Computational aspects of psychometrics taught with `R` and `Shiny`

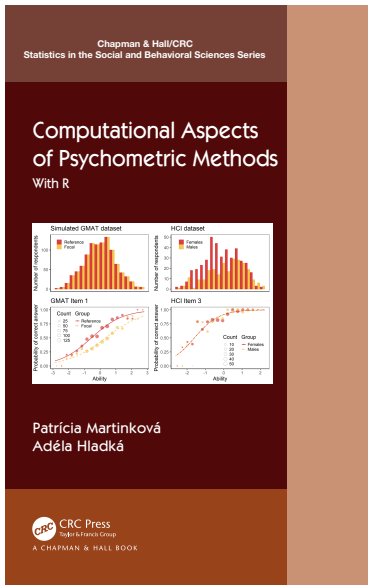
Patrícia Martinková[1,2]

useR!2021, July 5-9

[1] Institute of Computer Science, Czech Academy of Sciences
[2] Faculty of Education, Charles University

Outline

1. Introduction: Teaching psychometrics
2. `ShinyItemAnalysis`
3. Real and simulated datasets
   - Reliability and measurement error
   - Differential item functioning
4. Book in preparation
5. Discussion and conclusion



Psychometrics

- Psychometrics deals with the advancement of quantitative measurement practices in psychology, education, health, and many other fields
- Psychometric Society https://www.psychometricsociety.org/
- Covers a number of statistical methods that are useful for the behavioral and social sciences, such as:
  - estimation of reliability to deal with the omnipresence of measurement error
  - detailed description of item functioning encompassed in item response theory (IRT) models
- Number of existing R packages, see CRAN task View https://CRAN.R-project.org/view=Psychometrics

Teaching psychometrics

- Graduate course at University of Washington (2015)
- Graduate courses at Charles University, Prague
  - NMST570 Selected topics in psychometrics
  - NMST571 Seminar in psychometrics
- Pre-conference workshops, seminars
- Heterogeneous groups of students/participants
  - Students of psychology, education, ... and statistics
  - Researchers, practitioners from test companies
- Participants of various levels of `R` proficiency
- Participants of various levels of statistical focus and proficiency

Teaching psychometrics with R and ShinyItemAnalysis

Goals:

- Explain psychometric models and methods
  - in context of statistics and data science
- Illustrate important computational aspects
  - Real and simulated data from various fields
- Provide toolbox of R functions and packages
  - Similarities/differences across different packages
- Make procedures and concepts better available
  - Interactive application of the ShinyItemAnalysis package

ShinyItemAnalysis

Software for psychometric analysis of educational tests, psychological assessments, health-related and other types of multi-item measurements

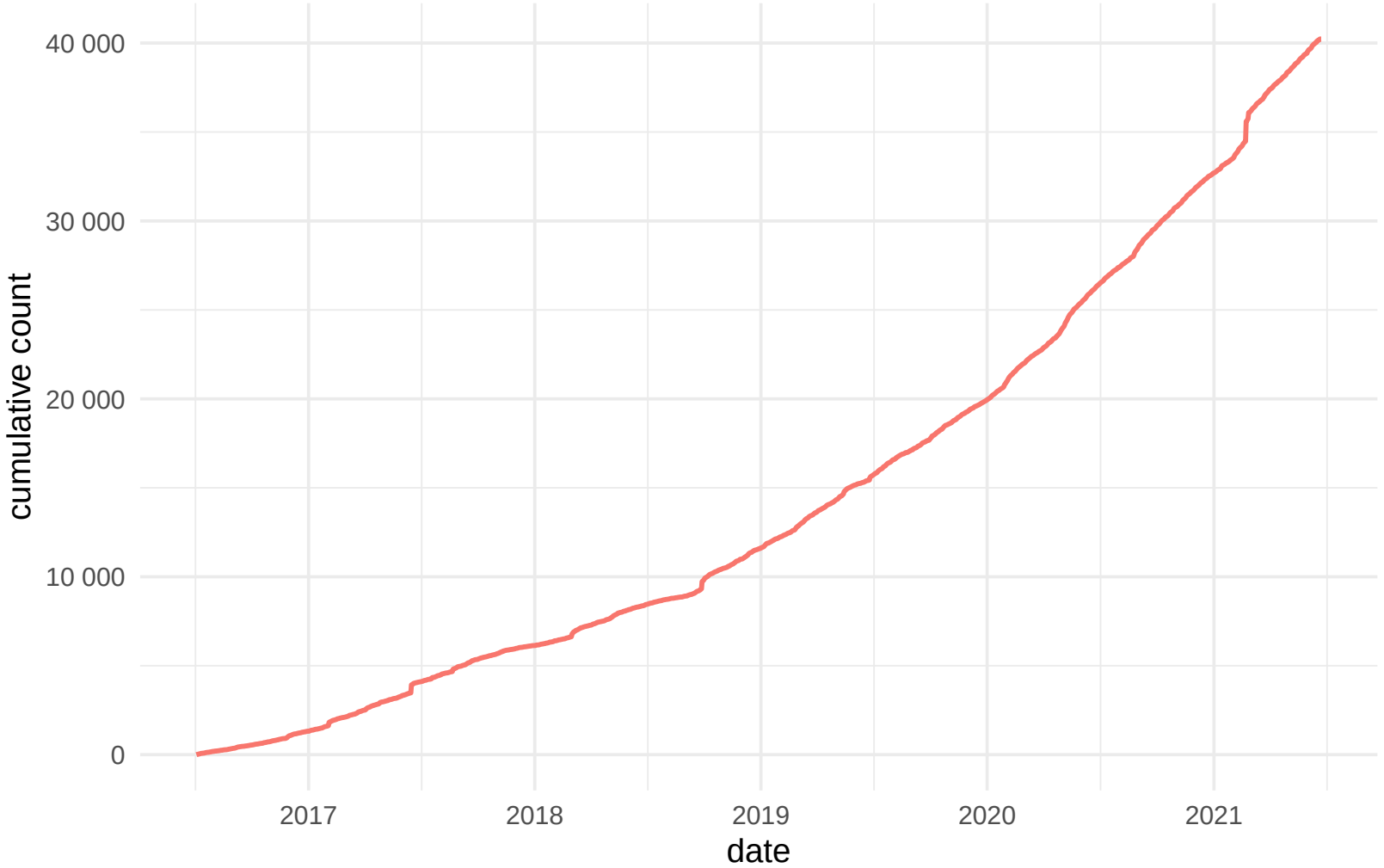
- R package
  - Version 1.3.7 on ▸ CRAN, newest version on ▸ GitHub
- Interactive `shiny` application
  - Accessible locally from `R` with `startShinyItemAnalysis()`
  - Online at ICS server and shinyapps.io

https://shiny.cs.cas.cz/ShinyItemAnalysis/

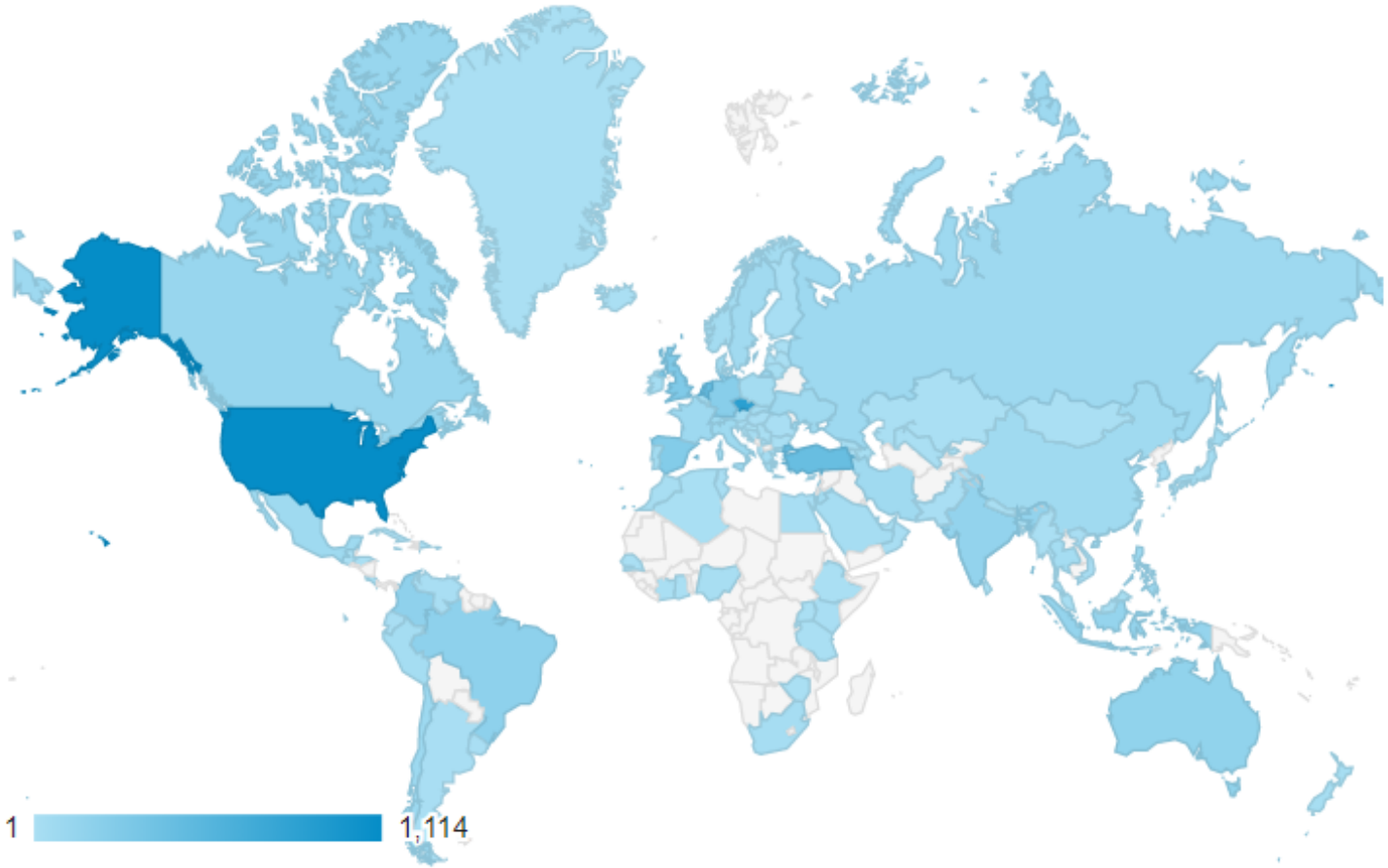
https://cemp.shinyapps.io/ShinyItemAnalysis/

Martinková and Drabinová (2018). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal.* 10(2), 503–515. https://doi.org/10.32614/RJ-2018-074

R package ShinyItemAnalysis downloads from CRAN



ShinyItemAnalysis online app is used worldwide!



Interactive application

Welcome

Welcome to ShinyItemAnalysis!

ShinyItemAnalysis is an interactive online application for the psychometric analysis of educational tests, psychological assessments, health-related and other types of multi-item measurements, or ratings from multiple raters, built on R and shiny. You can easily start using the application with the default toy dataset. You may also select from a number of other toy datasets or upload your own in the **Data** section. Offered methods include:

- Exploration of total and standard scores in the **Summary** section
- Analysis of measurement error in the **Reliability** section
- Correlation structure and criterion validity analysis in the **Validity** section
- Item and distractor analysis in the **Item analysis** section
- Item analysis with regression models in the **Regression** section
- Item analysis by item response theory models in the **IRT models** section
- Detection of differential item functioning in the **DIF/Fairness** section

All graphical outputs and selected tables can be downloaded via the download button. Moreover, you can automatically generate a HTML or PDF report in the **Reports** section. All offered analyses are complemented by selected R codes which are ready to be copied and pasted into your R console, therefore a similar analysis can be run and modified in R.

Visit the **www.ShinyItemAnalysis.org** webpage to learn more about ShinyItemAnalysis!

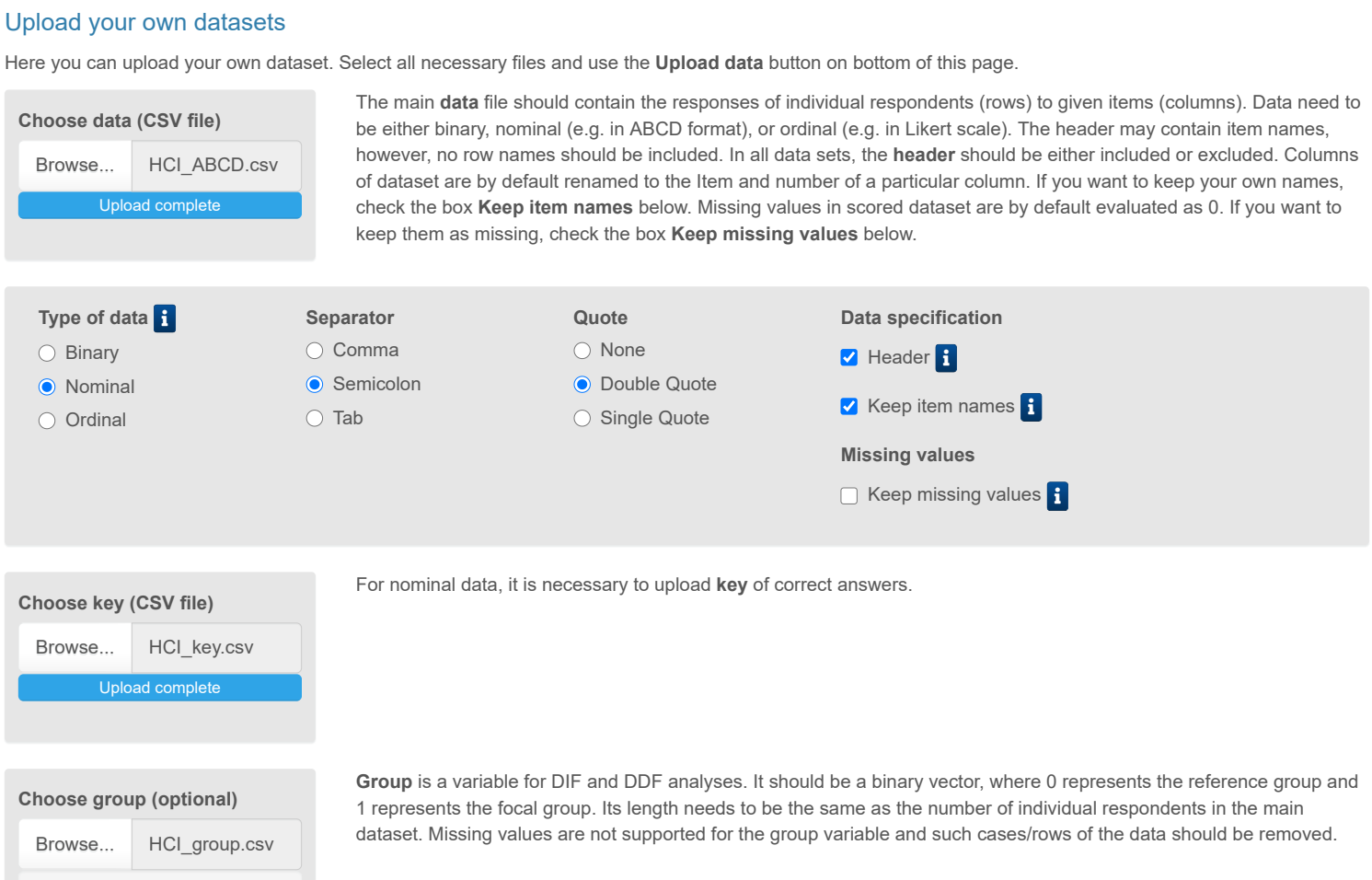
ShinyItemAnalysis: Newest developments

- New features of the interactive application
  - New toy data, new data types allowed for one's own upload
  - Validity: New corrplot, dendrograms, factor analysis
  - Reliability: Inter-rater reliability in restricted samples
  - Traditional item analysis: Item criterion validity
  - Regression models: Models for polytomous data
  - IRT models: reorganized
  - DIF: polytomous data, uploaded matching criterion

  - Interactive training sections with exercises
  - All plots interactive, created with `plotly`
  - Downloadable plots, tables and reports
  - Sample `R` code

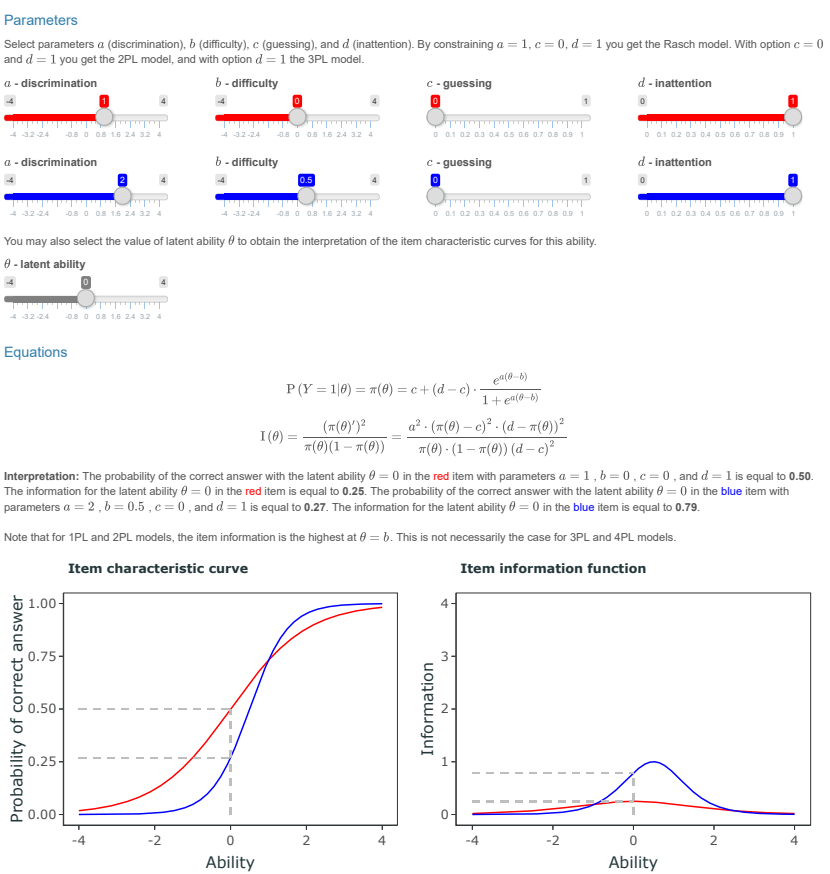
Toy datasets

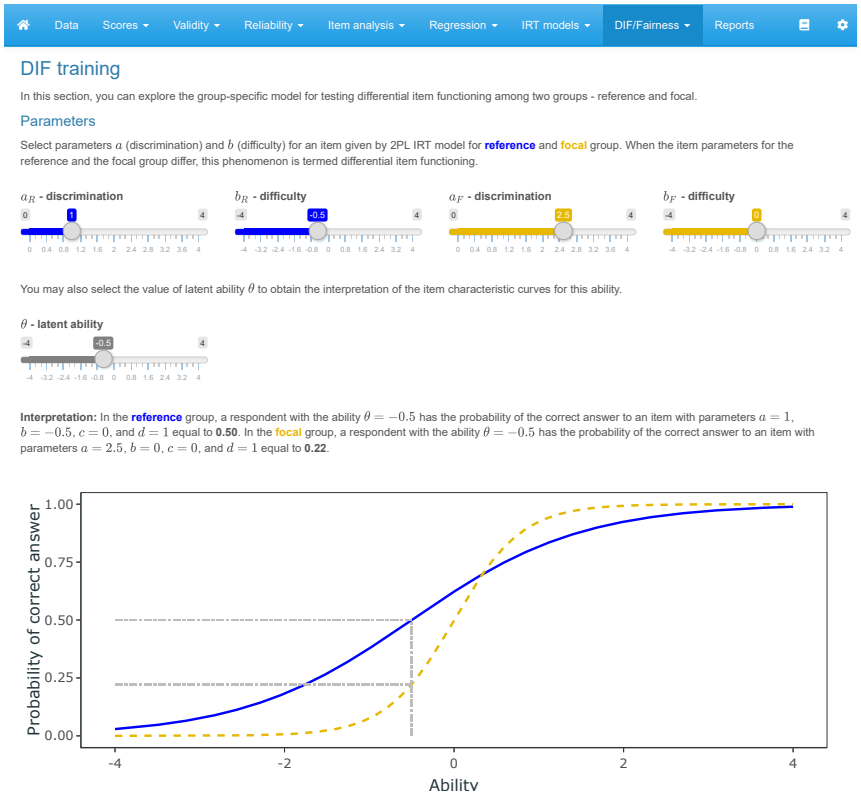
- Number of toy datasets, upload of one's own data is possible



Interactive training sections

- Interactive training sections for IRT models and DIF





Interactive training sections – check your understanding

- Interactive quizzes for IRT models and DIF

Exercise 1

Consider the following 2PL items with parameters
Item 1: $a = 2.5, b = -0.5$
Item 2: $a = 1.5, b = 0$
For these items fill in the following exercises with an accuracy of up to 0.05, then click on the **Submit answers** button. If you need a hint, click on the blue button with a question mark.

- Sketch the item characteristic and information curves. ? ✗
- Calculate the probability of a correct answer for latent abilities $\theta = -2, -1, 0, 1, 2$. ?

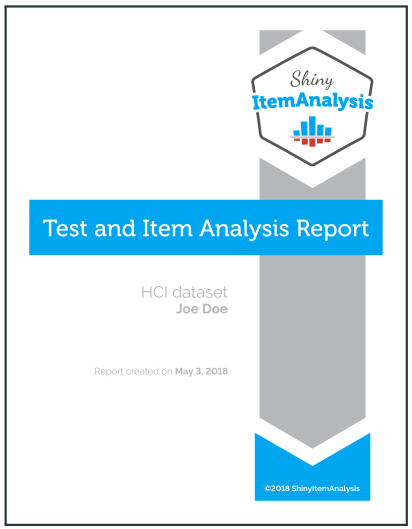
	$\theta = -2$	$\theta = -1$	$\theta = 0$	$\theta = 1$	$\theta = 2$
Item 1:	✓	✗	✗	✗	✗
	0	0	0	0	0
Item 2:	✓	✗	✗	✗	✗
	0	0	0	0	0

- For what level of ability θ are the probabilities equal? ?
  θ = ? ✗
  0
- Which item provides more information for weak ($\theta = -2$), average ($\theta = 0$) and strong ($\theta = 2$) students? ?
  $\theta = -2$ ◉ Item 1 ○ Item 2 ✓
  $\theta = 0$ ◉ Item 1 ○ Item 2 ✓
  $\theta = 2$ ◉ Item 1 ○ Item 2 ✗

27% correct. Try again.
Submit answers

Automatic report generation

- Generating PDF/HTML reports for uploaded data
- using `R Markdown`

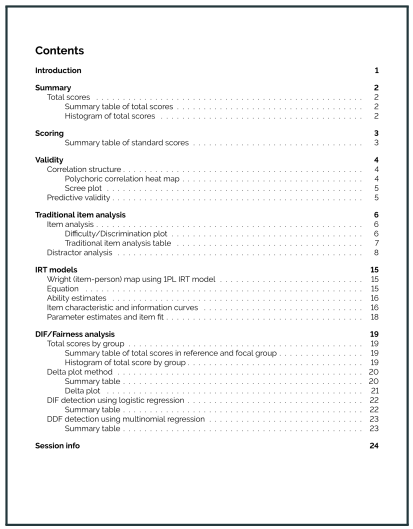

Test and Item Analysis Report

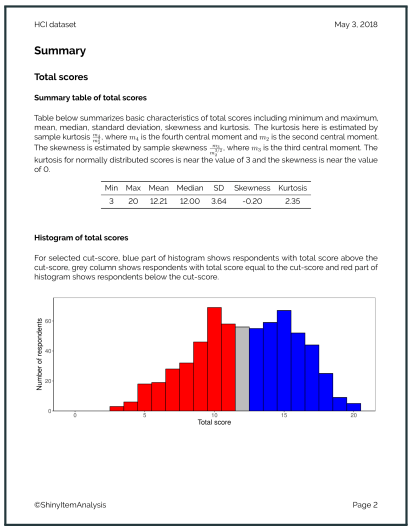
HCI dataset

Joe Doe

Report created on May 3, 2018

©2018 ShinyItemAnalysis


Contents



Summary

Total scores

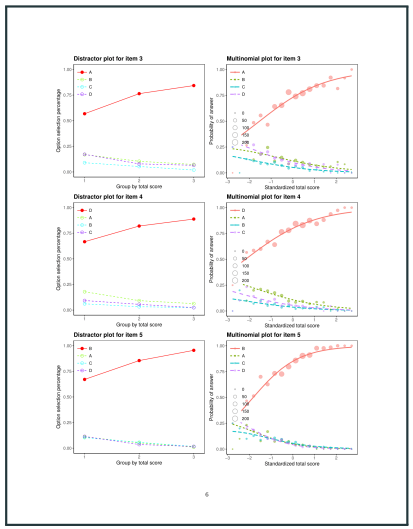
Summary table of total scores

Table below summarizes basic characteristics of total scores including minimum and maximum, mean, median, standard deviation, skewness and kurtosis. The kurtosis here is estimated by sample kurtosis $\frac{m_4}{m_2^2}$, where m_4 is the fourth central moment and m_2 is the second central moment. The skewness is estimated by sample skewness $\frac{m_3}{m_2^{3/2}}$, where m_3 is the third central moment. The kurtosis for normally distributed scores is near the value of 3 and the skewness is near the value of 0.

Min	Max	Mean	Median	SD	Skewness	Kurtosis
3	20	12.21	12.00	3.64	-0.20	2.35

Histogram of total scores

For selected cut-score, blue part of histogram shows respondents with total score above the cut-score, grey column shows respondents with total score equal to the cut-score and red part of histogram shows respondents below the cut-score.



ShinyItemAnalysis: Newest developments

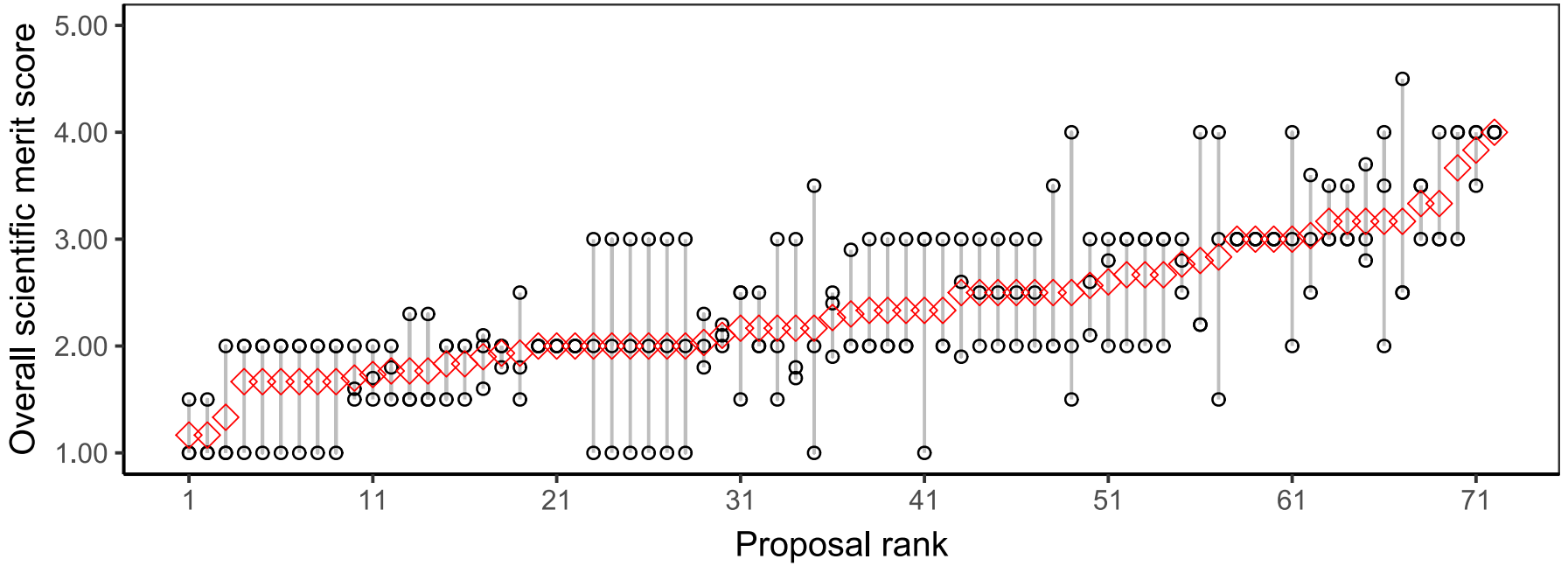
- New ShinyItemAnalysis package functions and functionalities
  - startShinyItemAnalysis() now using rstudioapi, runs as "Local job" in Jobs RStudio IDE pane, keeping the console available for trying sample R code
  - Testing of the online app on collection of datasets, unit tests using testthat
  - Refactoring the code using shiny modules, following the best practices with golem
  - Dealing with high number of dependencies

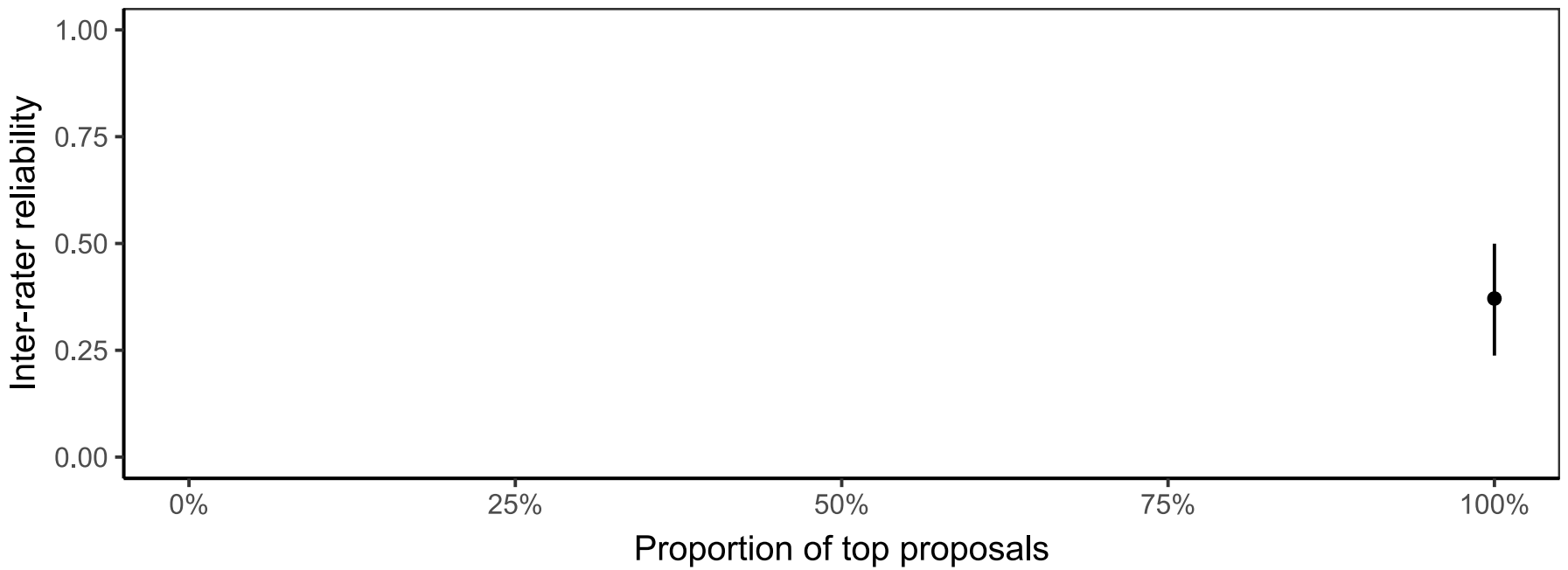
Datasets demonstrating computational aspects: IRR

- Why zero inter-rater reliability estimates are plausible under restricted range
- Statistical explanation: When proposal range is restricted by perceived quality, the between-proposal variance of peer review scores τ^2 decreases.
- Interactive illustration offered in `ShinyItemAnalysis` with the AIBS dataset.
- Animation created with the `gganimate` package.

Erosheva, Martinkova, and Lee (2021). When zero may not be zero: A cautionary note on the use of inter-rater reliability in evaluating grant peer review. *JRSS – A*. doi 10.1111/rssa.12681

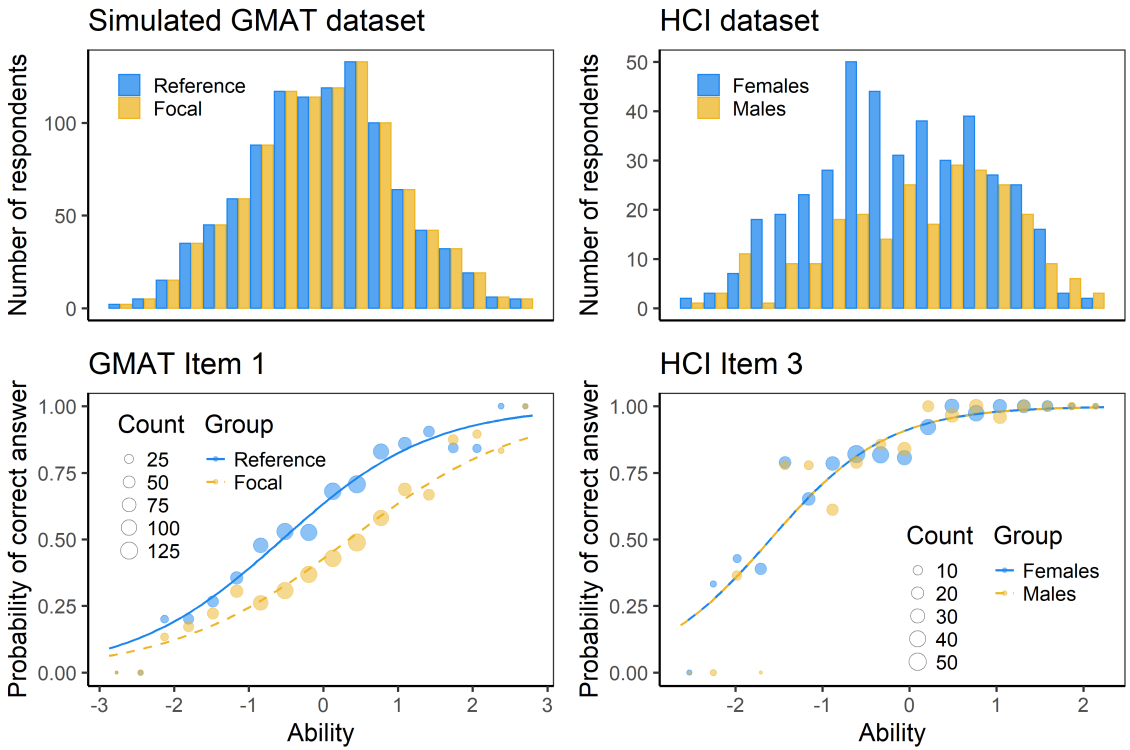
IRR in restricted range: Animation





Datasets demonstrating computational aspects: DIF

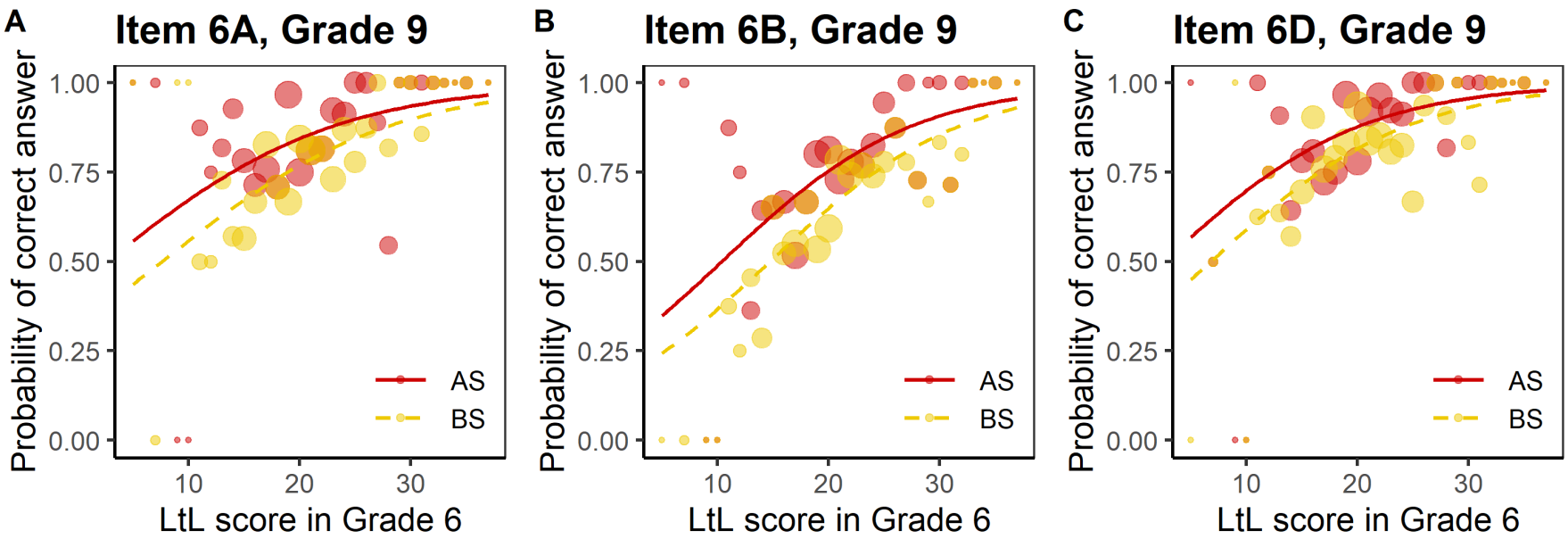
- Differential item functioning (DIF) analysis may provide deeper understanding to test functioning among groups.



Martinková et al. (2017). Checking Equity: Why DIF Analysis should be a Routine Part of Developing Conceptual Assessments. CBE-LSE, 16(2), rm2. doi 10.1187/cbe.16-10-0307

DIF in longitudinal designs

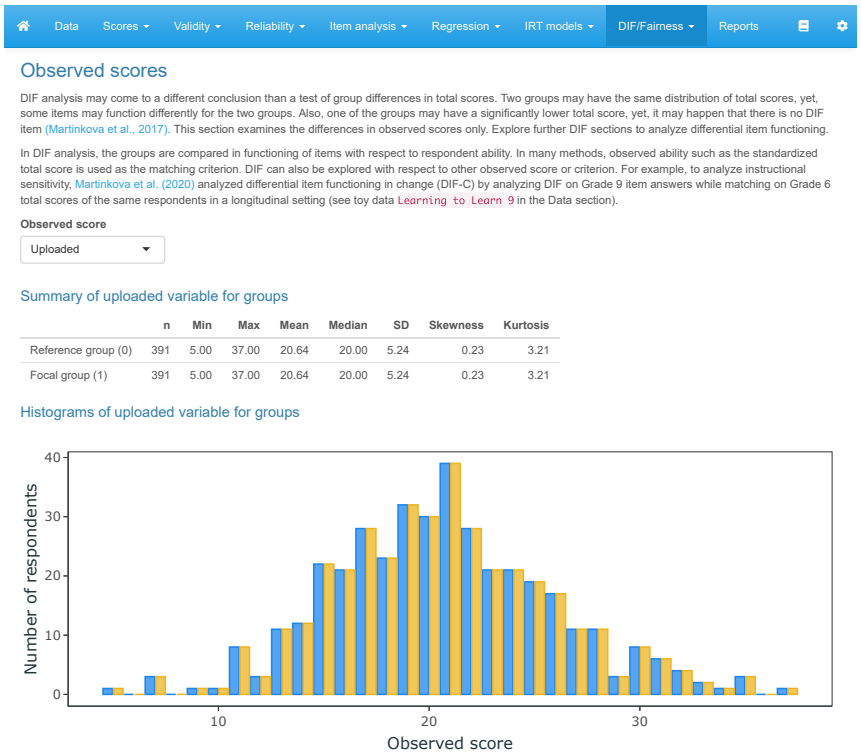
DIF-C can provide proof of instructional sensitivity, even when differences in change are not visible in total scores.

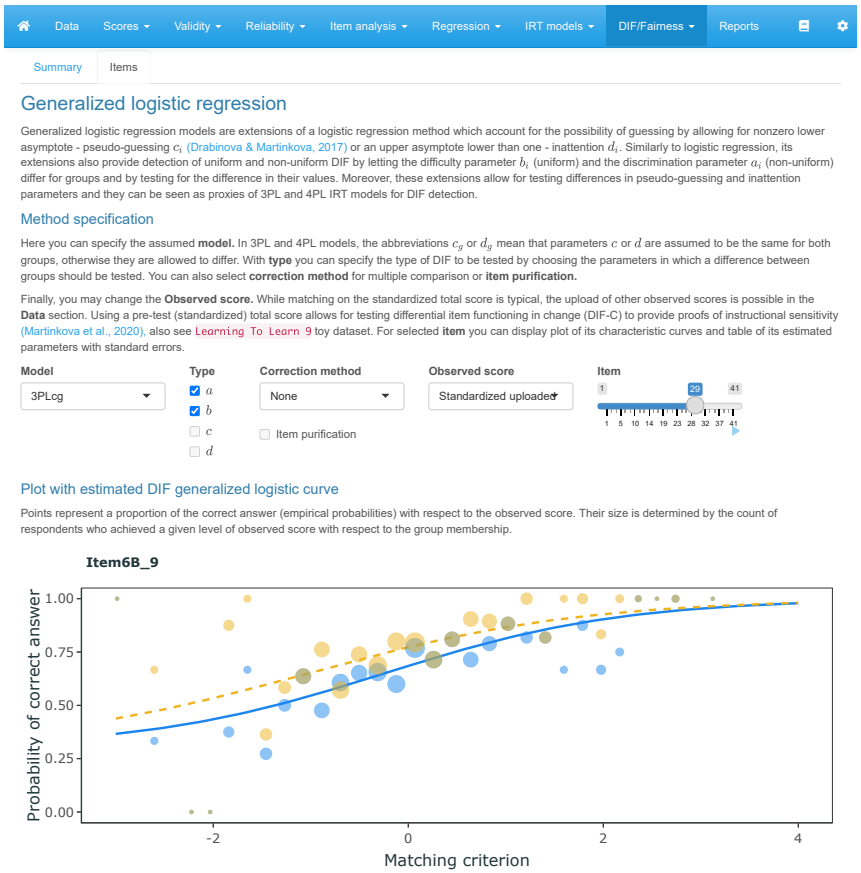


Martinková, Hladká, and Potužníková (2020). Is academic tracking related to gains in learning competence? Using propensity score matching and differential item change functioning analysis for better understanding of tracking implications. *Learning and Instruction*, 66, 101286. doi: 10.1016/j.learninstruc.2019.101286

DIF and DIF-C analysis available in ShinyItemAnalysis

- DIF and DIF-C analysis with `difNLR` package

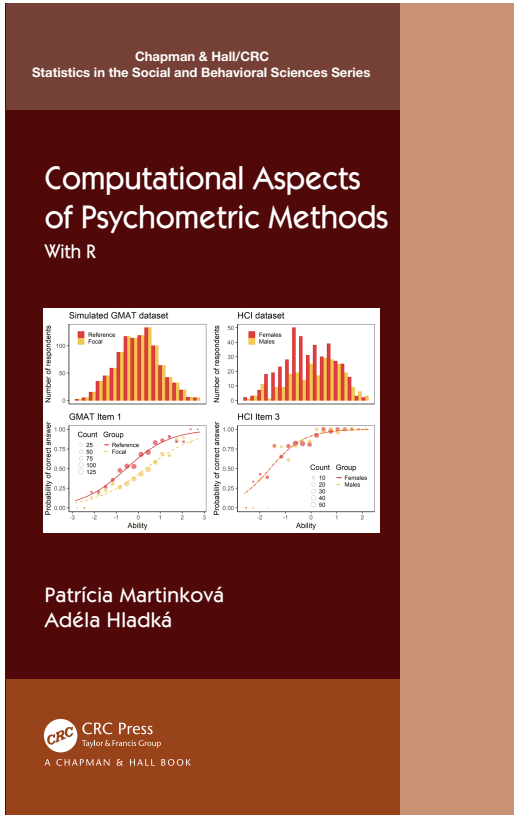




Hladká and Martinková (2020). `difNLR`: Generalized logistic regression models for DIF and DDFdetection. *The R Journal*, 12(1), 300–323. doi: 10.32614/RJ-2020-014

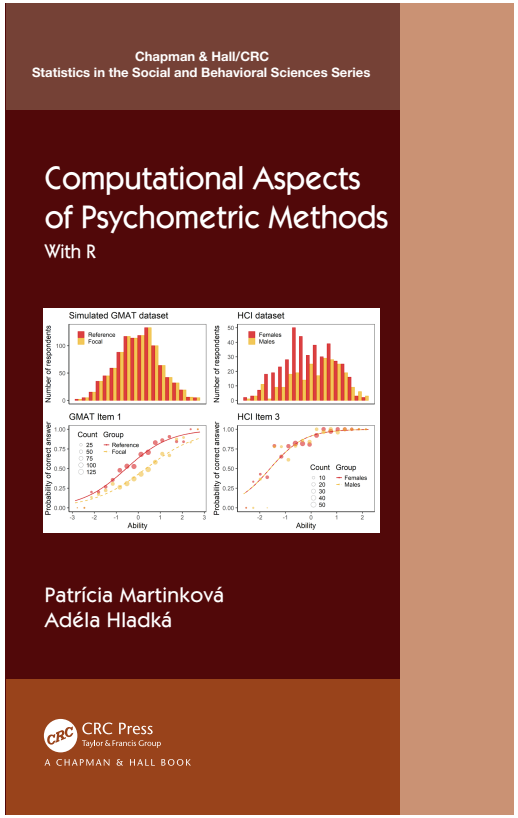
Book planned for publication in 2022

- Deeper understanding to psychometric models and methods
- For a wide audience
- Accompanied by sample R code, practical examples and datasets
- Each chapter includes a section presenting the analysis with respective tab of the `ShinyItemAnalsyis` interactive application



Discussion

- Teaching computational aspects of psychometrics with R and ShinyItemAnalysis
- Demonstrating the power of R
- Importance of sample R code within the Shiny app
- Importance of relevant simulated and real data examples

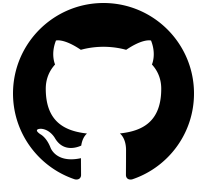


- Stay tuned for the new book!

Thank you for your attention!

www.cs.cas.cz/martinkova





Acknowledgements:

- Czech Science Foundation grant 21-03658S
- Technology Agency of the Czech Republic grant TL05000008
- Computational Psychometrics Group: https://www.cs.cas.cz/comps/

References

Erosheva, E. A., Martinkova, P., & Lee, C. J. (2021). When zero may not be zero: A cautionary note on the use of inter-rater reliability in evaluating grant peer review. *Journal of the Royal Statistical Society – Series A*. doi: 10.1111/rssa.12681

Hladká, A., & Martinková, P. (2020). difNLR: Generalized logistic regression models for DIF and DDF detection. *The R Journal*, *12*(1), 300–323. doi: 10.32614/RJ-2020-014

Martinková, P., & Drabinová, A. (2018). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal*, *10*(2). doi: 10.32614/RJ-2018-074

Martinková, P., Drabinová, A., Liaw, Y.-L., Sanders, E. A., McFarland, J. L., & Price, R. M. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE—Life Sciences Education*, *16*(2), rm2. doi: 10.1187/cbe.16-10-0307

Martinková, P., Hladká, A., & Potužníková, E. (2020). Is academic tracking related to gains in learning competence? Using propensity score matching and differential item change functioning analysis for better understanding of tracking implications. *Learning and Instruction*, *66*, 101286. doi: j.learninstruc.2019.101286