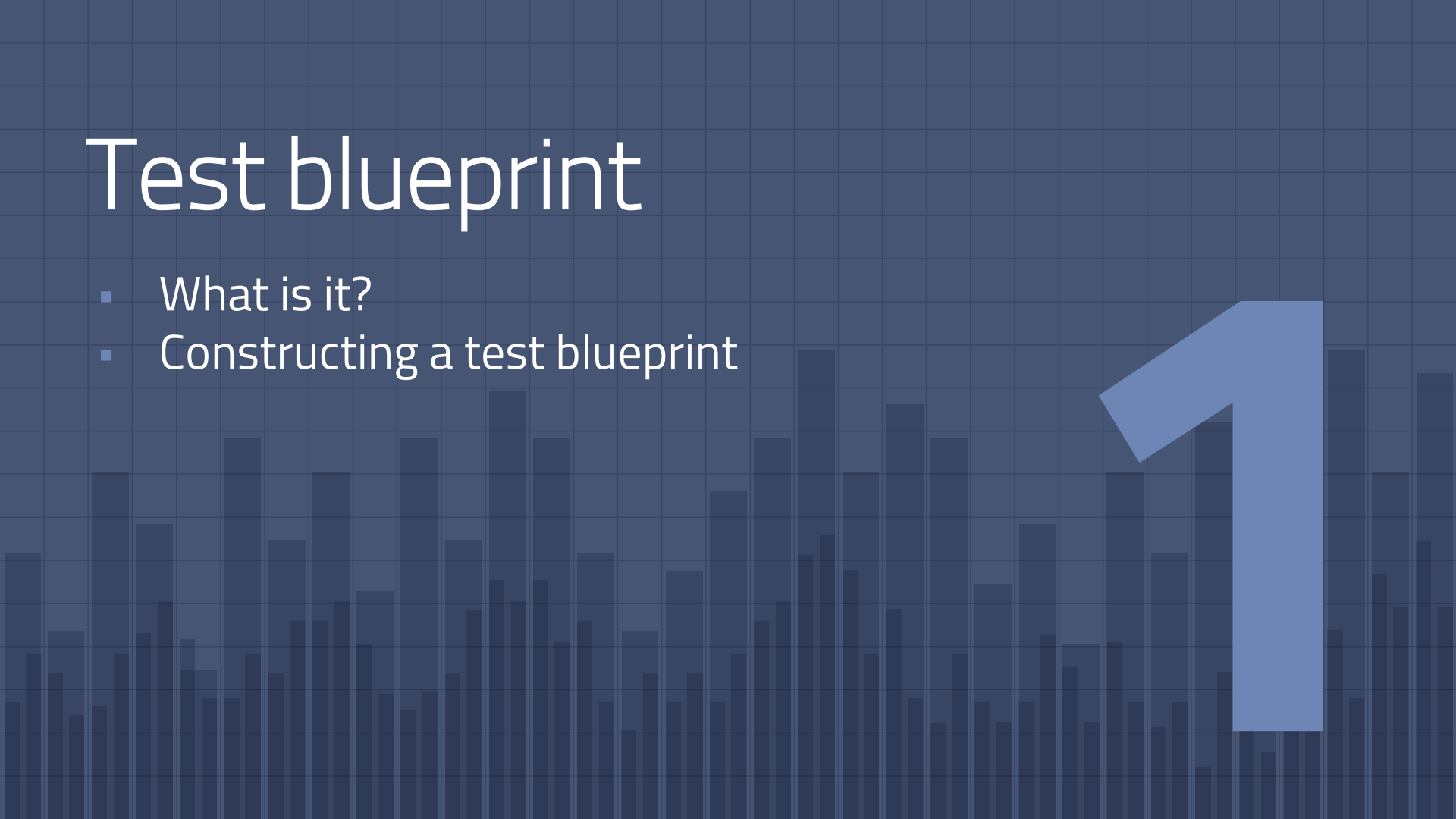
Parames Laosinchai

Outline

1) Test blueprint
2) Objective assessment
3) Item analysis and reliability

Test blueprint

- What is it?
- Constructing a test blueprint



Test blueprint

- Also called test specification
- Describe the elements of a test
  - Content to be covered
  - Amount of emphasis

Constructing a test blueprint

- Start from learning objectives
- Choose the method to assess each objective
- Specify the weight for each objective/sub-objective

Methods of assessment

- Written test
  - Objective test
  - Subjective test
- Performance-based assessment
  - Simulated situation
  - Real situation

Choosing method of assessment

- Which domain of learning?
  - Cognitive: written test
  - Affective/psychomotor: performance-based
- Which level within the cognitive domain?
  - Lower levels: objective test
  - Higher levels: subjective test

Considerations

- Validity of the score interpretation
  - Can the method assess the desired objective?
- Reliability of scores produced by the method
  - Does the method yield consistent results?
- Practical constraints
  - Testing and grading time, budget, logistic, etc.

Specifying weights

- Weigh each (sub)objective according to its importance
- Determine the amount of work for each assessment task
  - Make sure that each task can be completed within the allotted time
- Verify that each task is sufficient to support the claims in the learning objectives

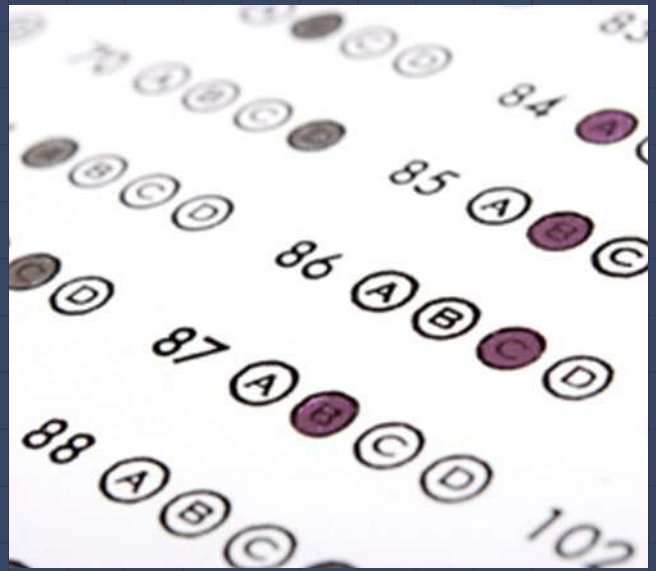
Objective assessment

- What is it?
- Objective vs subjective test
- Types of questions in objective assessment
- True-false question
- Multiple-choice question

2

Objective assessment

An assessment whose correct answers to the questions must be predetermined.



https://www2.le.ac.uk/projects/social-worlds/all-articles/education/multiple-choice

OBJECTIVE VS. SUBJECTIVE TEST

Objective

- short answer
- closed response
- mostly recognition, limited production
- difficult to write well
- quick and easy to grade
- reliable
- workload “up front”

Subjective

- long answer
- open response
- emphasis on production
- relatively easy to write
- difficult to grade
  - time-consuming
  - inter-rater reliability
- not as reliable
- workload post test

http://ethan-lucas.blogspot.com/2016/11/blog-4-testing.html

Types of questions in objective assessment

- True-false question
- Multiple true-false question (MTF)
- Multiple-choice question (MCQ)
- Two-tier MCQ
- Data sufficiency question
- Assertion-reason question
- Matching question
- Fill-in-the-blank question

Writing true-false question

- Focus on one key idea
- Use simple words and statements
- Clearly true or false
- Good mix of true and false answers
- Avoid negative statements
- Avoid qualifying words: always, never, every, …
- Keep the length short

Scoring multiple true-false question

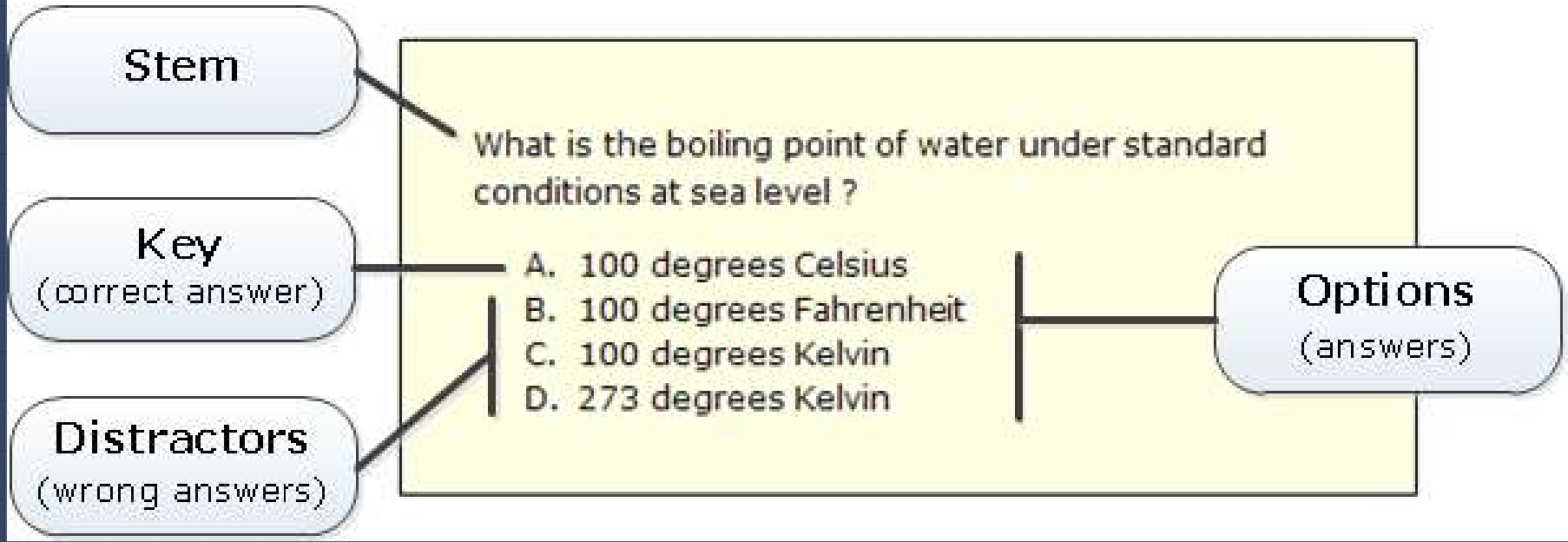
- All or nothing
- One point each
- 2 points if all are correct or
  1 point if
  - Half the answers are correct
  - 3/4 of the answers are correct

A virus:

*T F	a. Can cause disease.
T *F	b. Can reproduce by itself.
T *F	c. Is composed of large living cells.
*T F	d. Lives in plant and animal cells.

https://www.slideshare.net/katya923271/alternative-response

The Anatomy of an MCQ



http://www.intelligentassessment.com/examonline/examonline-resources/writing-good-mcqs/

Writing multiple-choice question

- Instruct students to select the best answer
- Express the full situation in the stem
- Make the distractors appealing and plausible
- Avoid conflicting and overlapping choices
- Avoid all and none of the above choices

Two-Tier MCQ

- First tier: normal MCQ
- Second tier: reason
  - Multiple choices
  - Open-ended
- Scoring
  - All or nothing
  - One point each
  - Coherence
  - Second tier higher

Topic: Respiration

Item 1

In plants the process of respiration takes place in:

A) Only in the cells of the leaves ☐

B) All the cells in a plant ☐

C) Only in the cells of the roots ☐

The reason for my answer is only because:

1. Every living cell needs energy to live. ☐
2. The stomata which are only on the leaves of the plant allow for gas exchange. ☐
3. The stomata which are only on the roots of the plant allow for gas exchange. ☐
4. Only roots need energy; which enables them to absorb water. ☐

Adapted from Mann & Treagust 2000

Data Sufficiency Question

- 1 question
- 2–3 statements
- 5 choices
  a) Statement 1 alone is sufficient
  b) Statement 2 alone is sufficient
  c) Statement 1 or 2 is sufficient
  d) Statements 1+2 are sufficient
  e) None of the above

- Who is the tallest among A, B, C, D, and E?
- Statement 1: B is only taller than D
- Statement 2: A is shorter than E but taller than C

https://byjus.com/govt-exams/data-sufficiency-questions-reasoning/

Assertion-Reason Question

- 2 statements
  - Assertion
  - Reason
- 4–5 choices
  a) Both true: R explains A
  b) Both true: R does not explain A
  c) A is true but R is false
  d) A is false but R is true
  e) Both are false

- Assertion: Increased consumption of sugary beverages is linked to a higher risk of obesity
- Reason: Sugary beverages are high in empty calories and can lead to an overconsumption of calories without providing significant nutritional value

Generated by ChatGPT

Matching Question

- 2 lists
- One to one
- One to many

- Thai
- Joey
- Elephant
- Lisa

- CD
- Blackpink
- Lottery
- Kancore Club

Item analysis and reliability

- Item analysis
- Item difficulty index
- Item discriminability index
- Internal reliability

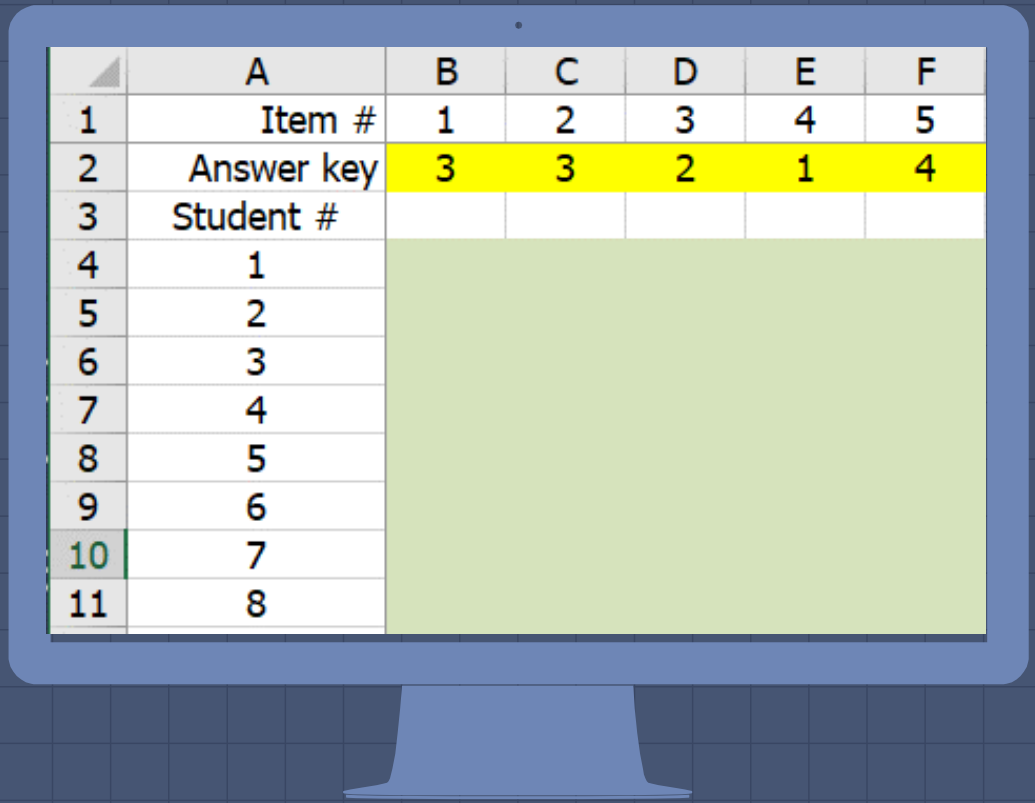
3

Answer keys

- The correct answer for each item
- Row 2

Student answers

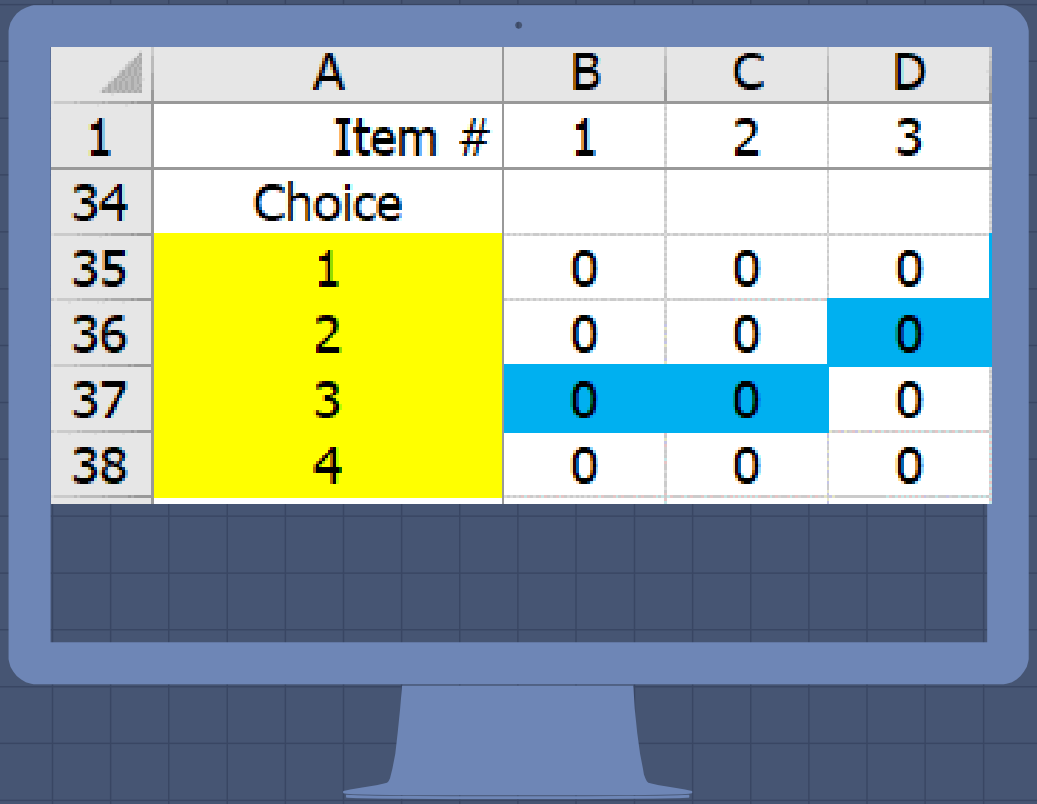
- From cell B4



	A	B	C	D	E	F
1	Item #	1	2	3	4	5
2	Answer key	3	3	2	1	4
3	Student #					
4	1					
5	2					
6	3					
7	4					
8	5					
9	6					
10	7					
11	8					

Answer count

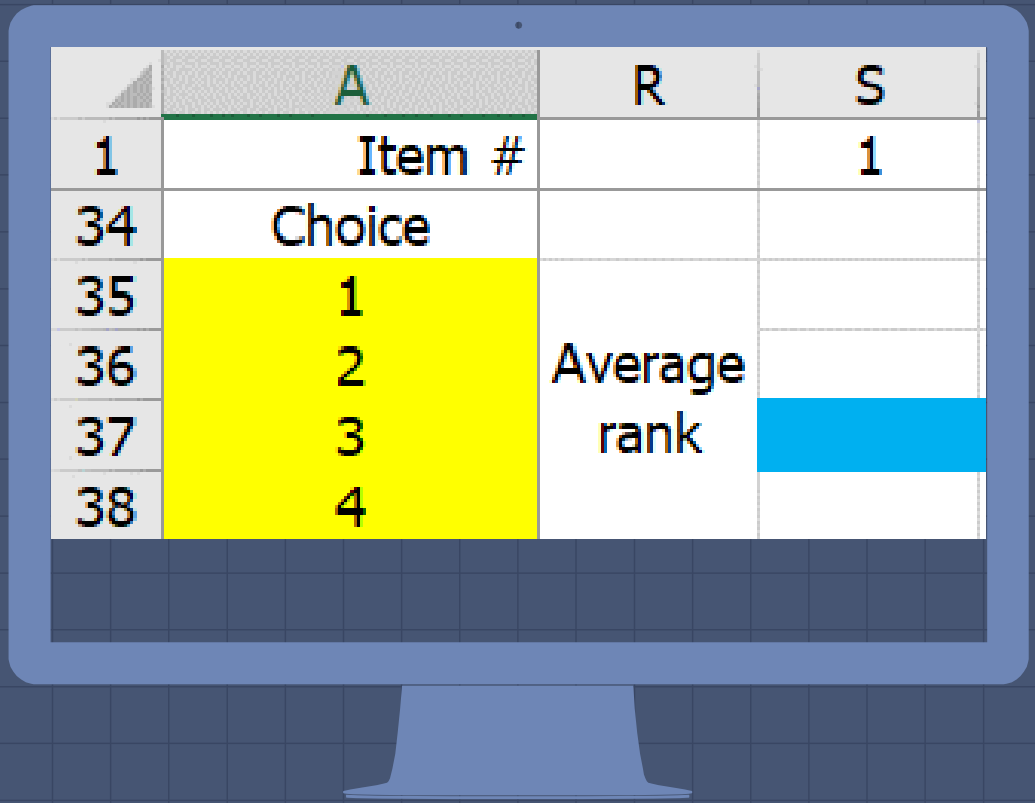
- The number of students who select that choice
- The correct choice is highlighted



	A	B	C	D
1	Item #	1	2	3
34	Choice			
35	1	0	0	0
36	2	0	0	0
37	3	0	0	0
38	4	0	0	0

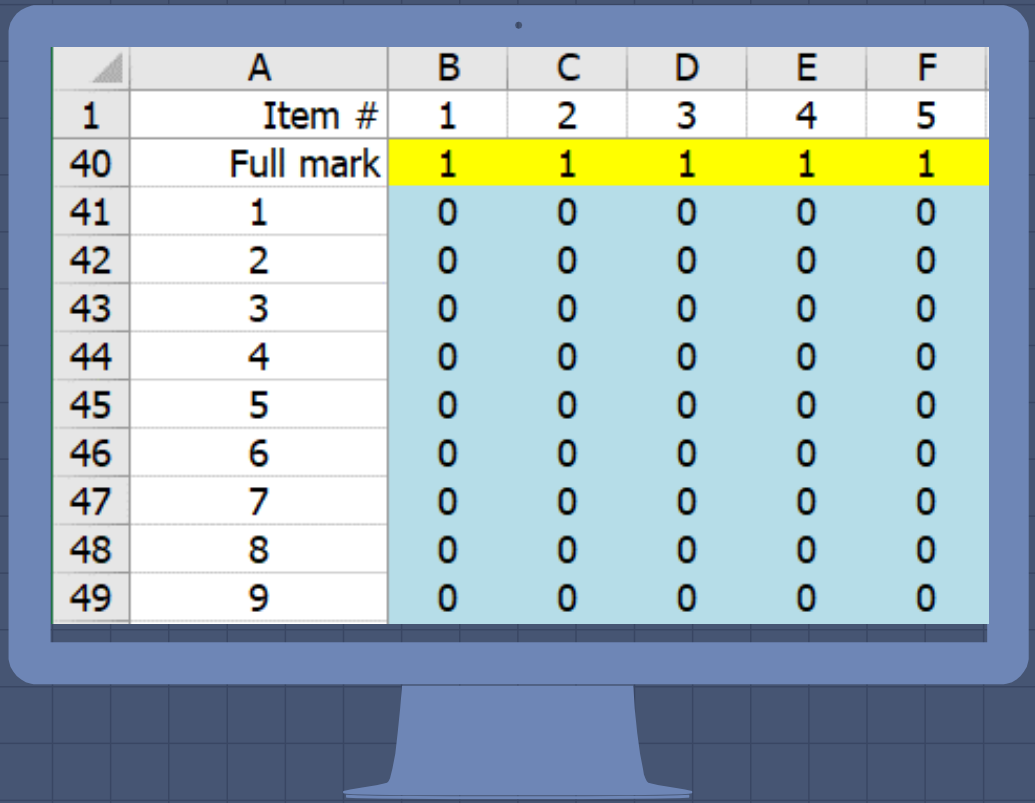
Average rank

- On average, how good are the students who select that choice?
- Rank 1 has the highest total score



Student scores

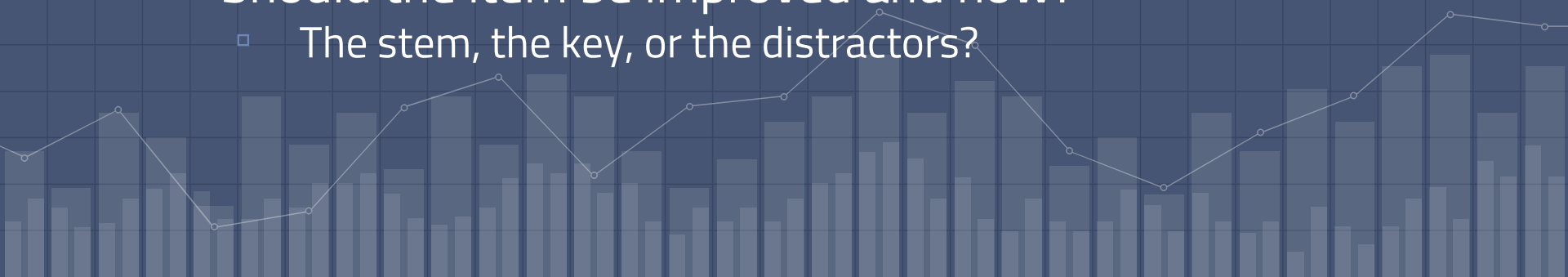
- The score of each student for each item

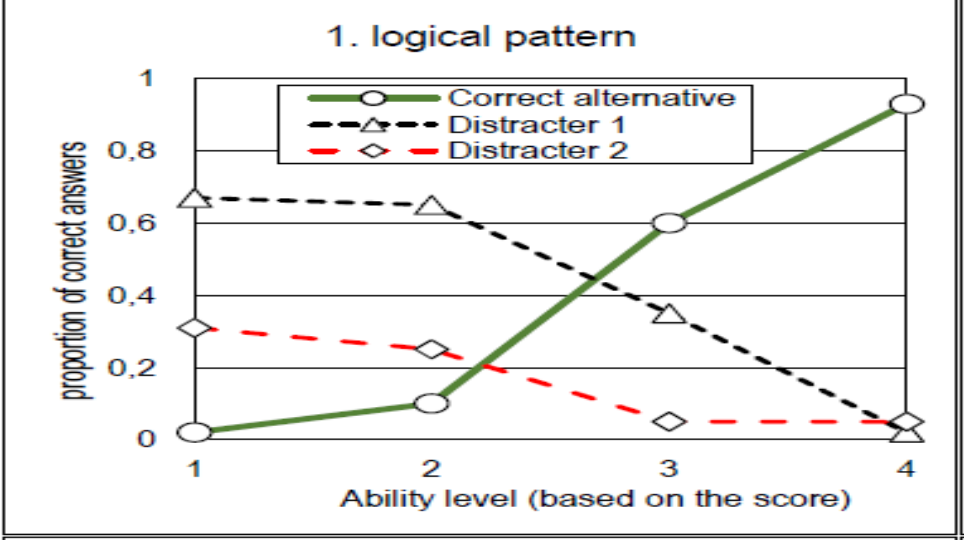


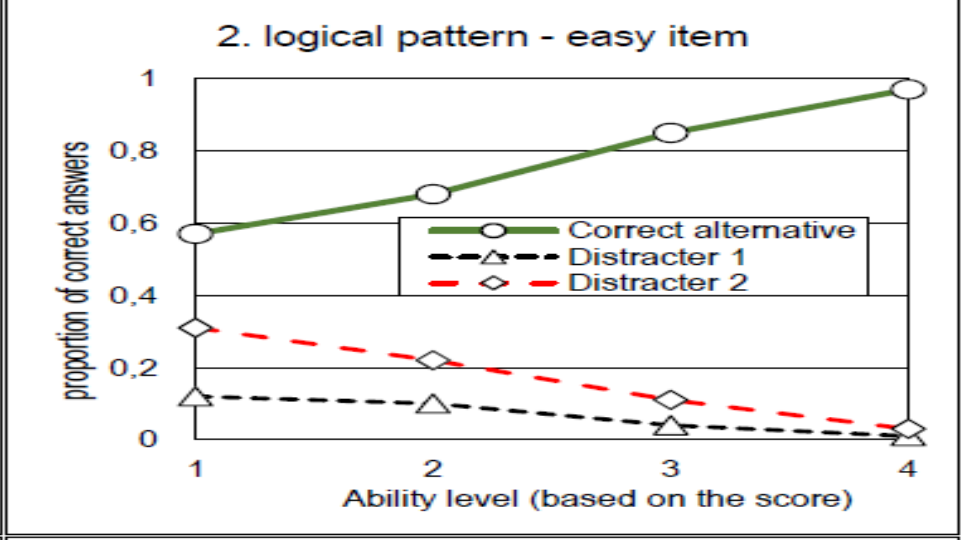
	A	B	C	D	E	F
1	Item #	1	2	3	4	5
40	Full mark	1	1	1	1	1
41	1	0	0	0	0	0
42	2	0	0	0	0	0
43	3	0	0	0	0	0
44	4	0	0	0	0	0
45	5	0	0	0	0	0
46	6	0	0	0	0	0
47	7	0	0	0	0	0
48	8	0	0	0	0	0
49	9	0	0	0	0	0

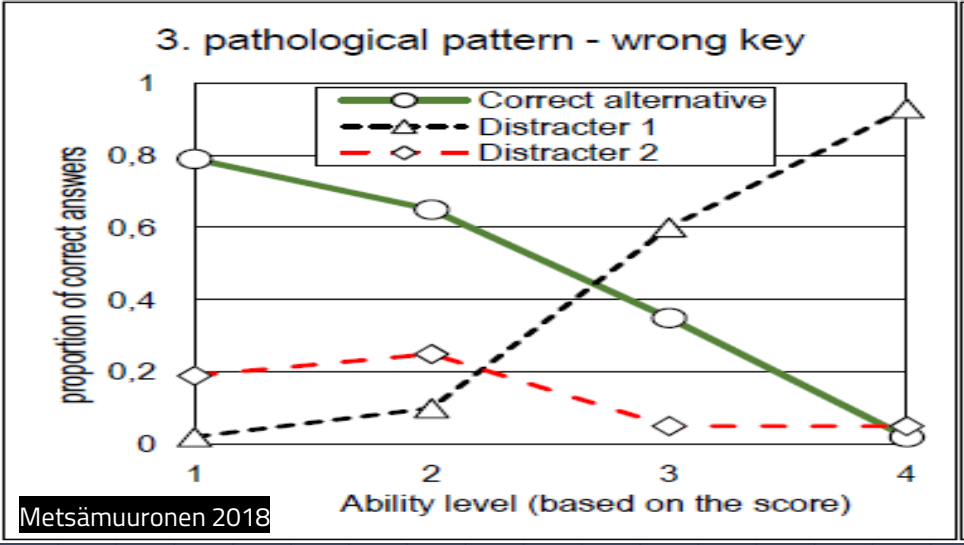
What Does the Excel Sheet Tell Us?

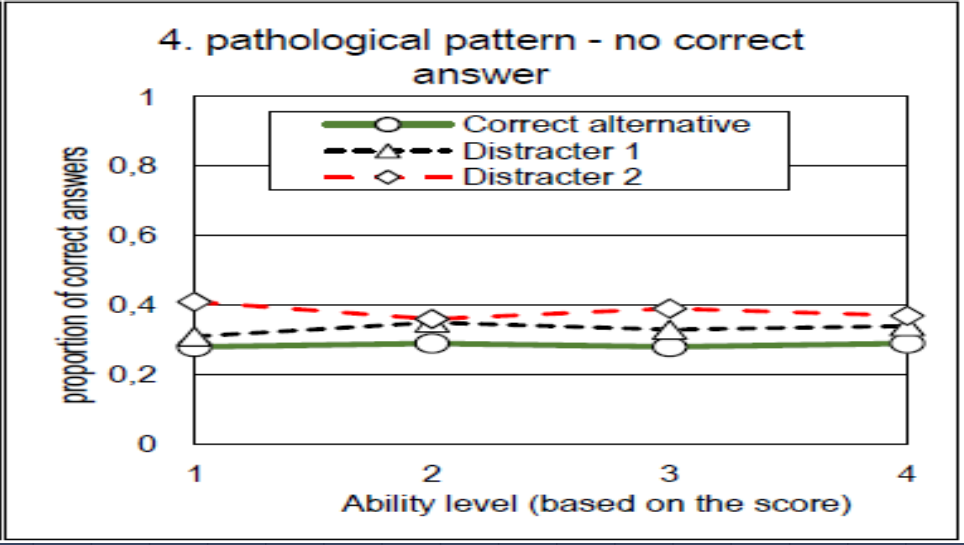
- How many students select each choice?
  - Is the answer key selected most?
  - Which choice is not selected at all?
- How good are those students?
  - Who does a choice appeal to?
  - Should that be expected?
- Should the item be improved and how?
  - The stem, the key, or the distractors?











Metsämuuronen 2018

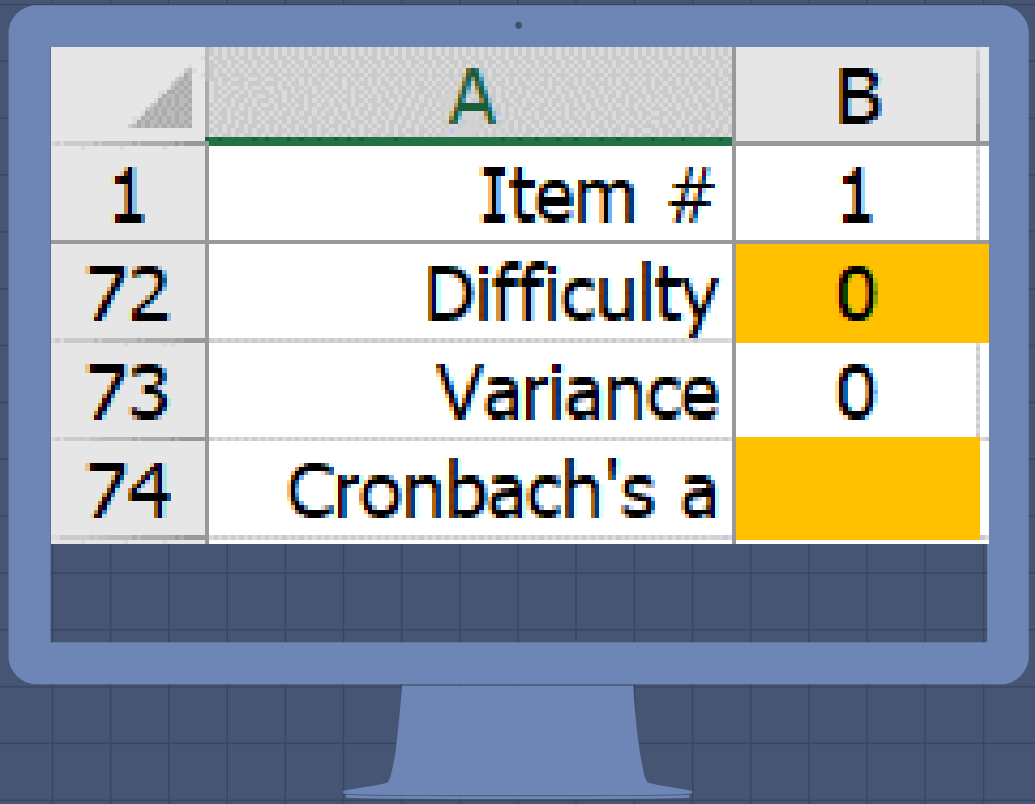
Item difficulty index

- The knowledge/skill level required to answer an item (or passing a test)
- Usually measured by Item Difficulty Index
  - The average score (0–1) of students for that item
  - Should be called Item Easiness Index
  - Acceptable range depends on the type of the item
  - For a dichotomous item (0 or full mark), it is the proportion of correct answers

Difficulty

- Item Difficulty Index
- The Difficulty Index for the whole test is on the right (same row)

(Cronbach's alpha is the measure of reliability)



	A	B
1	Item #	1
72	Difficulty	0
73	Variance	0
74	Cronbach's a	

Interpreting Item Difficulty Index

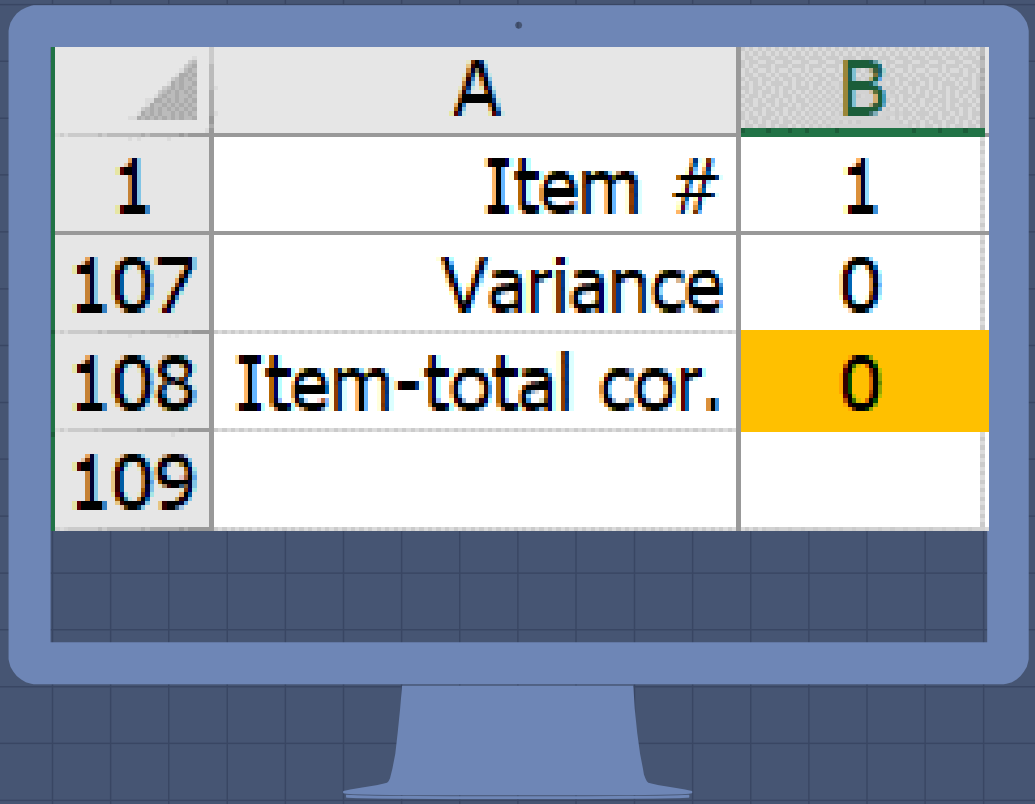
No choice	True false	4 choices	5x5 pairs	Difficulty
> 0.80	> 0.9	> 0.85	> 0.84	Too easy
0.61–0.80	0.81–0.90	0.71–0.85	0.69–0.84	Easy
0.40–0.60	0.70–0.80	0.55–0.70	0.52–0.68	Medium
0.20–0.39	0.60–0.69	0.40–0.54	0.36–0.51	Hard
< 0.20	< 0.60	< 0.40	< 0.36	Too hard

Item discrimination index

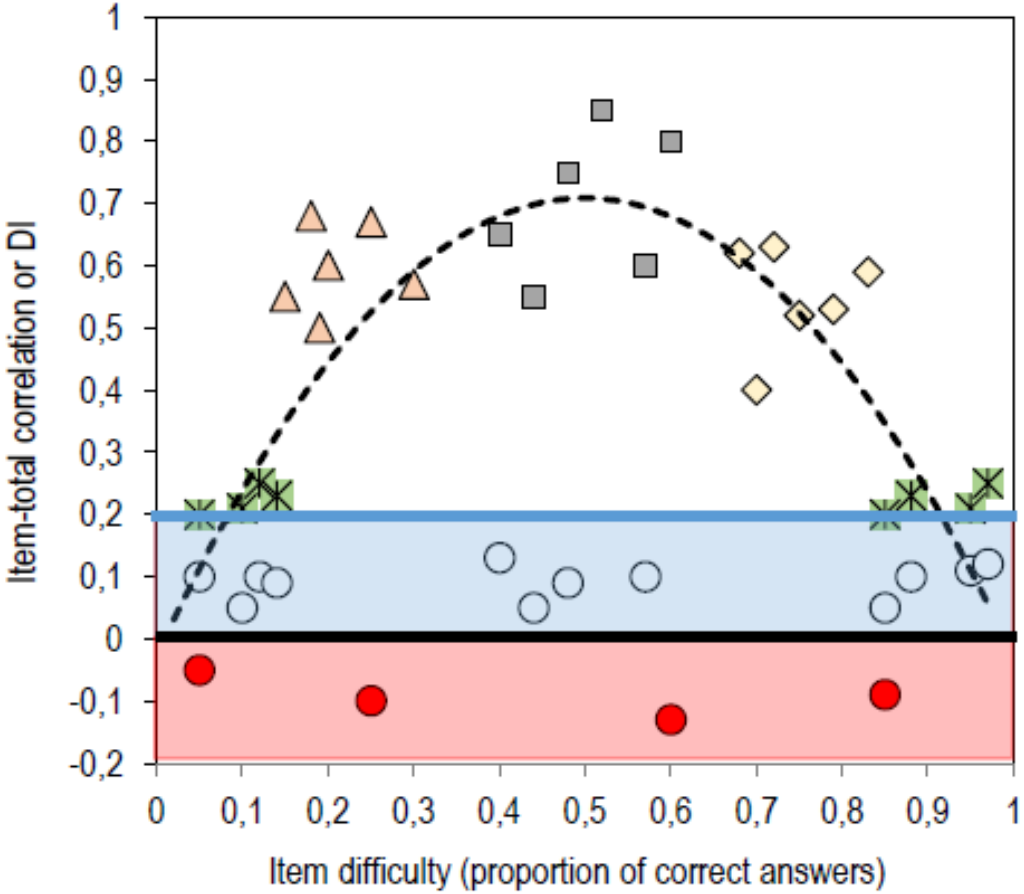
- The effectiveness of an item at discriminating those who know the content from those who do not OR
- The extent to which the success on an item corresponds to the success on the whole test
- Usually measured by Item Discrimination Index
  - The correlation between the score for an item and the score for all other items (corrected item-total correlation)

Discriminability

- Corrected item-total correlation
- The average correlation for the whole test is on the right (same row)



	A	B
1	Item #	1
107	Variance	0
108	Item-total cor.	0
109		



- Items discriminating the lower- and higher-than-medium-scoring ones
- Items discriminating the high-scoring ones from even higher-scoring ones
- Items discriminating the low-scoring ones from even lower-scoring ones
- items with low discrimination power because of technical reasons
- items with too LOW discrimination power
- items with patologically NEGATIVE discrimination

THANKS!

Any questions?

You can find me at

- parames.lao@mahidol.edu