# Examining Exams with the psycho* Family of R Packages

Achim Zeileis

http://eeecon.uibk.ac.at/~zeileis/

# Overview

**"Best of"** of previous Psychoco contributions:

- The *psycho** family of R packages
  - *psychotools*: Data infrastructure and psychometric models
  - *psychotree*: Model-based recursive partitioning
  - *psychomix*: Mixture models
- The R package *exams*
  - Flexible generation of e-learning and large lecture exams
  - Based on exercise templates in `Sweave` format (R & LaTeX)
- Examining exams
  - IRT modeling of recent mathematics exam (Universität Innsbruck)
  - Assessment of measurement invariance and beyond

**Many collaborators:** Basil Abou El-Komboz, Hannah Frick, Bettina Grün, Torsten Hothorn, Julia Kopf, Friedrich Leisch, Edgar C. Merkle, Carolin Strobl, Nikolaus Umlauf, Florian Wickelmaier.

# The psycho* family

**Original motivation:** “No frills” implementation for lean and fast estimation of psychometric models (first Bradley-Terry, then Rasch) in model-based recursive partitioning (and later on mixture models).

**Now:**

- Suite of packages for psychometric modeling.
- Special emphasis on assessment of measurement invariance.
- But also lots of standard methods for visualization, inference, etc.
- Classes and methods for data handling.

**psychotools:** Joint infrastructure, basic modeling functions, data sets.

**psychotree:** Trees based on *partykit*.

**psychomix:** Mixture models based on *flexmix*.

# The psycho* family

**Psychometric models:**

| | | *psychotools* `model` | *psychotree* `tree` | *psychomix* `mix` |
|---|---|---|---|---|
| Rasch | `rasch` | × | × | × |
| Rating scale | `rs` | × | × | |
| Partial credit | `pc` | × | × | |
| Bradley-Terry | `bt` | × | × | × |
| Multin. proc. tree | `mpt` | × | × | |

**Data classes:** `itemresp()`, `paircomp()`.

**Inference:** `anchortest()`.

**IRT infrastructure:** `itempar()`, `personpar()`, `theshpar()`, ...

**Visualization:** `profileplot()`, `regionplot()`, `curveplot()`, ...

# Flexible generation of exams

**Package:** *exams* for e-learning and large-lecture exams.

**Basis:** Exercises in `Sweave` format with R code and LaTeX text.

**Output formats:**

- PDF (e.g., for classical written exams).
- HTML
- Moodle XML.
- QTI 1.2 (especially OLAT, Blackboard under development).
- QTI 2.1 (e.g., for ONYX).

**Further functionality:** In package *c403*, not yet publicly available.

- PDF exams with extensive infrastructure, e.g., automatic scanning of sheets and student report generation.
- ARSnova for quizzes during large-scale lectures including peer instruction elements.

# Mathematics 101 at Universität Innsbruck

**Course:** Mathematics for first-year business and economics students at Universität Innsbruck.

**Format:** Biweekly online tests (conducted in OpenOLAT) and two written exams for about 1,000 students per semester.

**Here:** Individual results from an end-term exam.

- 729 students (out of 941 registered).
- 13 single-choice items with five answer alternatives, covering the basics of analysis, linear algebra, financial mathematics.
- Two groups with partially different item pools (on the same topics). Individual versions of items generated via *exams*.
- Correctly solved items yield 100% of associated points. Items without correct solution can either be unanswered (0%) or with an incorrect answer ($-25\%$). Mostly only considered as binary here.

# Examining exams

**Packages:**

```
R> library("psychotools")
R> library("psychotree")
R> library("psychomix")
```

**Data:** Load and exclude extreme scorers.

```
R> load("MathExam.rda")
R> mex <- subset(MathExam, nsolved > 0 & nsolved < 13)
```

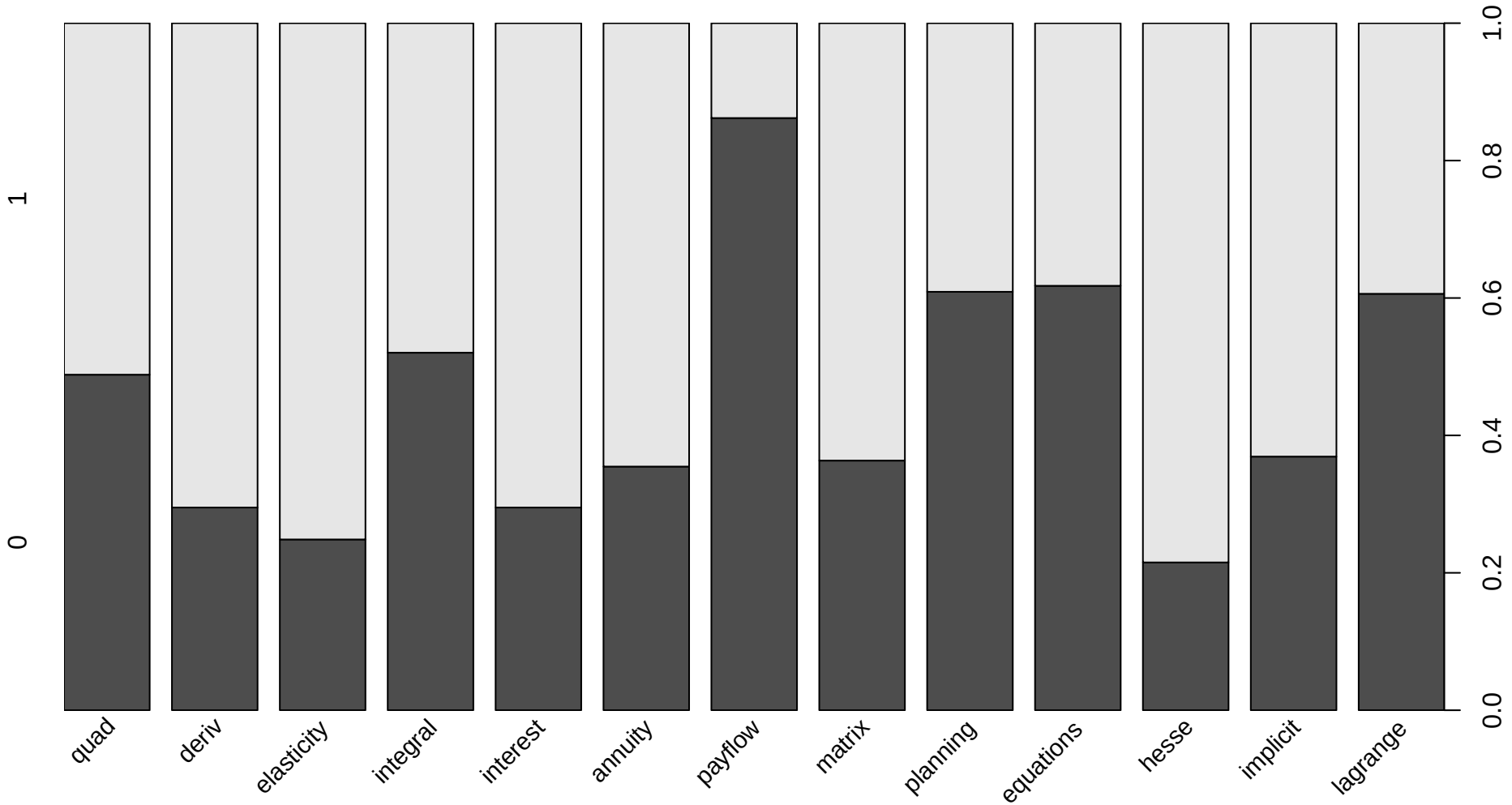
**Overview:** Data classes, see `str(mex)` for more details.

```
R> sapply(mex, function(x) class(x)[1])
```

```
    solved    credits    nsolved      tests      study   semester
"itemresp" "itemresp"  "numeric"  "integer"   "factor"  "integer"
   attempt      group    solved2   credits2
 "ordered"   "factor" "itemresp" "itemresp"
```

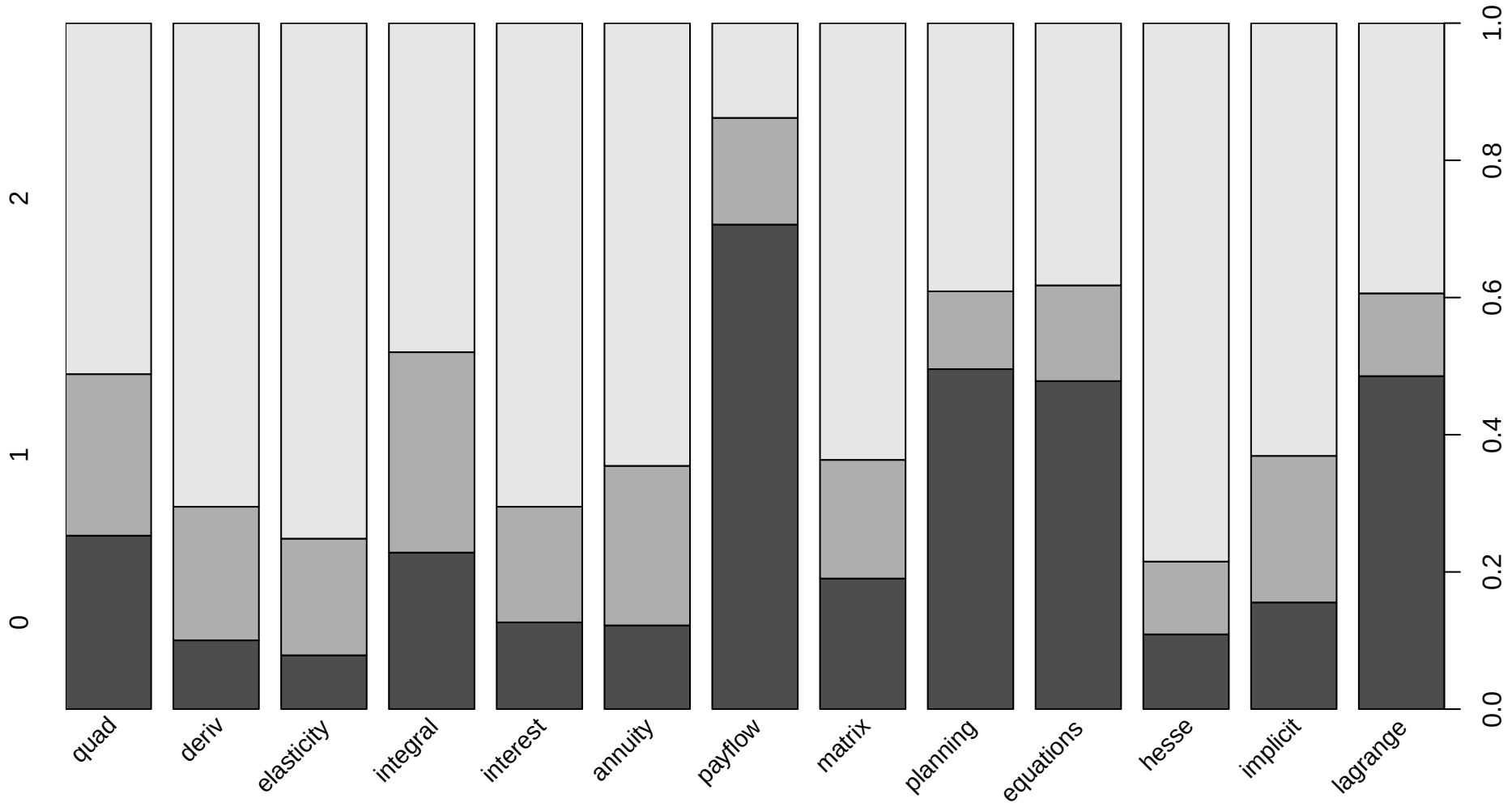
# Examining exams

```
R> plot(mex$solved)
```



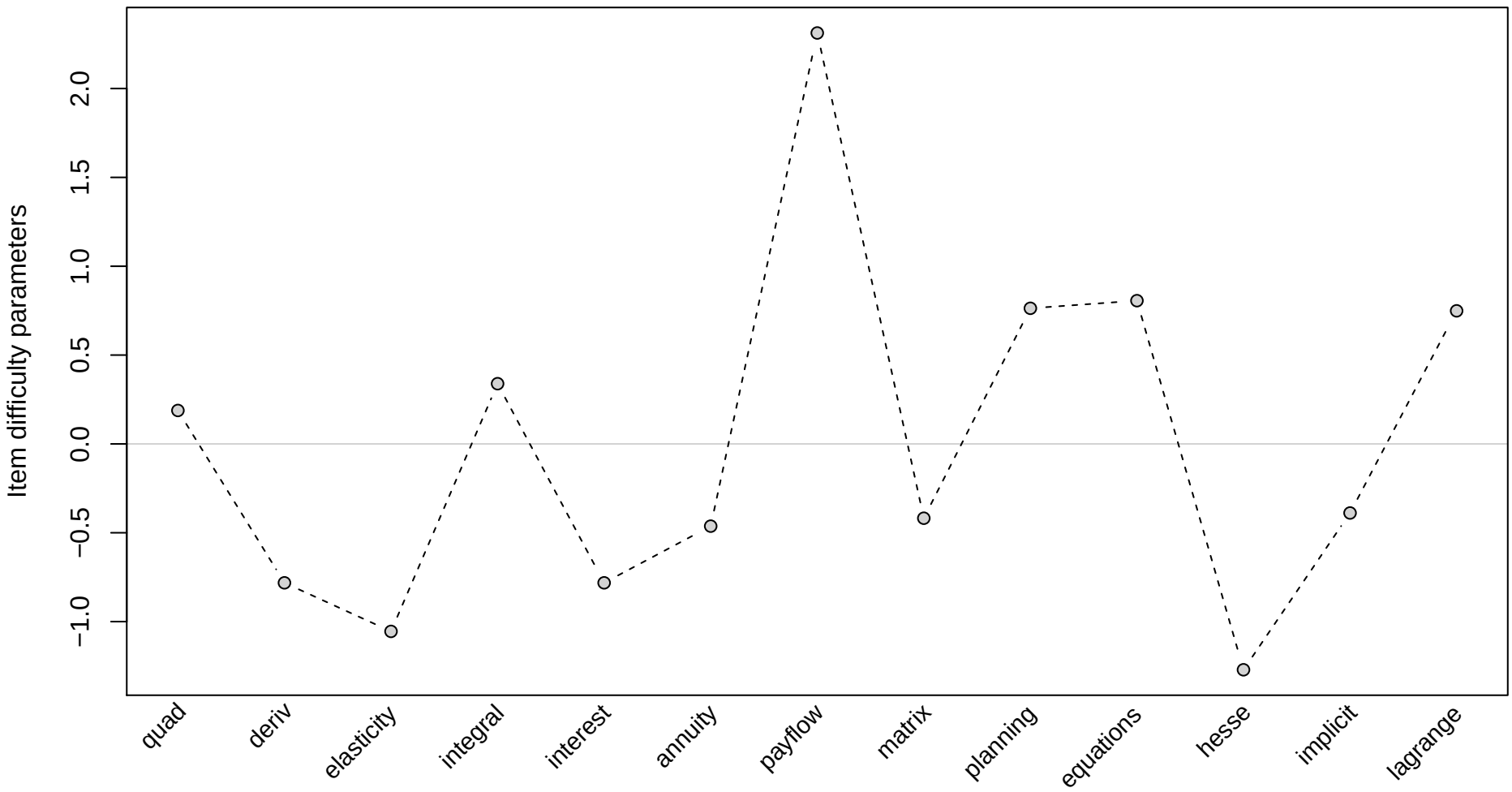
# Examining exams

```
R> plot(mex$credits)
```



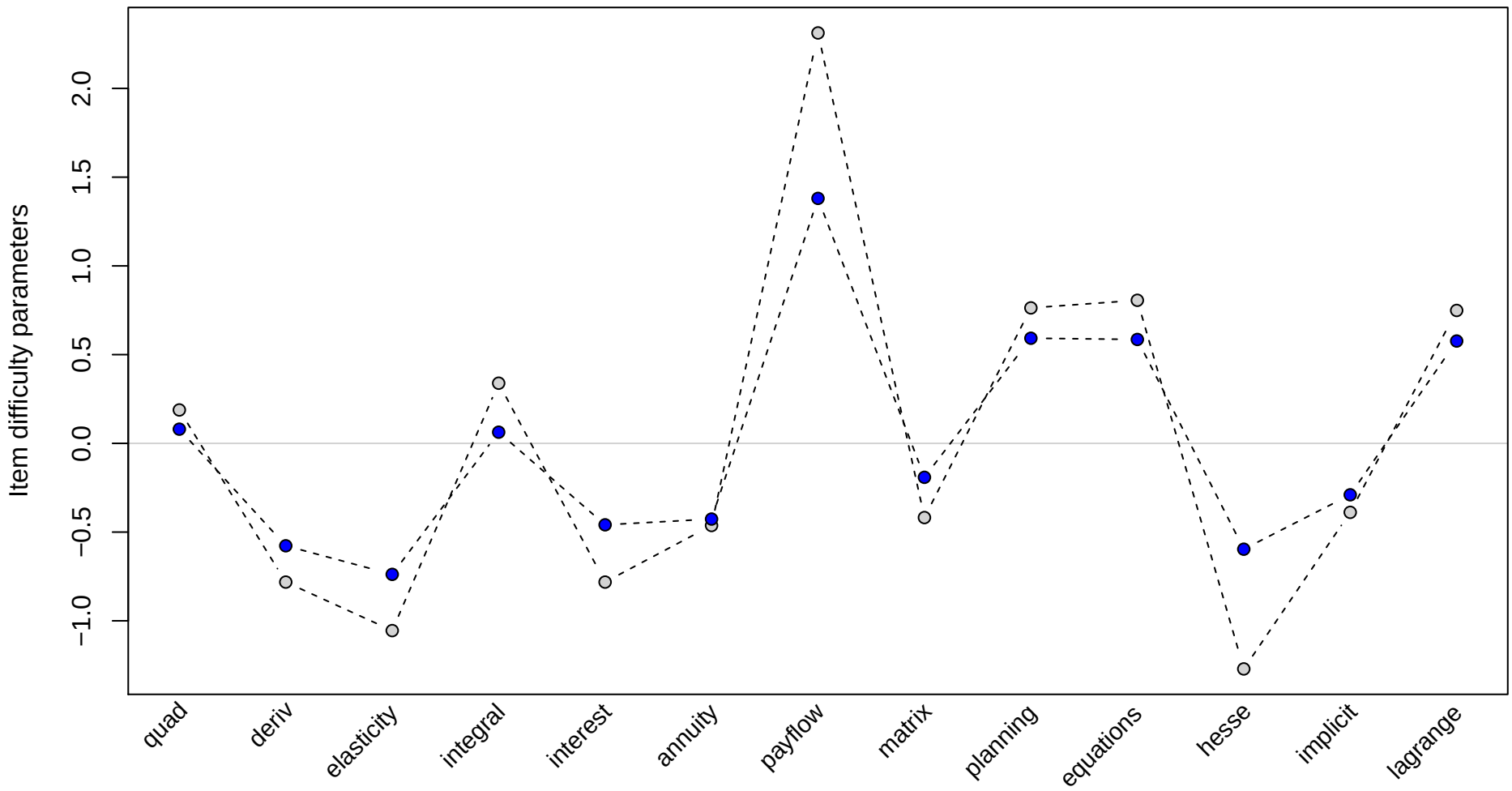
# Examining exams

```
R> ram <- raschmodel(mex$solved)
R> plot(ram, type = "profile")
```



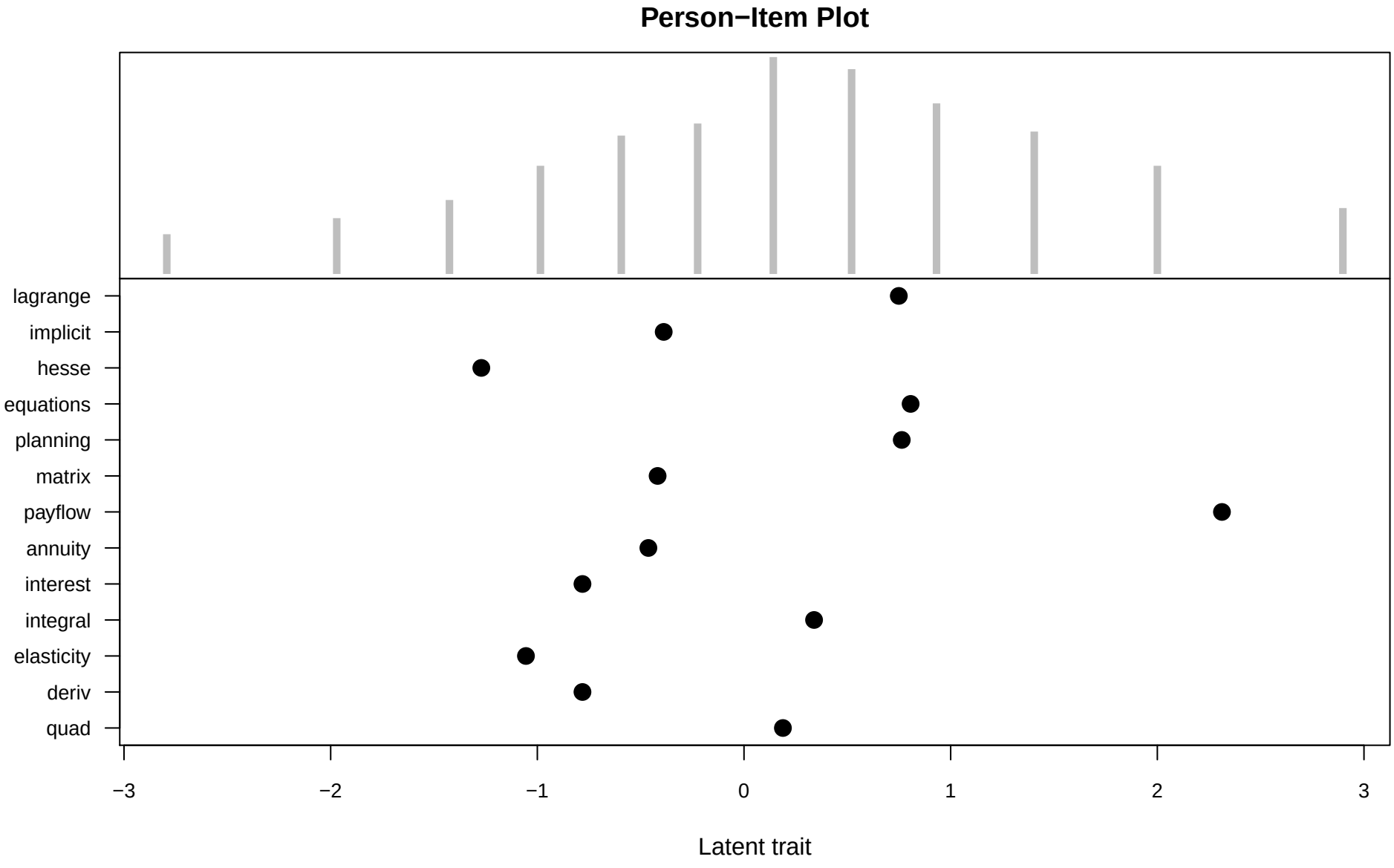
# Examining exams

```
R> pcm <- pcmodel(mex$credits)
R> plot(pcm, type = "profile", add = TRUE, col = "blue")
```



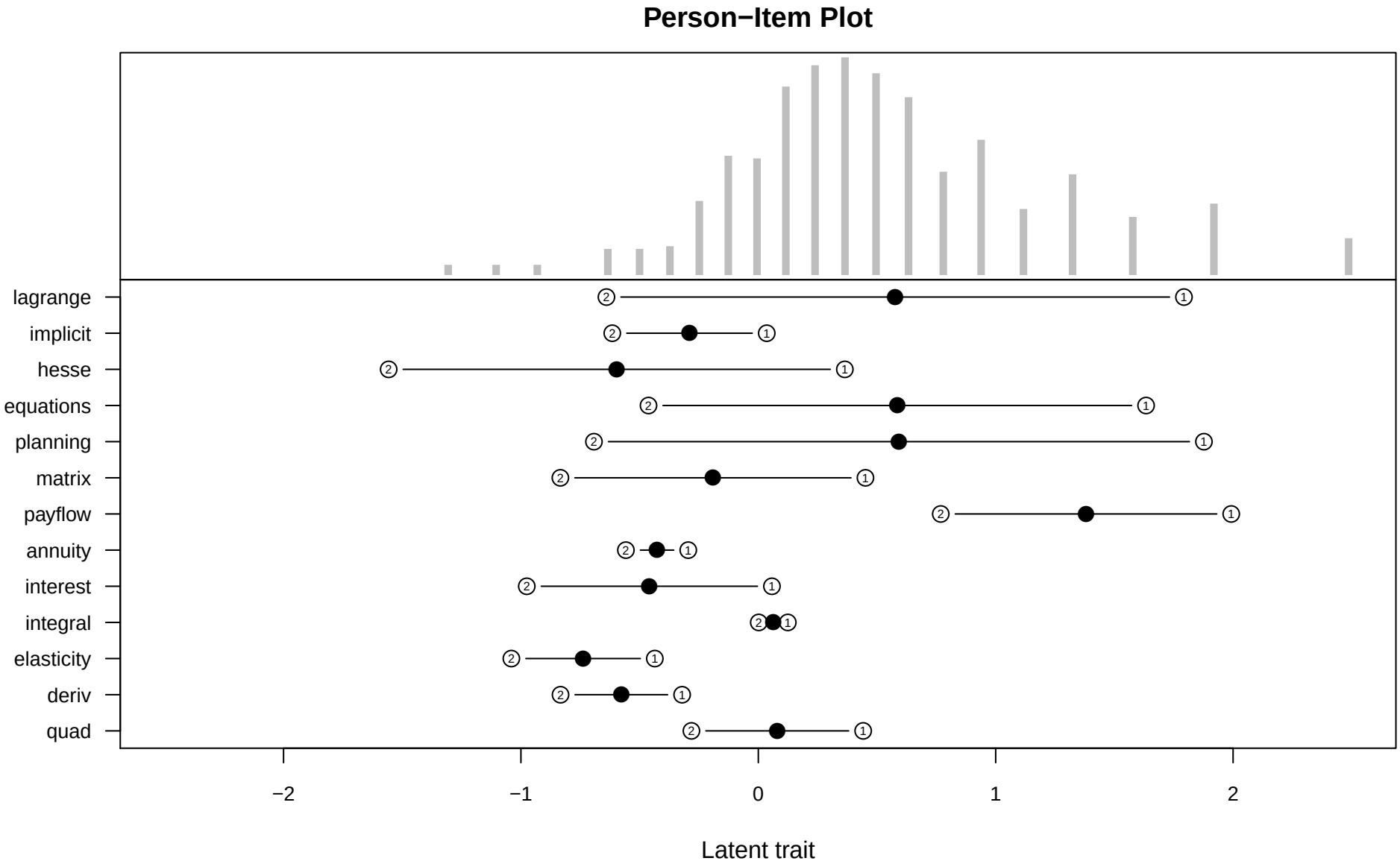
# Examining exams

```
R> plot(ram, type = "piplot")
```



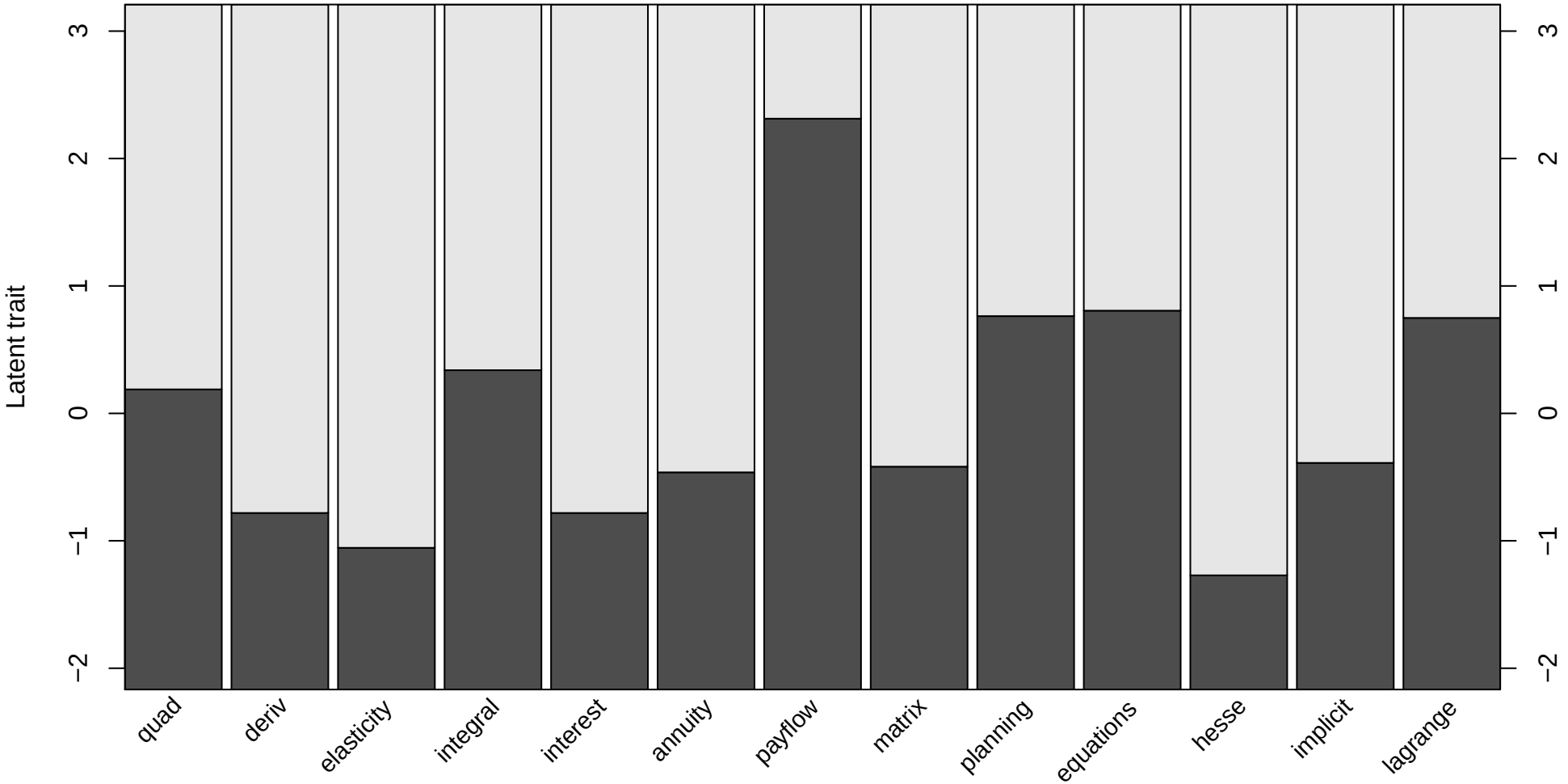
# Examining exams

```
R> plot(pcm, type = "piplot")
```



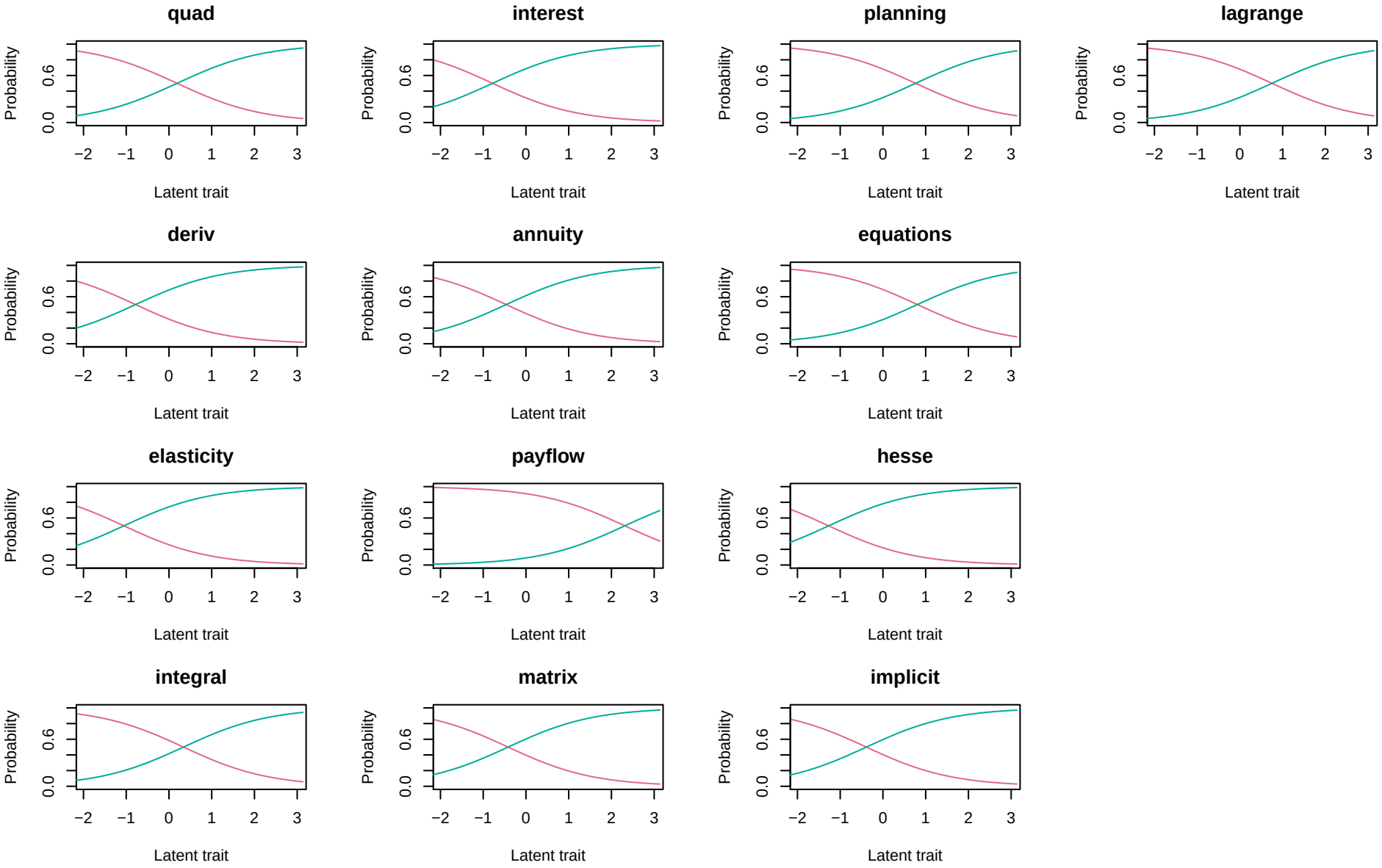
# Examining exams

```
R> plot(ram, type = "regions")
```



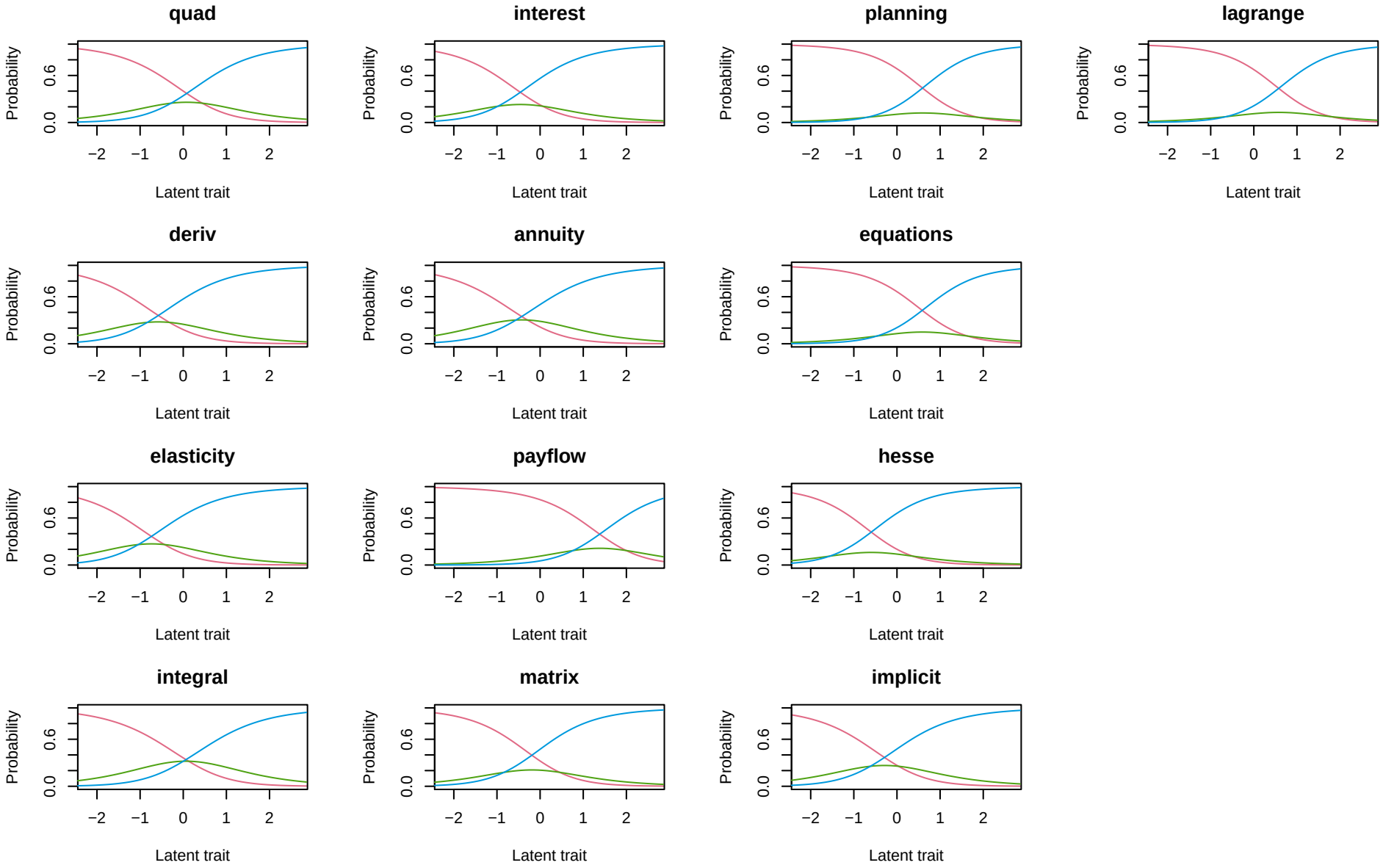
# Examining exams

```
R> plot(ram, type = "curves")
```



# Examining exams

```
R> plot(pcm, type = "curves")
```



# Examining exams

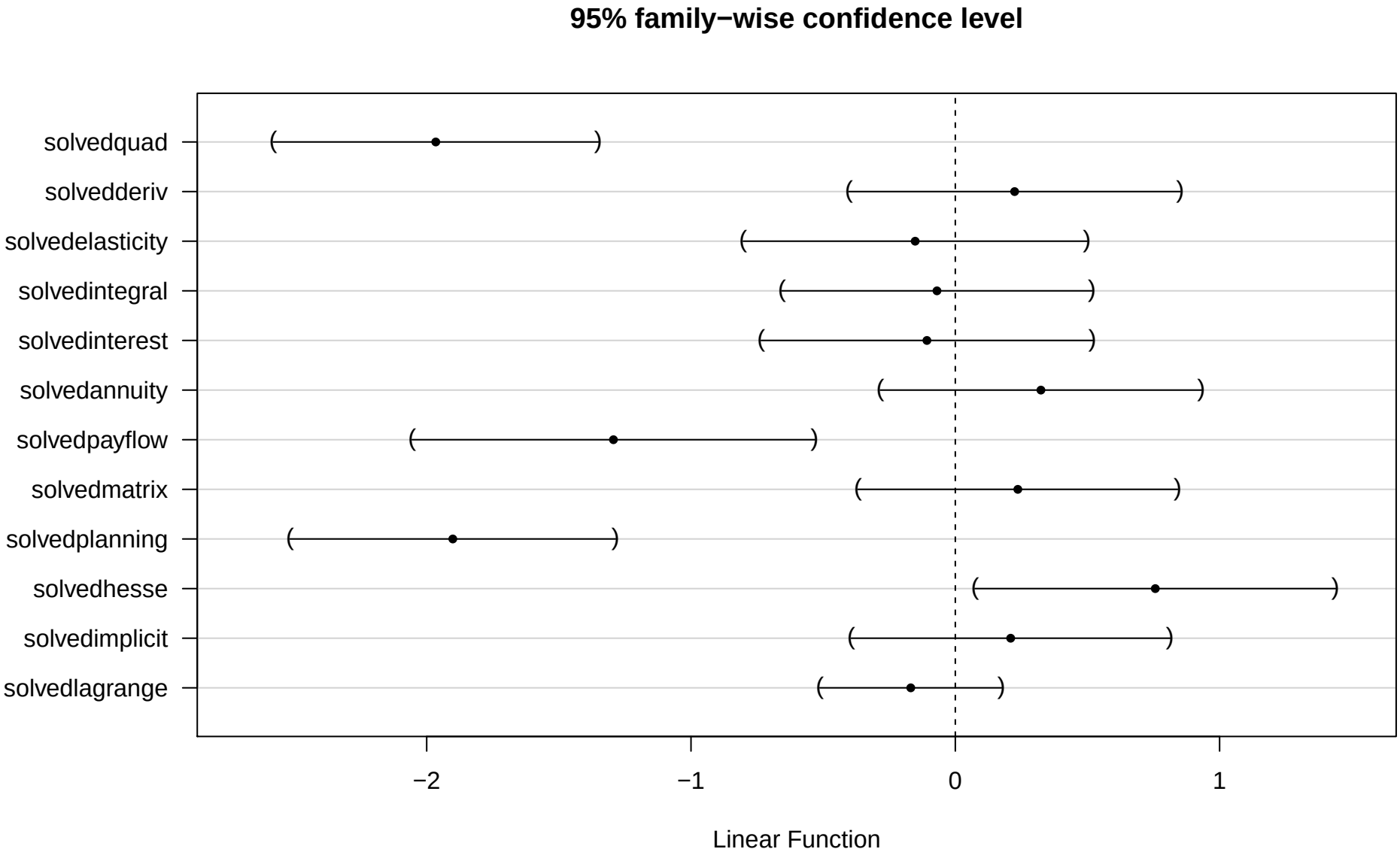
```
R> at1 <- anchortest(solved ~ group, data = mex, select = c(10, 13))
R> at1

Anchor items:
[1] 10 13

Anchored item parameters:
      solvedquad_1      solvedderiv_1 solvedelasticity_1
        -1.69396506         -1.46526218         -1.94031996
  solvedintegral_1   solvedinterest_1    solvedannuity_1
        -0.48379709         -1.63984169         -1.08836162
   solvedpayflow_1     solvedmatrix_1   solvedplanning_1
         0.91949870         -1.08836162         -1.00947248
     solvedhesse_1   solvedimplicit_1   solvedlagrange_1
        -1.71220935         -1.07252012         -0.12303457
      solvedquad_2      solvedderiv_2 solvedelasticity_2
         0.27162967         -1.68965916         -1.78848532
  solvedintegral_2   solvedinterest_2    solvedannuity_2
        -0.41440939         -1.53242411         -1.41227186
   solvedpayflow_2     solvedmatrix_2   solvedplanning_2
         2.21290599         -1.32491395          0.89174766
     solvedhesse_2   solvedimplicit_2   solvedlagrange_2
        -2.46893320         -1.28201701          0.04542323
```

# Examining exams

```
[...]
Final DIF tests:

 Simultaneous Tests for General Linear Hypotheses

Linear Hypotheses:
                       Estimate Std. Error z value Pr(>|z|)
solvedquad == 0       -1.96559    0.21708  -9.055  < 2e-16 ***
solvedderiv == 0       0.22440    0.22129   1.014  0.31057
solvedelasticity == 0 -0.15183    0.22962  -0.661  0.50845
solvedintegral == 0   -0.06939    0.20709  -0.335  0.73758
solvedinterest == 0   -0.10742    0.22125  -0.486  0.62731
solvedannuity == 0     0.32391    0.21443   1.511  0.13090
solvedpayflow == 0    -1.29341    0.26881  -4.812  1.5e-06 ***
solvedmatrix == 0      0.23655    0.21347   1.108  0.26781
solvedplanning == 0   -1.90122    0.21738  -8.746  < 2e-16 ***
solvedhesse == 0       0.75672    0.24081   3.142  0.00168 **
solvedimplicit == 0    0.20950    0.21294   0.984  0.32520
solvedlagrange == 0   -0.16846    0.12130  -1.389  0.16490
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Univariate p values reported)
```

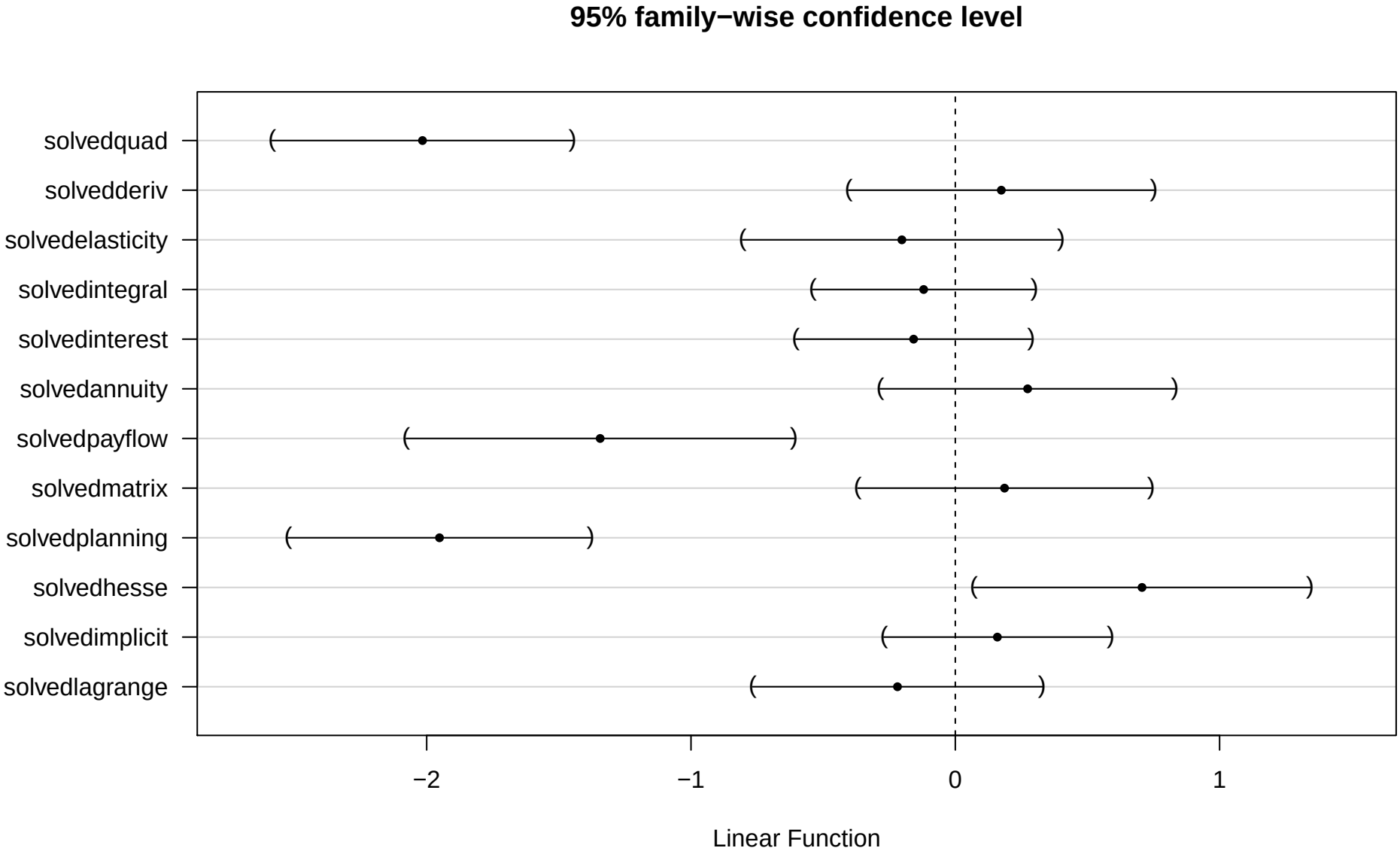
# Examining exams

```
R> plot(at1$final_tests)
```



# Examining exams

```
R> at2 <- anchortest(solved ~ group, data = mex, adj = "single-step")
R> plot(at2$final_tests)
```

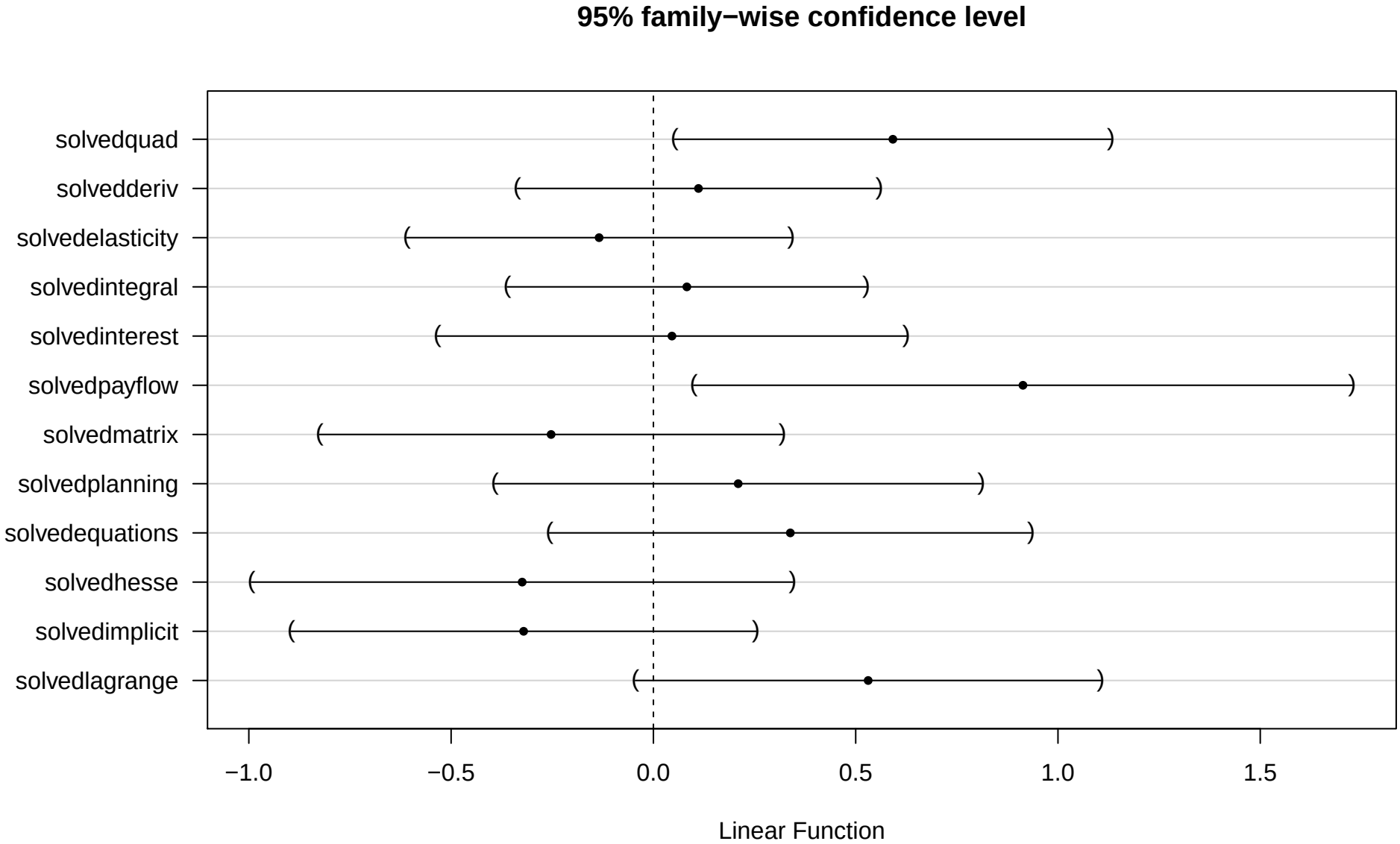


# Examining exams

```
R> at2

Anchor items:
[1] 10  4 12  5

[...]

Ranking order:
 [1] 10  4 12  5  6  2  8 13  3  9  1  7 11

Criterion values (not sorted):
 [1] -1 -4 -3 -4 -4 -4  0 -4 -1 -6  0 -4 -3

[...]

(Adjusted p values reported -- single-step method)
```

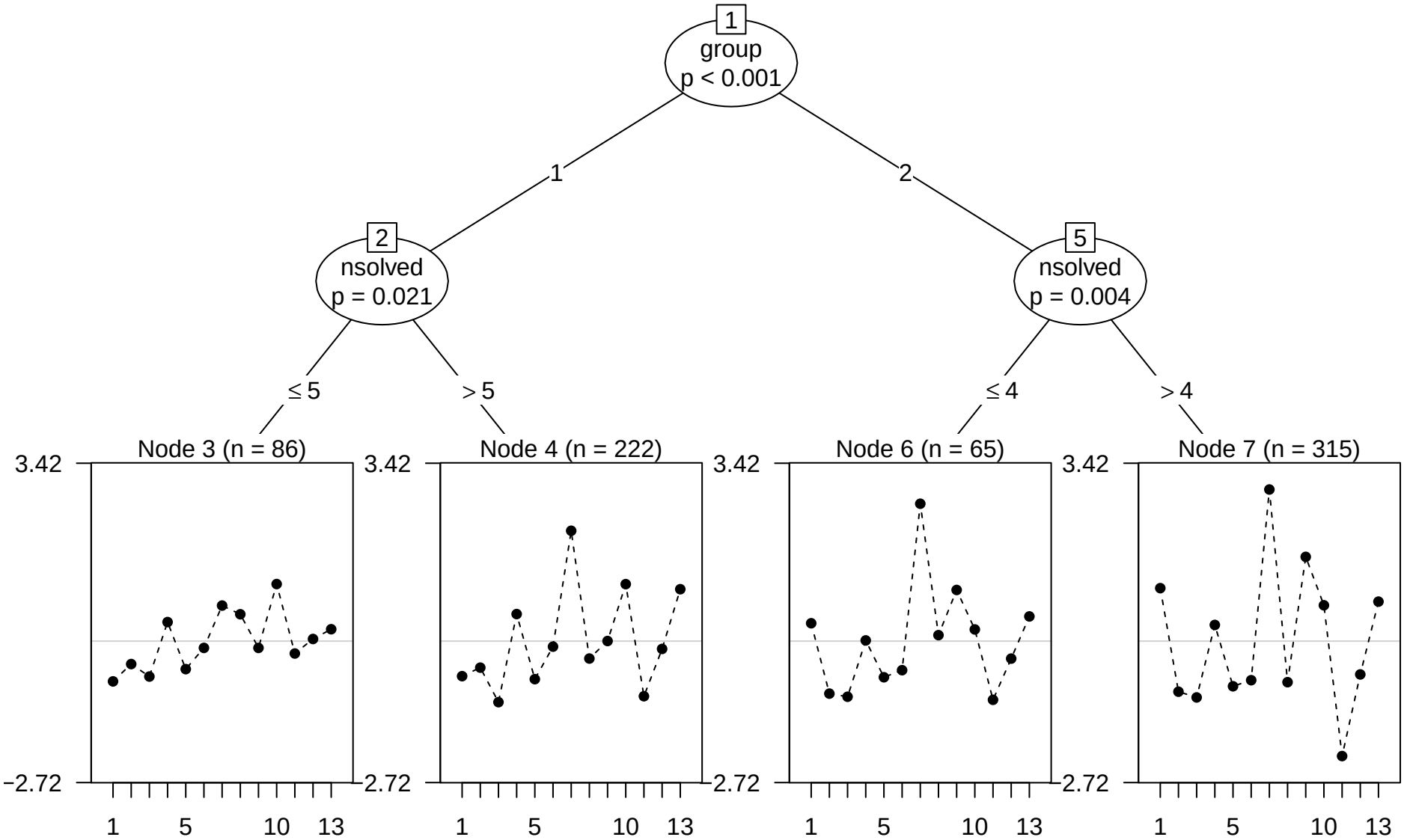
# Examining exams

```
R> at3 <- anchortest(solved ~ factor(nsolved < 7), data = mex,
+    adjust = "single-step")
```



# Examining exams

```
R> rt <- raschtree(solved ~ group + tests + attempt + nsolved + study,
+    data = mex, minsize = 50)
```



# Examining exams

**Assess DIF:** Mixture models with 1, 2, 3 clusters and score model restricted across clusters.

```
R> set.seed(1)
R> ramm <- raschmix(mex$solved, k = 1:3, scores = "meanvar",
+    restricted = TRUE)
```

**Model selection:** Some evidence for DIF as BIC selects 2 clusters.

```
R> BIC(ramm)
```

```
       1        2        3
10621.18 10618.88 10656.90
```

**Refit:** Given 2 clusters, refit the model with unrestricted score model.

```
R> ramm2 <- raschmix(mex$solved, k = 2, scores = "meanvar")
```

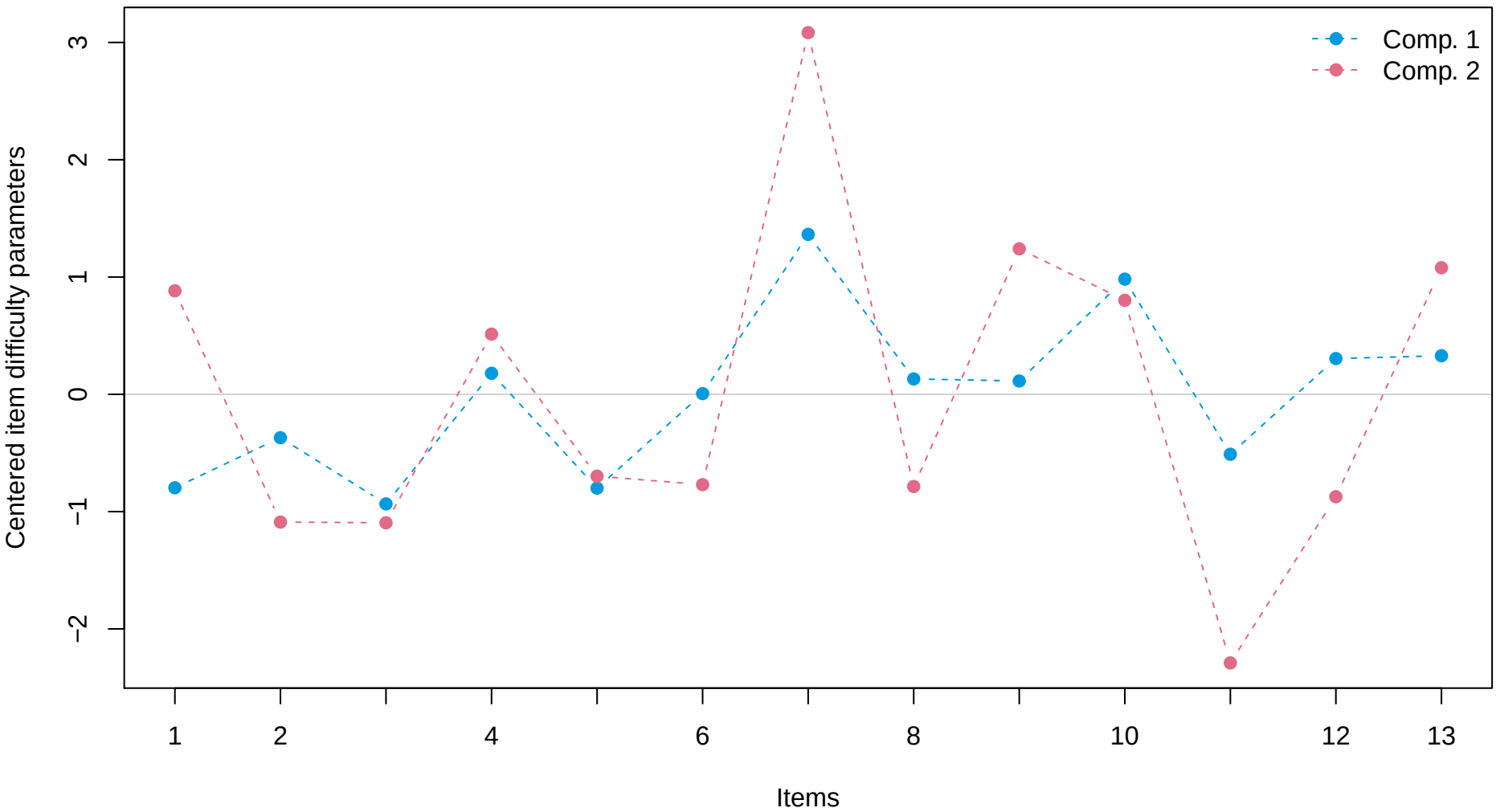
**Model selection:** Some further improvement.

```
R> BIC(ramm2)
```

```
[1] 10597.99
```

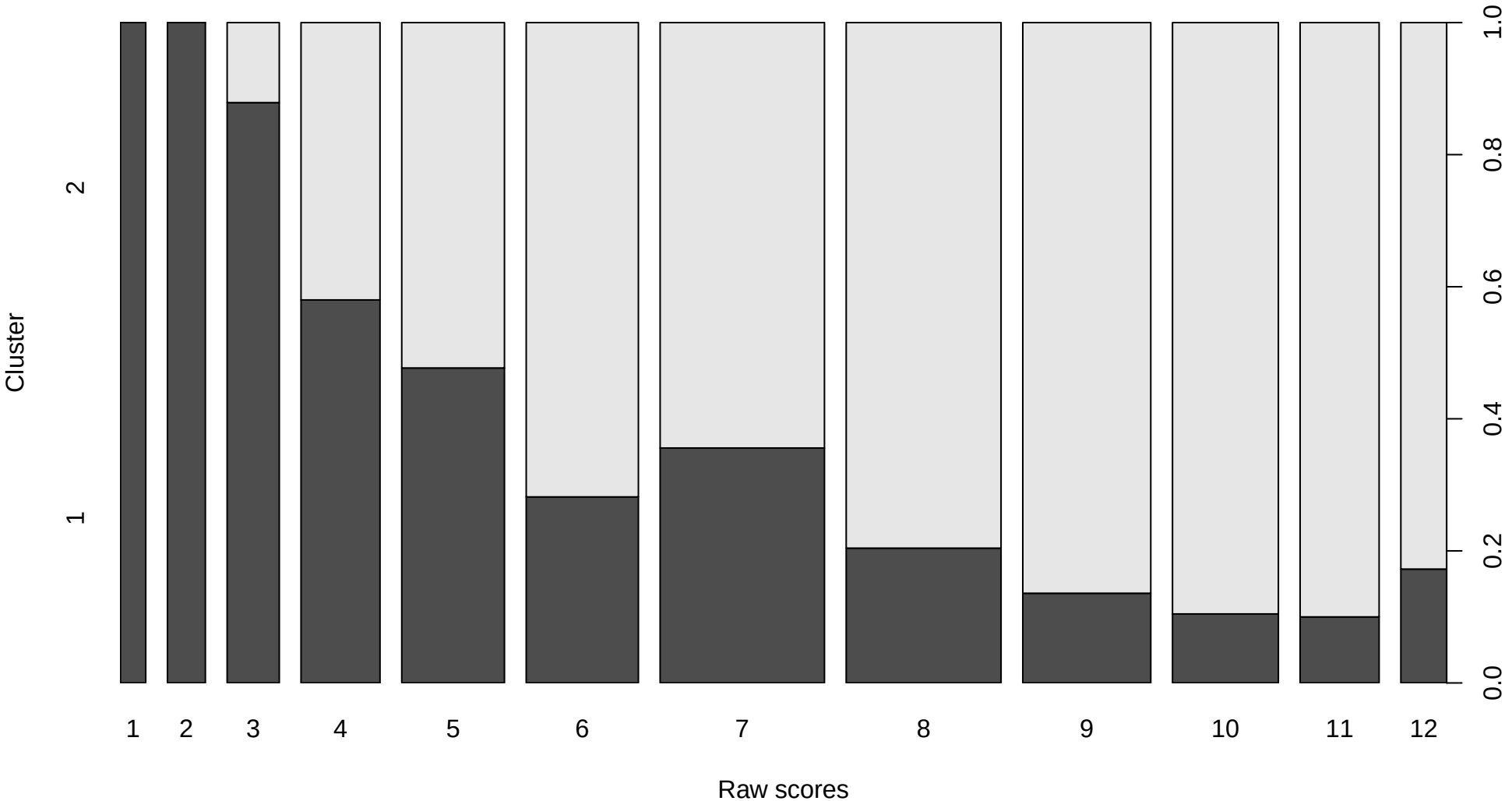
# Examining exams

```
R> plot(ramm2)
```



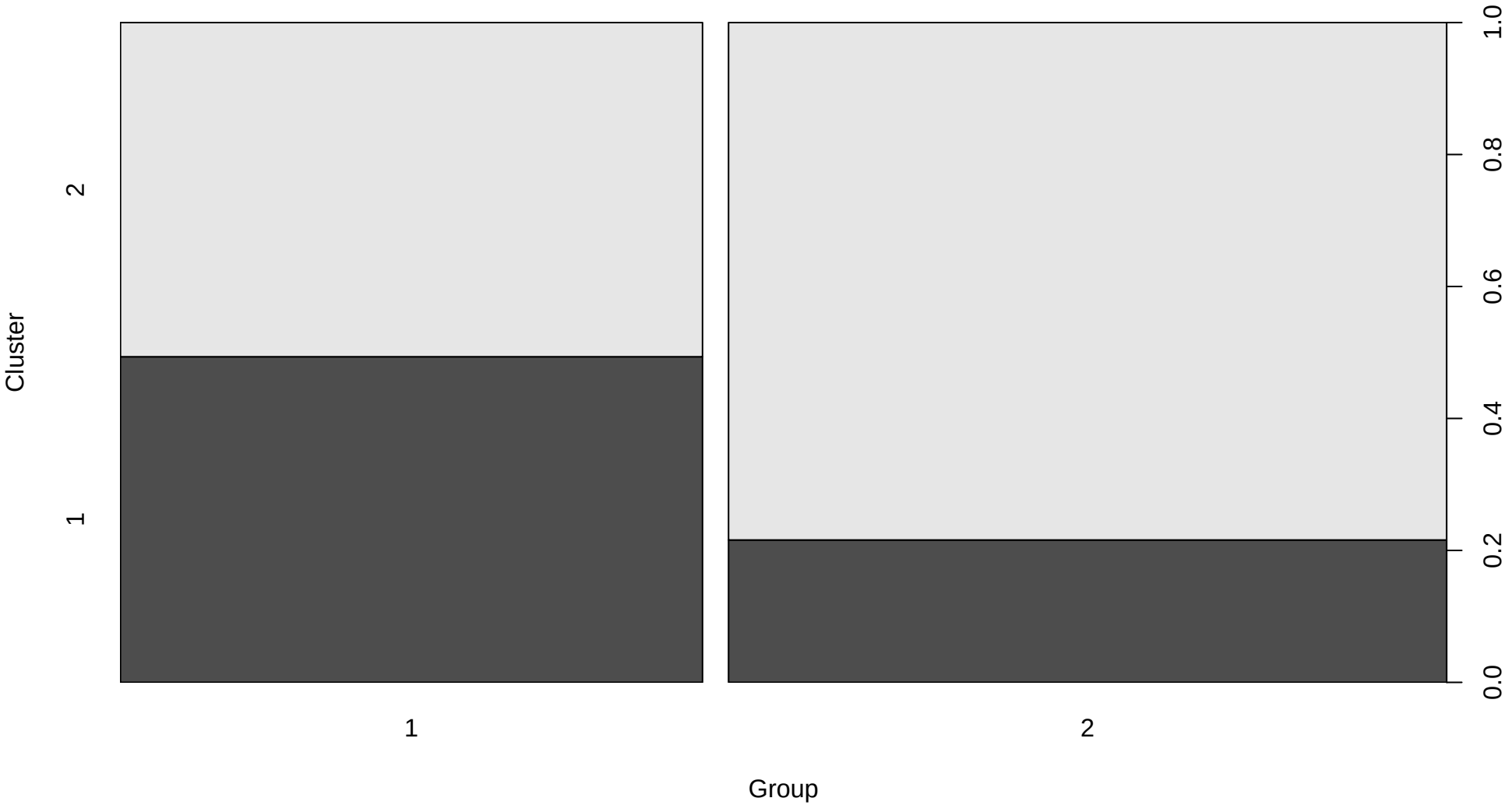
# Examining exams

```
R> plot(factor(clusters(ramm2)) ~ factor(nsolved), data = mex,
+    xlab = "Raw scores", ylab = "Cluster")
```



# Examining exams

```
R> plot(factor(clusters(ramm2)) ~ group, data = mex,
+    xlab = "Group", ylab = "Cluster")
```



# References


Abou El-Komboz B, Zeileis A, Strobl C (2014). "Detecting Differential Item and Step Functioning with Rating Scale and Partial Credit Trees." *Technical Report 152*, Department of Statistics, Ludwig-Maximilians-Universität München. http://epub.ub.uni-muenchen.de/17984/

Frick H, Strobl C, Leisch F, Zeileis A (2012). "Flexible Rasch Mixture Models with Package psychomix." *Journal of Statistical Software*, **48**(7), 1–25. http://www.jstatsoft.org/v48/i07/

Frick H, Strobl C, Zeileis A (2015). "Rasch Mixture Models for DIF Detection: A Comparison of Old and New Score Specifications." *Educational and Psychological Measurement*. Forthcoming. doi:10.1177/0013164414536183

Kopf J, Zeileis A, Strobl C (2015). "A Framework for Anchor Methods and an Iterative Forward Approach for DIF Detection." *Applied Psychological Measurement*, **39**(2), 83–103. doi:10.1177/0146621614544195

Kopf J, Zeileis A, Strobl C (2015). "Anchor Selection Strategies for DIF Analysis: Review, Assessment, and New Approaches." *Educational and Psychological Measurement*, **75**(1), 22–56. doi:10.1177/0013164414529792

Merkle EC, Zeileis A (2013). "Tests of Measurement Invariance without Subgroups: A Generalization of Classical Methods." *Psychometrika*, **78**(1), 59–82. doi:10.1007/s11336-012-9302-4

Strobl C, Wickelmaier F, Zeileis A (2011). "Accounting for Individual Differences in Bradley-Terry Models by Means of Recursive Partitioning." *Journal of Educational and Behavioral Statistics*, **36**(2), 135–153. doi:10.3102/1076998609359791

Strobl C, Julia Kopf, Zeileis A (2015). "Rasch Trees: A New Method for Detecting Differential Item Functioning in the Rasch Model." *Psychometrika*. Forthcoming. doi:10.1007/s11336-013-9388-3

Zeileis A, Umlauf N, Leisch F (2014). "Flexible Generation of E-Learning Exams in R: Moodle Quizzes, OLAT Assessments, and Beyond." *Journal of Statistical Software*, **58**(1), 1–36. http://www.jstatsoft.org/v58/i01/