Lesson 4: Traditional Item Analysis

Patrícia Martinková

Department of Statistical Modelling
Institute of Computer Science, Czech Academy of Sciences

Institute for Research and Development of Education
Faculty of Education, Charles University, Prague

NMST570, October 23, 2018

Table of contents

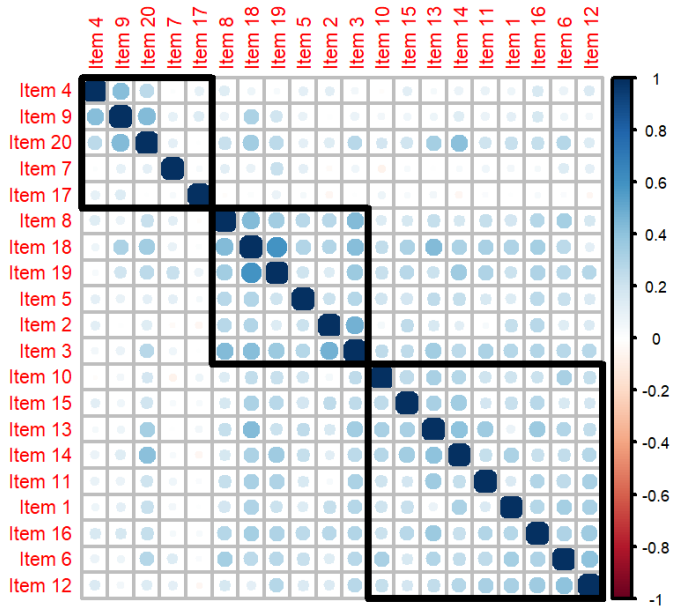
Review – Reliability and validity

- Reliability
  - Test-retest reliability
  - Alternate forms
  - Internal consistency
- Validity
  - Content-related
    - Construct validity
    - Content validty
    - Face validity
  - Criterion-related
    - Concurrent validity
    - Predictive validity
    - Incremental validity
- Correction for range restriction
- Correction for unreliability

Review – Validation studies

Homeostasis concept inventory (HCI)

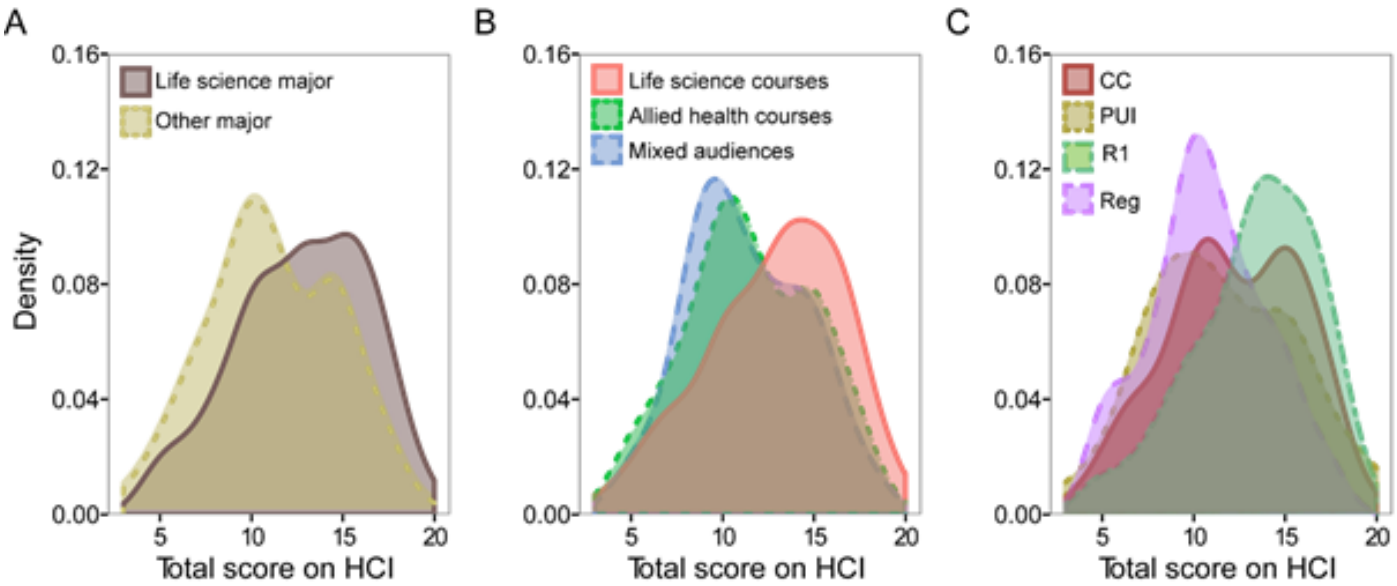
- HCI developed based on Homeostasis conceptual framework (HCF)
- Items test knowledge of individual elements in HCF



McFarland, Price, Wenderoth, Martinková, et al. Development and Validation of the Homeostasis Concept Inventory. CBE Life Sciences Education, 16(2), ar35, 2017.

Review – Validation studies

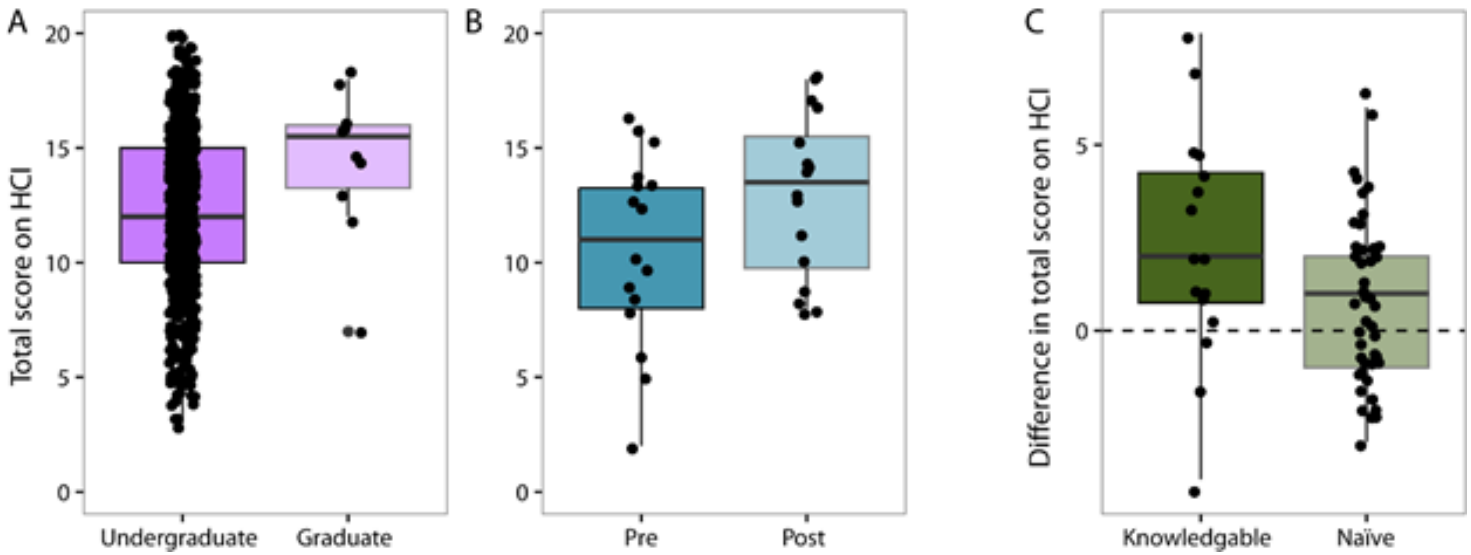
Homeostasis concept inventory (HCI)



McFarland, Price, Wenderoth, Martinková, et al. Development and Validation of the Homeostasis Concept Inventory. CBE Life Sciences Education, 16(2), ar35, 2017.

Review – Validation studies

Homeostasis concept inventory (HCI)



McFarland, Price, Wenderoth, Martinková, et al. Development and Validation of the Homeostasis Concept Inventory. CBE Life Sciences Education, 16(2), ar35, 2017.

Item analysis

Analysis of individual items should be part of test validation process

Conventional/traditional methods use ratios and correlations to describe following item properties:

- Difficulty
  - How hard was the item, how many students answered correctly?
- Discrimination
  - How well does the item discriminate between low ability and high ability students?
- Functioning of distractors
  - How attractive are distractors offered in multiple-choice items?
  - What are the most common misconceptions?
- Unanswered items
  - How often was the item not reached?

Item difficulty

For binary items

$$Y_{ij} \in \{0, 1\}$$

$i = 1, \ldots, n$ students

$j = 1, \ldots, m$ items

Difficulty of item j:

$p_j \ldots$ ratio of those who answered item j correctly

$\bar{Y}_j \ldots$ average number of points in item j

$$p_j = \bar{Y}_j = \frac{\sum_{i=1}^{n} Y_{ij}}{n}$$

- Depends on ability of tested students
  - An item can appear difficult in group of low ability students
  - An item can appear easy in group of high ability students

Item difficulty

For ordinal items

$$Y_{ij} \in 0, 1, \ldots, k_j$$

- Binarizing the data:
  $p_j \ldots$ ratio of those who received maximal score k_j in item j

$$\bar{Y}_j = \frac{\sum_{i=1}^{n} I(Y_{ij} = k_j)}{n}$$

- Using all information:
  $\bar{Y}_j \ldots$ average number of points in item j

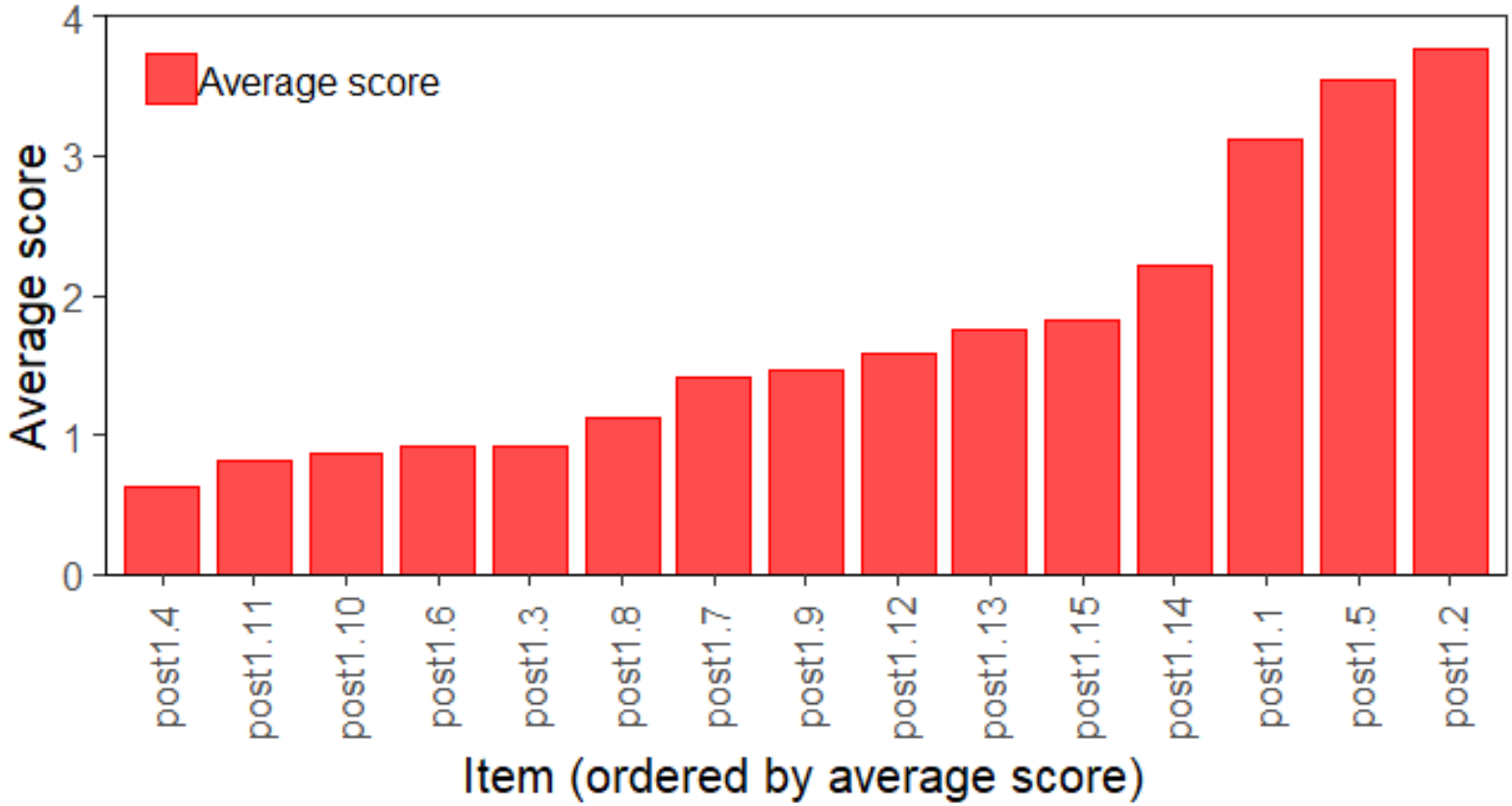
$$\bar{Y}_j = \frac{\sum_{i=1}^{n} Y_{ij}}{n}$$

  $p_j \ldots$ average number of points in item j scaled to interval $[0, 1]$

$$p_j = \frac{\bar{Y}_j}{k_j}$$

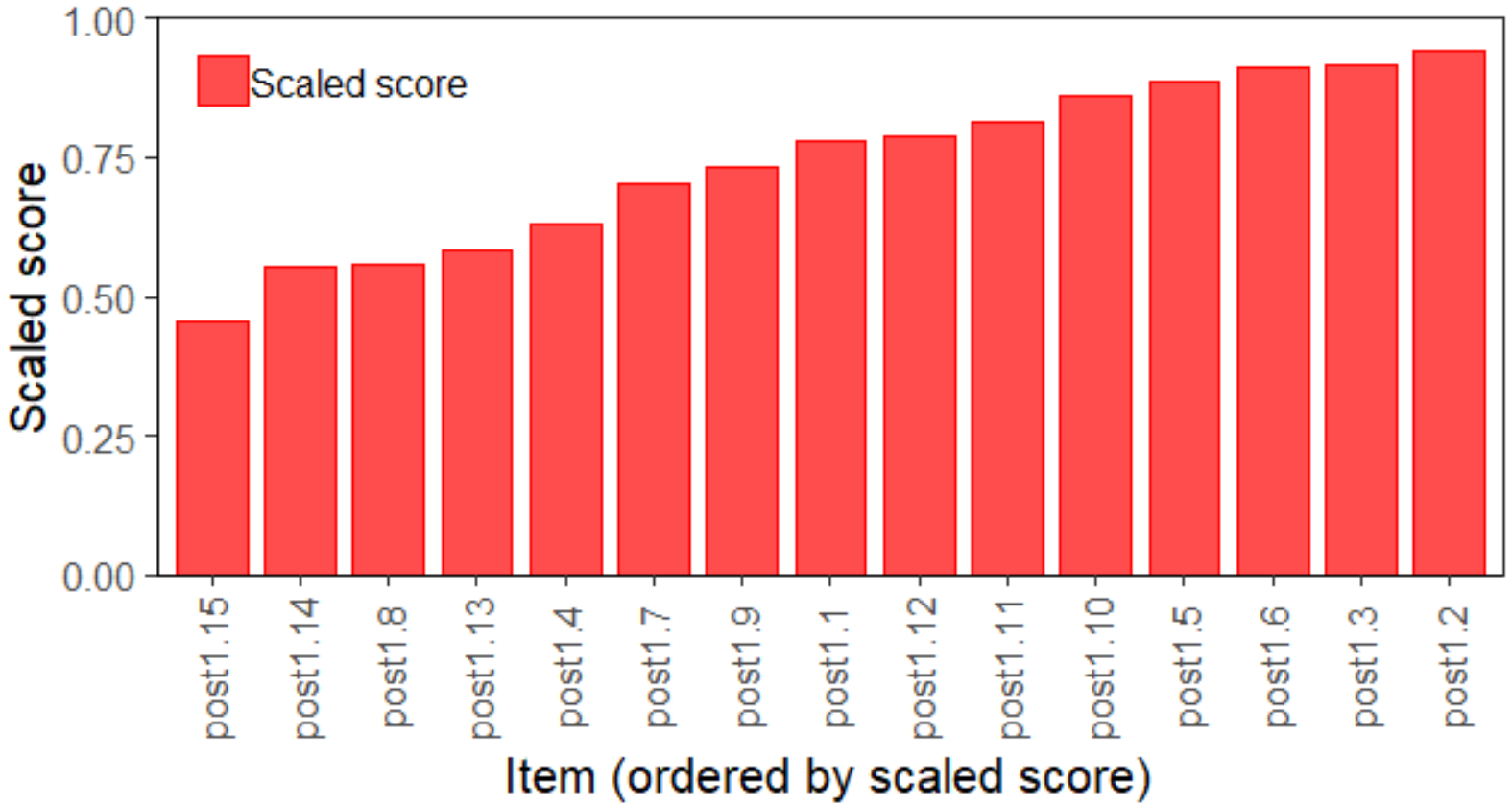
Item difficulty

Average number of points



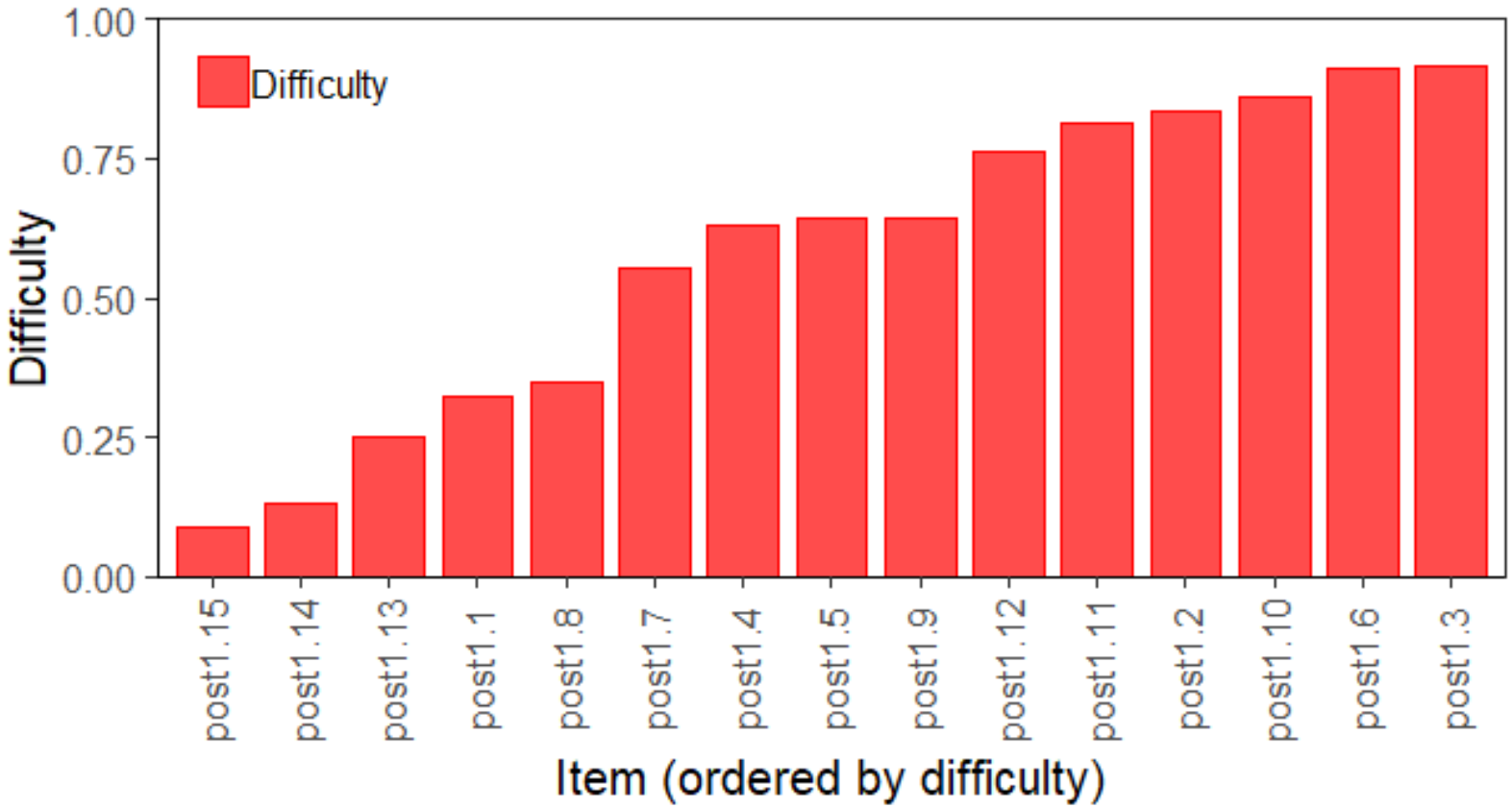
Item difficulty

Average number of points scaled to $[0, 1]$



Item difficulty

Binary approach: ratio of those who received full points



Item Discrimination

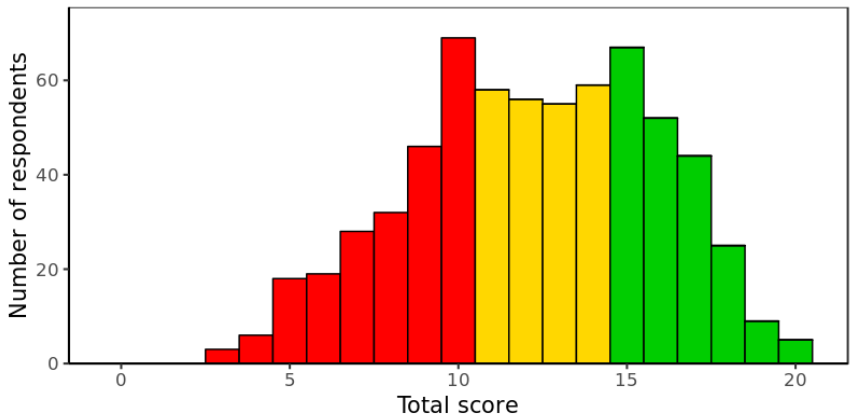
How well does the item discriminate between low ability and high ability students?

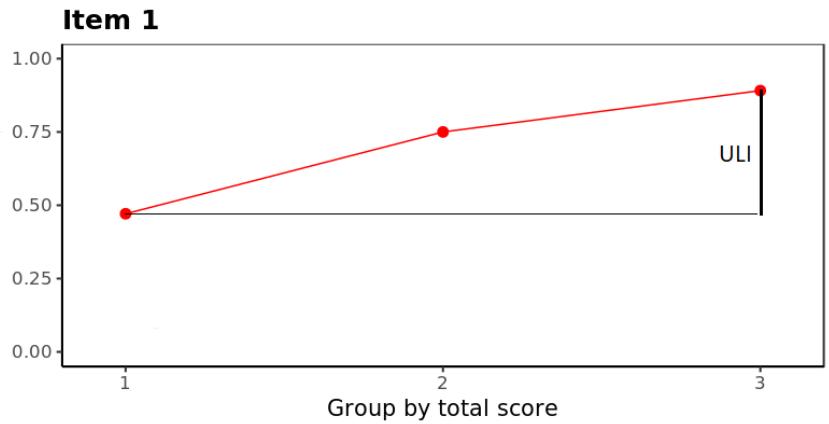
Conventional approaches to item discrimination:

- Upper-lower index (ULI)
- Generalized ULI
- Correlation Item – Test (RIT)
- Correlation Item – Rest (RIR)
- Cronbach's alpha without item

Upper - lower index (ULI)

- Respondents are divided into 3 groups by total score
- Item difficulty (percent correct) is calculated for upper third p_{jU} and lower third p_{jL}
- $ULI = p_{jU} - p_{jL}$





- Rule of thumb: ULI should be >0.2

  (except for very easy and very hard items)

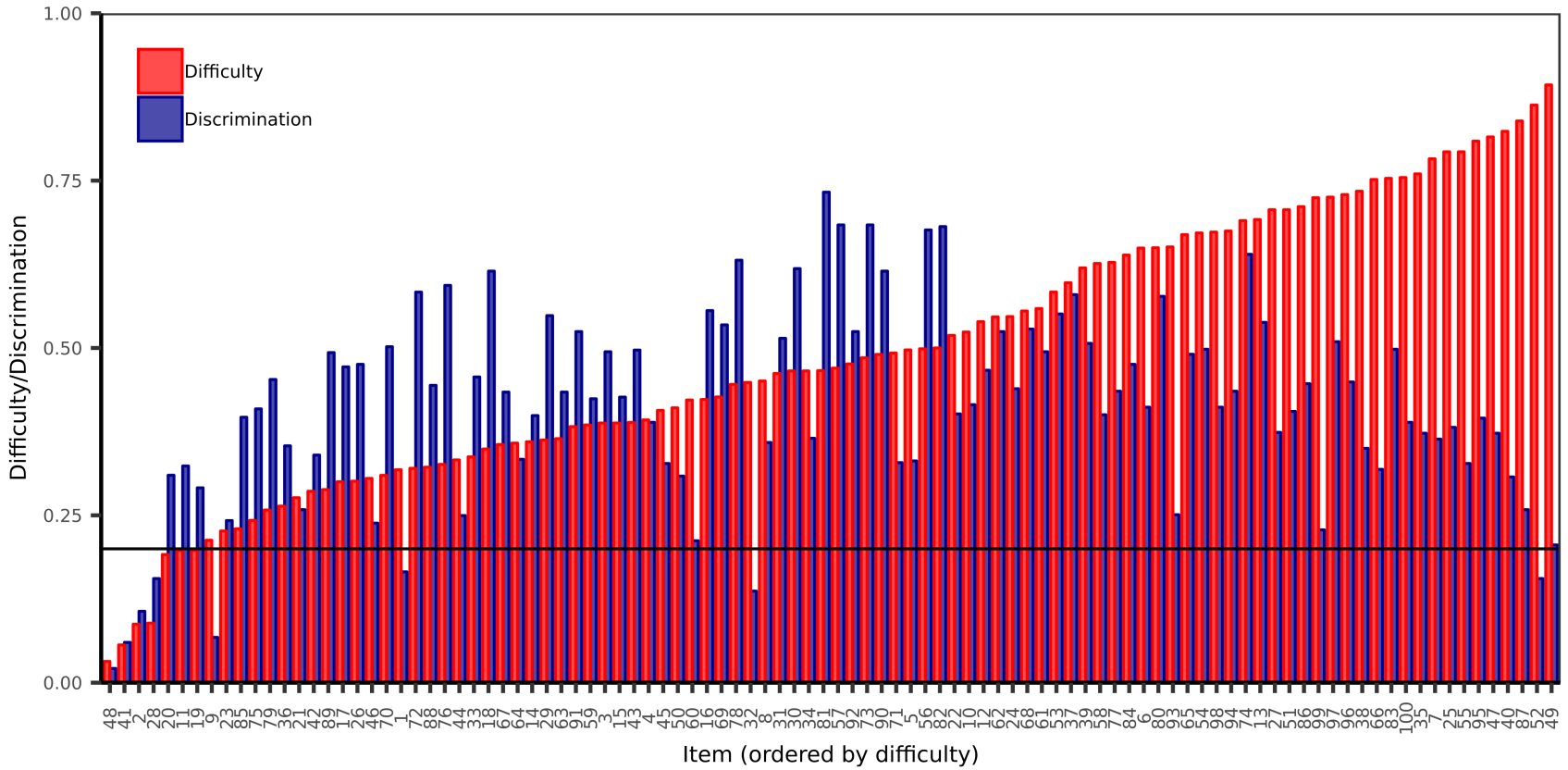
Generalized ULI

- Allows also other number of groups $(2, 3, 4, 5, \dots)$
- Compares any two of these groups
  - E.g., item discrimination between 4^{th} and 5^{th} fifths may be of interest if only one fifth of students is admitted

Martinková P, Štěpánek L, Drabinová A, Houdek J, Vejražka M, Štuka Č. Semi-real-time analyses of item characteristics for medical school admission tests. Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 11, pages 189–194, 2017. doi 10.15439/2017F380

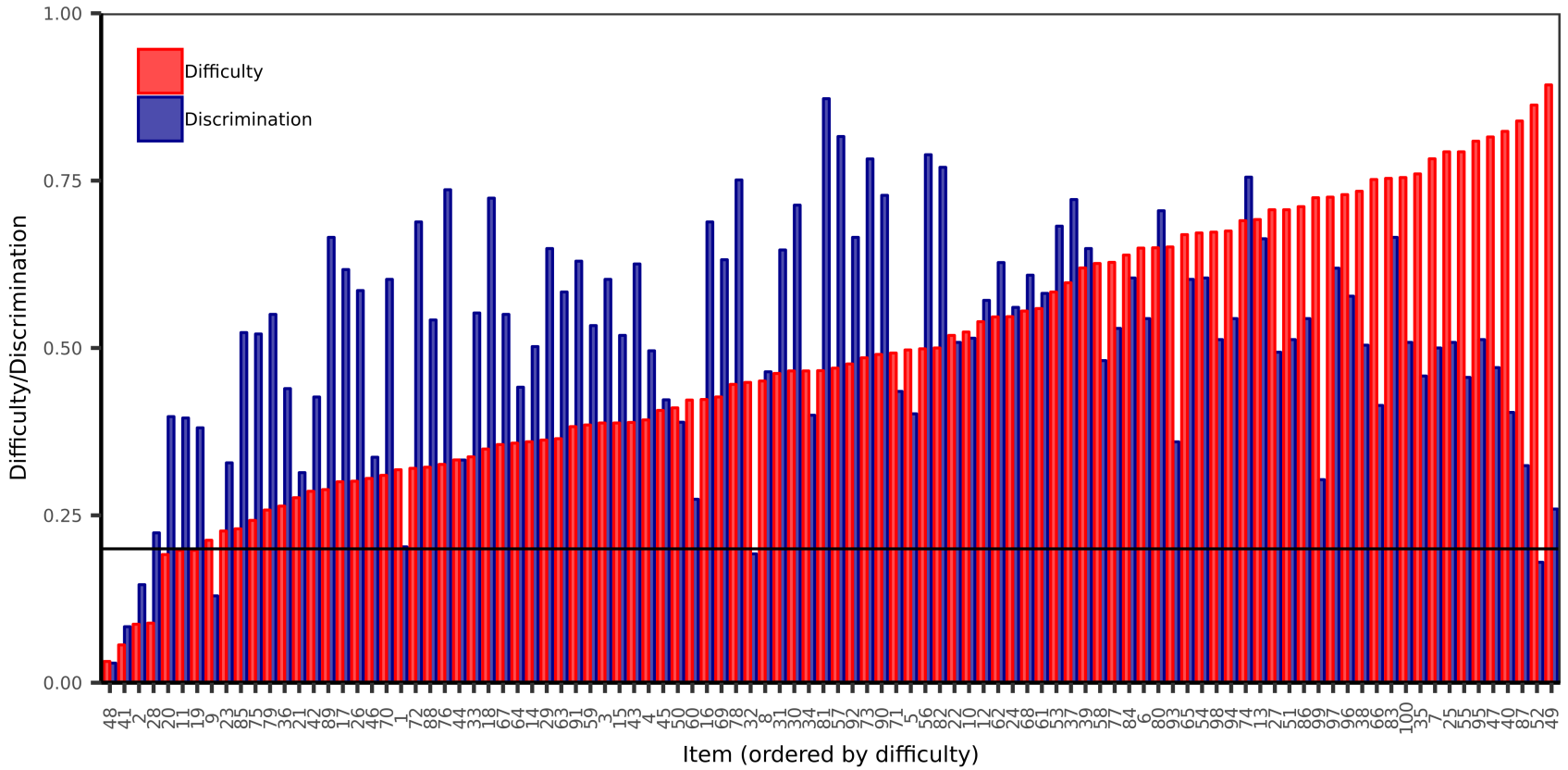
Difficulty - Discrimination Plot

Discrimination: Difference between first and last third (ULI)



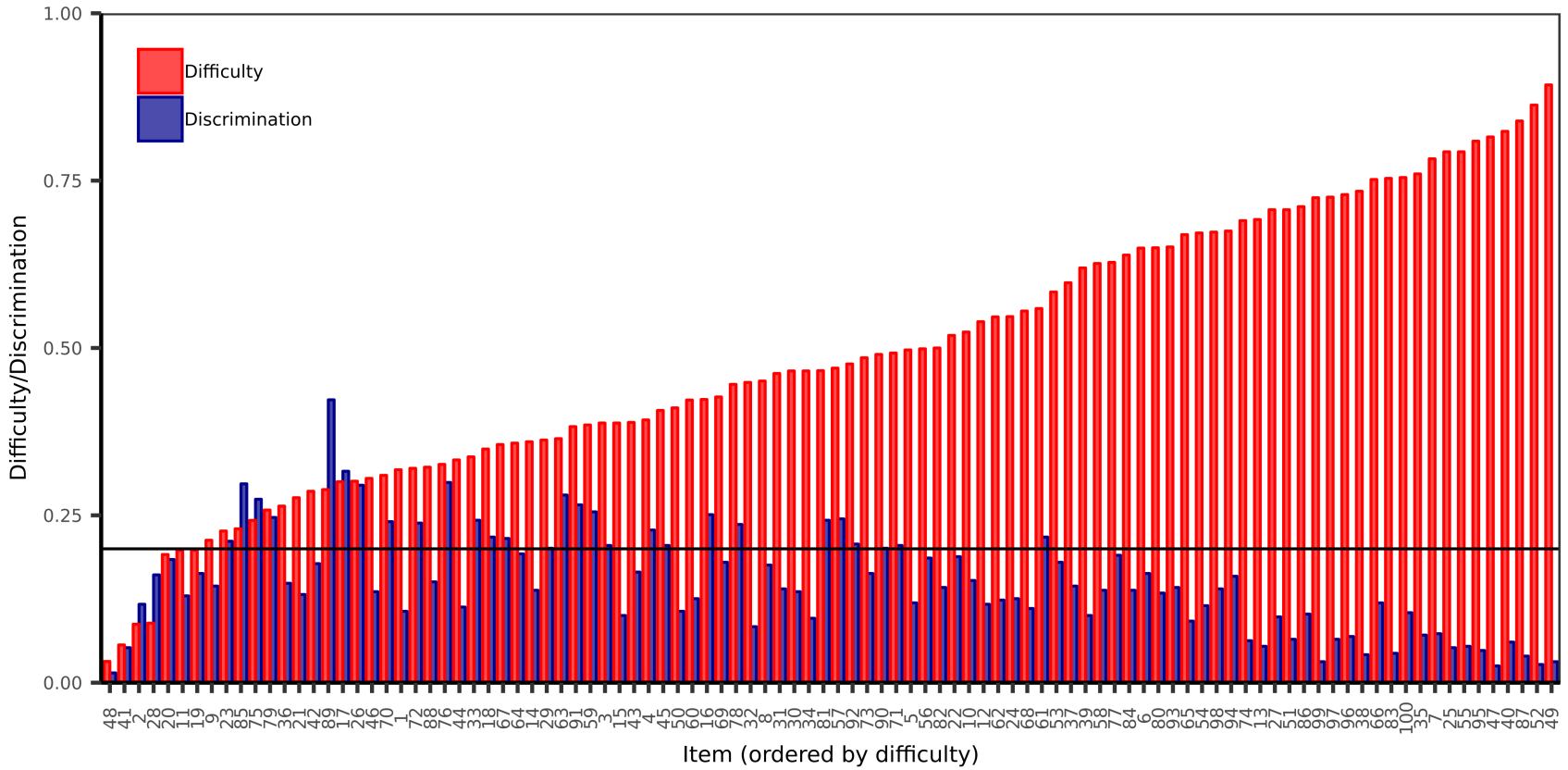
Difficulty - Discrimination Plot

Discrimination: Difference between first and last fifth (generalized ULI)



Difficulty - Discrimination Plot

Discrimination: Difference between fourth and last fifth (generalized ULI)



RIT, RIR

- Correlation Item – Test

$$RIT = \text{cor}\left(Y_j, \sum_{k=1}^{m} Y_k\right)$$

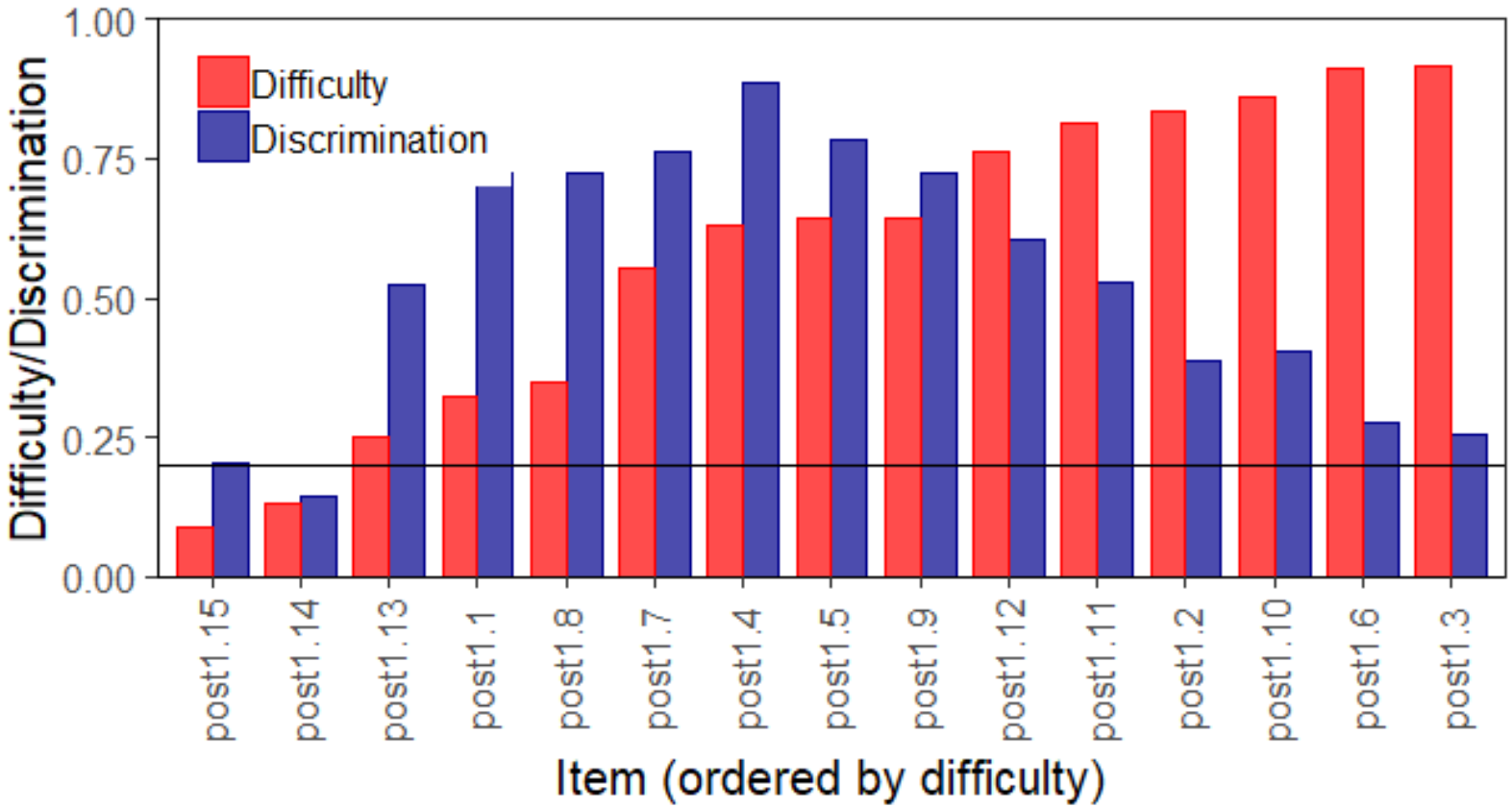
- Correlation Item – Rest of items

$$RIR = \text{cor}\left(Y_j, \sum_{k \neq j} Y_k\right)$$

Low correlation (< 0.2) signals item unconsistent with rest of the test.

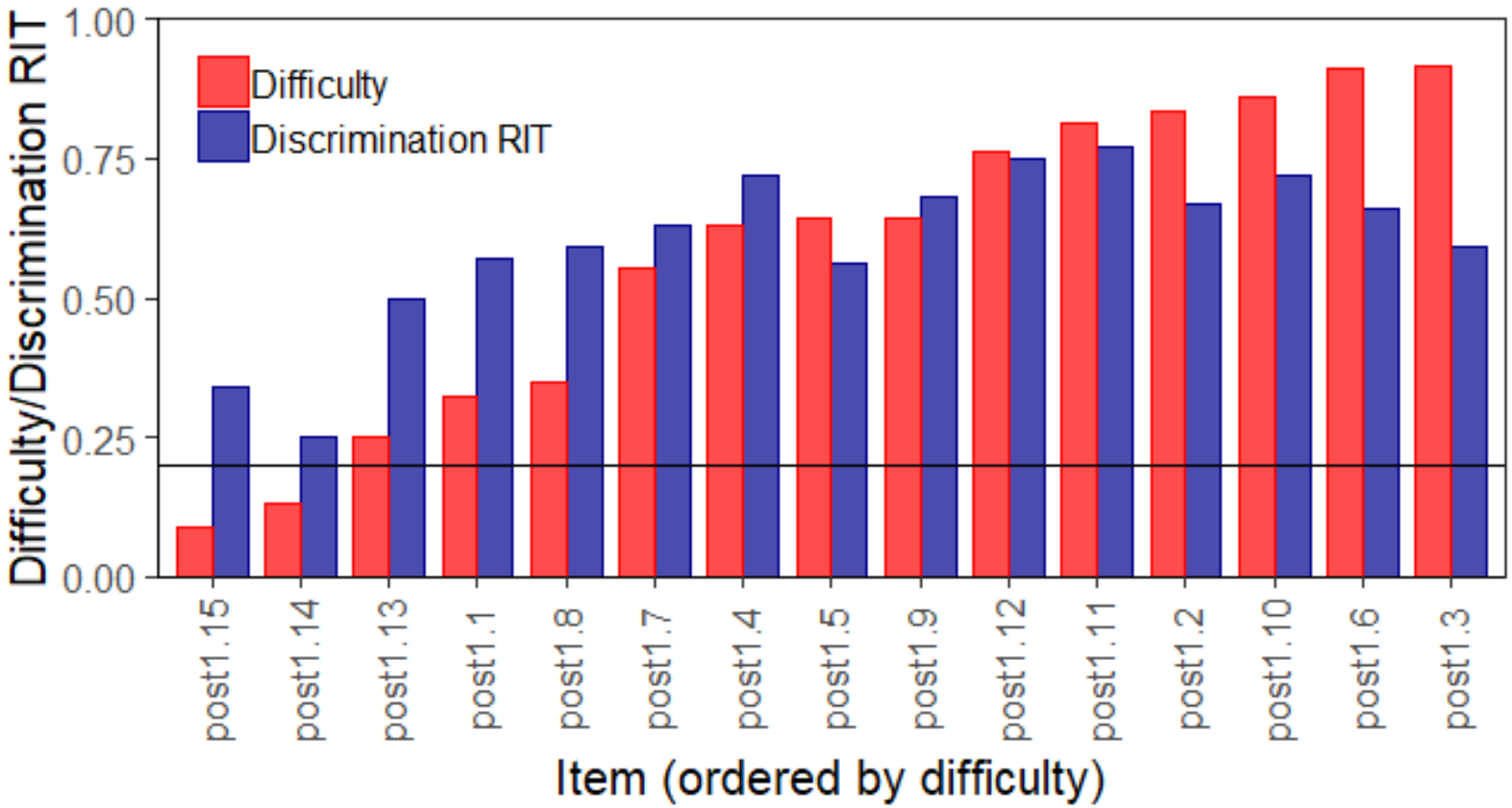
Difficulty – Discrimination Plot with ULI

Item discrimination calculated by ULI



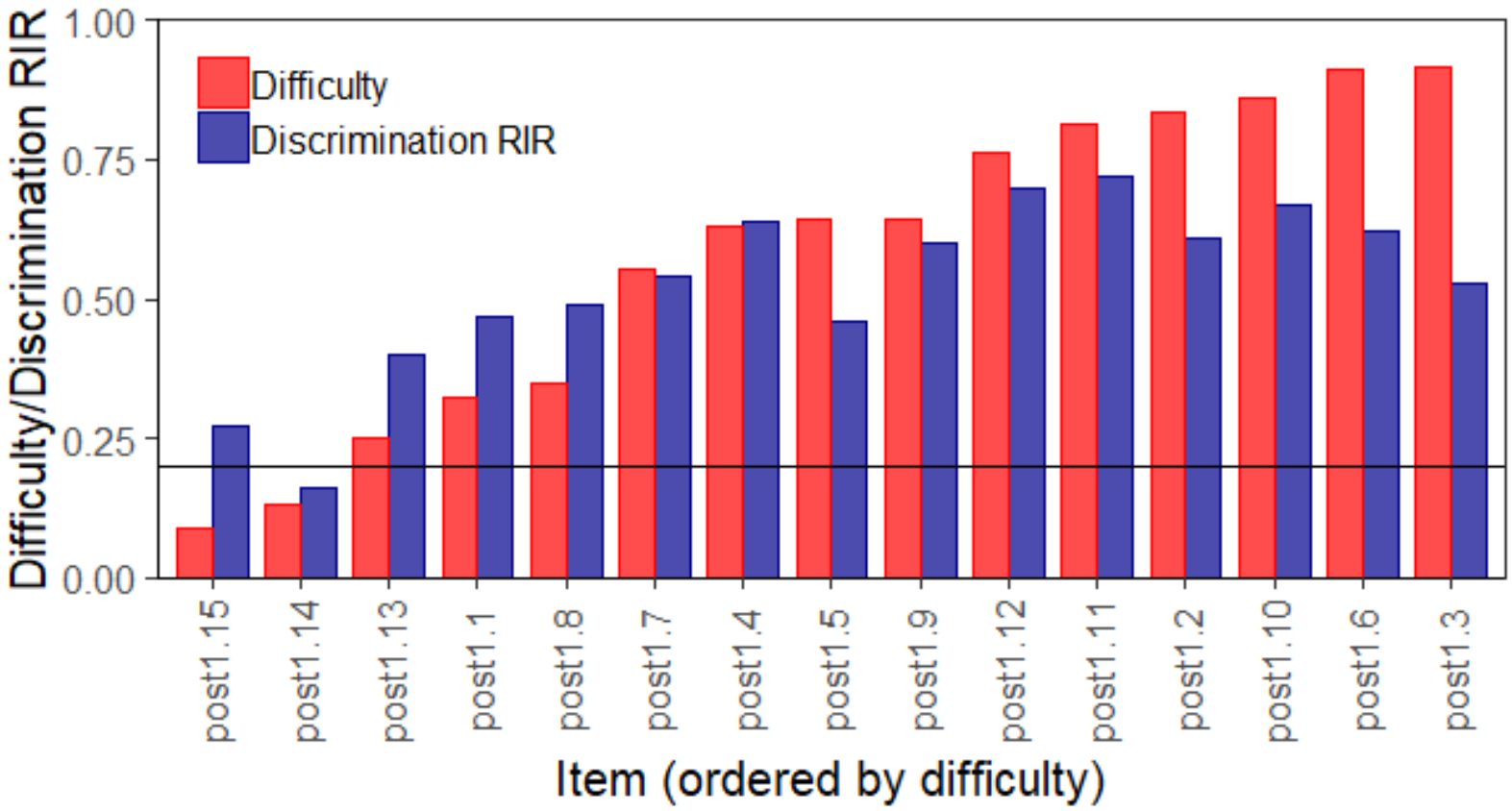
Difficulty – Discrimination Plot with RIT

Item discrimination calculated by RIT



Difficulty – Discrimination Plot with RIR

Item discrimination calculated by RIR



Cronbach's alpha without item

Cronbach's alpha:

$$\alpha = \frac{m}{m-1}\frac{\sum\sum_{j\neq k}\mathrm{cov}\left(Y_j, Y_k\right)}{\mathrm{var}\left(\sum_j Y_j\right)} = \frac{m}{m-1}\left(1 - \frac{\sigma^2_{Y_1} + \cdots + \sigma^2_{Y_m}}{\sigma^2_Y}\right)$$

Cronbach's alpha without item l:

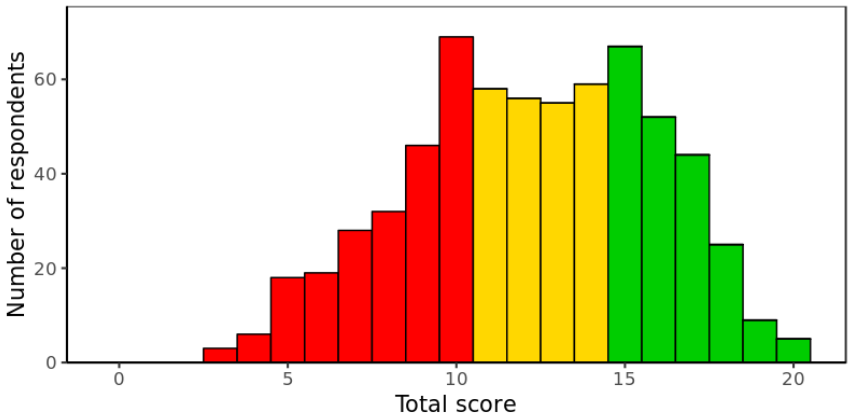
$$\alpha_l = \frac{m}{m-1}\left(1 - \frac{\sum_{j\neq l}\sigma^2_{Y_j}}{\mathrm{var}\left(\sum_{j\neq l} Y_j\right)}\right)$$

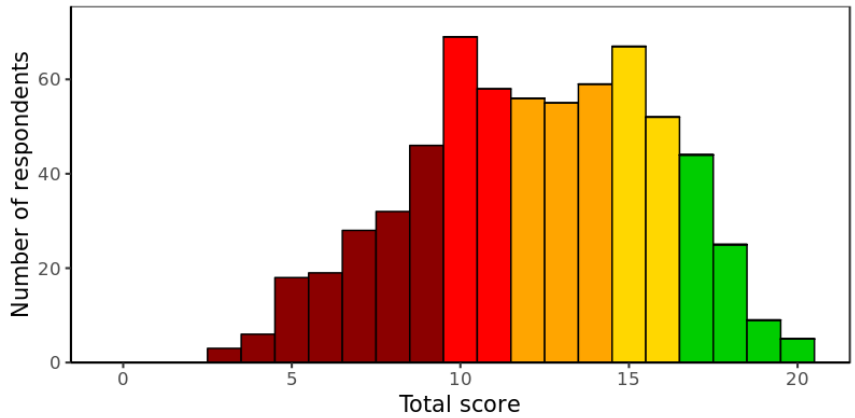
Increase of Cronbach's alpha after item removal signals item unconsistent with rest of the test.

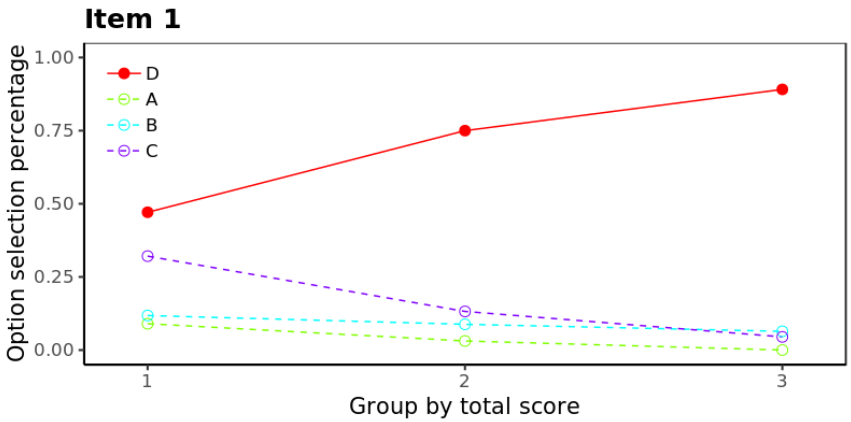
Distractor analysis

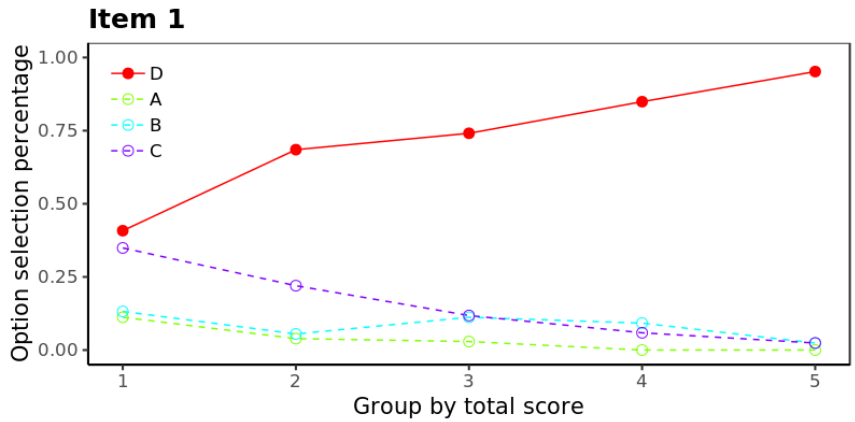
Distractor analysis

- Respondents are divided into (3 or more) groups by total score
- Option selection is displayed with respect to group



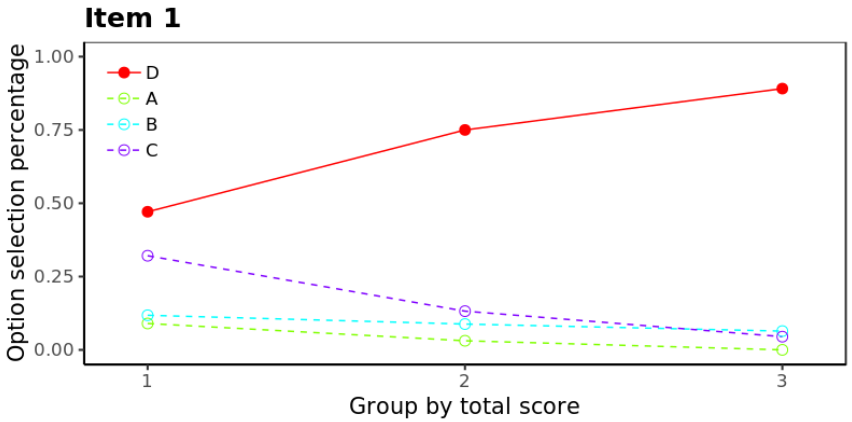


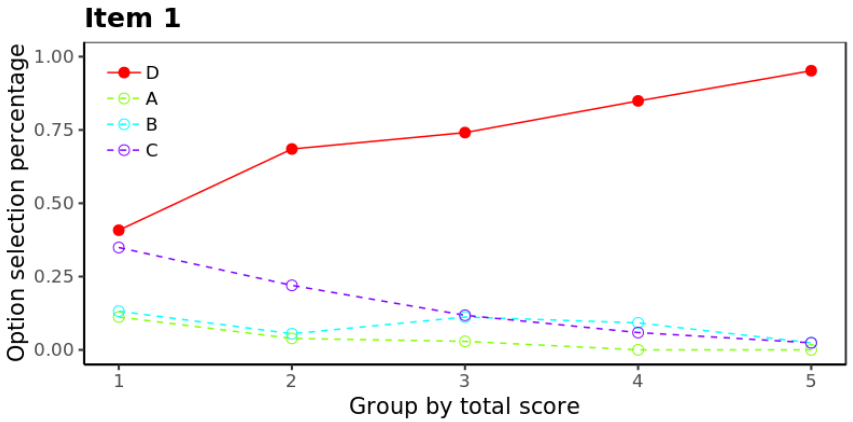




Distractor analysis

- Checking for misconceptions
- In-attractive distractors may need to be replaced

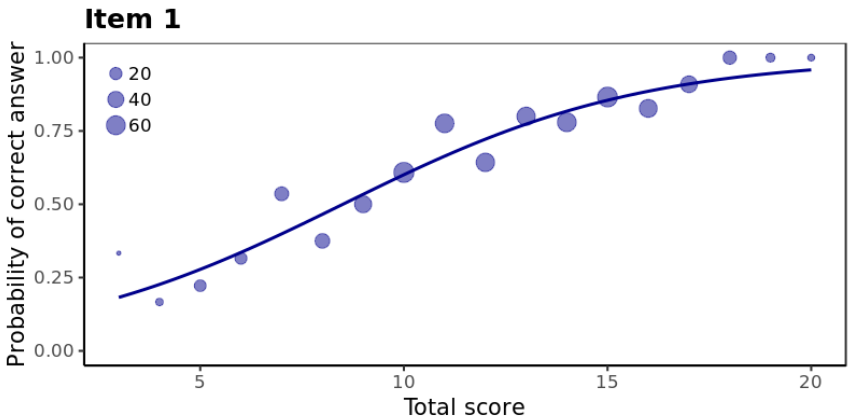


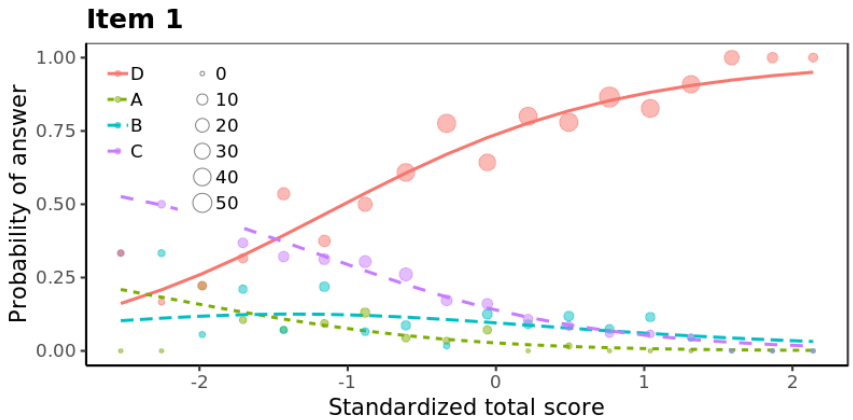


Towards regression models for item description

Models describing mean item score or probability of seclection of given answer with respect to total (standardized total) score

- Logistic regression
- Nonlinear regression
- Multiomial regression





Further issues

- Item discrimination with respect to criterion
  - Item criterion validity
  - E.g. with respect to IQ, GPA, grade, etc.
  - See ShinyItemAnalysis
- Analysis of missing answers
  - Percentage of missing answers
- Analysis of unreached items
  - Missed items such that all subsequent items are missed
  - If percentage of unreached is high, consider shortening the test

Conclusion

In this presentation, we have presented conventional/traditional approaches to description of item properties

- Item difficulty
  - Ratio of correct answers
  - (Scaled) average item score
- Item discrimination
  - Upper-lower index (ULI)
  - Generalized ULI
  - Correlation Item – Test (RIT)
  - Correlation Item – Rest (RIR)
  - Cronbach's alpha without item
- Functioning of discractors

Thank you for your attention!

www.cs.cas.cz/martinkova

References

- McFarland, Price, Wenderoth, Martinková, et al. (2017). Development and Validation of the Homeostasis Concept Inventory. CBE Life Sciences Education, 16(2), ar35. doi 10.1187/cbe.16-10-0305
- Martinková P, Štěpánek L, Drabinová A, Houdek J, Vejražka M, Štuka Č. Semi-real-time analyses of item characteristics for medical school admission tests. Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 11, pages 189–194, 2017. doi 10.15439/2017F380

Vocabulary

- Item analysis
  - Difficulty
    - Ratio of correct answers
    - Average item score
    - Scaled average item score
  - Discrimination
    - Upper-lower index (ULI)
    - Generalized ULI
    - Correlation Item – Test (RIT)
    - Correlation Item – Rest (RIR)
    - Cronbach's alpha without item
  - Distractor analysis
  - Analysis of non-reached items