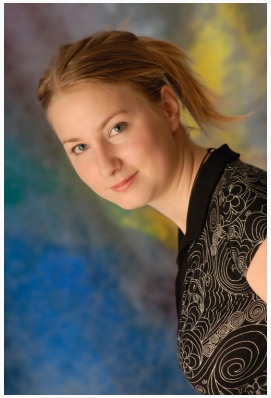
Teaching Psychometrics and Analysing Educational Tests with `ShinyItemAnalysis`

Patricia Martinkova et al.

useR!2017, Brussels, July 7

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

Joint work...


Patricia Martinkova[1] Adela Drabinova[1,2] Jakub Houdek[1,3] Ondrej Leder[2] Lubomir Stepanek[3,4]

[1]Department of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences
[2]Department of Probability and Mathematical Statistics, Charles University, Prague
[3]Faculty of Informatics and Statistics, University of Economics, Prague
[4]First Faculty of Medicine, Charles University, Prague

Acknowledgement: Czech Science Foundation grant GJ15-15856Y

Table of contents

Introduction

Motivation

- Admission tests development
  - to select the best applicants
  - need for valid, reliable and fair tests, well functioning items
- Development of tests for classroom use
  - to keep students interested in the subject
  - to avoid discrimination of minorities
- To teach psychometric models and concepts
- To promote our own psychometric research

Need for user-friendly and freely available tool

ShinyItemAnalysis

Interactive (and step by step) analysis of educational tests and their items

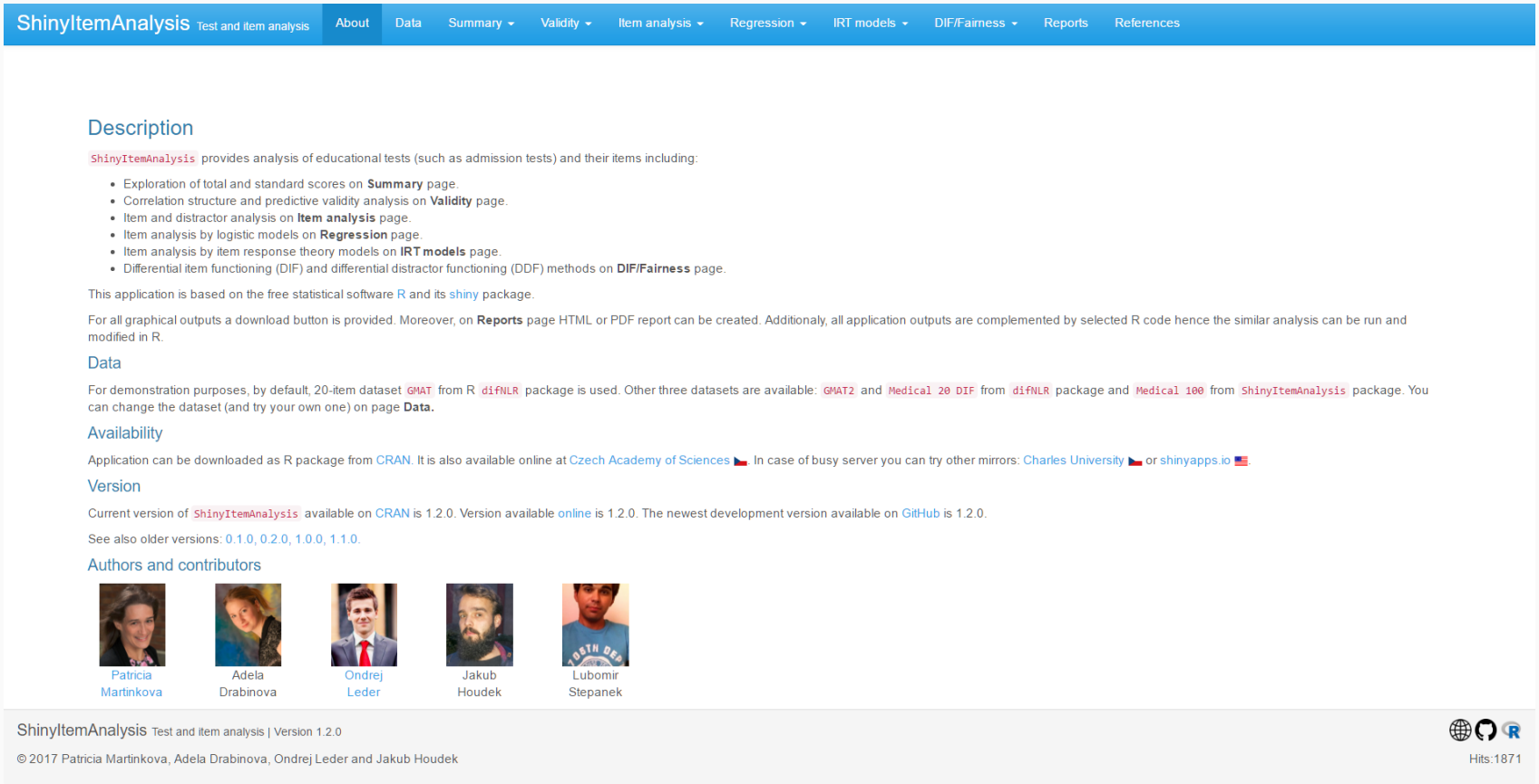
Available as:

- R package
  - Version 1.2.0 now on ▸ CRAN
  - Newest version on ▸ GitHub
- Online shiny application
  - ICS server in Prague, CZ:

    https://shiny.cs.cas.cz/ShinyItemAnalysis/
  - shinyapps.io:

    https://patriciamar.shinyapps.io/ShinyItemAnalysis/

ShinyItemAnalysis Application



`ShinyItemAnalysis` for TEACHING

ShinyItemAnalysis for Teaching

Who do we teach:

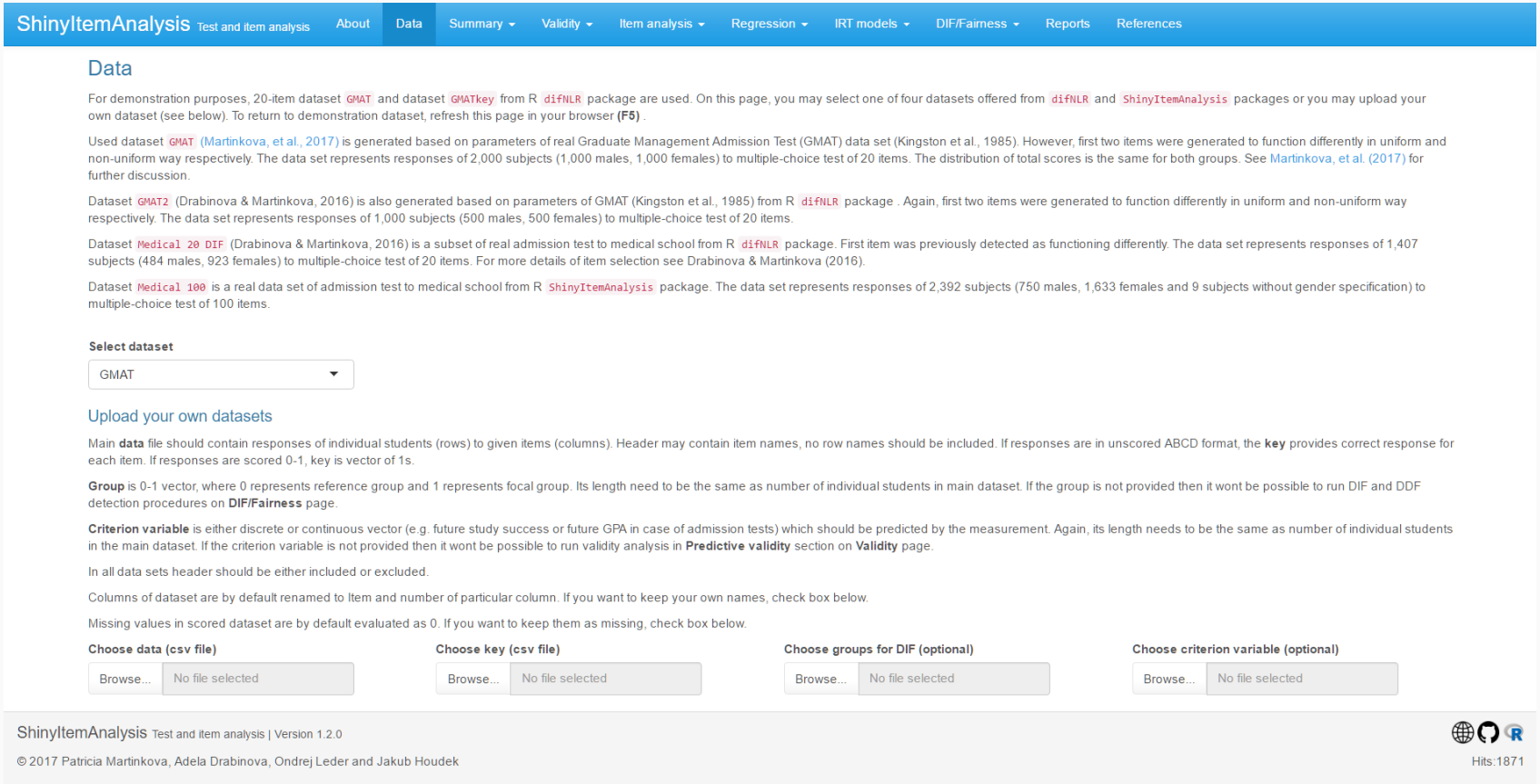
- students of educational measurement
- faculties, university stakeholders

Some helpful features:

- Interactive plots
- Example datasets, allows to upload own data
- Shows model equations, provides interpretation of results
- Allows to download plots, generate extensive reports
- Provides sample R code

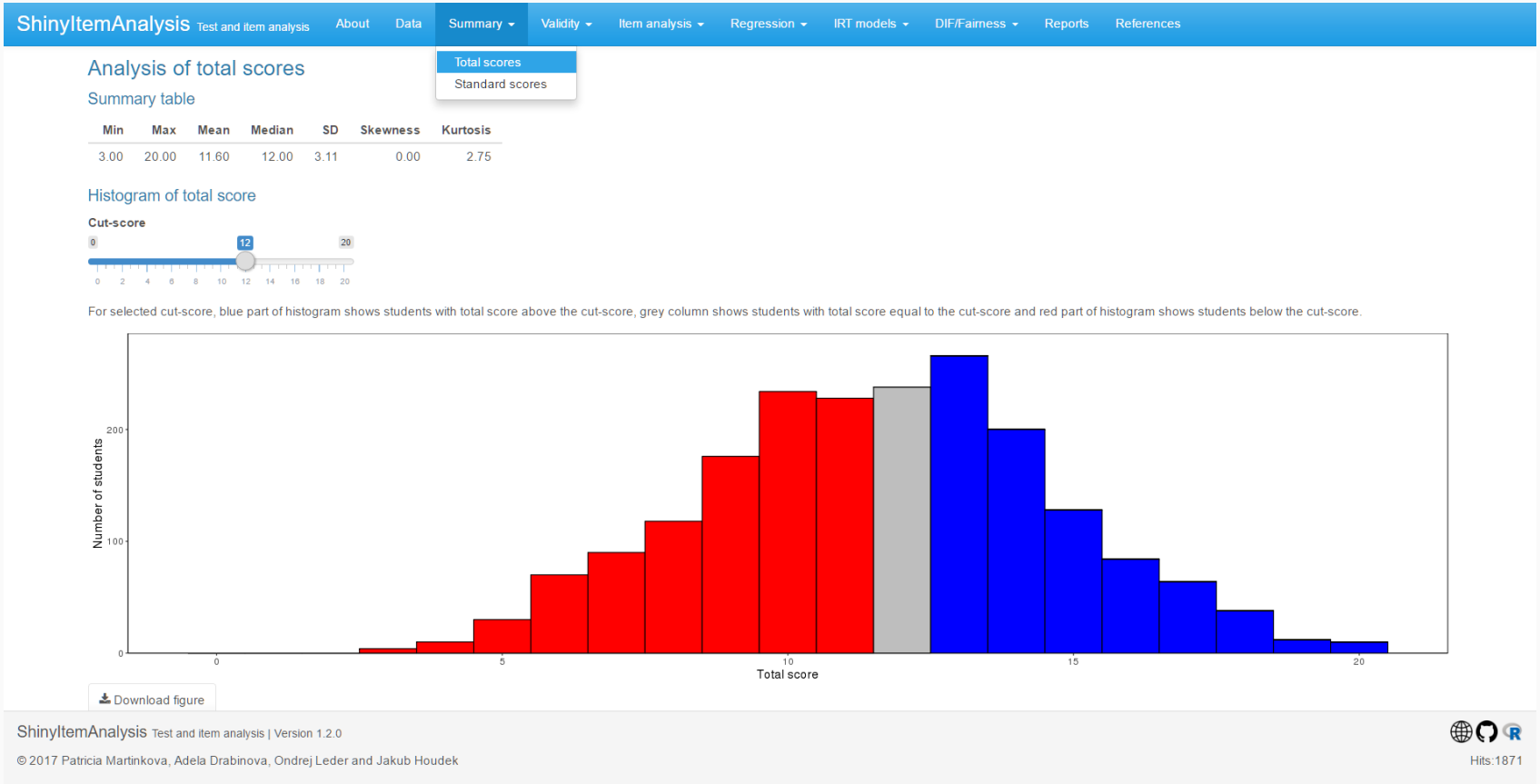
Datasets

- Four toy datasets are available
- Allows to upload one's own dataset



Summary of Total Scores

- Summary statistics
- Interactive histogram

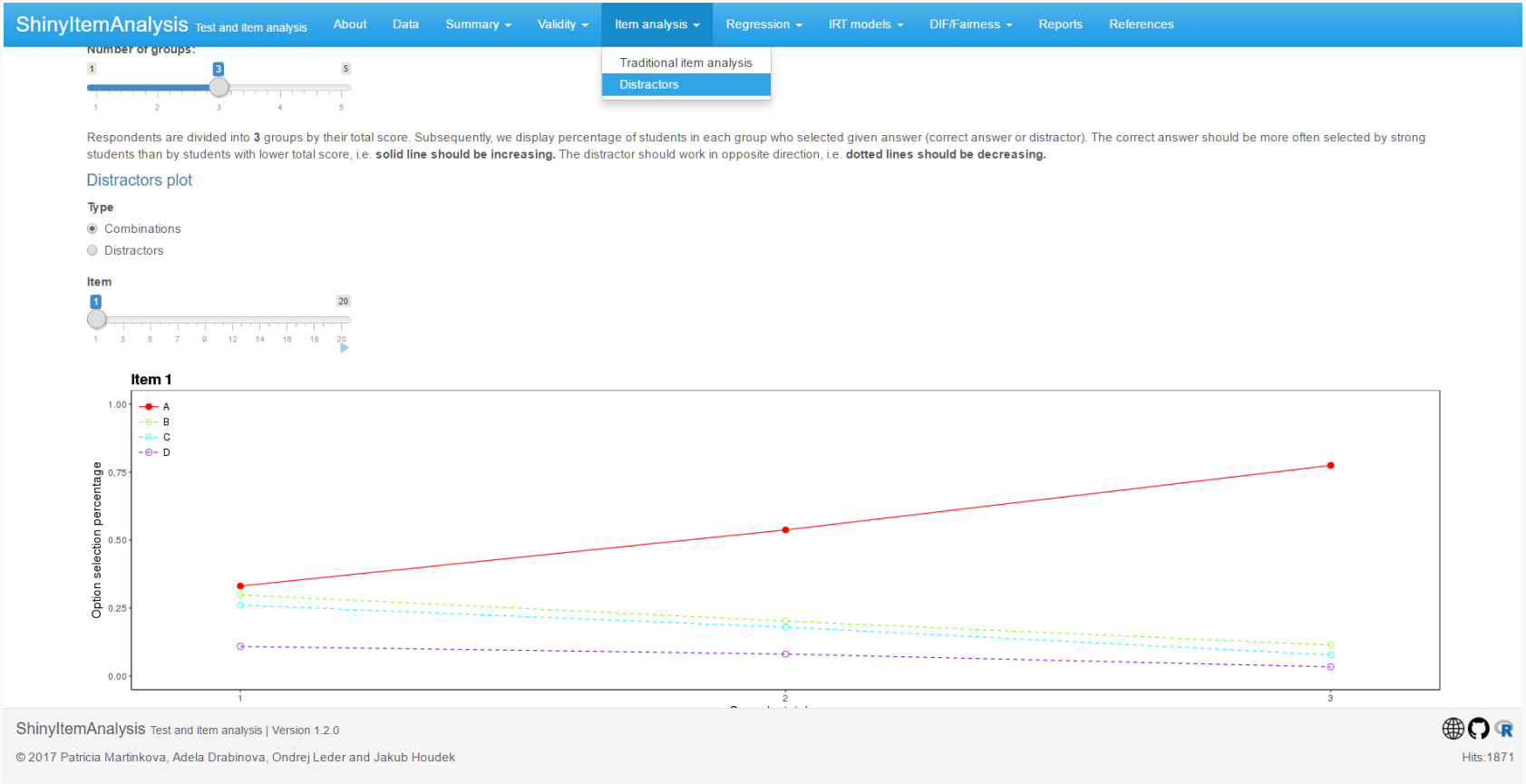


Traditional Item Analysis

- Difficulty, discrimination
- Cronbach's alpha w/o item, index RIT, RIR, etc.

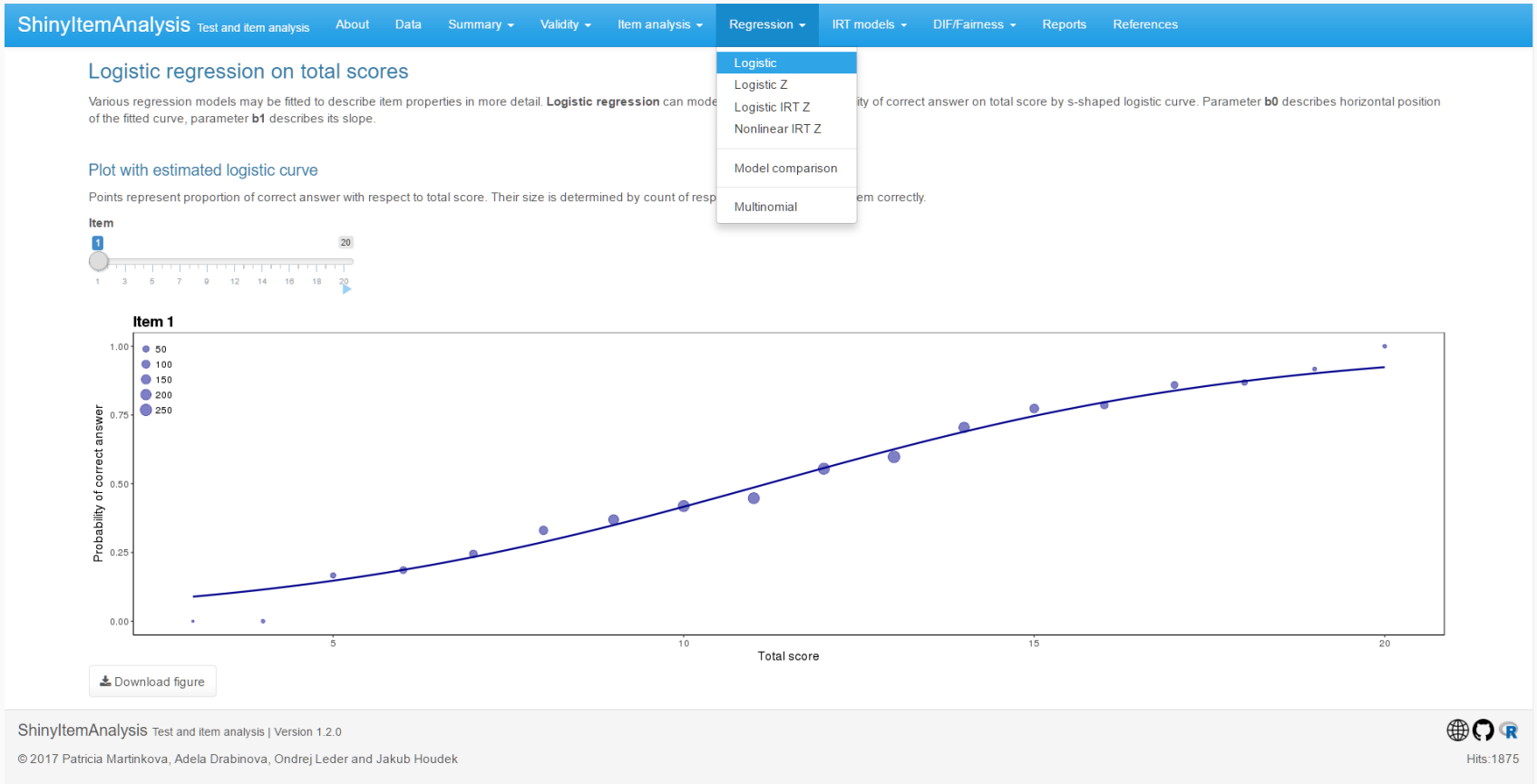
Distractor Analysis

- Displays option selection percentage by total score group
- Number of groups can be changed



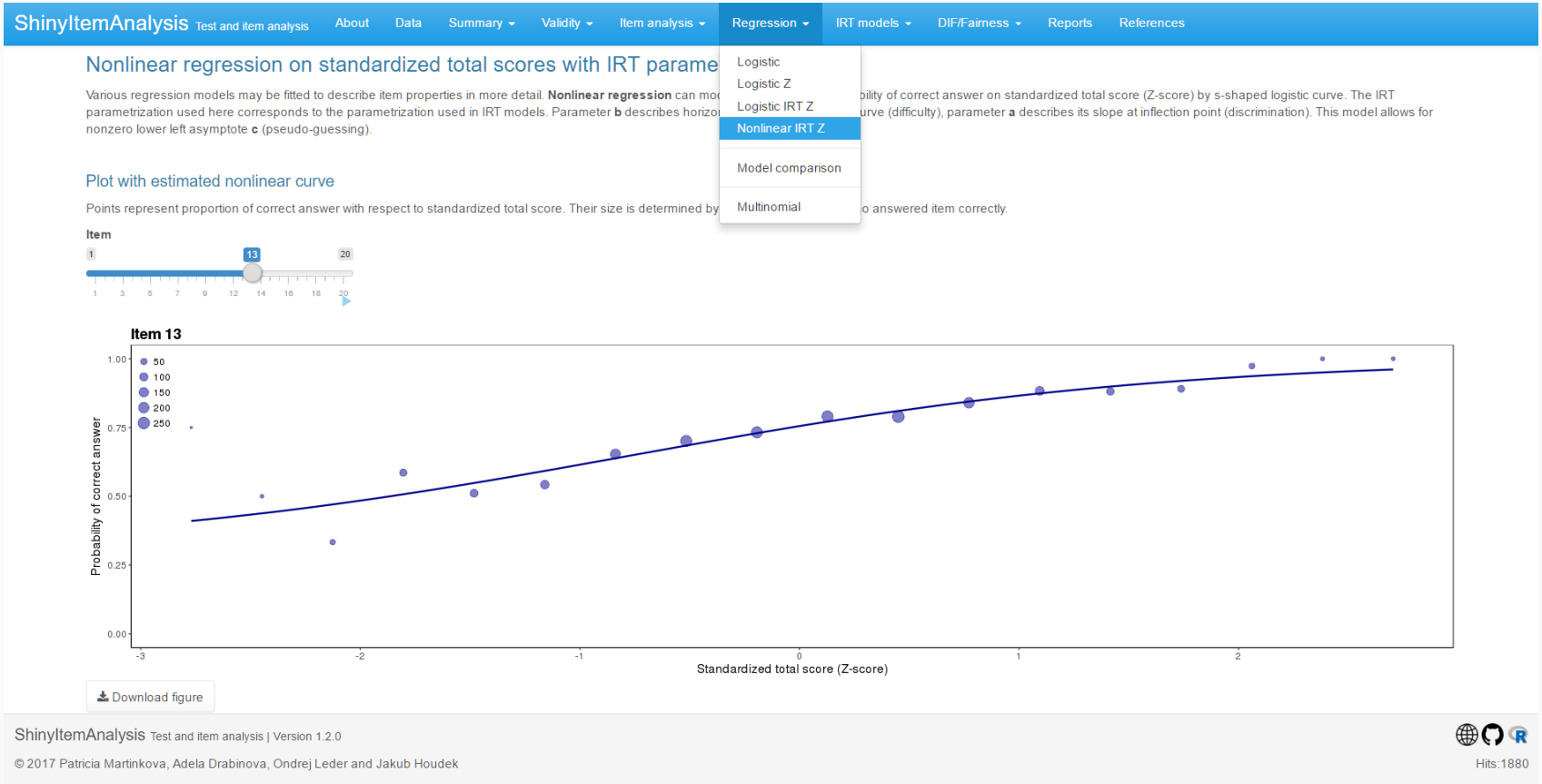
Logistic Regression

- Displays probability of correct answer by total score
- Parameterization can be changed (Z scores, IRT parameterization)



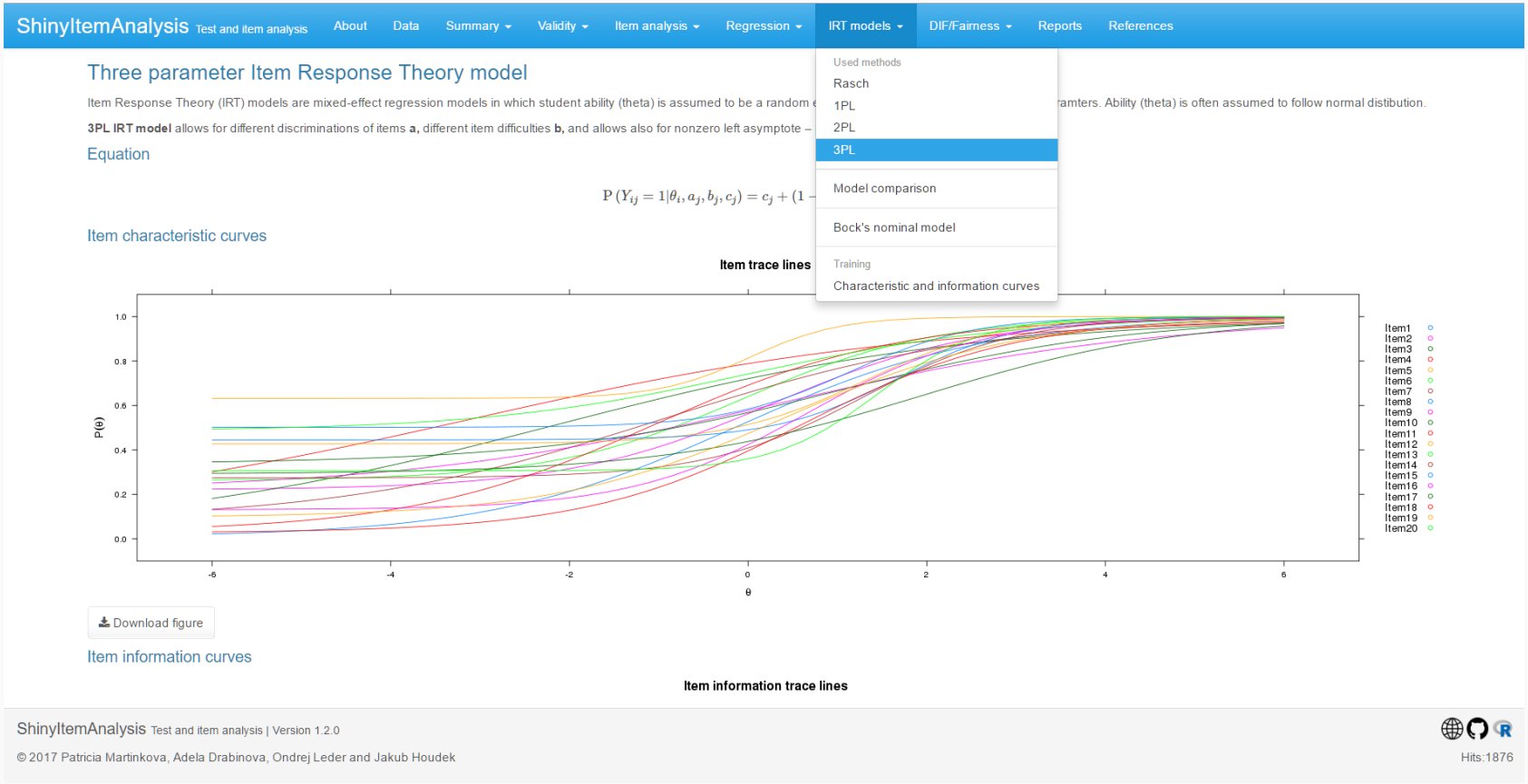
Nonlinear Regression

- Allows for guessing (and inattention)



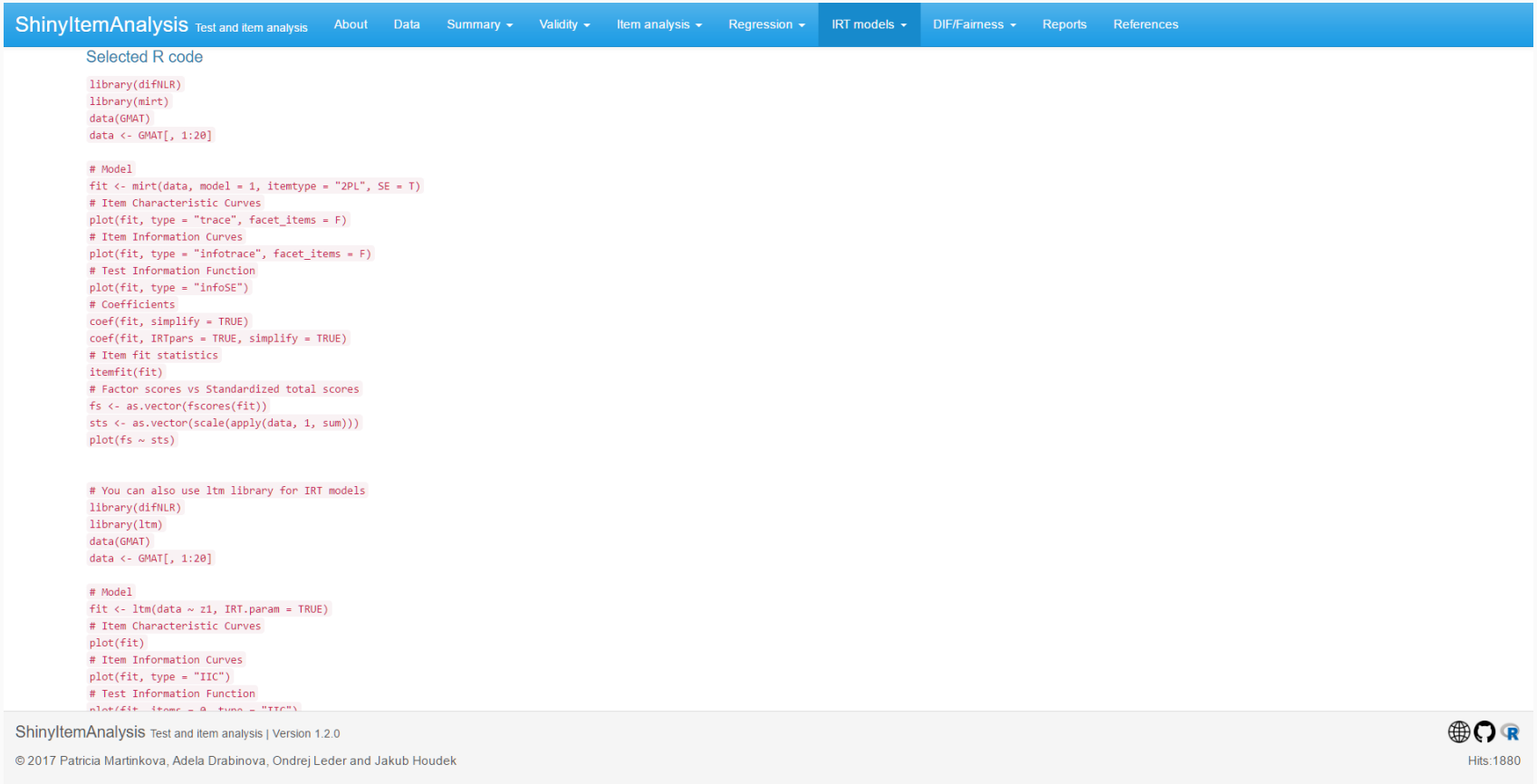
IRT Models

- Conceptualized as nonlinear mixed effect models
- More precise ability estimation



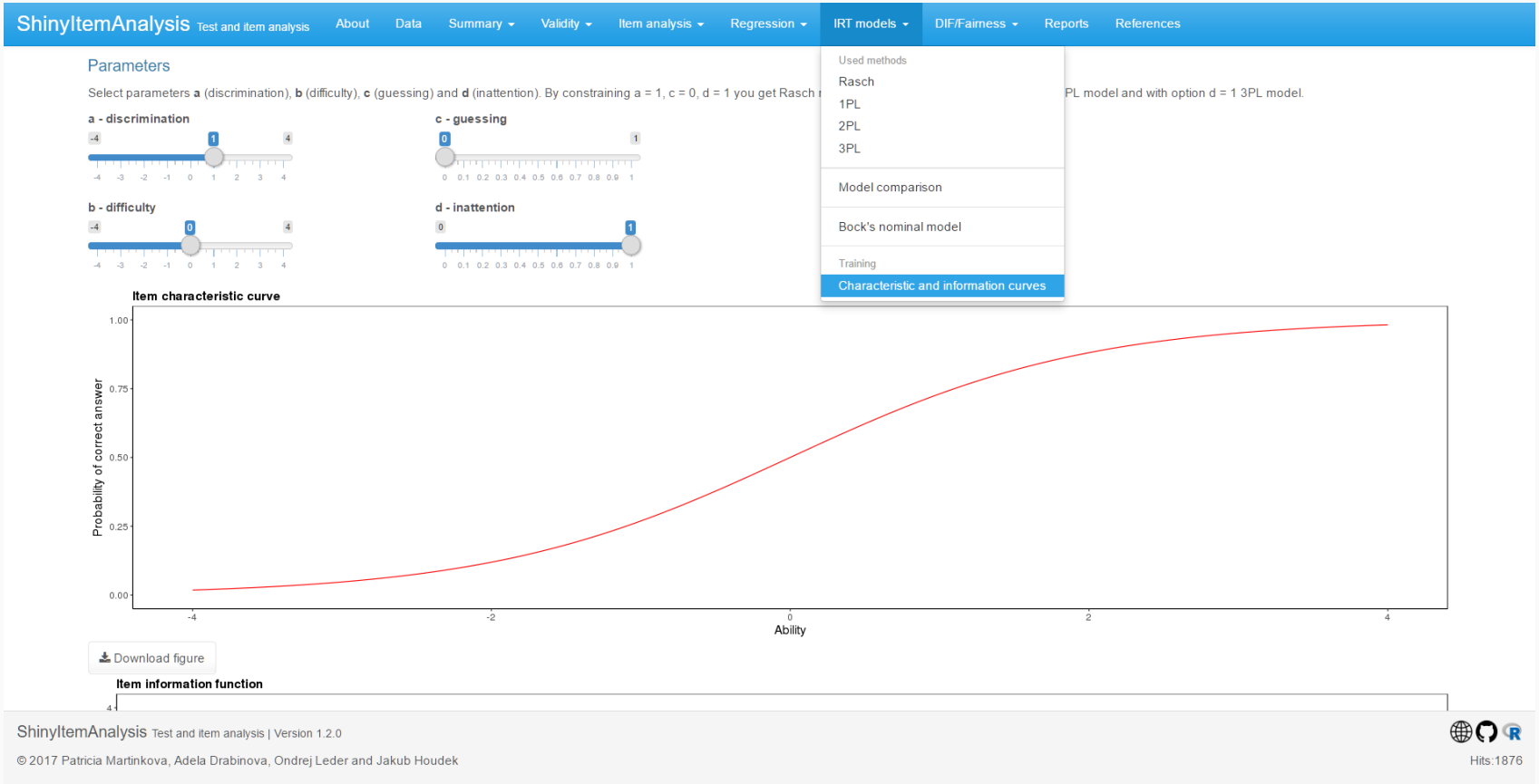
Selected R Code

- Sample R code may be run in separate R session



IRT Models - construct your own item

- Plots Item Characteristic and Information Curves (ICC and IIC) based on selected parameters



`ShinyItemAnalysis` for RESEARCH

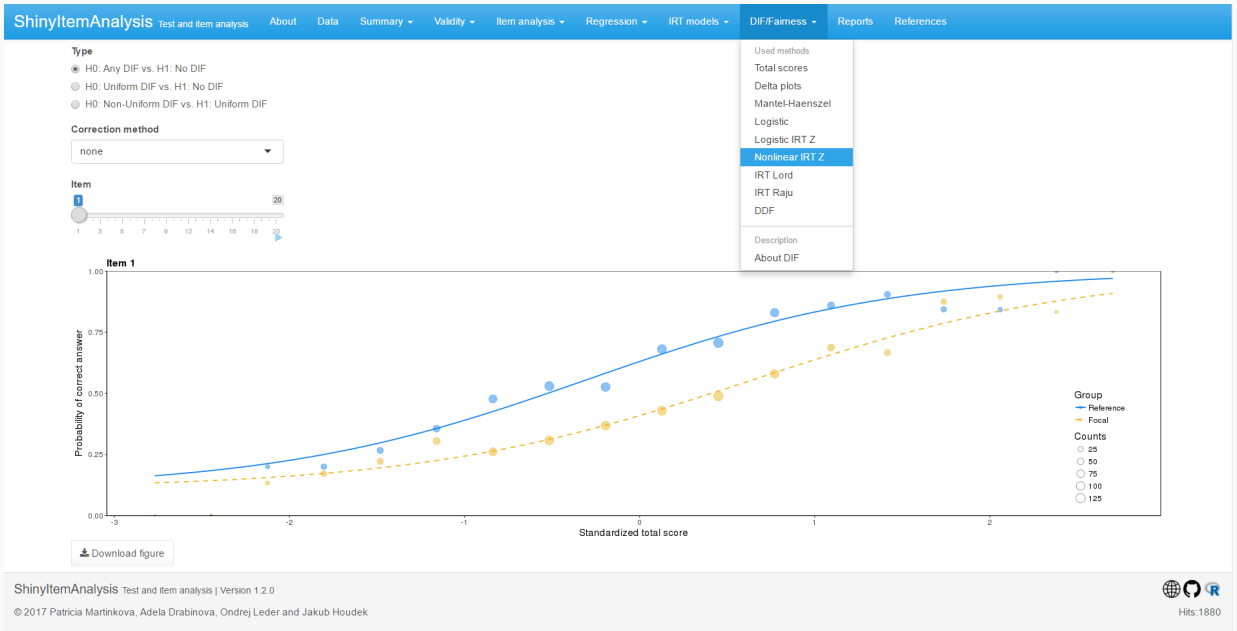
ShinyItemAnalysis for Research

To widespread novel methodology:

- write scientific paper
- ... and provide R code
- ... and provide dataset
- ... and write accompanying R package
- ... and **prepare shiny application**

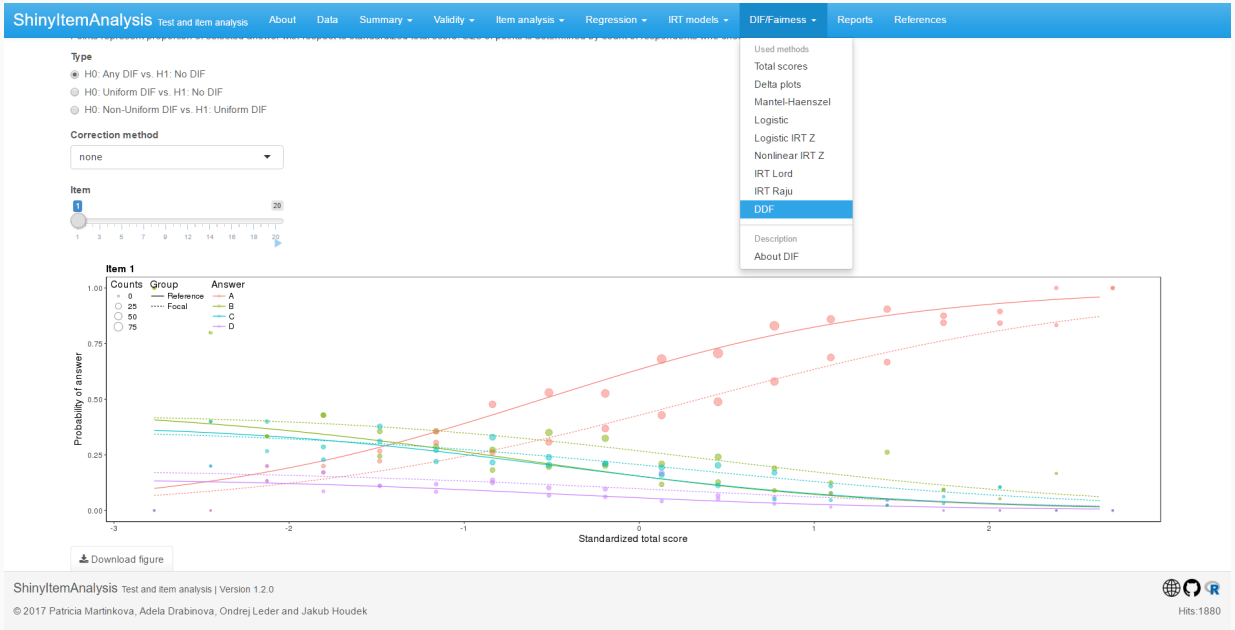
Differential Item Functioning (DIF)

DIF: Students from two groups and *with the same underlying latent ability* have different probability of answering the item correctly.



Differential Distractor Functioning (DDF)

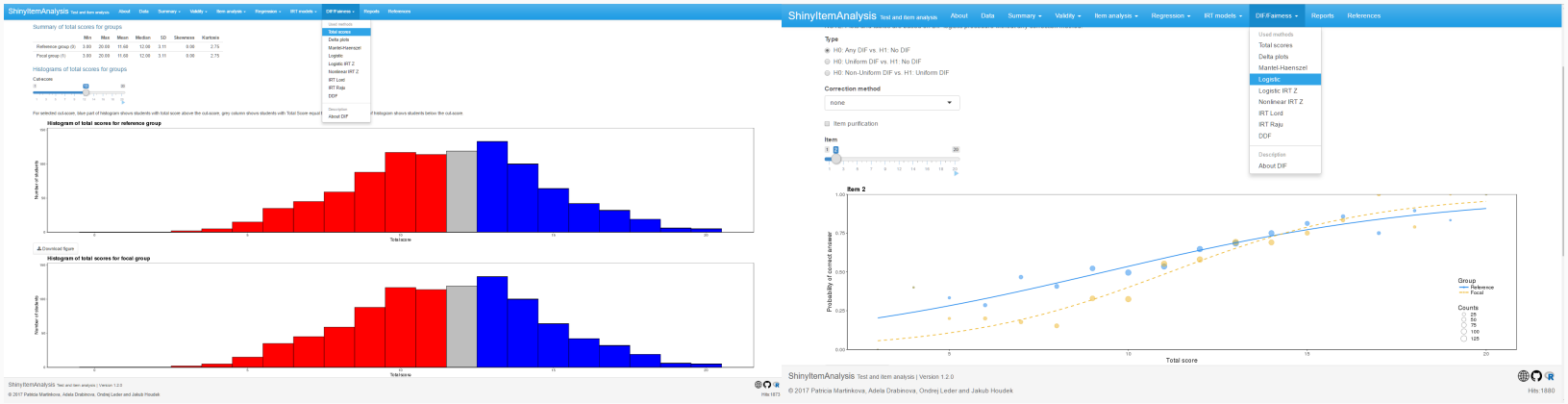
DDF: Students from two groups and *with the same underlying latent ability* have different probability of selecting given options.



To learn more, go to
Adela Drabinova: `difNLR`: Detection of potentional gender/minority bias with extensions of logistic regression (Thursday, July 6) ▶ UseR!2017

Why DIF Analysis Should Be Analyzed Routinely?

- Simulated GMAT data: total scores may have exactly the same distribution, yet there may be DIF present in some items!

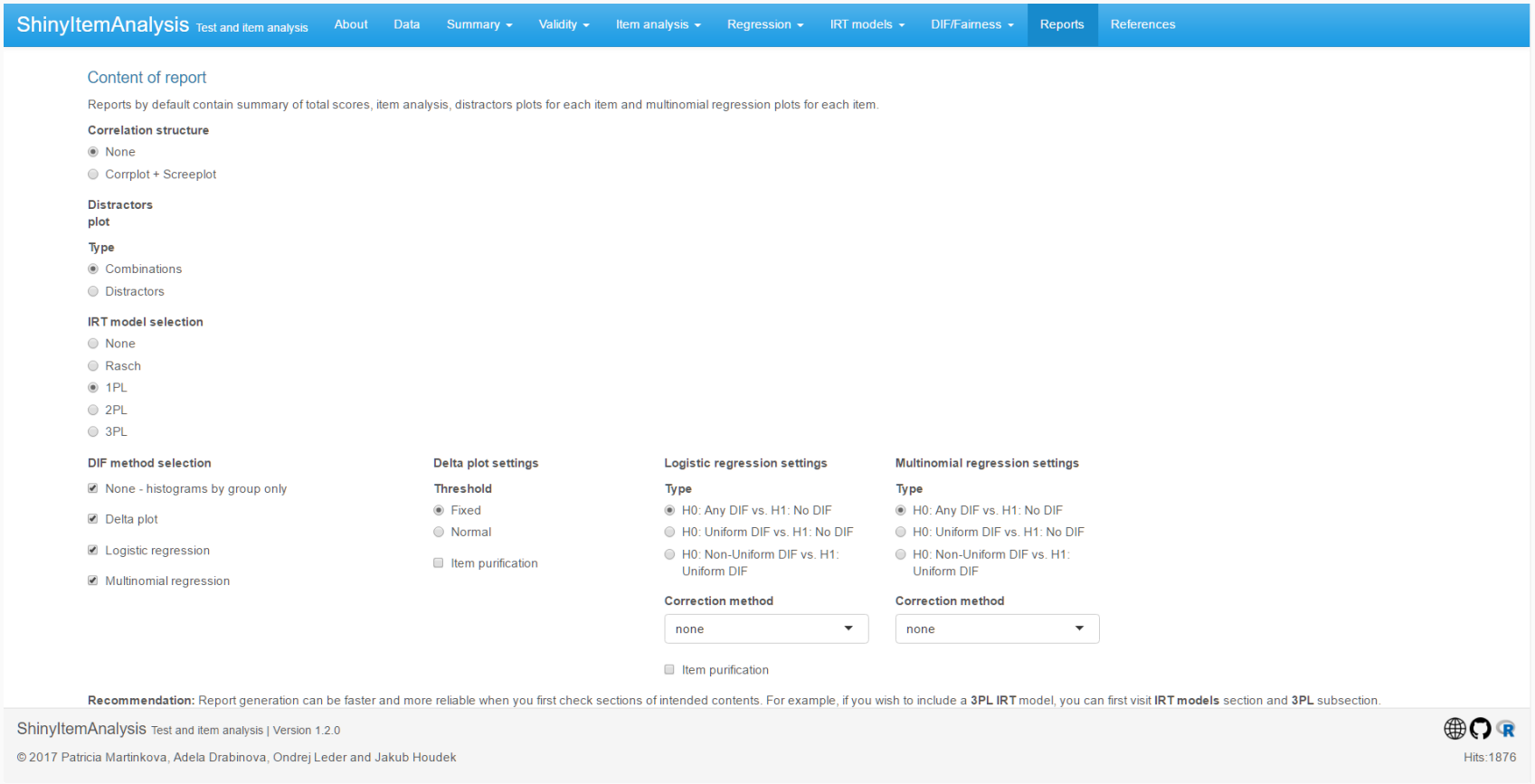


Martinkova et al. (2017): Checking Equity: Why DIF Analysis should be a Routine Part of Developing Conceptual Assessments. CBE-Life Sciences Education, 16(2), rm2. ▶ Online

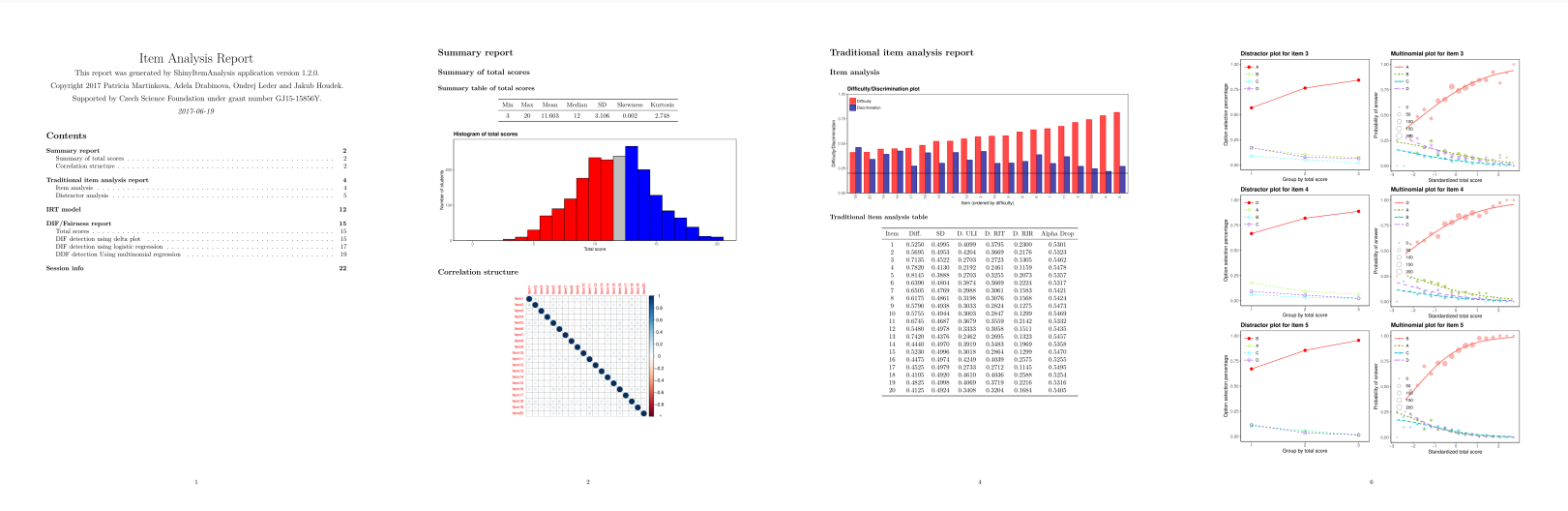
ShinyItemAnalysis for PRODUCTION

Report Generation

- Upload your own data
- Generate PDF/HTML report



PDF Reports



Item Analysis Report

This report was generated by ShinyItemAnalysis application version 1.2.0.

Copyright 2017 Patricia Martinkova, Adela Drabinova, Ondrej Leder and Jakub Houdek.

Supported by Czech Science Foundation under grant number GJ15-15856Y.

2017-06-19

Contents

Summary report 2
Summary of total scores 2
Correlation structure 2
Traditional item analysis report 4
Item analysis 4
Distractor analysis 5
IRT model 12
DIF/Fairness report 15
Total scores 15
DIF detection using delta plot 15
DIF detection using logistic regression 17
DDF detection Using multinomial regression 19
Session info 22

Summary report

Summary of total scores

Summary table of total scores

Min	Max	Mean	Median	SD	Skewness	Kurtosis
3	20	11.603	12	3.106	0.002	2.748

Histogram of total scores

Correlation structure

Traditional item analysis report

Item analysis

Difficulty/Discrimination plot

Traditional item analysis table

Item	Diff.	SD	D. ULI	D. RIT	D. RIR	Alpha Drop
1	0.5250	0.4995	0.4099	0.3795	0.2300	0.5301
2	0.5695	0.4953	0.4204	0.3669	0.2176	0.5323
3	0.7135	0.4522	0.2703	0.2723	0.1305	0.5462
4	0.7820	0.4130	0.2192	0.2461	0.1159	0.5478
5	0.8145	0.3888	0.2703	0.3255	0.2073	0.5357
6	0.6390	0.4804	0.3874	0.3669	0.2224	0.5317
7	0.6505	0.4769	0.2888	0.3061	0.1583	0.5421
8	0.6175	0.4861	0.3198	0.3076	0.1568	0.5424
9	0.5790	0.4938	0.3033	0.2824	0.1275	0.5473
10	0.5755	0.4944	0.3003	0.2847	0.1299	0.5469
11	0.6745	0.4687	0.3679	0.3559	0.2142	0.5332
12	0.5480	0.4978	0.3333	0.3058	0.1511	0.5435
13	0.7420	0.4376	0.2462	0.2695	0.1323	0.5457
14	0.4440	0.4970	0.3919	0.3483	0.1969	0.5358
15	0.5230	0.4996	0.3018	0.2861	0.1299	0.5470
16	0.4475	0.4974	0.4240	0.4039	0.2575	0.5255
17	0.4525	0.4979	0.2733	0.2712	0.1145	0.5495
18	0.4105	0.4920	0.4610	0.4036	0.2588	0.5254
19	0.4825	0.4998	0.4069	0.3719	0.2216	0.5316
20	0.4125	0.4924	0.3408	0.3204	0.1684	0.5405

Distractor plot for item 3

Multinomial plot for item 3

Distractor plot for item 4

Multinomial plot for item 4

Distractor plot for item 5

Multinomial plot for item 5

IRT model

Available models:

Rasch model

$$P(Y_{ij}=1|\theta_i,b_j)=\frac{e^{(\theta_i-b_j)}}{1+e^{(\theta_i-b_j)}}$$

1PL IRT model

$$P(Y_{ij}=1|\theta_i,a,b_j)=\frac{e^{a(\theta_i-b_j)}}{1+e^{a(\theta_i-b_j)}}$$

2PL IRT model

$$P(Y_{ij}=1|\theta_i,a_j,b_j)=\frac{e^{a_j(\theta_i-b_j)}}{1+e^{a_j(\theta_i-b_j)}}$$

3PL IRT model

$$P(Y_{ij}=1|\theta_i,a_j,b_j,c_j)=c_j+(1+c_j)\cdot\frac{e^{a_j(\theta_i-b_j)}}{1+e^{a_j(\theta_i-b_j)}}$$

Wright (item-person) map using 1PL IRT model

Selected model: 3PL IRT model

Item trace lines

Item information trace lines

Test Information and Standard Errors

DIF/Fairness report

Total scores

Summary table of total scores in reference and focal group

	Min	Max	Mean	Median	SD	Skewness	Kurtosis
Reference group (0)	3	20	11.600	12	3.110	0.000	2.750
Focal group (1)	3	20	11.600	12	3.110	0.000	2.750

Histogram of total scores for reference group

Histogram of total scores for focal group

DIF detection using delta plot

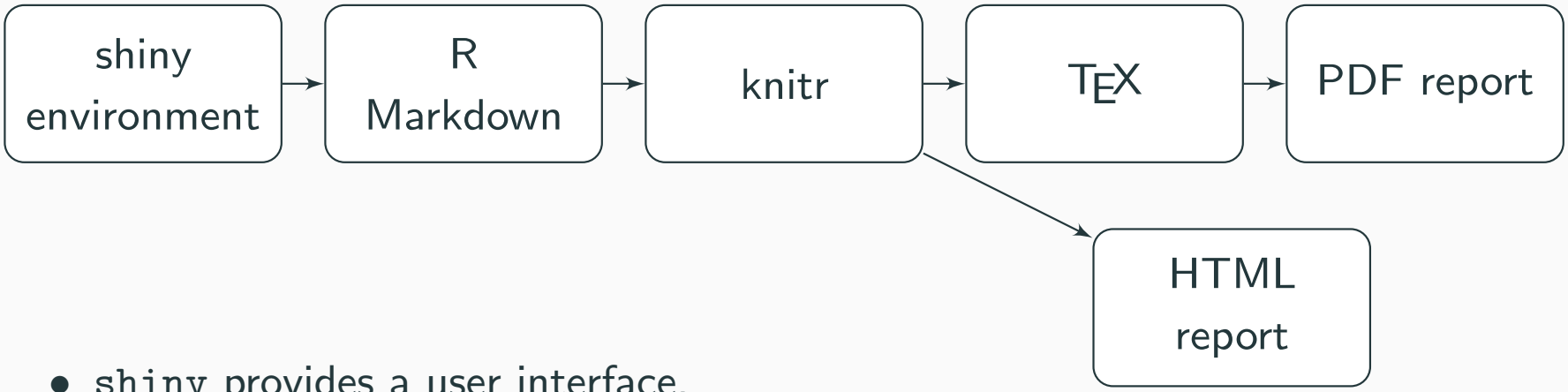
```
##
## Detection of Differential Item Functioning using Angoff's Delta method
##   without item purification
##
## Extreme proportions adjusted by constraining to [0.001; 0.999]
##
```

```
## Item12  6.9577      0.3248
## Item13  3.6414      0.7261
## Item14  5.9761      0.4259
## Item15  2.2061      0.8999
## Item16  2.2730      0.8930
## Item17  3.9542      0.6815
## Item18  8.0746      0.2327
## Item19  7.4243      0.2834
## Item20  5.0249      0.5406
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Items detected as DDF items:
##  Item1
##  Item2
##  Item5
```

DDF multinomial plot for item 1

DDF multinomial plot for item 2

Report Generation



- `shiny` provides a user interface.
- `rmarkdown` for creating templates for PDF/HTML report generation.
- `knitr` for compiling R markdown syntax into HTML/PDF
- TEX for creating PDF reports (latest distribution of TEX is needed).

To learn more see UseR!2017 poster by Jakub Houdek et al. ▸ Online

Conclusion

Conclusion

ShinyItemAnalysis is a shiny application for interactive and step-by-step analysis of educational tests. It is useful for:

- TEACHING of psychometrics and educational measurement
  - offers example datasets, upload of new datasets
  - visualization, interpretation of results
  - sample R Code
- PRODUCTION
  - generates extensive reports for supplied data

ShinyItemAnalysis also promotes our RESEARCH in DIF/DDF detection

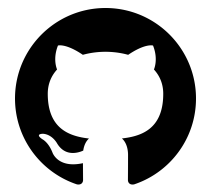
https://shiny.cs.cas.cz/ShinyItemAnalysis/

martinkova@cs.cas.cz

Thank you for your attention!

www.cs.cas.cz/martinkova

martinkova@cs.cas.cz



This research was supported by Czech Science Foundation grant GJ15-15856Y

References

- Martinkova, Drabinova, Leder & Houdek (2017): `ShinyItemAnalysis`: Test and Item Analysis with Shiny.
  https://shiny.cs.cas.cz/ShinyItemAnalysis/
  https://CRAN.R-project.org/package=ShinyItemAnalysis
- Drabinova, Martinkova & Zvara (2017): `difNLR`: Detection of Dichotomous DIF by Non-linear Regression.
  https://CRAN.R-project.org/package=difNLR
- Drabinova & Martinkova (under review): Detection of DIF Based with Non-Linear Regression: Non-IRT Approach Accounting for Guessing.
- Martinkova, Drabinova, Liaw, Sanders, McFarland & Price (2017): Checking Equity: Why DIF Analysis should be a Routine Part of Developing Conceptual Assessments. CBE-Life Sciences Education, 16(2), rm2.
  www.lifescied.org/content/16/2/rm2
- Martinkova, Drabinova & Houdek (2017): ShinyItemAnalysis: Analysis of admission and other educational and psychological tests. Testforum, Accepted.