

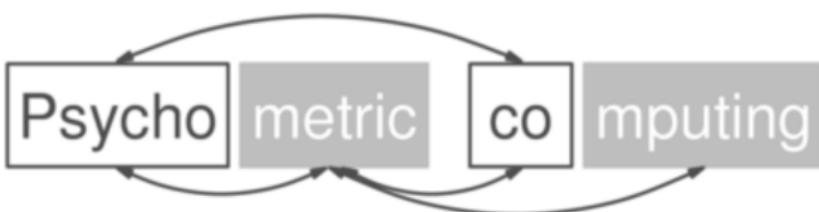
ShinyItemAnalysis for Psychometric Training and to Enforce Routine Analysis of Educational Tests

Patrícia Martinková

Dept. of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences
College of Education, Charles University in Prague

R meetup Warsaw, May 24, 2018

Announcement 1: Save the date for Psychoco 2019!



International Workshop on Psychometric Computing

[Psychoco 2019](#)

February 21 - 22, 2019

Charles University & Czech Academy of Sciences, Prague

www.psychoco.org

Since 2008, the international Psychoco workshops aim at bringing together researchers working on modern techniques for the analysis of data from psychology and the social sciences (especially in R).

Announcement 2: Job offers

Job offers at Institute of Computer Science:

- CAS-ICS Postdoctoral position (deadline: August 30)
 - ICS Doctoral position (deadline: June 30)
 - ICS Fellowship for junior researchers (deadline: June 30)
 - ... further possibilities to participate on grants

E-mail at martinkova@cs.cas.cz if interested in position in the area of

- Computational psychometrics
 - Interdisciplinary statistics
 - Other related disciplines

Outline

1. Introduction
2. ShinyItemAnalysis
3. Teaching psychometrics
4. Routine analysis of tests
5. Discussion

Motivation

- To teach psychometric concepts and methods
 - Graduate courses "IRT models", "Selected topics in psychometrics"
 - Workshops for admission test developers
 - Active learning approach w/ hands-on examples
- To enforce routine analyses of educational tests
 - Admission tests to Czech Universities
 - Physiology concept inventories
 - ... tests of various purposes across the world
- Promotion of own psychometrics research
 - Detection of Differential Item Functioning (DIF)

Need for user-friendly and freely available tool

ShinyItemAnalysis Application



ShinyItemAnalysis

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

Available as:

- R package

- Version 1.2.7 now on [► CRAN](#)
- Newest version on [► GitHub](#)

```
startShinyItemAnalysis()
```

- Online shiny application

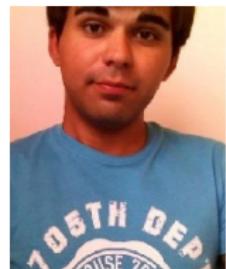
- ICS server in Prague, CZ:

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

- shinyapps.io:

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

Authors and contributors

Patrícia Martinková^{1,2}Adéla Drabinová^{1,3}Jakub Houdek^{1,4}Ondřej Leder³Lubomír Štěpánek^{4,5}

¹Department of Statistical Modelling, Institute of Computer Science, Czech Academy of Sciences

²College of Education, Charles University, Prague

³Department of Probability and Mathematical Statistics, Charles University, Prague

⁴Faculty of Informatics and Statistics, University of Economics, Prague

⁵First Faculty of Medicine, Charles University, Prague

ShinyItemAnalysis application

ShinyItemAnalysis Test and Item analysis About Data Summary Validity Item analysis Regression IRT models DIF/Fairness Reports References

Description

`ShinyItemAnalysis` provides analysis of educational tests (such as admission tests) and their items including:

- Exploration of total and standard scores on [Summary](#) page
- Correlation structure and predictive validity analysis on [Validity](#) page
- Item and distractor analysis on [Item analysis](#) page
- Item analysis by logistic models on [Regression](#) page
- Item analysis by item response theory models on [IRT models](#) page
- Differential item functioning (DIF) and differential distractor functioning (DDF) methods on [DIF/Fairness](#) page

This application is based on the free statistical software R and its shiny package.

For all graphical outputs a download button is provided. Moreover, on [Reports](#) page HTML, or PDF report can be created. Additionally, all application outputs are complemented by selected R code hence the similar analysis can be run and modified in R.

Data

For demonstration purposes, by default, 20-item dataset `GHAT` from R `difNLR` package is used. Other four datasets are available: `GHAT2` and `HSAT-B` from `difNLR` package and `Medical_100` and `HCI` from `ShinyItemAnalysis` package. You can change the dataset (and try your own one) on page [Data](#).

Availability

Application can be downloaded as R package from [CRAN](#). It is also available online at [Czech Academy of Sciences](#) and [shinyapps.io](#).

Version

Current version of `ShinyItemAnalysis` available on [CRAN](#) is 1.2.7. Version available [online](#) is 1.2.7. The newest development version available on [GitHub](#) is 1.2.7. See also older versions: 0.1.0, 0.2.0, 1.0.0, 1.1.0, 1.2.3, 1.2.6.

Authors and contributors



Patricia
Martinkova



Adela
Drabnova



Ondrej
Leder



Jakub
Houdek



Lubomir
Stepanek

List of packages used

<code>library(corrplot)</code>	<code>library(ggplot2)</code>	<code>library(moments)</code>	<code>library(rmarkdown)</code>
<code>library(CT)</code>	<code>library(grid)</code>	<code>library(msn)</code>	<code>library(shiny)</code>



ShinyItemAnalysis Test and Item analysis | Version 1.2.7
© 2018 Patricia Martinkova, Adela Drabnova, Ondrej Leder and Jakub Houdek



Hits: 5776

R package ShinyItemAnalysis downloads from CRAN

Package CRAN downloads over time

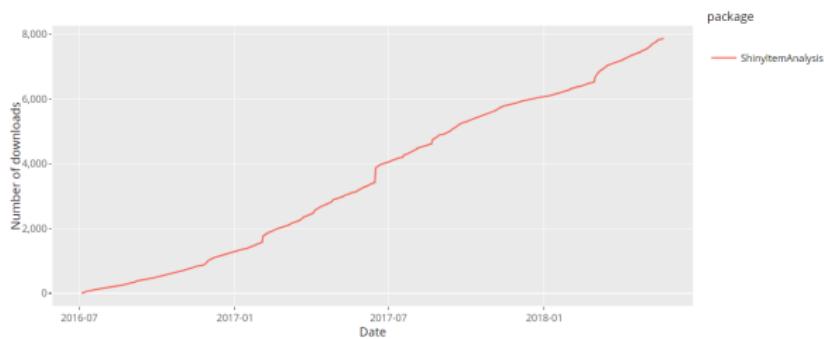
Enter an R package to see the # of downloads over time from the RStudio CRAN Mirror. You can enter multiple packages to compare them

Packages:

Data Transformation:
 Daily
 Weekly
 Cumulative

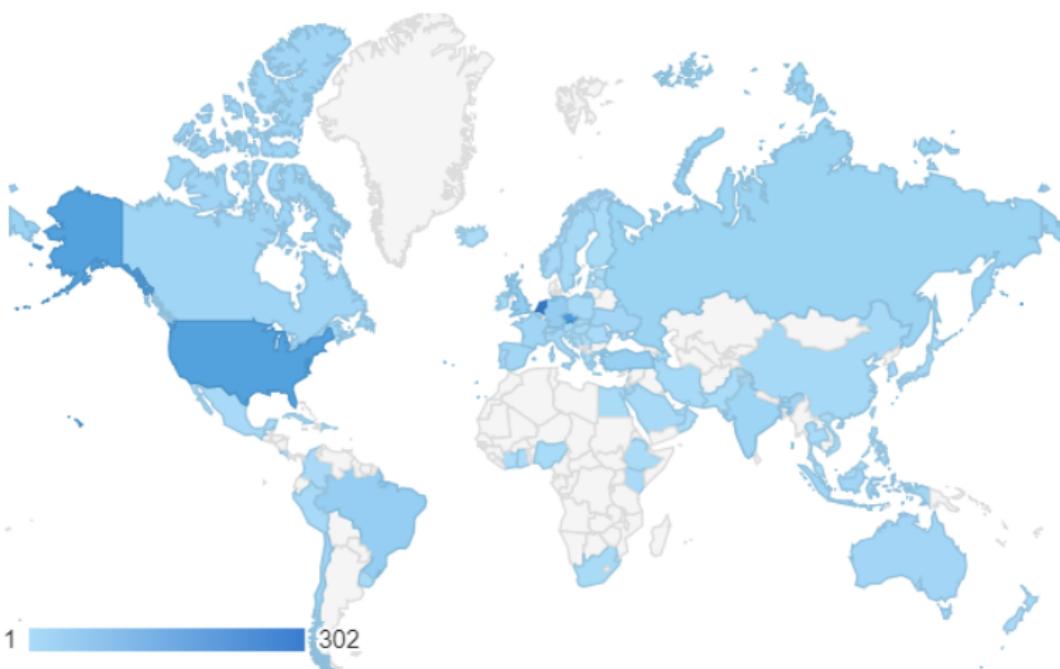
Date range: yyyy-mm-dd
 to

Created using the [cranlogs](#) package. This app is not affiliated with RStudio or CRAN. You can find the code for the app [here](#), or read more about it [here](#).



ShinyItemAnalysis
Title: Test and Item Analysis via Shiny
Latest version: 1.2.7

ShinyItemAnalysis online app is used worldwide!



ShinyItemAnalysis for teaching psychometrics

Who do we teach:

- Graduate students of different fields (Psychometrics ▶ NMST570)
- Faculties, university stakeholders

Some helpful features:

- Example datasets, allows to upload own data
- Building models in a step-by-step way
- Models, estimates, interactive interpretation of results
- Interactive training and exercises
- Provides sample R code

Datasets

- Five toy datasets are available
- Allows to upload and preview one's own dataset

ShinyItemAnalysis Test and Item analysis About Data Summary Validity Item analysis Regression IRT models DIF/Fairness Reports References

HCI (McFarland et al. 2017) is a real dataset of Homeostasis Concept Inventory from ShinyItemAnalysis R package. The dataset represents responses of 681 subjects (405 males, 246 females) to multiple-choice test of 20 items. HCI contains criterion variable - indicator whether student plans to major in the life sciences.

Select dataset
GMAT

Upload your own datasets

Main data file should contain responses of individual respondents (rows) to given items (columns). Header may contain item names, no row names should be included. If responses are in unscored ABCD format, the key provides correct response for each item. If responses are scored 0-1, key is vector of 1s.

Group is 0-1 vector, where 0 represents reference group and 1 represents focal group. Its length need to be the same as number of individual respondents in main dataset. If the group is not provided then it won't be possible to run DIF and DDF detection procedures on DIF/Fairness page.

Criterion variable is either discrete or continuous vector (e.g. future study success or future GPA in case of admission tests) which should be predicted by the measurement. Again, its length needs to be the same as number of individual respondents in the main dataset. If the criterion variable is not provided then it will not be possible to run validity analysis in Predictive validity section on Validity page.

In all data sets header should be either included or excluded. Columns of dataset are by default renamed to item and number of particular column. If you want to keep your own names, check box Keep item names below. Missing values in scored dataset are by default evaluated as 0. If you want to keep them as missing, check box Keep missing values below.

Choose data (csv file)
Browse... HCI_ABCD.csv Upload complete

Choose key (csv file)
Browse... HCI_key.csv Upload complete

Choose groups for DIF (optional)
Browse... HCI_group.csv Upload complete

Choose criterion variable (optional)
Browse... HCI_major.csv Upload complete

Submit Data

Your data were successfully uploaded. Check them in Data exploration tab.

Data specification

Header
 Keep item names
 Keep missing values

Comma
 Semicolon
 Tab

Double Quote
 Single Quote

 ShinyItemAnalysis Test and item analysis | Version 1.2.7
© 2018 Patričia Martinková, Adéla Draženová, Ondřej Leder and Jakub Housák

 Hits 5776

Summary of Total Scores

- Summary statistics
- Interactive histogram

ShinyItemAnalysis Test and item analysis About Data Summary Validity Item analysis Regression IRT models DIF/Fairness Reports References

Analysis of total scores

Summary table

Min	Max	Mean	Median	SD	Skewness	Kurtosis
3.00	20.00	11.60	12.00	3.11	0.00	2.75

Total scores
Standard scores

Histogram of total score

Cut-score

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

Number of students

Total score

Download figure

ShinyItemAnalysis Test and item analysis | Version 1.2.0
© 2017 Patricia Martinková, Adela Drabinová, Ondřej Leder and Jakub Houdek

Hits 1871

Criterion validity

- Only when criterion variable is available (study success, GPA, etc.)
- Available for total score as well as for items



Correlation structure

- Correlations between items
- Item clusters

ShinyItemAnalysis Test and item analysis About Data Summary Validity Item analysis Regression IRT models DIF/Fairness Reports References

Correlation structure Predictive validity

Polychoric correlation heat map

Polychoric correlation heat map is a correlation plot which displays a polychoric correlations of items. The size and shade of circles indicate how much the items are correlated (larger and darker circle means larger correlation). The color of circles indicates in which way the items are correlated - blue color shows positive correlation and red color shows negative correlation.

Polychoric correlation heat map can be reorder using hierarchical clustering method below. Ward's method aims at finding compact clusters based on minimizing the within-cluster sum of squares. Ward's n. 2 method used squared dissimilares. Single method connects clusters with the nearest neighbours, i.e. the distance between two clusters is calculated as the minimum of distances of observations in one cluster and observations in the other clusters. Complete linkage with farthest neighbours, i.e. maximum of distances. Average linkage method used the distance based on weighted average of the individual distances. With McQuitty method used unweighted average. Median linkage calculates the distance as the median of distances between an observation in one cluster and observation in the other cluster. Centroid method used distance between centroids of clusters.

With number of clusters larger than 1, the rectangles represent clusters are drawn.

Number of clusters Clustering method

3 Ward's

Download figure

ShinyItemAnalysis Test and item analysis | Version 1.2.6
© 2018 Patricia Martinková, Adéla Drábová, Ondřej Leder and Jakub Houděk

R HHS 6062

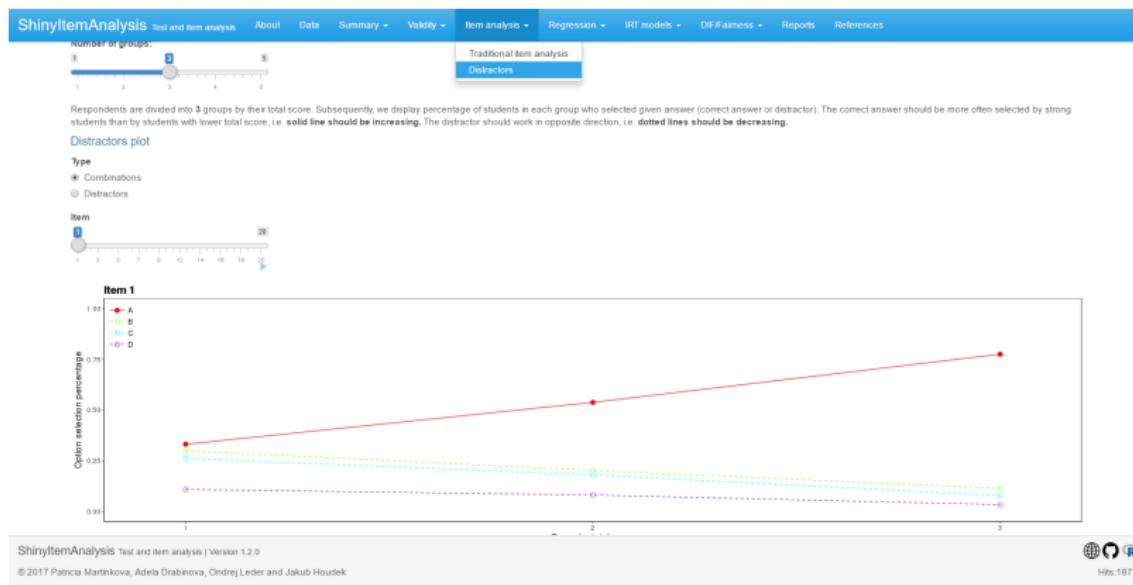
Traditional Item Analysis

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



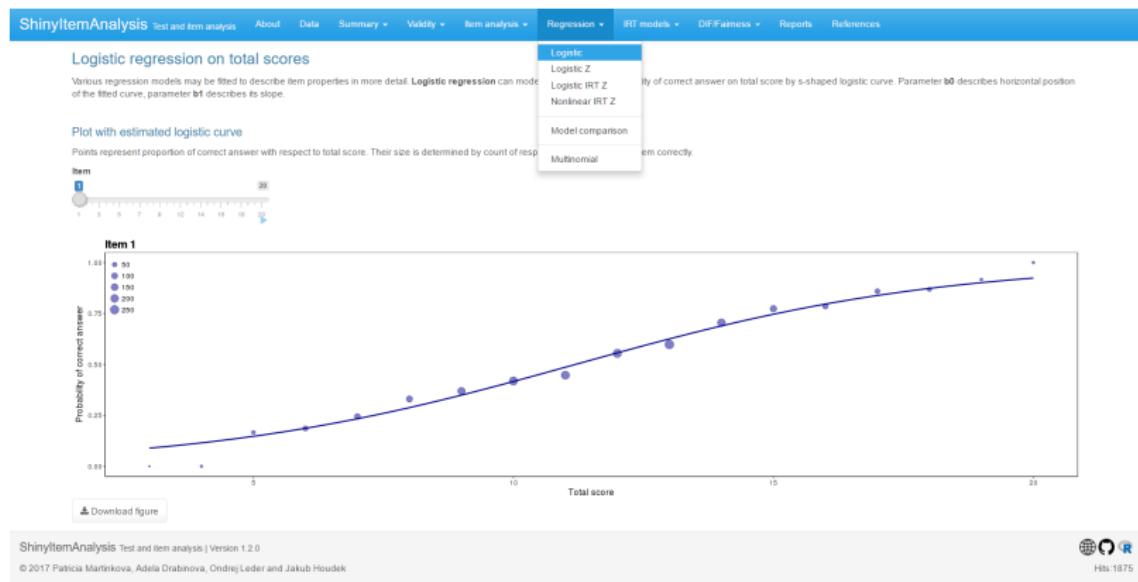
Distractor Analysis

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



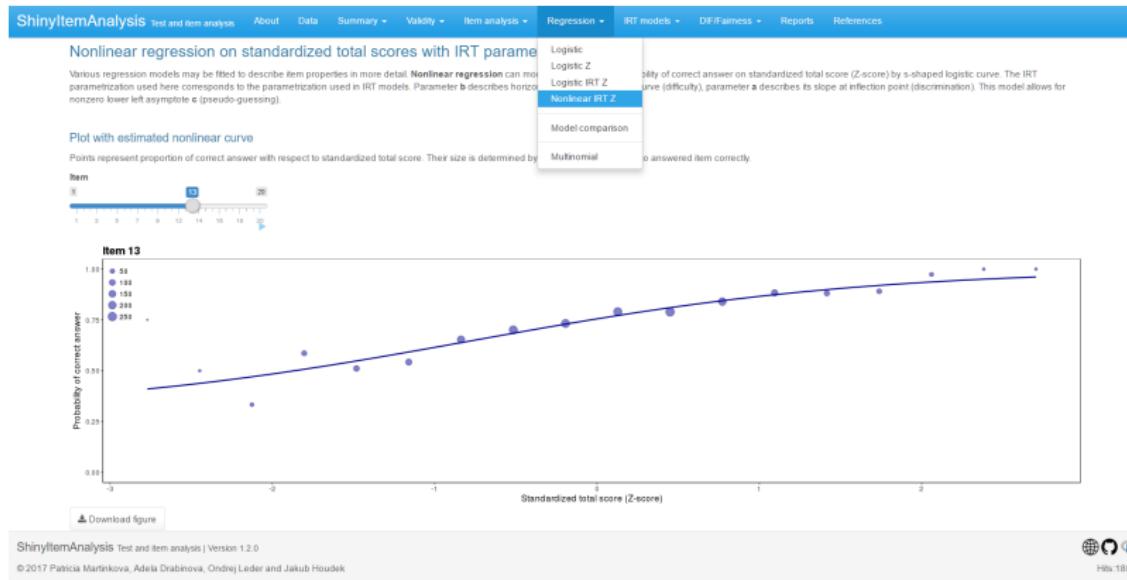
Logistic Regression

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



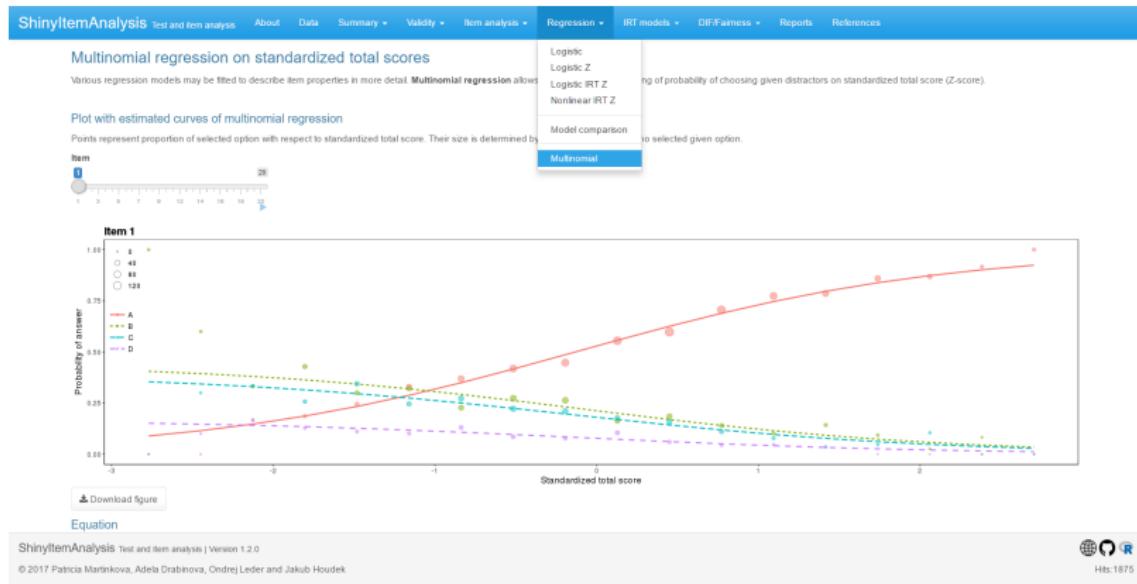
Nonlinear Regression

- Allows for guessing (and inattention)



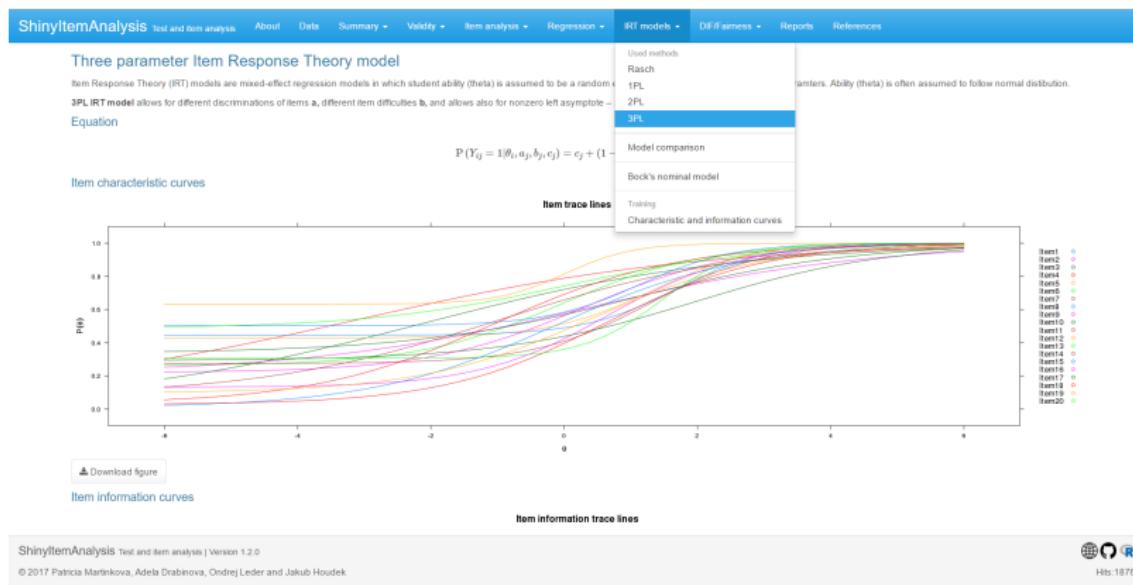
Multinomial Regression

- Allows for joint modeling of distractors



IRT Models

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



Dichotomous IRT Models - interactive training

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

ShinyItemAnalysis Test and item analysis About Data Summary Validity Item analysis Regression IRT models DIF/Invariance Reports References

Parameters

Select parameters a (discrimination), b (difficulty), c (guessing) and d (inattention). By constraining $a = 1$, $c = 0$, $d = 1$ you get Rasch model. With option $c = 0$ and $d = 1$ you get 2PL model and with option $d = 1$ 3PL model.

When different curve parameters describe properties of the same item but for different groups of respondents, this phenomenon is called Differential Item Functioning (DIF). See further section for more information.

a - discrimination

b - difficulty

c - guessing

d - inattention

Select also the value of latent ability if to see the interpretation of the item characteristic curves.

θ - latent ability

Equations

$$P(\theta|a, b, c, d) = c + (d - c) \cdot \frac{e^{a(\theta-b)}}{1 + e^{a(\theta-b)}}$$

$$I(\theta|a, b, c, d) = a \cdot (d - c) \cdot \frac{e^{a(\theta-b)}}{\left[1 + e^{a(\theta-b)}\right]^2}$$

Interpretation: The probability of correct answer with latent ability $\theta = 0$ in red item with parameters $a = 1$, $b = 0$, $c = 0$, $d = 1$ is equal to 0.59. The probability of correct answer with latent ability $\theta = 0$ in blue item with parameters $a = 2$, $b = 0.5$, $c = 0$, $d = 1$ is equal to 0.77.

Item characteristic curve

Item information function

ShinyItemAnalysis 0.6 Test and item analysis | Version 1.2.7
© 2018 Patrik Martinková, Adéla Chaloupečka, Ondřej Látek and Jakub Housák

RRS 5776

Polytomous IRT Models - interactive training

- Plots Category Response Curves and Expected Item Score based on selected parameters

ShinyItemAnalysis Test and item analysis About Data Summary Validity Item analysis Regression IRT models DIF/Invariance Reports References

Parameters
Select number of responses and difficulty for cumulative probabilities b and common discrimination parameter a. Cumulative probability $P(Y \geq 0)$ is always equal to 1 and it is not displayed, corresponding category probability $P(Y = 0)$ is displayed with black color.

Highest score: 4
a - discrimination: 0.5
b1 - difficulty: 0.2 (red)
b2 - difficulty: 0.4 (yellow)
b3 - difficulty: 0.6 (green)
b4 - difficulty: 0.8 (blue)

Equations

$$\pi_{k+} = P(Y \geq k|\theta, a, b_k) = \frac{e^{a(\theta-k)}}{1 + e^{a(\theta-k)}}$$
$$\pi_k = P(Y = k|\theta, a, b_k, b_{k+}) = \pi_{k+} - \pi_{k+1+}$$
$$E(Y|\theta, a, b_1, \dots, b_K) = \sum_{k=0}^K k\pi_k$$

Plots

Cummulative probabilities

Category probabilities

Expected item score

ShinyItemAnalysis 5.0. Test and item analysis | Version 1.2.7
© 2018 Patrícia Martinková, Alena Drahoninová, Ondřej Lálek and Jakub Hrušek

RRS 5776

Dichotomous IRT Models - check your understanding

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 the following exercises with an accuracy of up to 0.05. Then click on **Submit answers** button. If you need a hint, click on blue button with question mark.

- Sketch item characteristic and information curves

- Calculate probability of correct answer for latent abilities $\theta = -2, -1, 0, 1, 2$

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

0

0.3

0.4

0.9

0.9

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

0

0.2

0.5

0.8

0.9

- For what level of ability θ are the probabilities equal?

$\theta = ?$

0.2

- 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

Help

Look at the figure on the right side.
Which curve does have larger value for desired level of ability θ ?

40% correct. Try again

Submit answers

Exercise 2

Consider now 2 items with following parameters

Item 1: $a = 1.5, b = 0, c = 0, d = 0$

Item 2: $a = 1.5, b = 0, c = 0.2, d = 0$

For these items fill the following exercises with an accuracy of up to 0.05. Then click on **Submit answers** button.

- What is the lower asymptote for items?

Item 1:



Selected R Code

- Sample R code may be run and modified in separate R session

ShinyItemAnalysis Test and item analysis About Data Summary • Validity • Item analysis • Regression • IRT models • Diff/Fairness • Reports References

Selected R code

```
library(difR)
library(ltm)
data(WSAT)
data(GMAT)

# Model
fit <- mirt(data, model = 1, itemtype = "2PL", SE = T)
# Item Characteristic Curves
plot(fit, type = "trace", facet_items = F)
# Item Information Curves
plot(fit, type = "Iinfotrace", facet_items = F)
# Test Information Function
plot(fit, type = "lInfofct")
# Coefficients
coef(fit, simplify = TRUE)
coef(fit, IRTpars = TRUE, simplify = TRUE)
# Item fit statistics
itemfit(fit)
# Factor scores vs Standardized total scores
fs <- as.vector(fit$scores)
sts <- as.vector(scale(apply(data, 1, sum)))
plot(fs ~ sts)

# You can also use ltm library for IRT models
library(difR)
library(ltm)
data(WSAT)
data(GMAT)

# Model
fit <- ltm(data = x1, IRT.param = TRUE)
# Item Characteristic Curves
plot(fit)
# Item Information Curves
plot(fit, type = "IIC")
# Test Information Function
#itemfit(fit, IRTpars = TRUE)
```

ShinyItemAnalysis Test and item analysis | Version 1.2.0

© 2017 Patricia Martinková, Adela Drabinová, Ondřej Leder and Jakub Houdek

 Hits:1880

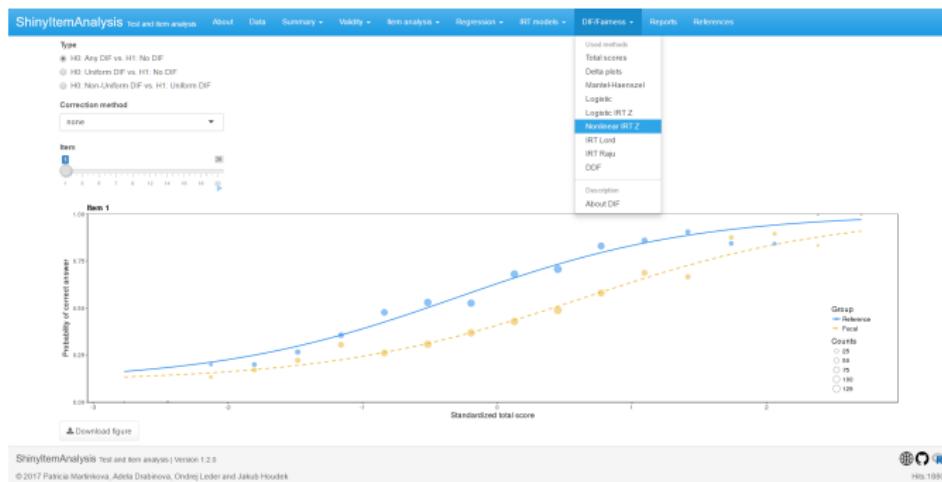
ShinyItemAnalysis to promote our research

App promotes methods and research of our team:

- Detection of Differential Item Functioning (DIF)
- Detection of Differential Distractor Functioning (DDF)
- Why DIF/DDF analysis should be routine part of test development
- etc.

Differential Item Functioning (DIF)

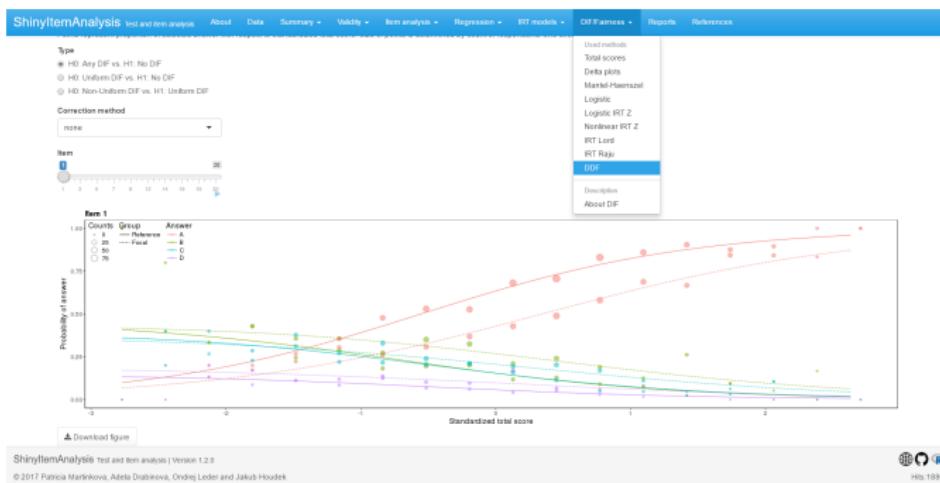
DIF: Students from two groups and *with the same underlying latent ability* have different probability of answering the item correctly.



Drabinová & Martinková (2017): Detection of DIF with Non-Linear Regression: Non-IRT Approach Accounting for Guessing. Journal of Educational Measurement, 54(4), pp. 498-517. doi [10.1111/jedm.12158](https://doi.org/10.1111/jedm.12158)

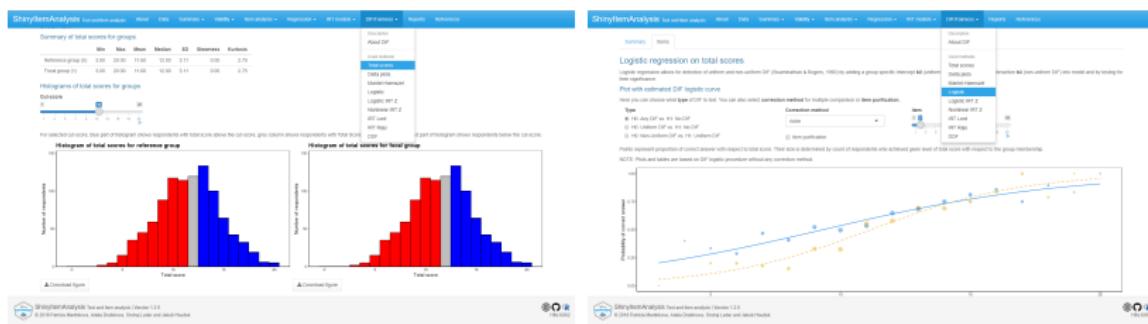
Differential Distractor Functioning (DDF)

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



Why DIF Analysis Should Be Analyzed Routinely?

- Dataset HCI: significant difference in total score between males and females, yet no DIF item!
- Simulated GMAT data: total scores may have exactly the same distribution, yet there may be DIF present in some items!



Martinková, Drabinová, 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. doi [10.1187/cbe.16-10-0307](https://doi.org/10.1187/cbe.16-10-0307)

Routine validation of educational tests

Supporting tool for routine validation of educational tests:

- Upload your own data
- Generate PDF/HTML report
- Local or online version

Report generation - settings

- Chose methods, customize settings
- Chose report format (PDF/HTML)

ShinyItemAnalysis Test and Item analysis About Data Summary Validity Item analysis Regression IRT models DIF/Fairness Reports References

Download report

Settings of report

ShinyItemAnalysis offers an option to download a report in HTML or PDF format. PDF report creation requires latest version of **MiK^TeX** (or other TeX distribution). If you don't have the latest installation, please, use the HTML report.

There is an option whether to use customize settings. By checking the **Customize settings** local settings will be offered and use for each selected section of report. Otherwise the settings will be taken from pages of application. You can also include your name into report as well as the name of dataset which was used.

Format of report HTML PDF Customize settings Author Joe Doe Dataset HCI dataset

Content of report

Reports by default contain summary of total scores, table of standard scores, item analysis, distractors plots for each item and multinomial regression plots for each item. Other analyses can be selected below.

Validity Correlation structure

Number of clusters Clustering method 1 Wards

Predictive validity

Difficulty/discrimination plot

Number of groups: 5 Which two groups to compare: 1 2 3 4 5 1 2 3 4 5

Distractors plots

Type Combinations Distractors Number of groups: 5 1 2 3 4 5

ShinyItemAnalysis Test and Item analysis | Version 1.2.7
© 2018 Patricia Martinková, Adela Drabinová, Ondřej Leder and Jakub Houšek

ShinyItemAnalysis R Hits 5776

Report generation

- Generate report (run analyses)
- Download report (compile text into HTML/PDF)

ShinyItemAnalysis Test and Item analysis About Data Summary ▾ Validity ▾ Item analysis ▾ Regression ▾ IRT models ▾ DIF/Fairness ▾ Reports References

Number of groups:

Which two groups to compare:

Distractors plots

Type
 Combinations
 Distractors

IRT model selection

None
 Rasch
 1PL
 2PL
 3PL
 4PL

DIF method selection

None - histograms by group only
 Delta plot
 Logistic regression
 Multinomial regression

Delta plot settings

Threshold
 Fixed
 Normal
 Item purification

Logistic regression settings

Type
 H0: Any DIF vs. H1: No DIF
 H0: Uniform DIF vs. H1: No DIF
 H0: Non-Uniform DIF vs. H1: Uniform DIF

Multinomial regression settings

Correction method
 BH
 BH

Item purification

Logistic regression Loading

Recommendation: Report generation can be faster and more reliable when you first check sections of intended contents. For example, if you wish to include a SPL IRT model, you can first visit IRT models section and SPL subsection.

Generate report

ShinyItemAnalysis Test and Item analysis | Version 1.2.7
© 2018 Patricia Martinková, Adéla Drábníková, Ondřej Leder and Jakub Houdek

Creating content

Sample PDF report

A screenshot of a Shiny Item Analysis application. The title bar features a blue header with the text "Test and Item Analysis Report". Below the header is a logo consisting of a hexagon containing the text "Shiny ItemAnalysis" above a bar chart icon. The main content area has a light gray background. In the center, there is a white rectangular box containing the text "HCI dataset" and "Joe Doe". At the bottom of the page, there is a footer bar with the text "Report generated on May 8, 2020" and the "RStudio Shiny" logo.

Contents	
Introduction	
Summary	
Total scores	1
Summarized value of total scores	1
Designs of total scores	1
Scaling	
Summary statistic of standard scores	2
Validity	
Construct structure	4
Effect-size correlation load map	4
Reliability	4
Transformations	
Basic analysis	4
General linear transformation	4
Dimensional manipulation factor	4
IRT models	
Item response theory (IRT) vs. IRT model	5
Item parameters	5
Reliable measures	5
Reliable and valid measures	5
Parameter estimation and item fit	5
CFI/Pathmodel analysis	
Summary var. of total scores in reference anchorage group	10
Summary var. of total scores in target anchorage group	10
Summary statistic	10
CFI detection along sample regression	10
GOF detection along multilevel regression	10
Model fit	10
General tests	
	24

ICC statistics

Summary

Total scores

Summary table of total scores

Table below summarizes basic characteristics of total scores including sample size, mean, median, standard deviation, range, minimum and maximum. The row shows the first three moments (mean, median, SD), while the last four central moments show the second moment (range), third moment (skewness) and fourth moment (kurtosis). The last two rows show the minimum and maximum values observed in the data set. The last two rows show the minimum and maximum values observed in the data set.

	1983	1984	Median	SD	Range	Kurtosis	SKEWNESS
n	2	25	25	32.00	30.00	40.00	-0.26

Histogram of total scores

For selected cut-score: Blue part of histogram shows respondents with total score above the cut-score, while red part shows respondents with total score below the cut-score and not put of histogram if it were to prevent below the cut-score

Number of respondents

Total score

© 2014 Pearson Analytics

Page 2

The figure consists of six subplots arranged in a 3x2 grid, each showing the performance of three optimization methods (Newton, BFGS, and CG) on three different datasets (I, II, and III). The x-axis for all plots is 'Number of iterations' ranging from 0 to 100, and the y-axis is 'Relative error' ranging from 0 to 1.0.

- Top Row:**
 - Dataset I:** Shows a steady increase in relative error for all methods as iterations increase. Newton (red squares) reaches the highest error (~0.85 at 100 iterations), while CG (green circles) reaches the lowest (~0.25 at 100 iterations).
 - Dataset II:** Shows a similar trend to Dataset I, with CG (green circles) maintaining the lowest relative error (~0.25 at 100 iterations) and Newton (red squares) reaching the highest (~0.85 at 100 iterations).
- Middle Row:**
 - Dataset I:** Shows a steady increase in relative error. CG (green circles) reaches the lowest error (~0.25 at 100 iterations), followed by BFGS (blue triangles) (~0.45 at 100 iterations), and Newton (red squares) reaches the highest (~0.85 at 100 iterations).
 - Dataset II:** Shows a steady increase in relative error. CG (green circles) reaches the lowest error (~0.25 at 100 iterations), followed by BFGS (blue triangles) (~0.45 at 100 iterations), and Newton (red squares) reaches the highest (~0.85 at 100 iterations).
- Bottom Row:**
 - Dataset I:** Shows a steady increase in relative error. CG (green circles) reaches the lowest error (~0.25 at 100 iterations), followed by BFGS (blue triangles) (~0.45 at 100 iterations), and Newton (red squares) reaches the highest (~0.85 at 100 iterations).
 - Dataset II:** Shows a steady increase in relative error. CG (green circles) reaches the lowest error (~0.25 at 100 iterations), followed by BFGS (blue triangles) (~0.45 at 100 iterations), and Newton (red squares) reaches the highest (~0.85 at 100 iterations).

ACI-estimated

Score plot

Score plots display the observations associated with an component as a function of descending order of the size of the component in ACI.

Component number

Predictive validity

Not for association between total score and children's behaviour. Appearance to us: The null hypothesis is that the total score is not associated with children's behaviour. If we accept this null hypothesis, then we expect that most hypotheses about scores and children's behaviour are predicting correctly.

Score

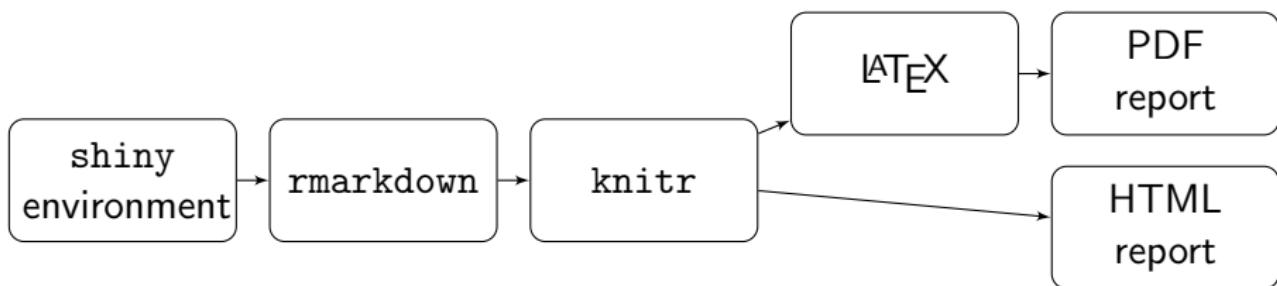
Total score

The figure consists of six scatter plots arranged in a 3x2 grid. Each plot has 'Searched hit items' on the x-axis (ranging from 0 to 10) and 'Retrieval hits' on the y-axis (ranging from 0 to 10). The plots are titled as follows:

- Top Row:**
 - Left: Discriminator analysis
 - Right: Multimodel plot for Item 1
- Middle Row:**
 - Left: Discriminator plot for Item 2
 - Right: Multimodel plot for Item 2
- Bottom Row:**
 - Left: Discriminator plot for Item 3
 - Right: Multimodel plot for Item 3

In all plots, the data points are color-coded by method: red for 'Discriminatory analysis', green for 'Multimodel', and blue for 'Retrieval hits'. The plots show varying degrees of correlation between the two variables, with some methods showing a strong positive trend while others show a more scattered or negative relationship.

Report generation workflow



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

Conclusion and Discussion

ShinyItemAnalysis is an R package and online 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
- ROUTINE VALIDATION OF EDUCATIONAL TESTS
 - 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 under grant GJ15-15856Y

References

- Martinková, Drabinová, Leder & Houdek (2017). ShinyItemAnalysis: Test and Item Analysis with Shiny.
<https://shiny.cs.cas.cz/ShinyItemAnalysis/>
<https://CRAN.R-project.org/package=ShinyItemAnalysis>
- Martinková, Drabinová & Houdek (2017). ShinyItemAnalysis: Analýza přijímacích a jiných znalostních či psychologických testů. TESTFÓRUM, č.9, str. 16-35. [doi 10.5817/TF2017-9-129](https://doi.org/10.5817/TF2017-9-129)
- 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](https://doi.org/10.1187/cbe.16-10-0305)
- Martinková, Drabinová, 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. [doi 10.1187/cbe.16-10-0307](https://doi.org/10.1187/cbe.16-10-0307)
- Drabinová & Martinková (2017). Detection of DIF with Non-Linear Regression: Non-IRT Approach Accounting for Guessing. Journal of Educational Measurement, 54(4), pp. 498-517. [doi 10.1111/jedm.12158](https://doi.org/10.1111/jedm.12158)