eat: An R Package for Automation of Data Preparation and IRT Modeling

Karoline Sachse, Martin Hecht, Sebastian Weirich,
Nicole Haag, Malte Jansen, Sebastian Wurster,
Christiane Penk, Anna Lenski, Thilo Siegle

Institute for Educational Quality Improvement
Humboldt-University, Berlin

February 10, 2012

Overview

1. eat - History
   - The Institute for Educational Quality Improvement
   - ACER ConQuest
   - The Idea
2. eat - Concept
   - Overview
   - Data
3. eat - Examples
   - Data Preparation
   - Unidimensional 1PL model with automateModels
   - Grouping options in automateModels
   - automateModels
4. Discussion
   - Outlook

The Institute

- Independent research and test institute founded by the 16 federal states in 2004
- Nationwide Educational Standards Assessments in German, the first foreign language, Mathematics and Science which allow comparison of federal states ($N \approx 30,000$)
- Assessment tests in German, Mathematics and the first foreign language in the 8th grade at secondary school (once a year)
- Assessment tests in German and Mathematics in the 3rd grade at primary school (once a year)

ConQuest

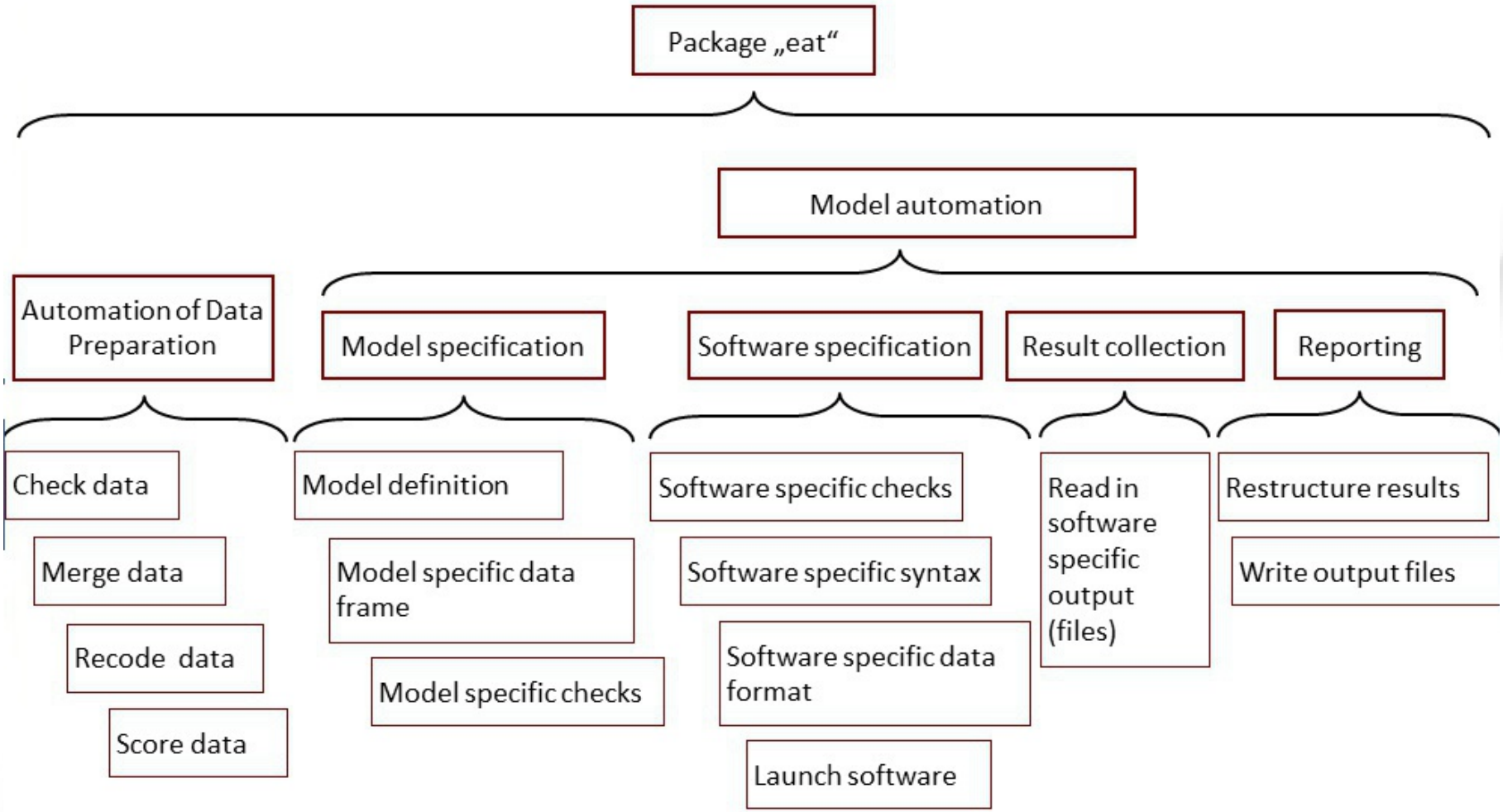
- Commercial Software developed by ACER (Wu, Adams & Wilson, 1997)
- Major scaling tool of the Organisation for Economic Co-operation and Development's Programme for International Student Assessment (PISA)
- Fits a large number of different item response models
  - Rasch, partial credit, rating scale, facets, ...
  - Latent regression
  - Multidimensionality
- Estimation
  - Marginal Maximum Likelihood
  - Gaussian quadrature/ Monte Carlo approximations
  - Person parameter estimation: EAP, MLE, WLE, Plausible values

Automation of Data Preparation and Analysis

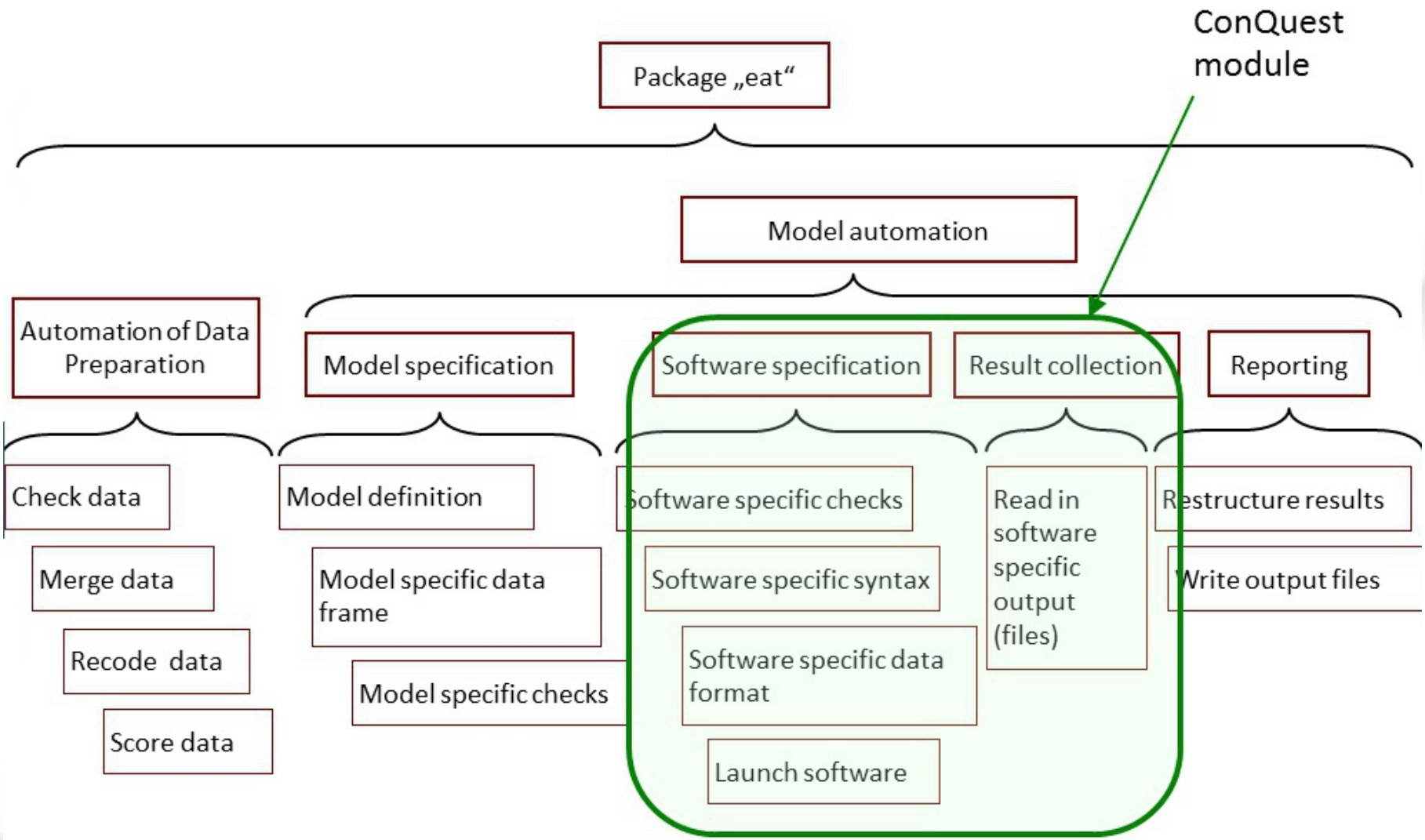
- automate data preparation
  1. read in & check SPSS-files
  2. merge data frames (booklets)
  3. recode & dichotomize data
- automate IRT calibration
  1. write ConQuest syntax, generate appropriate data input
  2. execute ConQuest
  3. read in ConQuest output
- facilitate reporting
  1. write out results (graphics, tables, ...)

$\Rightarrow$ "eat" ("Educational Assessment Tools")

Implemented Modules



Wrapping ConQuest



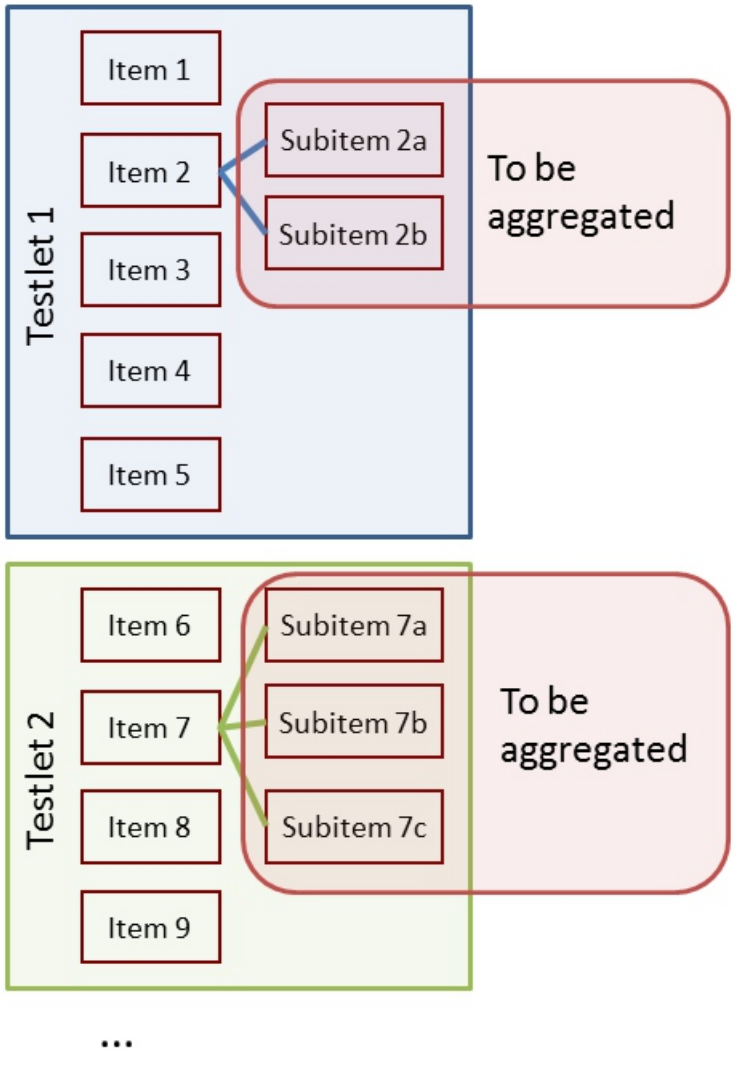
Typical Items

	0. Place to hire a bike	Cambridge Station
item 1	1. Standard equipment of all bikes (Name one.)	lights
item 2	2. Extras you can ask for	a)
		b)
item 3	3. Price for 4 hours	£ 6
item 4	4. Price for one month	£ 38
item 5	5. Students' offer for 2 weeks	£ 25

Typical Items - Scores

0. Place to hire a bike	Cambridge Station		
1. Standard equipment of all bikes (Name one.)	lights	3	1
2. Extras you can ask for	a)	9	mbi
	b)	9	mbi
3. Price for 4 hours	£ 6	0	0
4. Price for one month	£ 38	1	1
5. Students' offer for 2 weeks	£ 25	1	1
		Trained Rater	Recoded

Data Structure



1. Recode all Items & Subitems
2. Aggregate Subitems
3. Recode aggregated Subitems

automateDataPreparation

```
dataset <- automateDataPreparation ( inputList = inputList, path = path,
      loadSav = TRUE,
      checkData = TRUE,
      mergeData = TRUE,
      recodeData = TRUE,
      aggregateData = TRUE,
      scoreData = TRUE,
      writeSpss = TRUE )
```

Data Preparation Input

```
> inputList$units
      unit                    unitLabel unitType unitAggregateRule
1   ER52501          Bikes ER52501       TI
2   ER52502          Bikes ER52502       TI               SUM
3   ER52503          Bikes ER52503       TI
4   ER52504          Bikes ER52504       TI
5   ER52505          Bikes ER52505       TI
> inputList$subunits
      unit  subunit                    subunitLabel subunitRecoded subunitLabelRecoded
1   ER52501  ER52501          Bikes ER52501       ER52501R     Recoded ER52501
2   ER52502 ER52502a          Bikes ER52502a      ER52502aR    Recoded ER52502a
3   ER52502 ER52502b          Bikes ER52502b      ER52502bR    Recoded ER52502b
4   ER52503  ER52503          Bikes ER52503       ER52503R     Recoded ER52503
5   ER52504  ER52504          Bikes ER52504       ER52504R     Recoded ER52504
6   ER52505  ER52505          Bikes ER52505       ER52505R     Recoded ER52505
> inputList$values
    subunit value valueRecode valueType                valueLabel
9  ER52502a     1           1        vc                    gloves
10 ER52502a     2           0        vc                     other
11 ER52502a     6         mnr       mnr       missing not reached
12 ER52502a     7         mci       mci missing coding impossible
13 ER52502a     8         mir       mir        invalid response
14 ER52502a     9         mbi       mbi      missing by intention
15 ER52502b     1           1        vc              water bottle
16 ER52502b     2           0        vc                     other
17 ER52502b     6         mnr       mnr       missing not reached
18 ER52502b     7         mci       mci missing coding impossible
19 ER52502b     8         mir       mir        invalid response
20 ER52502b     9         mbi       mbi      missing by intention
> inputList$unitRecodings
      unit value valueRecode valueType valueLabelRecoded
1  ER52502     0           0        vc          ER52502R
2  ER52502     1           0        vc          ER52502R
3  ER52502     2           1        vc          ER52502R
```

Data Preparation Logfile

```
1  ------------------------------------------------------------------------------
2  | 2012-01-20 17:10:28 | USER212 | PC017 | R version 2.14.0 (2011-10-31) /i386 |
3  ------------------------------------------------------------------------------
4  automateDataPreparation: Starting automateDataPreparation 2012-01-20 17:10:28
5
6  automateDataPreparation: Load .sav Files
7
8  automateDataPreparation: Check data...
9  .checkData.checkID: Only valid codes in ID variable.
10 .checkData.checkID: No duplicated entries in ID variable.
11 .checkData.checkVars: No duplicated variable names.
12 .checkData.checkCodes: Found no invalid codes.
13 .checkData.checkMissings: Case(s)23 contain(s) only missing values.
14
15 automateDataPreparation: Start merging
16 mergeData_0.4.0: Start merging of dataset 1.
17 mergeData_0.4.0: Start merging of dataset 2.
18 mergeData_0.4.0: Start merging of dataset 3.
19
20 automateDataPreparation: Start recoding
21 recodeData_1.0.1: ER52501 has been recoded.
22 recodeData_1.0.1: ER52502a has been recoded.
23 recodeData_1.0.1: ER52502b has been recoded.
24 recodeData_1.0.1: ER52503 has been recoded.
25 recodeData_1.0.1: ER52504 has been recoded.
26 recodeData_1.0.1: ER52505 has been recoded.
27
28 automateDataPreparation: Start aggregating
29 aggregateData_1.2.0: Aggregate unit ER52502.
30
31 automateDataPreparation: Start scoring
32 recodeData_1.0.1: ER52502 has been recoded.
33
34 automateDataPreparation: Writing dataset in last transformation status to disk
35 writeSpss_0.2.0:  Data values written to c:/temp/zkddata.txt
36 writeSpss_0.2.0:  Syntax file written to c:/temp/readZkdData.sps
37
38 automateDataPreparation: terminated successfully! 2012-01-20 17:10:39
```

Simple Unidimensional 1PL Model

```
results01 <- automateModels( dataset = dataset , folder = folder )
```

ConQuest Dataset & Label File Creation

labels

```
===> item
1 ER22201
2 ER22202
3 ER22203
4 ER22204
5 ER22205
6 ER30401
...
233 ER33401
234 ER33402
235 ER33403
236 ER33404
237 ER19501
238 ER19502
239 ER20503
240 ER26301
===> dimensions
1 all.i
```

dataset

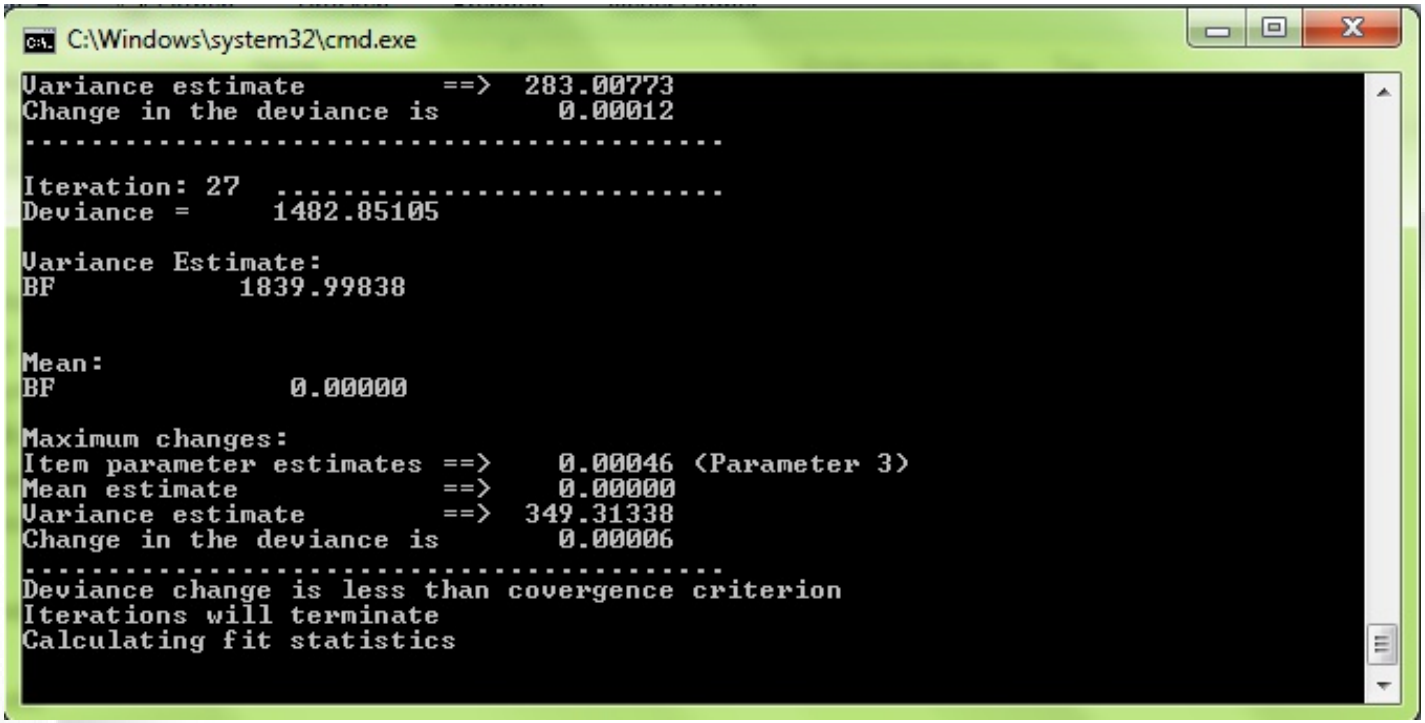
```
341611011111111111110111111.....10100111010010101100100.
341611021111111111100111111.....11111001111011000010000.
341611031111111101010111111.....11000000011010111011110.
341611041111111101010111111.....11111111111010100111101.
341611051111101010111111.....11101111010000010111101.
341611061111111101010011111.....00111111111011111000000.
341611071111111100000111111.....10000001010011011001011.
341611081111111101010011111.....11111111111011100110000.
341611091111101010010111111.....01111111111010010111111.
341611101111111111110011111.....11111011111010100101001.
341611111111101111101111111.....01101111010010001110000.
341611121111111101000111111.....01100101010010101011101.
341611131111111101001111000001.....
341611141111111101001111001011.....
341611151111111111111111100000.....
341611161111111101011111110101.....
341611171111111111001101100100.....
341611181110011011011111011100.....
341611191111111000011111001100.....
341611201110101010101111001010.....
341611211110101010011111110010.....
341611221111111101001111000001.....
351612031111101010011111.....10110111111011000000101.
351612041110111110011111.....00111111010010000001000.
```

ConQuest Syntax Creation

```
title = Analysis name: all.i__all.p, User: USER212, Computername: PC017,
R version 2.14.1 (2011-12-22), Time: Sun Jan 29 15:18:50 2012;
export logfile >> all.i__all.p.log;
datafile all.i__all.p.dat;
Format pid 1-8 responses 9-248;
codes 1,0;
labels  << all.i__all.p.lab;
score ( 0 1 ) ( 0 1 ) ! items( 1-240 );
set constraints=cases;
set warnings=no,update=yes,n_plausible=5,p_nodes=2000,f_nodes=2000;
export par    >> all.i__all.p.prm;
model item;
estimate ! method=gauss,iter=1000,nodes=15,converge=0.0001,deviancechange=0.0001,
stderr=quick,distribution=normal;
Itanal >> all.i__all.p.itn;
show cases! estimates=latent >> all.i__all.p.pvl;
show cases! estimate=wle >> all.i__all.p.wle;
equivalence wle >> all.i__all.p.equ;
show >> all.i__all.p.shw;
descriptives !estimates=pv >> all.i__all.p_pvl.dsc;
descriptives !estimates=wle >> all.i__all.p_wle.dsc;
quit;
```

ConQuest Run

- ConQuest runs due to automatic creation and execution of batch files



ConQuest Output Files

- Item parameter estimates
  - .shw, .itn, ...
- Person parameter estimates
  - .wle, .mle, .eap, .pvl, ...

$\Rightarrow$ Many different output files

ConQuest Item Parameter Output

all.i__all.p.itn

```
FF================================================================================
= Analysis name: all.i__all.p, User: USER"!", ComputernameSun Jan 29 15:20 2012
GENERALISED ITEM ANALYSIS
================================================================================
Item 1
------
item:1 (ER22201)
Cases for this item    247   Discrimination  0.11
Item Threshold(s):    -4.38   Weighted MNSQ   1.04
Item Delta(s):     -4.38
--------------------------------------------------------------------------------
 Label     Score     Count    % of tot  Pt Bis     t  (p)    PV1Avg:1 PV1 SD:1
--------------------------------------------------------------------------------
   0        0.00         4        1.62   -0.11    -1.71(.089) -0.32     1.35
   1        1.00       243       98.38    0.11     1.71(.089)  0.32     1.00
================================================================================
```

all.i__all.p.shw

```
FF================================================================================
= Analysis name: all.i__all.p, User: USER212, ComputernameSun Jan 29 15:20 2012
TABLES OF RESPONSE MODEL PARAMETER ESTIMATES
================================================================================
TERM 1: item
--------------------------------------------------------------------------------
   VARIABLES                                UNWEIGHTED FIT          WEIGHTED FIT
---------------                        ----------------------  ----------------------
     item            ESTIMATE  ERROR^  MNSQ       CI        T    MNSQ       CI        T
--------------------------------------------------------------------------------------
 1  ER22201           -4.378    0.513   2.34 ( 0.82, 1.18) 11.0   1.04 ( 0.07, 1.93)  0.2
 2  ER22202           -3.925    0.423   0.54 ( 0.82, 1.18) -6.1   0.96 ( 0.26, 1.74)  0.0
 3  ER22203           -5.781    1.007   0.58 ( 0.82, 1.18) -5.5   1.03 ( 0.00, 2.94)  0.4
 4  ER22204           -1.283    0.170   1.11 ( 0.82, 1.18)  1.2   1.06 ( 0.81, 1.19)  0.6
 5  ER22205           -4.142    0.461   1.85 ( 0.82, 1.18)  7.6   1.07 ( 0.18, 1.82)  0.3
```

ConQuest Person Parameter Output

all.i__all.p.wle

```
 1      1   01111101      37.00      48.00     1.06340     0.42056
 2      2   01111103      38.00      48.00     1.24321     0.43329
 3      3   01111104      40.00      48.00     1.64186     0.46594
 4      4   01111105      40.00      48.00     1.64186     0.46594
 5      5   01111106      45.00      48.00     3.09501     0.65467
 6      6   01111107      45.00      48.00     3.09501     0.65467
 7      7   01111108      46.00      48.00     3.56905     0.74690
 8      8   01111109      36.00      48.00     0.89348     0.40974
 9      9   01111110      41.00      48.00     1.86715     0.48683
10     10   01111111      40.00      48.00     1.64186     0.46594
```

all.i__all.p.pvl

```
 1   1  01111101
 2  1          0.52
 3  2          1.11
 4  3          1.04
 5  4          1.06
 6  5          0.99
 7  0.95723
 8  0.38437
 9   2  01111103
10  1          0.38
11  2          1.05
12  3          1.57
13  4          2.08
14  5          1.07
15  1.10824
16  0.39400
17   3  01111104
18  1          0.97
19  2          1.07
20  3          2.10
21  4          1.12
22  5          2.34
23  1.44900
24  0.43521
```

eat Reporting

Item parameter estimates

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	analysis	dimension	group	item	n.valid	p	a	b	c	d	b.se	infit	infit.ci.lb	infit.ci.ub	infit.t	outfit	outfit.ci.lb	outfit.ci.ub	outfit.t	pbc
2	all.i__all.p	all.i	all.p	ER22601	489	0,52		-0,096			0,228	0,82	0,75	1,25	-1,5	0,13	0,86	1,14	-21,1	0,97
3	all.i__all.p	all.i	all.p	ER22602	489	0,49		0,483			0,234	1,88	0,74	1,26	5,3	1	0,86	1,14	0	0,91
4	all.i__all.p	all.i	all.p	ER22603	489	0,53		-0,358			0,231	1,32	0,74	1,26	2,2	16,53	0,86	1,14	66,4	0,93
5	all.i__all.p	all.i	all.p	ER22604	489	0,52		-0,096			0,228	0,82	0,75	1,25	-1,5	0,14	0,86	1,14	-20,8	0,97
6	all.i__all.p	all.i	all.p	ER22605	489	0,49		0,538			0,236	1,47	0,74	1,26	3,1	0,36	0,86	1,14	-12,5	0,93

Person parameter estimates

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	analysis	dimension	group	person	n.solved	n.total	wle	wle.se	eap	eap.se	pv.1	pv.2	pv.3	pv.4	pv.5
2	all.i__all.p	all.i	all.p	1120011210	0	5	-2,36696	1,64599	-34,9682	25,24358	-66,44	-42,28	-81,59	-17	-30,93
3	all.i__all.p	all.i	all.p	1120019109	5	5	2,55778	1,65001	35,74393	26,45026	9,43	22,41	27,42	18,32	22,97
4	all.i__all.p	all.i	all.p	1140011218	5	5	2,55778	1,65001	35,74393	26,45026	46,4	9,92	24,47	3,59	19,81
5	all.i__all.p	all.i	all.p	1140019107	0	5	-2,36696	1,64599	-34,9682	25,24358	-2,44	-16,11	-22,83	-43,8	-21,21
6	all.i__all.p	all.i	all.p	1140019114	5	5	2,55778	1,65001	35,74393	26,45026	8,46	45,65	72,45	12,19	31,53
7	all.i__all.p	all.i	all.p	1140019127	0	5	-2,36696	1,64599	-34,9682	25,24358	-20,13	-26,22	-9,79	-19,92	-48,85
8	all.i__all.p	all.i	all.p	1211029102	5	5	2,55778	1,65001	35,74393	26,45026	63,45	43,11	36,9	34,25	18,16
9	all.i__all.p	all.i	all.p	1211029106	5	5	2,55778	1,65001	35,74393	26,45026	13,73	52,2	32,86	17,91	31,81
10	all.i__all.p	all.i	all.p	1211029111	5	5	2,55778	1,65001	35,74393	26,45026	2,81	8,9	58,77	47,88	42,22
11	all.i__all.p	all.i	all.p	1220021217	5	5	2,55778	1,65001	35,74393	26,45026	6,07	8,46	60,51	53,7	9,18
12	all.i__all.p	all.i	all.p	1220029115	0	5	-2,36696	1,64599	-34,9682	25,24358	-30,52	-31,62	-31,36	-19,04	-9,63
13	all.i__all.p	all.i	all.p	1221019101	5	5	2,55778	1,65001	35,74393	26,45026	16,54	25,97	13,25	2,08	26,03
14	all.i__all.p	all.i	all.p	1221019105	0	5	-2,36696	1,64599	-34,9682	25,24358	-34,27	-6,23	-56,56	-12,46	-25,5
15	all.i__all.p	all.i	all.p	1221019112	5	5	2,55778	1,65001	35,74393	26,45026	15,07	14,95	10,52	13,25	39,2

eat Log

- all objects (dataset, item.grouping, ...) will be archived into an .RData file
- an INFO file will be created

```
MODEL INFORMATION

Model Name:             all.i__all.p
Dimensions:             1 (unidim)
Dimension Names:        all.i
Groups:                 1 (singlegroup)
Group Names:            all.p
Measurement Model:      1pl
Software:               conquest
DIF:                    none
Regression:             none
Anchor:                 none
Missing Rule:           mvi = 0, mnr = 0, mci = NA, mbd = NA, mir = 0, mbi = 0
Deskr. Gruppenvar.:     none
Deskriptive Gruppen:    none

Generated:              2012-01-29 15:20:31
User:                   USER212
Userdomain:             user
Computername:           PC017
```

Multidimensional vs. Unidimensional Analysis

```
results02 <- automateModels( dataset = dataset , id = "id" , folder = folder ,
     item.grouping = item.grouping ,
     select.item.group = c ( "ER" , "EL" ) )

results03 <- automateModels( dataset = dataset , id = "id" , folder = folder ,
     item.grouping = item.grouping ,
     select.item.group = c ( "ER" , "EL" ) , cross="all")
```

```
> item.grouping
      item ER EL EW
1  ER22201  1  0  0
2  ER22202  1  0  0
3  ER22203  1  0  0
4  ER22204  1  0  0
5  ER22205  1  0  0
6  EL30101  0  1  0
7  EL30102  0  1  0
8  EL23401  0  1  0
9  EL23402  0  1  0
10 EL23403  0  1  0
11 EW00101  0  0  1
12 EW00102  0  0  1
13 EW02301  0  0  1
14 EW02302  0  0  1
15 EW02303  0  0  1
```

Person groups & weights

```
dataset <- cbind ( dataset , "weight1" = as.character(sample(c(0.8, 1, 1.2),
      nrow(dataset), replace=TRUE)), "weight2" = as.character(sample(c(1),
      nrow(dataset), replace=TRUE)), stringsAsFactors = FALSE )

results04 <- automateModels( dataset = dataset, folder = folder
     context.vars = c ( "weight1" , "weight2" ) ,
     item.grouping = item.grouping ,
     select.item.group = "ER" ,
     person.grouping = person.grouping ,
     select.person.group = list ( "gr.9" , "gr.10" ) ,
     weight = list ( "weight1" , "weight2" ) )
```

automateModels – Overview

```
automateModels(dataset, id = NULL, context.vars = NULL, items = NULL,
    item.grouping = NULL, select.item.group = NULL,
    person.grouping.vars = NULL, person.grouping.vars.include.all = FALSE,
    person.grouping = NULL, select.person.group = NULL,
    additional.item.props = NULL, folder, overwrite.folder = TRUE,
    analyse.name.prefix = NULL, analyse.name = NULL,
    analyse.name.elements = NULL, data.name = NULL, m.model = NULL,
    software = NULL, dif = NULL, weight = NULL, anchor = NULL,
    regression = NULL, adjust.for.regression = FALSE, q3 = FALSE,
    missing.rule = NULL, cross = NULL, subfolder.order = NULL,
    subfolder.mode = NULL, additionalSubFolder = NULL, run.mode = NULL,
    n.batches = NULL, run.timeout = 1440, run.status.refresh = 0.2,
    email = NULL, smtpServer = NULL, write.txt.dataset = FALSE,
    delete.folder.countdown = 5, conquestParameters = NULL )
```

Thank you

- Thank you for your attention!

  `http://r-forge.r-project.org/eat`

  `eat-commits@lists.r-forge.r-project.org`

- Special thanks to

  Alexander Robitzsch (Measurement Statistician, bifie)
  Martin Mechtel (IT Director, IQB)