TAM: An R Package for Item Response Modelling

Thomas Kiefer[1], Alexander Robitzsch[1], Margaret Wu[2]

[1]Federal Institute for Educational Research, Innovation and Development of the Austrian School System (BIFIE), Salzburg, Austria,

[2]Educational Measurement Solutions, Victoria, Australia

Bundesinstitut bifie
Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens

EMS
EDUCATIONAL MEASUREMENT SOLUTIONS

useR! 2015 Aalborg

July 2, 2015

Co-Authors

Alexander Robitzsch.
Measurement Statistician at the Federal Institute for Educational Research, Innovation and Development of the Austrian School System (BIFIE), Salzburg, Austria.
`https://www.bifie.at/user/robitzsch-alexander`

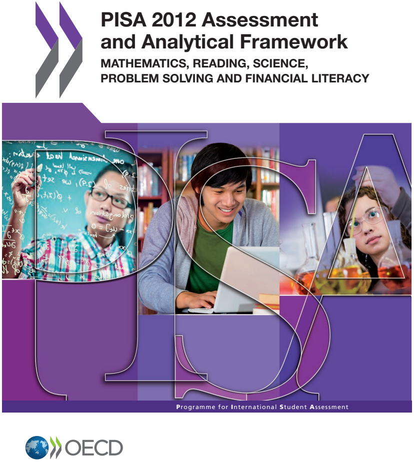
Professor Margaret Wu.
Long-term Statistician for the OECD, Scientist at the Work-based Education Research Centre of the Victoria University Melbourne, Australia and Co-Founder of Educational Measurement Solutions.
`http://www.edmeasurement.com.au`

Overview

Introduction

Area of Application: Psychometrics

- Psychometrics is the (statistical) field of measuring psychological concepts.
- A field of application is the educational large-scale assessment (LSA).
- The psychological concept in an LSA is defined in a competency construct.
- The competence construct is a – often very broad – definition of the students trait in question.



Area of Application: Psychometrics

- Psychometrics is the (statistical) field of measuring psychological concepts.
- A field of application is the educational large-scale assessment (LSA).
- The psychological concept in an LSA is defined in a competency construct.
- The competence construct is a – often very broad – definition of the students trait in question.

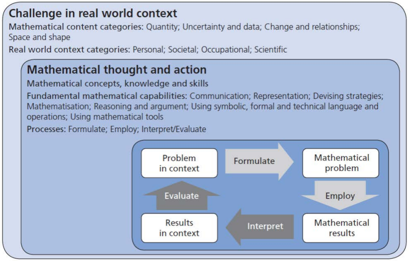


Figure: Model for the Competency construct of mathematical literacy in PISA

(OECD, 2013)

Area of Application: Educational Assessment

- Traits are measured using items.
- Items apply to a specific domain within the competence construct.
- Items are scored dichotomously (right / wrong) or polytomously (partial credit).
- Items are either closed response or constructed response format.
- A measurement is the students score to that or an equivalent item of the domain.

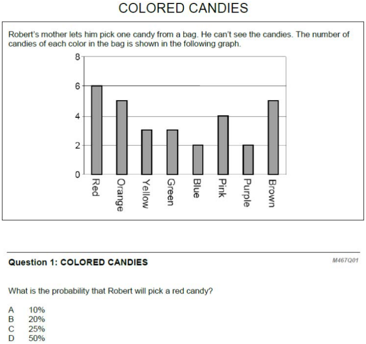
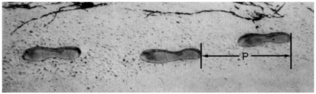


Figure: A typical item used in PISA

(`https://nces.ed.gov/surveys/pisa/releaseditems.asp`, June '15)

- Traits are measured using items.
- Items apply to a specific domain within the competence construct.
- Items are scored dichotomously (right / wrong) or polytomously (partial credit).
- Items are either closed response or constructed response format.
- A measurement is the students score to that or an equivalent item of the domain.

WALKING



The picture shows the footprints of a man walking. The pacelength P is the distance between the rears of two consecutive footprints.

For men, the formula, $\frac{n}{P} = 140$, gives an approximate relationship between n and P where,

n = number of steps per minute, and

P = pacelength in meters

Question 1: WALKING *M124Q01*

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

Figure: Another typical item used in PISA

(https://nces.ed.gov/surveys/pisa/releaseditems.asp, June '15)

Measurement

- To measure a domain sufficiently precise a large amount of items is used.
- Individual students are presented with a reasonably small representative sample (a booklet) of all possible items.
- Statistical Inference is obtained using Item Response Theory (IRT).

```
       idstud female migra M192Q01 M406Q01 M406Q02 M423Q01 M496Q01 M496Q02
1 90001500281      0     0       1       1       1       0       1       1
2 90001500290      0     0       0       1       0       1       0       0
3 90001500292      0     0       1       1       0       0       0       0
4 90001500294      0     0       1       0       0       1       0       1
5 90001500295      0     0       1       1       1       1       0       0
6 90001500297      0     0       1       0       0       1       0       1
  M564Q01 M564Q02 M571Q01 M603Q01 M603Q02
1       1       1       1       1       1
2       1       0       0       0       0
3       1       1       0       0       0
4       1       1       0       0       0
5       1       1       1       1       1
6       1       0       0       0       0
```

Measurement

- To measure a domain sufficiently precise a large amount of items is used.
- Individual students are presented with a reasonably small representative sample (a booklet) of all possible items.
- Statistical Inference is obtained using Item Response Theory (IRT).

```
    IDSTUD IDBOOK FEMALE R11F01M R11F02M R11F03M R21E01M R21E02M R21E03M
80   40105      1      1       1       0       0       1       1       1
194  80120      1      1       1       1       1       1       0       1
40   30101      2      0      NA      NA      NA       1       1       1
81   40106      2      0      NA      NA      NA       1       1       0
125  60204      3      1      NA      NA      NA      NA      NA      NA
235 100210      3      0      NA      NA      NA      NA      NA      NA
    R21Y08M R21Y09C R21Y10C R31M08M R31M09C R31M10C
80       NA      NA      NA      NA      NA      NA
194      NA      NA      NA      NA      NA      NA
40        0       2       0      NA      NA      NA
81        1       2       0      NA      NA      NA
125       1       2       1       1       1       1
235       1       2       0       1       1       1
```

Item Response Theory

IRT Models

IRT models are generalized nonlinear mixed effects models:

- the score $Y_{pi} \in \{0, 1\}$ of a student p to an item i is the dependent variable,
- given a randomly sampled student's trait, e.g. $\theta_p \sim N\left(\mu, \sigma^2\right)$, the responses are assumend to be independent Bernoulli distributed,
- given θ_p, the predictor $\eta_{pi} = \mathrm{logit}\left(P\left(Y_{pi} = 1\right)\right)$ is a linear combination of item characteristics

$$\eta_{pi} = \sum_{k=0}^{K} b_k X_{ik} + \theta_p + \varepsilon_{pi},$$

- let $X_{ik} = -1$, if $i = k$, and $X_{ik} = 0$, otherwise - thus obtain the Rasch model

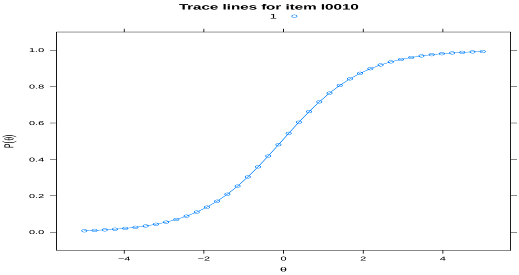
$$P\left(Y_{pi} = 1 \mid \theta_p\right) = \frac{\exp\left(\theta_p - b_i\right)}{1 + \exp\left(\theta_p - b_i\right)};$$

(De Boeck & Wilson, 2004; Lord & Novick, 1968)

Extended IRT Models: Item side

IRT models are extended towards different aspects:

- With respect to discriminatory power and guessing ratio of an item
- With respect to polytomous scores



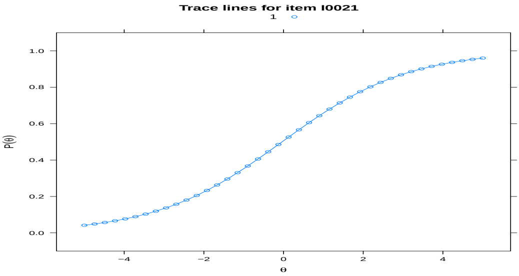
(Andersen, 1977; Birnbaum, 1968; Muraki, 1993; Rasch, 1960)

IRT models are extended towards different aspects:

- With respect to discriminatory power and guessing ratio of an item

$$P\left(Y_{pi}=1 \mid \theta_p\right)=\frac{\exp\left(a_i\left(\theta_p-b_i\right)\right)}{1+\exp\left(a_i\left(\theta_p-b_i\right)\right)}$$

- With respect to polytomous scores



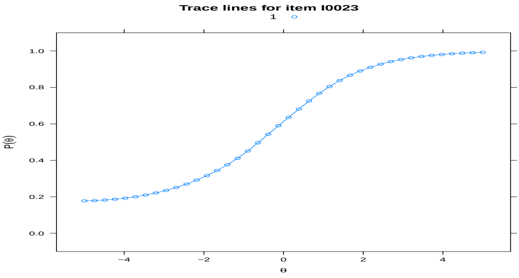
(Andersen, 1977; Birnbaum, 1968; Muraki, 1993; Rasch, 1960)

IRT models are extended towards different aspects:

- With respect to discriminatory power and guessing ratio of an item

$$P\left(Y_{pi} = 1 \mid \theta_p\right) = c_i + \left(1 - c_i\right) \frac{\exp\left(a_i\left(\theta_p - b_i\right)\right)}{1 + \exp\left(a_i\left(\theta_p - b_i\right)\right)},$$

- With respect to polytomous scores



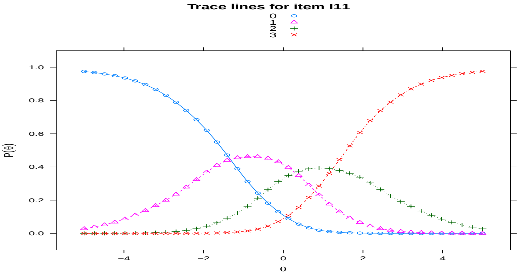
(Andersen, 1977; Birnbaum, 1968; Muraki, 1993; Rasch, 1960)

Extended IRT Models: Item side

IRT models are extended towards different aspects:

- With respect to discriminatory power and guessing ratio of an item
- With respect to polytomous scores

$$P\left(Y_{pi}=k \mid \theta_p\right)=\frac{\exp\left(a_{ik}\theta_p-b_{ik}\right)}{\sum_{k=0}^{K}\exp\left(a_{ik}\theta_p-b_{ik}\right)}.$$



(Andersen, 1977; Birnbaum, 1968; Muraki, 1993; Rasch, 1960)

Extended IRT Models: Person side

IRT models are extended towards different aspects:

- With respect to known student characteristics constituting the population (e.g., sex, migration status)
- With respect to construct dimensionality
- With respect to discrete skill classes (continuous distributions can be easily approximated by discrete ones)

```
....................................
Regression Coefficients
         V1
[1,] 0.704

Variance:
      [,1]
[1,] 1.613
```

(Adams, Wilson, & Wang, 1997; Xu & von Davier, 2007; ?)

Extended IRT Models: Person side

IRT models are extended towards different aspects:

- With respect to known student characteristics constituting the population (e.g., sex, migration status)

$$\theta_p \sim N\left(\mathbf{Z}\boldsymbol{\beta}, \sigma^2\right),$$

- With respect to construct dimensionality
- With respect to discrete skill classes (continuous distributions can be easily approximated by discrete ones)

```
....................................
Regression Coefficients
                V1
Intercept  1.0263
female     0.3342
migra     -0.7008

Variance:
      [,1]
[1,] 1.988
```

Extended IRT Models: Person side

IRT models are extended towards different aspects:

- With respect to known student characteristics constituting the population (e.g., sex, migration status)
- With respect to construct dimensionality

$$P\left(Y_{pi}=1 \mid \theta_p\right)=\frac{\exp(\sum_d a_{id}\theta_{pd}-b_i)}{1+\exp(\sum_d a_{id}\theta_{pd}-b_i)}, \quad \theta_p \sim N^d\left(\boldsymbol{\mu}, \Sigma\right)$$

- With respect to discrete skill classes (continuous distributions can be easily approximated by discrete ones)

```
.................................
Regression Coefficients
            [,1]    [,2]    [,3]
Intercept  1.4055  0.3483  1.2977
female     0.2428  0.3026  0.4821
migra     -0.6371 -0.5466 -0.7526

Variance:
      [,1]  [,2]  [,3]
[1,] 1.354 1.291 1.831
[2,] 1.291 1.327 1.812
[3,] 1.831 1.812 2.621
```

Extended IRT Models: Person side

IRT models are extended towards different aspects:

- With respect to known student characteristics constituting the population (e.g., sex, migration status)
- With respect to construct dimensionality
- With respect to discrete skill classes (continuous distributions can be easily approximated by discrete ones)

$$P\left(Y_{pi}=1 \mid \theta_{p(l)}\right)=\frac{\exp(\theta_{p(l)}-b_{i(l)})}{1+\exp(\theta_{p(l)}-b_{i(l)})}, \quad \theta_{p(l)} \in \left\{\theta_{p(1)},\ldots,\theta_{p(L)}\right\}.$$

```
....................................
Full Trait distribution
     Group1
[1,] 0.0667
[2,] 0.1730
[3,] 0.5206
[4,] 0.1730
[5,] 0.0667

SD Trait:
         [,1]
Group1 1.407
```

Why TAM?

Here's why!

- Open source solution for everyday work in an educational assessment context (such as BIFIE);
- Estimation processes at BIFIE prior to **TAM**:
  1. Data preparation in R,
  2. Call to third-party software for IRT analyses (e.g., **ConQuest**).
- **ConQuest**:
  - Absence of standard API,
  - commercial black-box software.
- R packages:
  - **mirt** recently became suitable for use in LSA; still lacks some flexibility in specifying dependencies among item parameters.
  - Other R packages (e.g., **eRm**, **ltm**, **psychotools**) lack model classes or processing speed (or both) required for population-sized context.
- **TAM** is flexible due to design matrices; yet reasonably fast.
- Bonus: gain some deeper understanding of the estimation processes.

(Adams & Wu, 2007; Chalmers, 2012; Wu, Adams, Wilson, & Haldane, 2007)

Model Syntax

IRT models can be set up using model syntax statements (based on **lavaan**).

- Relevant aspects for specifying IRT models in **TAM** group into four types.
- The Rasch model is specified by a minimaly complex input.
- Presented examples are necessarily limited; `tamaan` also allows for `MODEL PRIOR`, `DO` loops, and a lot more model classes.

```
> ## Toy example
> head(dat)


   A1 A2 A3 A4 B1 B2 B3 B4 C1 C2 C3 C4
2   1  1  1  1  1  1  1  1  1  1  1  0
22  1  1  0  0  1  0  1  1  1  0  1  0
23  1  1  0  1  1  0  1  1  1  1  1  1
41  1  1  1  1  1  1  1  1  1  1  1  1
43  1  0  0  1  0  0  1  1  1  0  1  0
63  1  1  0  0  1  0  1  1  1  1  1  1
```

Model Syntax

IRT models can be set up using model syntax statements (based on **lavaan**).

- Relevant aspects for specifying IRT models in **TAM** group into four types.
- The Rasch model is specified by a minimaly complex input.
- Presented examples are necessarily limited; `tamaan` also allows for `MODEL PRIOR`, `DO` loops, and a lot more model classes.

```
> ## Basic setup
> tammodel <- "
 ANALYSIS:

 LAVAAN MODEL:

 ITEM TYPE:

 MODEL CONSTRAINT:

 "
> ## estimate model
> # mod <- tamaan(tammodel, resp = dat)
```

Model Syntax

IRT models can be set up using model syntax statements (based on **lavaan**).

- Relevant aspects for specifying IRT models in **TAM** group into four types.
- The Rasch model is specified by a minimaly complex input.
- Presented examples are necessarily limited; `tamaan` also allows for `MODEL PRIOR`, `DO` loops, and a lot more model classes.

```
> ## Rasch model ----
> tammodel <- "
ANALYSIS:
  TYPE = TRAIT;
LAVAAN MODEL:
  F1 =~ A1__C4
  F1 ~~ F1
ITEM TYPE:
  ALL(Rasch);
    "
> # estimate model
> mod <- tamaan(tammodel, resp = dat, control = list(progress = FALSE))
> mod$variance


         V1
V1 1.190302
```

Model Syntax

IRT models can be set up using model syntax statements (based on **lavaan**).

- Relevant aspects for specifying IRT models in **TAM** group into four types.
- The Rasch model is specified by a minimaly complex input.
- Presented examples are necessarily limited; `tamaan` also allows for `MODEL PRIOR`, `DO` loops, and a lot more model classes.

```
> ## Rasch model ----
> tammodel <- "
ANALYSIS:
  TYPE = TRAIT;
LAVAAN MODEL:
  F1 =~ A1__C4
  F1 ~~ F1
ITEM TYPE:
  ALL(Rasch);
    "
> # estimate model
> mod <- tamaan(tammodel, resp = dat, control = list(progress = FALSE))
> mod$variance


         V1
V1 1.190302
```

Model Syntax II

Extending the **lavaan** syntax `tamaan` additionally implements convenient operators for specifications on the item side, such as sum over multiple entitites "__" and guessing parameters "?=".

```
> ## 3PL Model ----
> tammodel <- "
ANALYSIS:
  TYPE = TRAIT;
LAVAAN MODEL:
  F1 =~ A1__C4
  F1 ~~ 1 * F1

  A1 ?= g1
  B1 + C1 ?= gBC * g1
    "
> # estimate model
> mod <- tamaan(tammodel, resp = dat, control = list(progress = FALSE))
> round(mod$item$guess,2)

 [1] 0.68 0.00 0.00 0.00 0.24 0.00 0.00 0.00 0.24 0.00 0.00 0.00
```

Model Syntax II

Catchy definitions of model constraints for parameters are available.

```
> ## MODEL CONSTRAINTS ----
> tammodel <- "
ANALYSIS:
  TYPE = TRAIT;
LAVAAN MODEL:
  F1 =~ load1__load10 * A1__C2
  F1 ~~ 1 * F1
MODEL CONSTRAINT:
  load2 == 1.1*load1
  load3 == 0.9 * load1 + (-.1) * load0
  load8 == load0
  load9 == load0
"
> # estimate
> mod <- tamaan(tammodel , resp = dat, control = list(progress = FALSE))
> head(tamaanify(tammodel, dat)$L[, 1, ], 3)

   load1 load0 load4 load5 load6 load7 load10
A1   1.0   0.0     0     0     0     0      0
A2   1.1   0.0     0     0     0     0      0
A3   0.9  -0.1     0     0     0     0      0
```

Model Syntax II

Using the options in analysis TYPE, Latent Class Analysis (LCA) models can be specified.

```
> ## LCA Model ----
> tammodel <- "
ANALYSIS:
  TYPE=LCA;
  NCLASSES(3);
  NSTARTS(5, 20);
LAVAAN MODEL:
  F =~ A1__C4
    "
> # estimate model
> mod <- tamaan(tammodel, resp = dat,
+          control = list(progress = FALSE))
> head(mod$lcaprobs, 3)

  item itemno Cat    Class1    Class2       Class3
1   A1      1   0 0.3075322 0.1658559 0.0000458541
2   A1      1   1 0.6924678 0.8341441 0.9999541459
3   A2      2   0 0.4871546 0.4437222 0.0287977165
```

Processing Speed

Messy task: at multiple integration nodes θ_p, efficiently compute,

$$P\left(Y_{pi} = k \mid \theta_p\right) = \frac{\exp(\sum_d b_{ikd}\theta_{pd} + a_{ik}\xi_i)}{\sum_{k=0}^{k} \exp(\sum_d b_{ikd}\theta_{pd} + a_{ik}\xi_i)}, \forall i, k, p.$$

```
> calc_prob <- function(iIndex, A, AXsi, B, xsi, theta, nnodes, maxK){

+  AXsi.tmp <- array(tensor(A[iIndex, , , drop = FALSE], xsi, 3, 1),
+                    dim = c(length(iIndex), maxK, nnodes))
+  AXsi[iIndex,] = AXsi.tmp[,,1]

+  Btheta <- array(0, dim = c(length(iIndex) , maxK , nnodes) )
+  for( dd in 1:ncol(theta)){
+    Btheta <- Btheta + array(B[iIndex, , dd, drop = FALSE] %o% theta[, dd],
+                             dim = dim(Btheta))
+  }

+  rprobs <- (rr <- exp(Btheta + AXsi.tmp)) /
+    aperm(array(rep(colSums(aperm(rr ,c(2, 1, 3)), dims = 1, na.rm=TRUE), maxK),
+                dim = dim(rr)[c(1, 3, 2)]), c(1, 3, 2))

+  return(list("rprobs" = rprobs, "AXsi" = AXsi))
+}
```

Processing Speed II

Messy and time consuming task: efficiently compute the posterior distribution

$$f(\theta \mid Y) = \frac{f(Y \mid \theta)\, f(\theta)}{f(Y)}.$$

```
> # compute posterior distribution
> calc_posterior_TK <- function(rprobs, gwt, nitems){
+  fx <- gwt
+  for ( i in 1:nitems ){
+    r.ii <- rprobs[i ,, ]
+    fx <- fx * r.ii[ resp[,i] + 1 , ]
+  }
+  hwt <- fx / rowSums(fx)

+  return(hwt)
+ }
```

Processing Speed II

Messy and time consuming task: efficiently compute the posterior distribution

$$f(\theta \mid Y) = \frac{f(Y \mid \theta) f(\theta)}{f(Y)}.$$

```
for(i=0; i<nresp; i++){
  for(k=0; k<nnodes; k++){
    res[i+nresp*k] = REAL(sFx)[i+nresp*k];
  }
}

for(i=0; i<nitems; i++){
  // extract non-missing value list
  len = LENGTH(VECTOR_ELT(sRespIndList, i));
  ni = INTEGER(VECTOR_ELT(sRespIndList, i)) ; //ni indices in R, therefore '-1'

  //compute fx
  for(k=0; k<len; k++){
    for(l=0; l<nnodes; l++){
      res[ ni[k] + l*nresp - 1 ] = res[ ni[k] + l*nresp - 1] *
                    rii[ i + resp[ni[k]+i*nresp - 1]*nitems + l*nitems*ncats ];
    }
  }
}
```

		TAM	mirt	eRm	ConQuest*
N	I				
500	30	00:00,057	00:00,425	00:01,229	00:00,761
	60	00:00,088	00:00,320	00:05,550	00:00,900
	100	00:00,193	00:00,689	00:21,136	00:01,740
	200	00:00,579	00:01,876	02:31,158	00:03,661
3000	30	00:00,229	00:00,412	00:01,610	00:02,371
	60	00:00,297	00:00,639	00:06,447	00:03,303
	100	00:00,568	00:01,195	00:22,977	00:04,683
	200	00:01,422	00:03,041	02:31,919	00:10,779
10000	30	00:00,561	00:00,958	00:02,821	00:07,360
	60	00:00,656	00:01,447	00:08,969	00:09,405
	100	00:01,242	00:02,461	00:28,209	00:12,583
	200	00:03,325	00:06,053	02:47,708	00:24,396
70000	30	00:02,723	00:04,691	00:12,943	00:50,298
	60	00:04,240	00:09,244	00:29,783	08:03,661
	100	00:06,620	00:14,547	01:07,383	01:26,910
	2000	00:17,789	00:31,569	04:31,620	02:18,048

Framework of Packages

	TAM	CDM	sirt
Main functions	`tam`, `tam.mml`, `tam.mml.2pl/.3pl`	`din`, `gdina`, `gdm`	`supplementary`
Standard generics	`summary`, `plot`, `logLik`, `anova`, `residuals`		
Quasi-standard	`IRT.expectedCounts.*`, `IRT.factor.scores.*`, `IRT.irfprob.*`, `IRT.modelfit.*`, `IRT.posterior.*`		

Conclusion

Summary

- There already are several R packages for IRT analysis. None of which is suitable for LSA.
- **TAM** is unmatedly flexible (just a glimpse is presented) and competitive in means of processing speed.
- For convenient recovery of its flexibility, **TAM** offers and extends **lavaan**'s model syntax.
- **TAM** implements generic functions for objects from a wide range of packages (also beyond the scope of **TAM**, **CDM** and **sirt**).
- In the Future:
  - keep **TAM** competitive in terms of flexibility and processing speed,
  - extend and round up the model syntax,
  - extend the quasi standard generics,
  - provide a Vignette.

Thank you for your attention!

Thomas Kiefer

BIFIE Salzburg
t.kiefer@bifie.at
www.bifie.at/user/kiefer-thomas

Bundesinstitut bifie
Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens

References I

Adams, R. J., Wilson, M., & Wang, W.-C. (1997). The multidimensional random coefficient multinomial logit model. *Applied Psychological Measurement*, *21*, 1–23.

Adams, R. J., & Wu, M. L. (2007). The mixed-coefficients multinomial logit model: A generalized form of the Rasch model. In M. von Davier & C. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models* (p. 57-75). New York: Springer.

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, *42*, 69–78.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. Retrieved from `http://www.jstatsoft.org/v48/i06/`

De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

References II

Muraki, E. (1993, December). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, *17*(4), 351–363.

OECD (Ed.). (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from `http://www.jstatsoft.org/v48/i02/`

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). ACER Conquest Version 2.0 [Computer software manual]. Mulgrave.

Xu, X., & von Davier, M. (2007). *Fitting the Structured General Diagnostic Model to NAEP Data* (ETS Research Report). Princeton: ETS.