Basic Item Response Theory (IRT) with R source codes:

Rasch Models and Maximum Likelihood (ML) Parameter Estimation

Formosa Robotics Group

Oct. 10, 2021

Table of Binary Observations $X_{m,n}$

$(1 \le m \le M,\ 1 \le n \le N)$ *for M individuals and N items*

$X_{m,n}$ is the **Bernoulli (binary, dichotomous)** random variable denoting the response of individual m with respect to item n:

$X_{m,n}$=1 if the response is *correct*,
=0 if the response is **incorrect**.

$X_{2,3}$=1

→item n ↓individual m	1	2	3	4	5
1	0	0	1	1	0
2	1	0	1	0	0
3	1	1	1	0	1

$X_{1,5}$=0

M = 3 students (individuals), each answering $N = 5$ problems(items)

Problem (P1): How to build a simple model for the Probability of Correct Response $p_{m,n}(\theta_m, \lambda_n)$?

$p_{m,n}(\theta_m, \lambda_n) = P\left(X_{m,n} = 1 \middle| \Theta_m = \theta_m, \Lambda_n = \lambda_n\right)$ is the probability for individual m to make a **correct** response to item n, given $\Theta_m = \theta_m$ and $\Lambda_n = \lambda_n$, $1 \le m \le M$, $1 \le n \le N$

- Θ_m is the random variable denoting the ***ability*** level of individual m
- Λ_n is the random variable for the **parameter** of **item** n

item difficulty

$\Lambda_n = \lambda_n$

$n = 3$

individual ability

$\Theta_m = \theta_m$

$m = 2$

$\rightarrow$item n $\downarrow$individual m	1	2	3	4	5
1	$p_{1,1}(\theta_1, \lambda_1)$	$p_{1,2}\ (\theta_1, \lambda_2)$	$p_{1,3}(\theta_1, \lambda_3)$	$p_{1,4}(\theta_1, \lambda_4)$	$p_{1,5}(\theta_1, \lambda_5)$
2	$p_{2,1}(\theta_2, \lambda_1)$	$p_{2,2}(\theta_2, \lambda_2)$	$p_{2,3}(\theta_2, \lambda_3)$	$p_{2,4}(\theta_2, \lambda_4)$	$p_{2,5}(\theta_2, \lambda_5)$
3	$p_{3,1}(\theta_3, \lambda_1)$	$p_{3,2}(\theta_3, \lambda_2)$	$p_{3,3}(\theta_3, \lambda_3)$	$p_{3,4}(\theta_3, \lambda_4)$	$p_{3,5}(\theta_3, \lambda_5)$

$\Theta_m = \theta_m$: ability level of the mth student, $\Lambda_n = \lambda_n$: difficulty level of the nth problem

Answer to (P1): choose prototype of $p_{m,n}(\theta_m, \lambda_n)$ to be

$$\boxed{p(\theta, b) = \frac{1}{1+e^{-D(\theta-b)}}, \quad D \simeq 1.7}$$

(**Rasch Model**)

- Rasch model is a **one**-**p**arameter **l**ogistic (1PL) model with individual ability level θ and parameter b
- $p(\theta, b)$ is the probability for an individual to make a **correct** response to an item, given ability $\Theta = \theta$ and difficulty $\Lambda = b$
- $\boxed{q(\theta, b) = 1 - p(\theta, b) = \frac{1}{1+e^{D(\theta-b)}}}$ is the probability for an individual to make an **incorrect** response to an item, given $\Theta = \theta$ and $\Lambda = b$
- $0 \le p(\theta, b) \le 1, \quad 0 \le q(\theta, b) \le 1$: valid probability distributions
- e is Euler's number, a mathematical constant, $e \simeq 2.71828$, such that

$$e^t = \frac{d}{dt} e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} \qquad ln(t) = \log_e t, \qquad ln(e^t) = t$$

Rasch Model (1PL)

$$p(\theta,b)=\frac{1}{1+e^{-D(\theta-b)}}$$

$$0\le\ p(\theta,b)\ \le 1$$

monotonically increasing function of θ, for fixed *b*

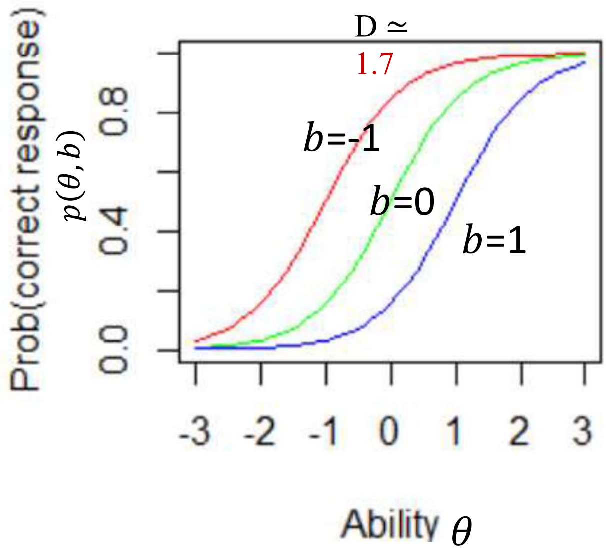
R source code for generating ICC

```
icc1 <- function(b,D,color) {
  theta <- seq(-3,3,.1)
  P1 <- 1/(1+exp(-D*(theta-b)))
  plot(theta, P1, type="l", xlim=c(-3,3), ylim=c(0,1),
       xlab="Ability", ylab="Prob(correct response)",
       col = color)
}

icc2 <- function(b,D,color) {
  theta <- seq(-3,3,.1)
  P1 <- 1/(1+exp(-D*(theta-b)))
  lines(theta, P1, col = color)
}

D <- 1.7
icc1(b=-1, D, "red")
icc2(b=0, D, "green")
icc2(b=1, D, "blue")
```

Item Characteristic Curve (ICC)



Rasch model greatly simplifies computations (1)

$$p=p(\theta,b)$$

$$q=q(\theta,b)=1-p$$

$$\frac{\partial p}{\partial\theta}=\frac{\partial}{\partial\theta}\ \frac{1}{1+e^{-D(\theta-b)}}=\frac{D\ e^{-D(\theta-b)}}{\left[1+e^{-D(\theta-b)}\right]^2}=Dpq$$

$$\frac{\partial q}{\partial b}=\frac{\partial}{\partial b}\ \frac{1}{1+e^{D(\theta-b)}}=\frac{D\ e^{D(\theta-b)}}{\left[1+e^{D(\theta-b)}\right]^2}=Dpq$$

$$\frac{\partial p}{\partial b}=\frac{\partial}{\partial b}\ \frac{1}{1+e^{-D(\theta-b)}}=\frac{-D\ e^{-D(\theta-b)}}{\left[1+e^{-D(\theta-b)}\right]^2}=-Dpq$$

$$\frac{\partial q}{\partial\theta}=\frac{\partial}{\partial\theta}\ \frac{1}{1+e^{D(\theta-b)}}=\frac{-D\ e^{D(\theta-b)}}{\left[1+e^{D(\theta-b)}\right]^2}=-Dpq$$

$$\frac{\partial p}{\partial\theta}=\frac{\partial q}{\partial b}=Dpq$$

$$\frac{\partial p}{\partial b}=\frac{\partial q}{\partial\theta}=-Dpq$$

$$\frac{d}{dt}\left(\frac{f}{g}\right)=\frac{f'g-g'f}{g^2}$$

Rasch model greatly simplifies computations (2)

$$\frac{\partial}{\partial\theta} ln(p) = \frac{\partial}{\partial\theta} ln\left(\frac{1}{1+e^{-D(\theta-b)}}\right) = -\frac{\partial}{\partial\theta} ln\left(1+e^{-D(\theta-b)}\right)$$
$$= \frac{-1}{1+e^{-D(\theta-b)}} \frac{\partial}{\partial\theta}\left(1+e^{-D(\theta-b)}\right) = \frac{D\ e^{-D(\theta-b)}}{1+e^{-D(\theta-b)}} = Dq = \frac{\partial}{\partial b} ln(q)$$

$$\frac{\partial}{\partial\theta} ln(q) = \frac{\partial}{\partial\theta} ln\left(\frac{1}{1+e^{D(\theta-b)}}\right) = -\frac{\partial}{\partial\theta} ln\left(1+e^{D(\theta-b)}\right)$$
$$= \frac{-1}{1+e^{D(\theta-b)}} \frac{\partial}{\partial\theta}\left(1+e^{D(\theta-b)}\right) = \frac{-D\ e^{D(\theta-b)}}{1+e^{D(\theta-b)}} = -Dp = \frac{\partial}{\partial b} ln(p)$$

$$\boxed{\frac{d}{dt} ln(g) = \frac{g'}{g}}$$

Problem (P2) How to estimate 1: Θ_m and 2: Λ_n?

Hint: Use **likelihood** function to find a ***good*** estimate $\hat{\theta}_m$ and $\hat{\lambda}_n$ for the true values of the (hidden) variables Θ_m and Λ_n.

→item n ↓individual m	1	2	3	4	5
1	$p_{1,1}(\theta_1,\lambda_1)$	$p_{1,2}\ (\theta_1,\lambda_2)$	$p_{1,3}(\theta_1,\lambda_3)$	$p_{1,4}(\theta_1,\lambda_4)$	$p_{1,5}(\theta_1,\lambda_5)$
2	$p_{2,1}(\theta_2,\lambda_1)$	$p_{2,2}(\theta_2,\lambda_2)$	$p_{2,3}(\theta_2,\lambda_3)$	$p_{2,4}(\theta_2,\lambda_4)$	$p_{2,5}(\theta_2,\lambda_5)$
3	$p_{3,1}(\theta_3,\lambda_1)$	$p_{3,2}(\theta_3,\lambda_2)$	$p_{3,3}(\theta_3,\lambda_3)$	$p_{3,4}(\theta_3,\lambda_4)$	$p_{3,5}(\theta_3,\lambda_5)$

$\hat{\theta}_2 \simeq \theta_2$

$\hat{\lambda}_3 \simeq \lambda_3$

The first step to solve (P2) is to define and find

$$r_{m,n}(x) = P\big(X_{m,n} = x \big| \Theta_m = \theta_m, \Lambda_n = \lambda_n\big), \quad x = 0, 1$$

- For simplicity, abbreviate $P\big(X_{m,n} = 1 \big| \Theta_m = \theta_m, \Lambda_n = \lambda_n\big)$ as $p_{m,n}$ and $q_{m,n} = P\big(X_{m,n} = 0 \big| \Theta_m = \theta_m, \Lambda_n = \lambda_n\big)$ =1- $p_{m,n}$
- We have $r_{m,n}(x) = [p_{m,n}]^x \, [q_{m,n}]^{1-x}$
  $= p_{m,n}$ if x = 1 (correct response)
  $q_{m,n}$ if x = 0 (incorrect response)

Note: $X_{m,n}$ is a binary **random variable**, and $x_{m,n}$ (or x) denotes its **value:** 0 or 1.

Answer to (P2.1)

Likelihood function of θ_m:

$$L_1(\boldsymbol{x}_{m,}|\theta_m, \boldsymbol{\lambda}) = \prod_{n=1}^{N} r_{m,n}(x_{m,n})$$

where $\boldsymbol{x}_{m,} = (x_{m,1}, x_{m,2}, \dots, x_{m,N})$; $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$

Maximum likelihood (ML) estimate of θ_m, given $\boldsymbol{\lambda}$:

$$\hat{\theta}_{m|\lambda} = \underset{\theta_m}{\arg\max} \; ln \; L_1(\boldsymbol{x}_{m,}|\theta_m, \boldsymbol{\lambda})$$

ln: monotonically increasing function

	Item 1	Item 2	…	Item N	
individual m	$r_{m,1}(x_{m,1})$	$r_{m,2}(x_{m,2})$	…	$r_{m,N}(x_{m,N})$	$\hat{\theta}_{m\|\lambda}$

Answer to (P2.2)

Likelihood function of λ_n:

$$L_2(\boldsymbol{x}_{,n}|\lambda_n,\boldsymbol{\theta}) = \prod_{m=1}^{M} r_{m,n}(x_{m,n})$$

where $\boldsymbol{x}_{,n} = (x_{1,n}, x_{2,n}, \dots, x_{M,n})$; $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$

Item n

$r_{1,n}(x_{1,n})$

.

.

$r_{M,n}(x_{M,n})$

$\hat{\lambda}_{n|\boldsymbol{\theta}}$

Maximum likelihood (ML) estimate of λ_n, given $\boldsymbol{\theta}$:

$$\hat{\lambda}_{n|\boldsymbol{\theta}} = \underset{\lambda_n}{\arg\max}\ ln\ L_2(\boldsymbol{x}_{,n}|\lambda_n,\boldsymbol{\theta})$$

Since $ln(t) = \log_e t$ is a monotonically increasing function of t,

$$\underset{\lambda_n}{\arg\max}\ L_2(\boldsymbol{x}_{,n}|\lambda_n,\boldsymbol{\theta}) = \underset{\lambda_n}{\arg\max}\ ln\ L_2(\boldsymbol{x}_{,n}|\lambda_n,\boldsymbol{\theta})$$

ML Estimates of Individual Ability $\hat{\theta}_{m|\boldsymbol{\lambda}}$ (1)

$$ln\ L_1(\boldsymbol{x}_{m,}|\theta_m,\boldsymbol{\lambda}) = \sum_{n=1}^{N} ln\ r_{m,n}(x_{m,n})$$

$$\frac{\partial}{\partial\theta_m} ln\ r_{m,n}(x_{m,n}) = \frac{\partial}{\partial\theta_m}\left[x_{m,n} ln(p_{m,n}) + (1-x_{m,n}) ln(q_{m,n})\right]$$

$$\frac{\partial}{\partial\theta_m} ln(p_{m,n}) = \frac{\partial}{\partial\theta_m} ln\left(\frac{1}{1+e^{-D(\theta_m-b_n)}}\right) = \frac{D\ e^{-D(\theta_m-b_n)}}{1+e^{-D(\theta_m-b_n)}} = Dq_{m,n} = D(1-p_{m,n})$$

From Rasch Model

$$\frac{\partial}{\partial\theta_m} ln(q_{m,n}) = \frac{\partial}{\partial\theta_m} ln\left(\frac{1}{1+e^{D(\theta_m-b_n)}}\right) = \frac{-D\ e^{D(\theta_m-b_n)}}{1+e^{D(\theta_m-b_n)}} = -Dp_{m,n}$$

$$\frac{\partial}{\partial\theta_m} ln\ r_{m,n}(x_{m,n}) = Dx_{m,n}(1-p_{m,n}) - D(1-x_{m,n})p_{m,n} = D(x_{m,n} - p_{m,n})$$

$$\frac{\partial}{\partial\theta_m} ln\ L_1(\boldsymbol{x}_{m,}|\theta_m,\boldsymbol{\lambda}) = D\sum_{n=1}^{N}(x_{m,n} - p_{m,n})$$

$$p_{m,n} = \frac{1}{1+e^{-D(\theta_m-b_n)}}$$

From Rasch Model

ML Estimates of Individual Ability $\hat{\theta}_{m|\boldsymbol{\lambda}}$ (2)

Newton-Raphson Method for root-finding:

The root x^* satisfying $f(t^*) = 0$ can be found by iterating

$\boxed{t_{k+1} = t_k - \frac{f(t_k)}{f'(t_k)},}$ $k = 0,1,2,3 \ldots$ until convergence.

usually, very fast convergence

Maximum likelihood solution $\hat{\theta}_{m|\boldsymbol{\lambda}} = \arg\max_{\theta_m} \; ln\; L_1(\boldsymbol{x}_{m,}|\theta_m, \boldsymbol{\lambda})$

occurs when $\frac{\partial}{\partial \theta_m} ln\, L_1(\boldsymbol{x}_{m,}|\theta_m, \boldsymbol{\lambda}) = \mathbf{0}$

Newton-Raphson iteration

$t = \theta_m$

for the ML ability estimate of θ_m

$m = 1, \ldots, M$

$$\boxed{\begin{aligned} f(\theta_m) &= \frac{\partial}{\partial \theta_m} ln\, L_1(\boldsymbol{x}_{m,}|\theta_m, \boldsymbol{\lambda}) = D \sum_{n=1}^{N} (x_{m,n} - p_{m,n}) \\ f'(\theta_m) &= \frac{\partial^2}{\partial \theta_m^2} ln\, L_1(\boldsymbol{x}_{m,}|\theta_m, \boldsymbol{\lambda}) = -D^2 \sum_{n=1}^{N} p_{m,n}\; q_{m,n} \end{aligned}}$$

ML Estimates of Item Parameters $\hat{\lambda}_{n|\boldsymbol{\theta}}$ (1)

Maximum likelihood (ML) estimate of λ_n, given $\boldsymbol{\theta}$:

$$\hat{\lambda}_{n|\boldsymbol{\theta}} = \arg\max_{\lambda_n} \; ln\; L_2(\boldsymbol{x}_{,n}|\lambda_n, \boldsymbol{\theta})$$

$$\frac{\partial}{\partial \lambda_n} ln\, L_2(\boldsymbol{x}_{,n}|\lambda_n, \boldsymbol{\theta}) = \frac{\partial}{\partial \lambda_n} \sum_{m=1}^{M} [x_{m,n} ln(p_{m,n}) + (1 - x_{m,n}) ln(q_{m,n})]$$

Rasch model: $\lambda_n = b_n$, $p_{m,n} = \frac{1}{1+e^{-D(\theta_m - b_n)}}$

$$\frac{\partial}{\partial b_n} ln(\, p_{m,n}) = \frac{\partial}{\partial b_n} ln\left(\frac{1}{1 + e^{-D(\theta_m - b_n)}}\right) = \frac{-D\; e^{-D(\theta_m - b_n)}}{1 + e^{-D(\theta_m - b_n)}} = -D q_{m,n} = D(p_{m,n} - 1)$$

$$\frac{\partial}{\partial b_n} ln(\, q_{m,n}) = \frac{\partial}{\partial b_n} ln\left(\frac{1}{1 + e^{D(\theta_m - b_n)}}\right) = \frac{D\; e^{D(\theta_m - b_n)}}{1 + e^{D(\theta_m - b_n)}} = D p_{m,n}$$

$$\Rightarrow \frac{\partial}{\partial \lambda_n} ln\, L_2(\boldsymbol{x}_{,n}|\lambda_n, \boldsymbol{\theta}) = \sum_{m=1}^{M} D[x_{m,n}(p_{m,n} - 1) + (1 - x_{m,n}) p_{m,n}] = D \sum_{m=1}^{M} (p_{m,n} - x_{m,n})$$

ML Estimates of Item Parameters $\hat{\lambda}_{n|\boldsymbol{\theta}}$ (2)

Apply Newton-Raphson iteration with

$$g(\lambda_n) = \frac{\partial}{\partial \lambda_n} ln\, L_2(\boldsymbol{x}_{,n}|\lambda_n, \boldsymbol{\theta}) = D \sum_{m=1}^{M} (p_{m,n} - x_{m,n})$$

$$g'(\lambda_n) = \frac{\partial^2}{\partial \lambda_n^2} ln\, L_2(\boldsymbol{x}_{,n}|\lambda_n, \boldsymbol{\theta}) = D^2 \sum_{m=1}^{M} p_{m,n}\, q_{m,n}$$

$$\lambda_{n,k+1} = \lambda_{n,k} - \frac{g(\lambda_{n,k})}{g'(\lambda_{n,k})}, \quad k = 0,1,2,3 \ldots$$

to find $\hat{\lambda}_{n|\boldsymbol{\theta}} = \lambda_{n,\infty} = \underset{\lambda_n}{\arg\max}\ \ ln\ L_2(\boldsymbol{x}_{,n}|\lambda_n, \boldsymbol{\theta})$ $\qquad n = 1,\ldots,N$

for the maximum likelihood (ML) parameter estimate of b_n (Rasch model, 1PL)

R source code for ML estimation of individual **ability**

```r
ability <- function(X, b) {
  D <- 1.7
  D2 <- D * D
  N <- length(b)
  MaxCnt <- 50
  theta <- 1.
  Tiny <- 0.001
  for (i in 1:MaxCnt) {
    numerator   <- 0
    denominator <- 0
    for (n in 1:N) {
      p <- 1/(1+exp(-D*(theta-b[n])))
      numerator   <- numerator   + D * (X[n] - p)
      denominator <- denominator - D2 * p * (1-p)
    }
    diff <- numerator / denominator
    theta <- theta - diff
    cat(paste("[", i, "] ability =", theta, "  diff=", diff, "\n")); flush.console()
    if (abs(diff) < Tiny) break
  }
}


b <- c(0.5, -0.5, 0.7, 0.6, -1, 0.9, 0, -0.5, 0) # difficulty level
X <- c(1, 0, 1, 1, 0, 1, 0, 1, 0) # individual responses, #individuals M = 1
ability(X,b)
```

$$p_{m,n} = \frac{1}{1 + e^{-D(\theta_m - b_n)}}$$

$$f(\theta_m) = D \sum_{n=1}^{N} (x_{m,n} - p_{m,n})$$

$$f'(\theta_m) = -D^2 \sum_{n=1}^{N} p_{m,n}\ q_{m,n}$$

$$t_{k+1} = t_k - \frac{f(t_k)}{f'(t_k)}$$

ability = 0.255586, convergence in 3 steps

Brief Review for Clarity (1)

A total of (M+N) hidden parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M)$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_N)$ are to be estimated and extracted form the M x N input data matrix $[x_{m,n}]$.

Column-n: item difficulty $\lambda_n = b_n$

Row-m: ability θ_m

→item n ↓individual m	λ_1	…	λ_n	…	λ_N
θ_1	$x_{1,1}$	…	$x_{1,n}$	…	$x_{1,N}$
. .	. .	… …	. .	… …	. .
θ_m	$x_{m,1}$	…	$x_{m,n}$	…	$x_{m,N}$
. .	. .	… …	. .	… …	. .
θ_M	$x_{M,1}$	…	$x_{M,n}$	…	$x_{M,N}$

Input data $x_{m,n} = 0,1$

Brief Review for Clarity (2)

Maximum Likelihood (ML) estimates of individual ability levels and item difficulty levels, respectively, are calculated efficiently by aggregating horizontally and vertically along the matrix.

→item n ↓individual m	λ_1	…	λ_n	…	λ_N	
θ_1	$r_{1,1}(x_{1,1})$	…	$r_{1,n}(x_{1,n})$	…	$r_{1,N}(x_{1,N})$	
. .	. .	… …	. .	… …	. .	
θ_m	$r_{m,1}(x_{m,1})$	…	$r_{m,n}(x_{m,n})$	…	$r_{m,N}(x_{m,N})$	$\hat{\theta}_{m\|\boldsymbol{\lambda}}$
. .	. .	… …	. .	… …	. .	
θ_M	$r_{M,1}(x_{M,1})$	…	$r_{M,n}(x_{M,n})$	…	$r_{M,N}(x_{M,N})$	
			$\hat{\lambda}_{n\|\boldsymbol{\theta}}$			

Rasch model: $p_{m,n} = \frac{1}{1+e^{-D(\theta_m - b_n)}} = 1 - q_{m,n}$
depends only on θ_m and b_n , so does $r_{m,n}(x)$, given x.

$r_{m,n}(x) = [p_{m,n}]^x \, [q_{m,n}]^{1-x}$,
$x_{m,n} = x \in \{0,1\}$

Conclusions

- We formulate the maximum likelihood (ML) estimates of individual ability and item parameters for IRT using Rasch model (1PL).
- R programs are developed to verify the validity of the Rasch model and ML estimate of individual ability θ_m.
- Newton-Raphson iteration method results in very fast convergence in computing the ML estimate of θ_m, usually < 10 steps.
- We reformulate the mathematical derivation of formulas in [1] to provide a succinct representation of basic IRT theory.
- We modify the simple R programs in [2] to yield our R source codes.
- Our analysis paves the way to Bayesian and Neural Networks.

References

[1] F.B. Baker & S.-H. Kim, *Item Response Theory: Parameter Estimation Techniques*, CRC Press, 2004 .

[2] F.B. Baker & S.-H. Kim, *The Basics of Item Response Theory using R*, Springer, 2017.

[3] P. Mair, *Modern Psychometrics with R*, Springer, 2018.

[4] N. Matloff, *The Art of R Programming*, No Starch Press, 2011.

[5]H. Wickham, *ggplot2, Elegant Graphics for Data Analysis*, Springer, 2016.

[6] J. Rust, M. Kosinski & D. Stillwell, *Modern Psychometrics*, Routledge, 2021.

[7] L. P. Price, *Psychometric Method, Theory into Practice*, Guilford Press, 2017.

[8] F.M. Lord, *Applications of Item Response Theory to Practical Testing Problems*, Routledge, 1980.

[9] J.-P. Fox, *Bayesian Item Response Theory*, Springer, 2010.

[10] C. D. Desjardins & O. Bulut, *Handbook of Educational Measurement and Psychometrics using R*, CRC Press, 2018.

[11] S. P. Reise & D. A. Revicki, *Handbook of Item Response Theory Modeling*, Routledge, 2015.