130 лет психометрике: от науки о тестах знаний к вычислительной науке о поведении

Дмитрий Аббакумов, *PhD KU Leuven in Educational Sciences*

Семинар Центра психометрики и измерений в образовании ИО НИУ ВШЭ

23 ноября 2020

Аннотация

Психометрика, наука о психологических и педагогических измерениях, отметила (и продолжает отмечать) 130-летие. За это время из науки, основным подходом которой был подсчет правильных ответов на тестовые задания, психометрика превратилась в сложную междисциплинарную область знания на стыке психологии, педагогики, математической статистики, науки о данных и машинного обучения. На лекции мы познакомимся с историей развития психометрики, задачами, которые решала и решает эта наука, ее достижениями последних лет и перспективными областями для новых исследований.

Psychometrics is a scientific discipline concerned with the construction of assessment tools, measurement instruments, and formalized models that may serve to connect observable phenomena to theoretical attributes.

prof. dr. Denny Borsboom, UvA

'. . . it is intelligible to speak of the mean judgment of competent critics as the true judgment; and deviations from that mean as errors'

(Edgeworth, 1888, p. 622)

Francis I. Edgeworth (1845–1926)

James McKeen Cattell (1860–1944) и его психометрическая группа в Кембридже (1888/89)

- Научный метод в психологии
- Измерения интеллекта
- Президент American Psychological Association в 1895 г.
- Cattell, J. M. (**1890**). Mental tests and measurements. *Mind*, *15*, 373-381

- Measurement of Attitudes, 1929
- Primary Mental Abilities, 1938
- Factorial Studies of Intelligence, 1941
- Multiple Factor Analysis, 1947
- Основатель и первый президент **Психометрического общества** (1935/36)

Louis L. Thurstone (1887–1955)

SCIENCE

VOL. 85 FRIDAY, MARCH 5, 1937 No. 2201

SCIENCE: A Weekly Journal devoted to the Advancement of Science, edited by J. MCKEEN CATTELL and published every Friday by

THE SCIENCE PRESS

New York City: Grand Central Terminal
Lancaster, Pa. Garrison, N. Y.

Annual Subscription, $6.00 Single Copies, 15 Cts.

SCIENCE is the official organ of the American Association for the Advancement of Science. Information regarding membership in the Association may be secured from the office of the permanent secretary, in the Smithsonian Institution Building, Washington, D. C.

PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE[1]

By Professor L. L. THURSTONE

THE UNIVERSITY OF CHICAGO

THE purposes of this society are not new, but they represent an emphasis and direction which have not hitherto received major consideration in psychological science. It seems proper that we should devote some share of our first program meeting to a consideration of our main objectives.

Our main purpose is briefly stated in the subtitle of the new journal, *Psychometrika*, namely, to encourage the development of psychology as a quantitative rational science. More briefly, this may be called mathematical psychology. We should justify our emphasis upon quantification and upon rationalization in science, as well as our conception of the fundamental nature of science.

I assume that we are in complete agreement that we can not suddenly quantify our comprehension of psychological phenomena over their entire range. As psychologists, we are as interested as ever in making exploratory studies of new psychological effects and in discovering hitherto unknown effects. At present, the range of psychological phenomena that can be profitably reduced to mathematical formulation is limited, and it is likely that every man who works on a problem of mathematical psychology will also concern himself with exploratory studies of other problems that are as yet too new for detailed rationalization.

After the discovery of a psychological effect, we naturally turn to the second phase of scientific inquiry, namely, to relate the new effect in a simple descriptive manner to what is already known. In this stage theories are devised to explain the experimentally known effects, and we try, of course, to make psychological theories less complicated than the effects that are to be explained. In this phase the descriptions of psycho-

[1] Abstract of address by the retiring president of the Psychometric Society at Hanover, N. H., September 4, 1936.

Dorothy C. Adkins (1912–1975)
Президент Психометрического общества (1949/50)

- Adkins, D. C. (1947). Construction and analysis of achievement tests.
- Adkins, D. C. (1958). **Measurement in relation to the educational process**. *Educational and Psychological Measurement*, *18*, 221-240.
- Flanagan, J. C., Adkins, D., & Cadwell, D. H. B. (1950). Major developments in examining methods.

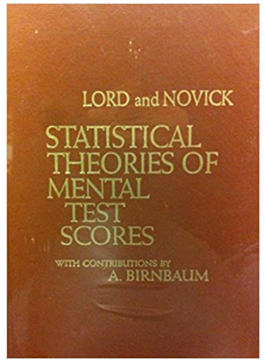
Классическая психометрическая теория

$$X = T + E$$

X – наблюдаемый балл
T – истинная подготовленность
E – ошибка измерения

Melvin R. Novick (1932–1986)
През. ПО (1979/80)

Frederik M. Lord (1912–2000)
През. ПО (1958/59)



Lord & Novick (1968)

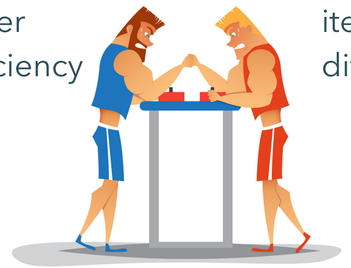
Frederik M. Lord (1912–2000)
Item Response Theory

Georg Rasch (1901–1980)
The Rasch Model

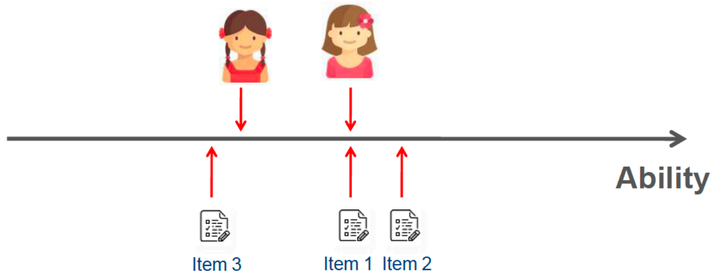
Современная психометрическая теория

Вероятность правильного ответа на задание описывается функцией разности уровня подготовленности студента и уровня трудности задания

learner proficiency



item difficulty



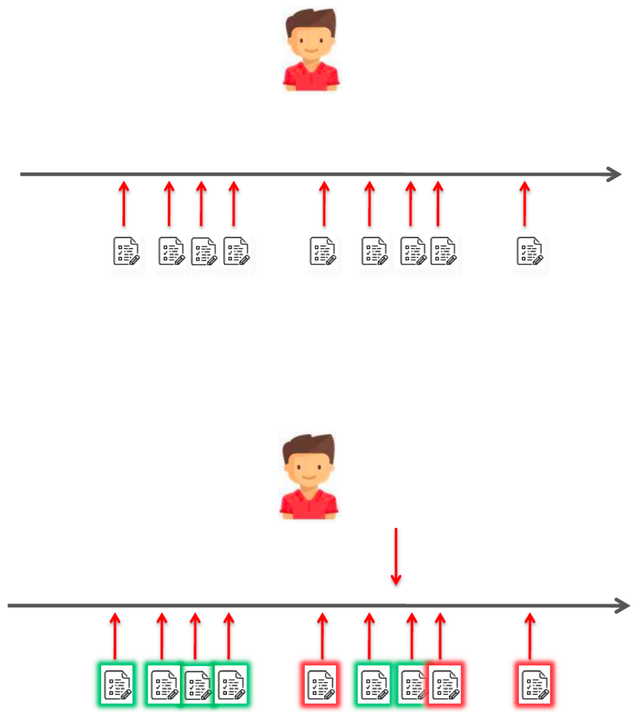
Rasch simple logistic model (SLM)

$$P_{ni} = \frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}}$$

One-parameter logistic IRT model

$$P_{ni} = \frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}}$$

Современная психометрическая теория



Jöreskog, K. G., & Sörbom, D. (1979). Advances in factor analysis and structural equation models.

Karl G. Jöreskog
През. ПО (1984/85)

Bloxom, B. (1985). Considerations in psychometric modeling of response time. *Psychometrika*, *50*(4), 383-397

Bruce Bloxom
Prez. PO (1984/85)

John M. Linacre

Linacre, J. (1992). Many-Facet Rasch Measurement

'. . . it is intelligible to speak of the mean judgment of competent critics as the true judgment; and deviations from that mean as errors'

(Edgeworth, 1888, p. 622)

Susan Embretson
През. ПО (1998/99)

Embretson, S. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika*, *64*(4), 407-433

van Der Linden, W., & Glas, C. (2000). Computerized Adaptive Testing: Theory and Practice.

Wim Van der Linden
През. ПО (1999/00)

Cees Glas
През. ПО (2017/18)

De Boeck, P., & Wilson, M. (2004). Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach.

Paul De Boeck
През. ПО (2007/08)

Mark Wilson
През. ПО (2011/12)

Frank Rijmen
Dissertation Prize 2003

A mixed and mixture model perspective on item response theory (prom. P. De Boeck)

Wim Van den Noortgate

Cross-classification multilevel logistic models in psychometrics (prom. P. De Boeck)

Психометрика в России

- в 1920-х бурно развивалась как часть экспериментальной педагогики и психологии
- в 1936 признана лженаукой (как и генетика, а после и кибернетика)
- в 1970-80–х начинается возрождение в рамках психодиагностики
- отставание не преодолено до сих пор, но…

Нидерландское психометрическое чудо

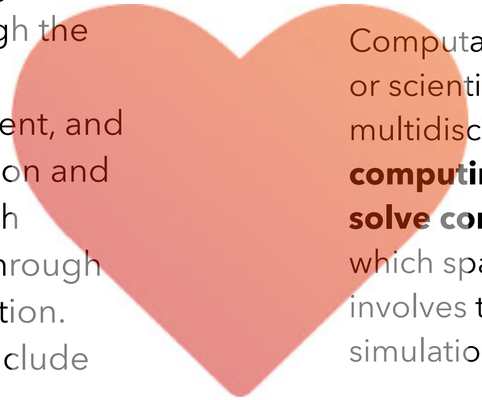
Van der Heijden, P. & Sijtsma, K. (1996). Fifty years of measurement and scaling in the Dutch social sciences. *Statistica Neerlandica*, *50*, 111-135

- от национальной к интернациональной ориентации, от изоляции к экспансии
- от нескольких публикаций в национальных журналах к осознанному увеличению публикационной активности на международном уровне
- от поддерживающей статистики к продвинутым разработкам в области методологии
- от малых наборов данных к большим данным

Сегодня Нидерланды и Фландрия глобальные инфлюенсеры в области психометрики

Computational Behavioral Science

Behavioral sciences **explore the cognitive processes** within organisms and **the behavioral interactions** between organisms in the natural world. It involves the systematic analysis and investigation of human and animal behavior through the study of the past, controlled and naturalistic observation of the present, and disciplined scientific experimentation and modeling. It attempts to accomplish legitimate, objective conclusions through rigorous formulations and observation. Examples of behavioral sciences include **psychology, psychobiology, anthropology, and cognitive science**. Generally, behavior science deals primarily with human action and often seeks to generalize about human behavior as it relates to society.



Computational science (also scientific computing or scientific computation (SC)) is a rapidly growing multidisciplinary field that uses **advanced computing capabilities to understand and solve complex problems**. It is an area of science which spans many disciplines, but at its core it involves the development of models and simulations to understand natural systems.

Combining explanatory IRT and psychological networks for understanding and modeling online learners' difficulties

Context and Problem

Massive open online courses (MOOCs) are scaled, internet-based courses developed by universities or industries in partnership with provider platforms such as Coursera (55 million learners, 3,800 courses). A MOOC consists of video lectures, reading assignments, assessments, and forums.

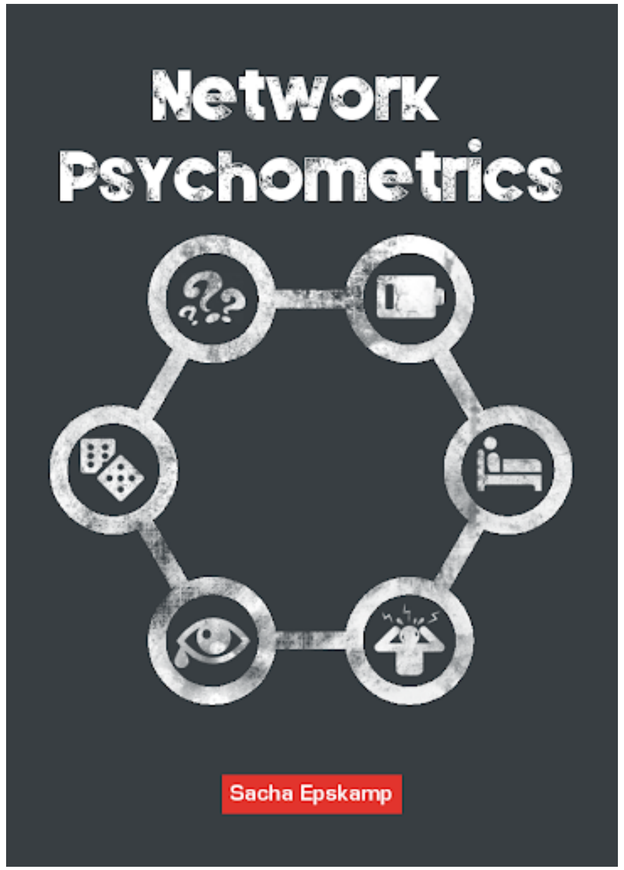
Context and Problem

Massive open online courses (MOOCs) are scaled, internet-based courses developed by universities or industries in partnership with provider platforms such as Coursera (55 million learners, 3,800 courses). A MOOC consists of video lectures, reading assignments, assessments, and forums.

MOOC developers need psychometric instruments for understanding learners' difficulties in assessments to provide better structure, content, and feedback in assessments.

- What is the structure of learners' difficulties (mistakes) in assessments? How are they connected?
- How are learners' different in terms of difficulties (mistakes)?
- Do teaching interventions work for decreasing the number of difficulties?

Networks in Psychometrics



The Ising Model (1925) in Psychometrics

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z}(\sum_{i} \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j),$$

$$\mathrm{Z} = \sum_{x} \exp(\sum_{i} \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j)$$

The Ising Model (1925) in Psychometrics

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z}(\sum_{i} \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j),$$

$$\mathrm{Z} = \sum_{x} \exp(\sum_{i} \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j)$$

Kruis and Maris (2016)

Epskamp, Maris, Waldorp, and Borsboom (2018)

$$\Pr(X_i \mid \mathbf{X}^{(-i)} = \mathbf{x}^{(-i)}) = \frac{\exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}{\sum_{x_i} \exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}$$

The Ising Model (1925) in Psychometrics

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z}(\sum_{i} \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j),$$

$$\mathrm{Z} = \sum_{x} \exp(\sum_{i} \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j)$$

$$\Pr(X_i \mid \mathbf{X}^{(-i)} = \mathbf{x}^{(-i)}) = \frac{\exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}{\sum_{x_i} \exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}$$

Kruis and Maris (2016)

Epskamp, Maris, Waldorp, and Borsboom (2018)

`IsingSampler` (Epskamp, 2020)

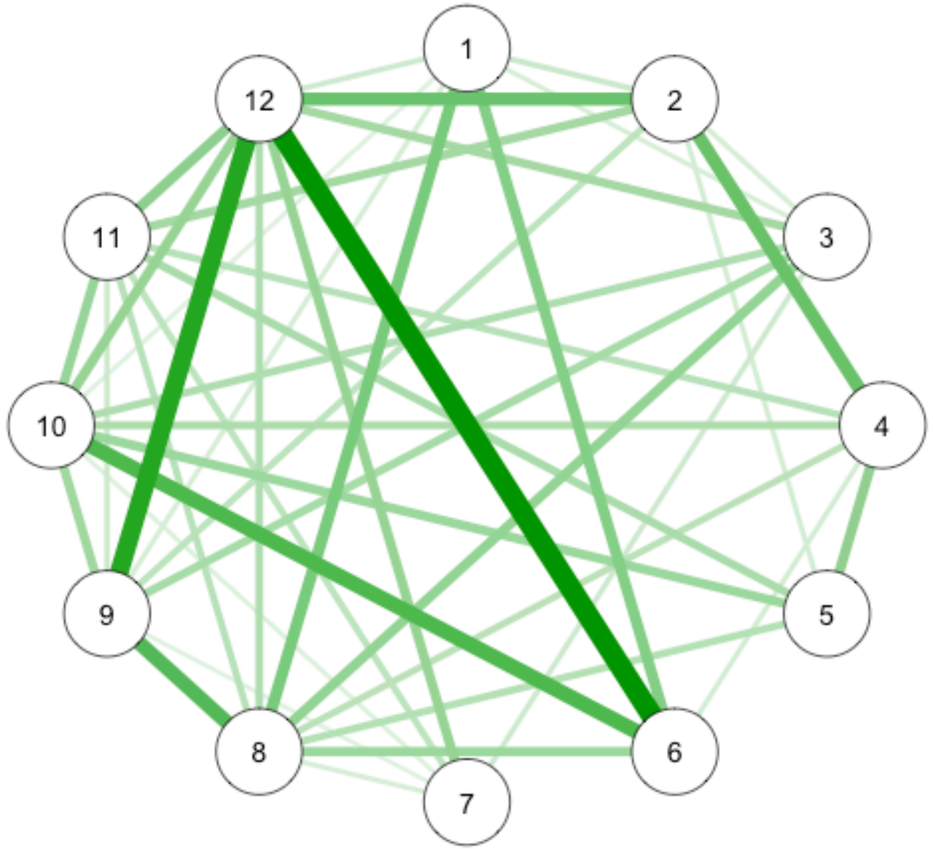
`IsingFit` (van Borkulo & Epskamp, 2016)

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1.05	-2.60	1.06	-0.33	-1.09	1.54	0.37	-0.28	-0.99	-0.20	0.39	0.60

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1.05	-2.60	1.06	-0.33	-1.09	1.54	0.37	-0.28	-0.99	-0.20	0.39	0.60	

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1.05	**-2.60**	1.06	-0.33	-1.09	1.54	0.37	-0.28	-0.99	-0.20	0.39	0.60

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
1.05	-2.60	1.06	-0.33	-1.09	1.54	0.37	-0.28	-0.99	-0.20	0.39	0.60



- What is the structure of learners’ difficulties (mistakes) in assessments? How are they connected?
- How are learners’ different in terms of difficulties (mistakes)?
- Do teaching interventions work for decreasing the number of difficulties?

- What is the structure of learners’ difficulties (mistakes) in assessments? How are they connected?

- How are learners’ different in terms of difficulties (mistakes)?

- Do teaching interventions work for decreasing the number of difficulties?

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z}(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j),$$

$$Z = \sum_x \exp(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j)$$

$$\Pr(X_i \,|\, \mathbf{X}^{(-i)} = \mathbf{x}^{(-i)}) = \frac{\exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}{\sum_{x_i} \exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}$$

Small tuning: $\tau_{i0} + \tau_{iC}$ and $\omega_{ij0} + \omega_{ijC}$, where C is a grouping variable (several grouping variables are possible)

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z}(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j),$$

$$\mathrm{Z} = \sum_x \exp(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j)$$

$$\Pr(X_i \,|\, \mathbf{X}^{(-i)} = \mathbf{x}^{(-i)}) = \frac{\exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}{\sum_{x_i} \exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}$$

Small tuning: $\tau_{i0} + \tau_{iC}$ and $\omega_{ij0} + \omega_{ijC}$, where C is a grouping variable (several grouping variables are possible)

`glmnet` (Friedman et al., 2020) **for C < 5**

$$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{Z}(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j),$$

$$Z = \sum_x \exp(\sum_i \tau_i x_i + \sum_{<ij>} \omega_{ij} x_i x_j)$$

$$\Pr(X_i \,|\, \mathbf{X}^{(-i)} = \mathbf{x}^{(-i)}) = \frac{\exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}{\sum_{x_i} \exp(x_i(\tau_i + \sum_j \omega_{ij} x_i))}$$

Small tuning: $\tau_{i0} + \tau_{iC}$ and $\omega_{ij0} + \omega_{ijC}$, where C is a grouping variable (several grouping variables are possible)

`glmmLasso` (Groll, 2017) **for C > 5**

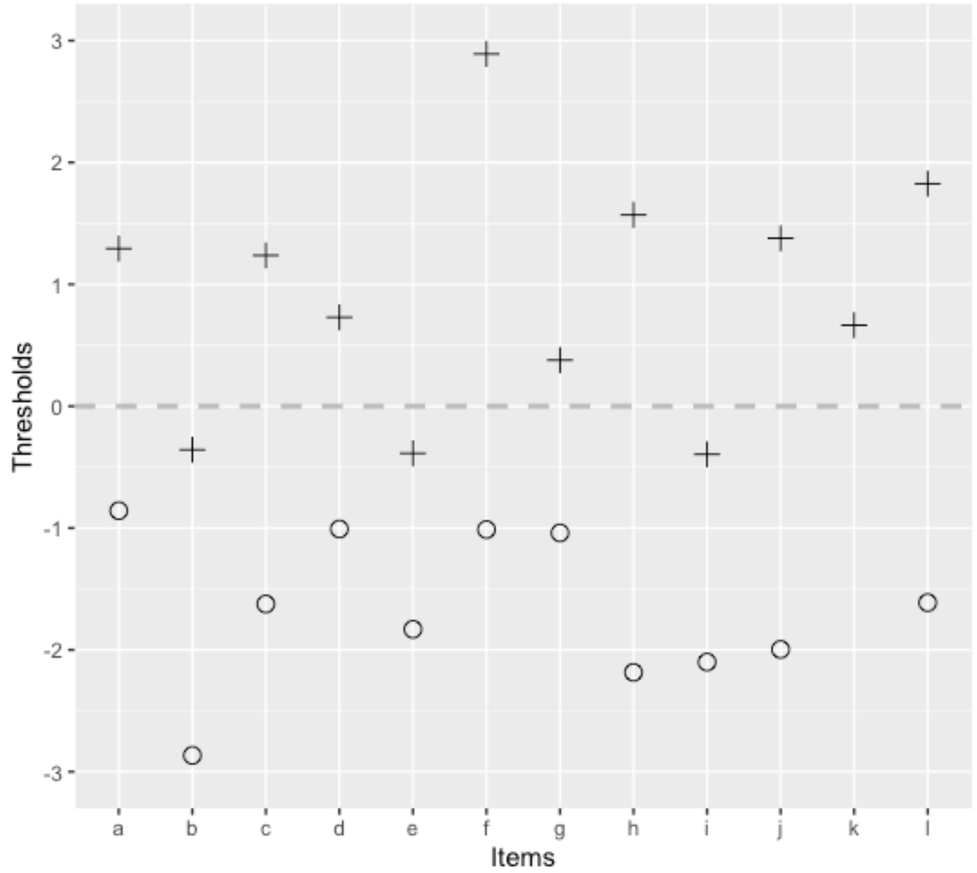
Examples of Grouping

1. Rewatching
2. Performance
3. Context

Illustration Data

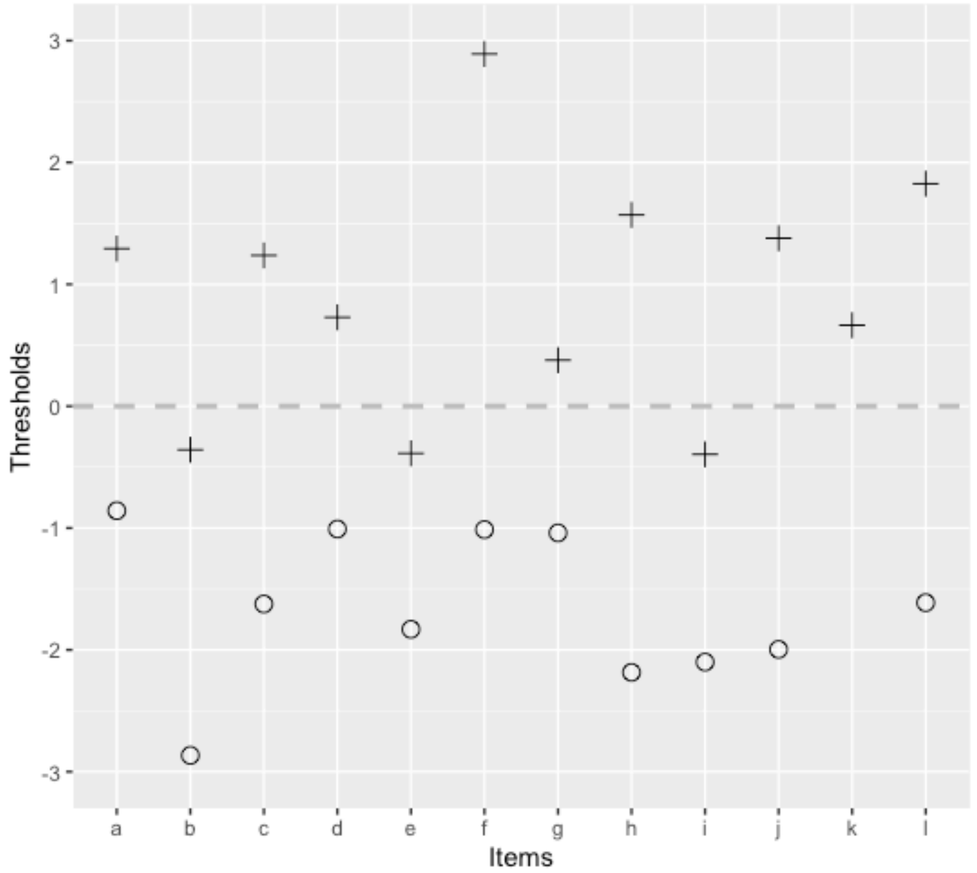
	Course #1
Field	Social Science
Discipline	Sociology
Language	Russian
Number of enrollments	11,259
Completion rate, %	16.9
Top five countries by number of students, %	Russia, 78 Ukraine, 6 Kazakhstan, 2 Belarus, 2 USA, 1
Female Students, %	54

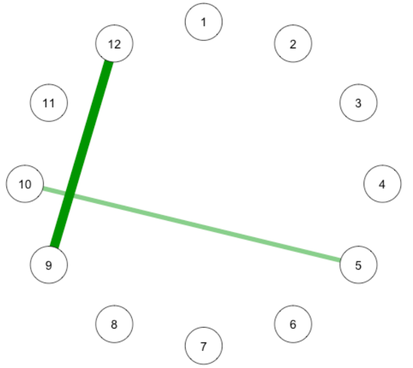
Example 1



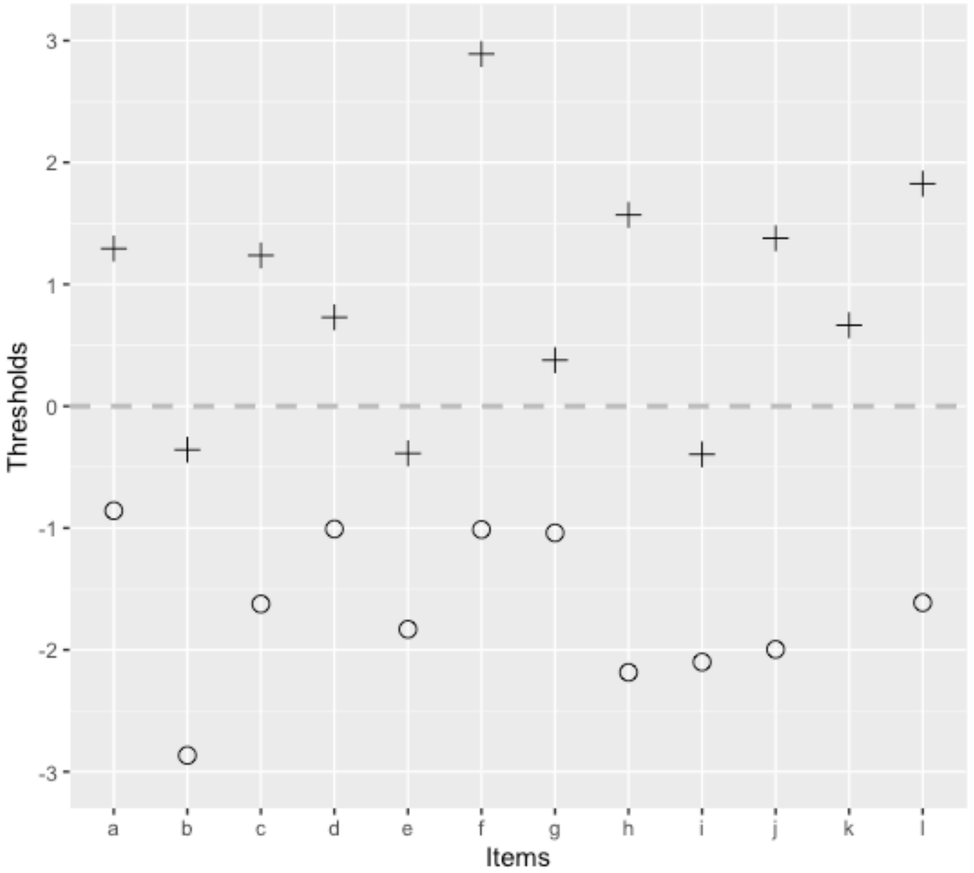
+ — rewatched video lectures
0 — did not rewatch video lectures

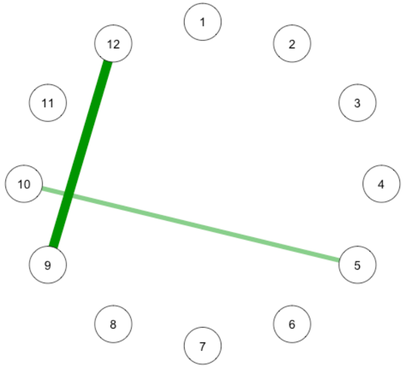
Example 1

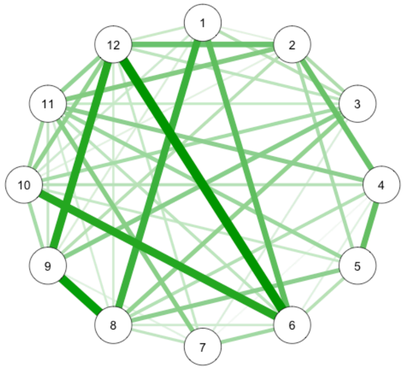




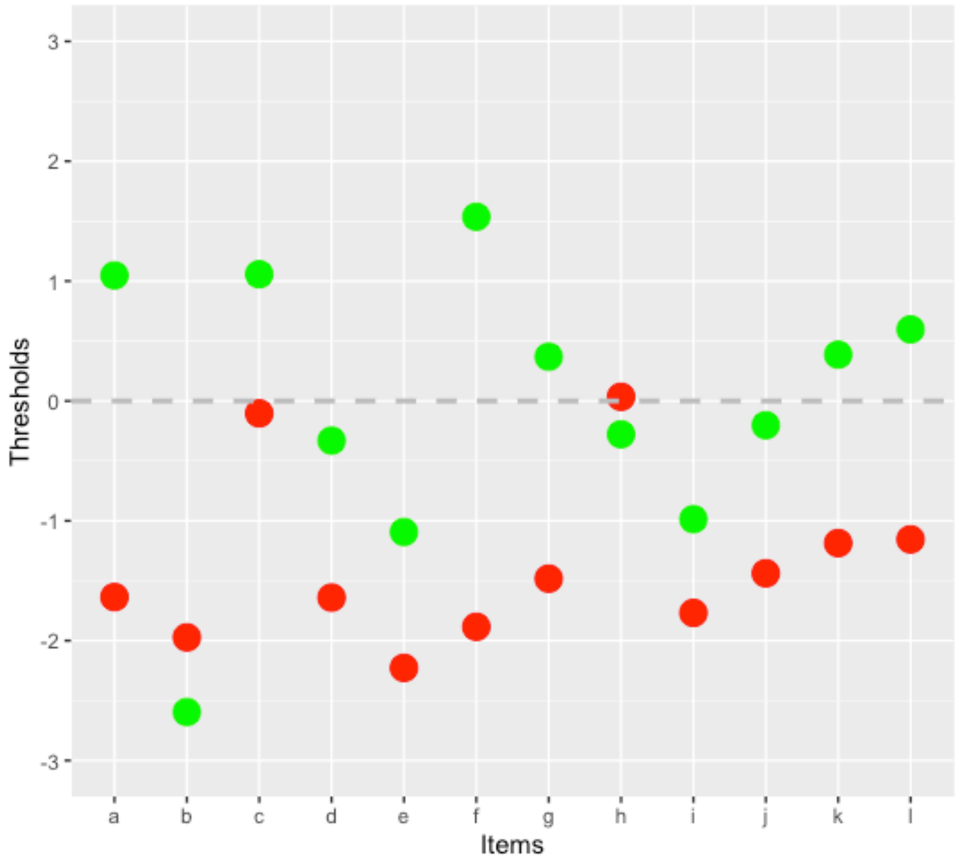
Example 1


Thresholds
3
2
1
0
-1
-2
-3
a
b
c
d
e
f
g
h
i
j
k
l
Items


1
2
3
4
5
6
7
8
9
10
11
12

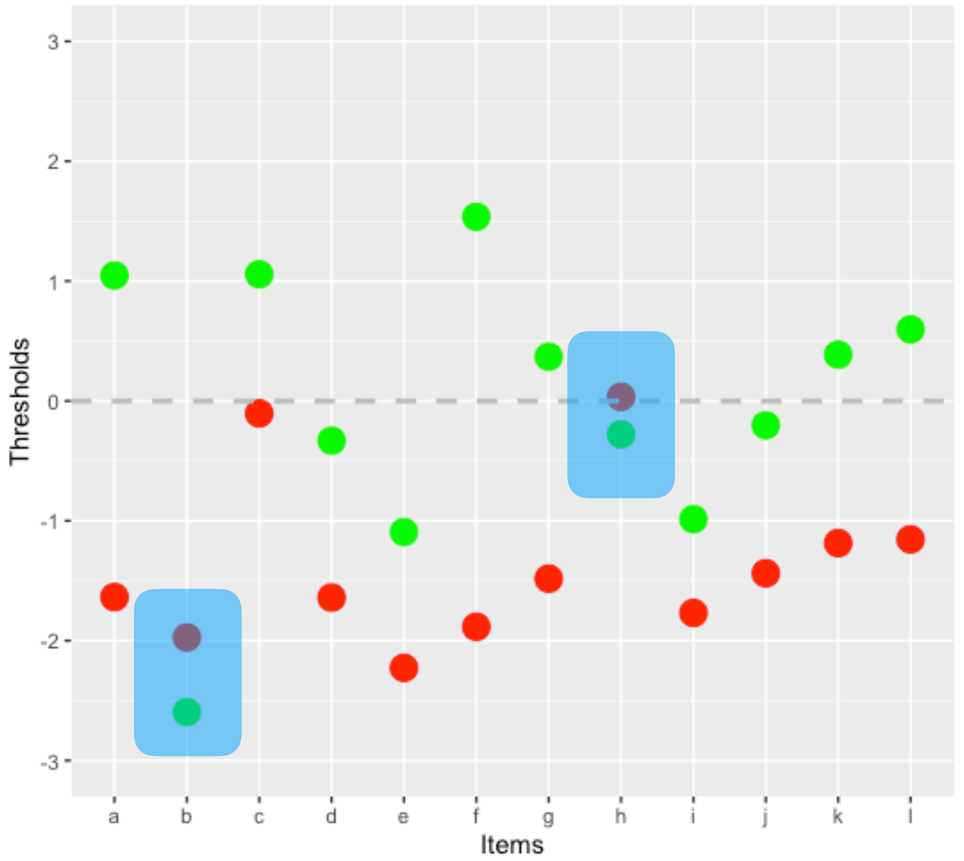

1
2
3
4
5
6
7
8
9
10
11
12

Example 2



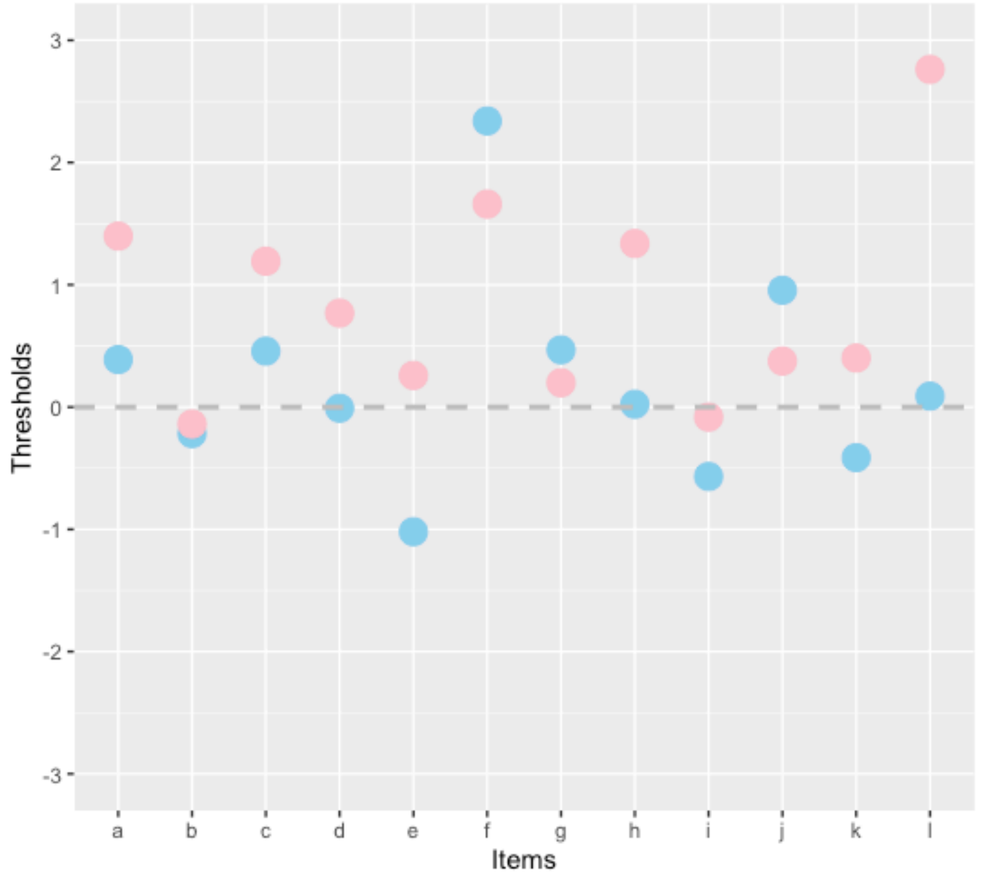
green — have high performance in the course
red — have low performance in the course

Example 2



green — have high performance in previous parts of the course
red — have low performance in previous parts of the course

Example 3



pink — course in curricula
blue —course on demand

Further Steps

- Moving from Revealing to **Explaining** the differences
- Elaborating the Effects of **Learning** and **Forgetting**

Impact to Practice

- Adaptive Learning
- Tuning Feedback

Спасибо!

www.abbakumov.com