

Code ▾

Golang Assessment

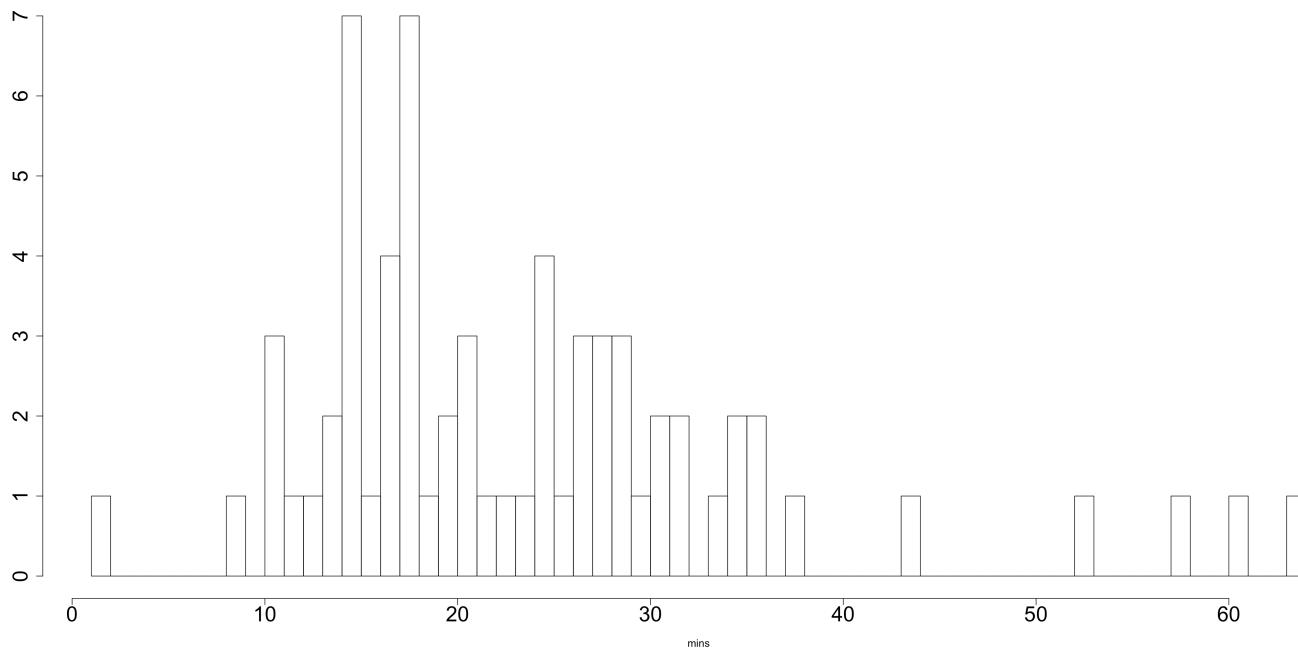
Calibration, Scoring, & Impact Analyses

Number of Questions Attempted

| 1 | 8 | 30 | NA |
|---|---|----|----|
| 2 | 1 | 67 | 0 |

Time spent on Golang Quiz

Golang 30-Question Quiz - Time to Complete (mins)



| n | min | q1 | avg | med | q3 | max |
|-------|------|--------|----------|------|--------|------|
| 67 | 1.78 | 15.865 | 23.79746 | 20.6 | 28.435 | 63.4 |
| 1 row | | | | | | |

Time spent on Quiz, by Claim

Time spent on Quiz, by Question

Analytic Sample

Exclusion Criteria:

- test/fake accounts
- attempted all questions
- total time to complete > 15min (30s per question)

The proportion of candidates who completed the Golang Quiz of those who attempted it is 0.957.

The proportion of candidates who have a valid attempted is 0.729 of those who attempted the quiz.

Classical Test Theory Analysis

Item mean is an indicator of item difficulty. Items with extremely low or high means are indicators of poorly performing items as they are overly difficult or easy, respectively, for test takers. (parallels IRT item difficulty parameter)

Item-total correlation is the correlation between a score on a particular item and the performance on the rest of the test. High item-total correlations would indicate that test takers who score well on the overall test generally also performed well on the individual item. Low or negative item-total correlations are indicators of poor item performance as they suggest that test takers who score well on the overall test score lower on the individual item. This kind of item performance would suggest that the item may be measuring something different than the other items on the test or the item may be keyed in the wrong direction. (parallels IRT item discrimination parameter)

Alpha-deleted is a measure of the test reliability (internal consistency) - a measure of domain sampling and the impact of a flawed items if a particular item is removed. Alpha gives an estimate of average inter-item correlation among the items. An indicator of poor item performance is if once an item is removed then the overall test reliability is greatly improved.

Items are flagged to be reviewed or removed if:

- *Item means* are outside of the range 0.20 and 0.80 - indicating really easy or difficult items
- *Item-total correlations* are less than 0.20; the larger the better - indicating may be measuring something different from the rest of the test
- *Alpha-deleted* increase if the item is removed from the test - indicating may be measuring something different from the rest of the test or is adding random noise (unsystematic variation) to the overall test.

Some items (1359 1375) were negatively correlated with the total scale and probably should be reversed.

To do this, run the function again with the 'check.keys=TRUE' option Overall Reliability is: 0.789

| df_blueprint\$question_id | flag | name | target | n_candidates | item_mean | item_total_correlation | alpha_deleted |
|---------------------------|------|------------|------------|--------------|-----------|------------------------|---------------|
| 1356 | | lang_go_01 | variable | 51 | 0.784 | 0.295 | 0.784 |
| 1357 | | lang_go_02 | defer | 51 | 0.765 | 0.252 | 0.785 |
| 1358 | | lang_go_03 | for | 51 | 0.490 | 0.376 | 0.780 |
| 1359 | <= | lang_go_04 | range | 51 | 0.569 | -0.087 | 0.802 |
| 1360 | <= | lang_go_05 | pointers | 51 | 0.529 | 0.155 | 0.791 |
| 1361 | <= | lang_go_06 | map | 51 | 0.961 | 0.297 | 0.786 |
| 1362 | <= | lang_go_07 | structs | 51 | 0.431 | 0.023 | 0.797 |
| 1363 | <= | lang_go_08 | embedding | 51 | 0.843 | 0.408 | 0.780 |
| 1364 | | lang_go_09 | slices | 51 | 0.235 | 0.363 | 0.781 |
| 1365 | | lang_go_10 | exports | 51 | 0.588 | 0.363 | 0.780 |
| 1366 | <= | lang_go_11 | methods | 51 | 0.961 | 0.234 | 0.787 |
| 1367 | | lang_go_12 | goroutines | 51 | 0.745 | 0.261 | 0.785 |
| 1368 | <= | lang_go_13 | errors | 51 | 0.451 | 0.119 | 0.793 |
| 1369 | <= | lang_go_14 | channels | 51 | 0.902 | 0.497 | 0.778 |
| 1370 | | lang_go_15 | interfaces | 51 | 0.725 | 0.481 | 0.775 |
| 1371 | <= | lang_go_16 | json | 51 | 0.882 | 0.450 | 0.779 |
| 1372 | | lang_go_17 | http | 51 | 0.745 | 0.466 | 0.776 |
| 1373 | | lang_go_18 | testing | 51 | 0.627 | 0.433 | 0.777 |
| 1374 | <= | lang_go_19 | reflect | 51 | 0.627 | 0.196 | 0.789 |

| df_blueprint\$question_id | flag | name | target | n_candidates | item_mean | item_total_correlation | alpha_deleted |
|---------------------------|------|------------|------------|--------------|-----------|------------------------|---------------|
| 1375 | <= | lang_go_20 | files | 51 | 0.412 | -0.055 | 0.801 |
| 1376 | <= | lang_go_21 | errors | 51 | 0.196 | 0.202 | 0.787 |
| 1377 | | lang_go_22 | gorm | 51 | 0.451 | 0.355 | 0.781 |
| 1378 | | lang_go_23 | mod | 51 | 0.608 | 0.331 | 0.782 |
| 1379 | | lang_go_24 | gin | 51 | 0.549 | 0.307 | 0.783 |
| 1380 | | lang_go_25 | logrus | 51 | 0.745 | 0.516 | 0.774 |
| 1381 | | lang_go_26 | defer-bug | 51 | 0.608 | 0.518 | 0.772 |
| 1382 | | lang_go_27 | waitgroups | 51 | 0.784 | 0.274 | 0.784 |
| 1383 | | lang_go_28 | time | 51 | 0.745 | 0.396 | 0.779 |
| 1384 | | lang_go_29 | jsontags | 51 | 0.373 | 0.547 | 0.771 |
| 1385 | | lang_go_30 | capitalize | 51 | 0.784 | 0.274 | 0.784 |

IRT Analysis

Unidimensional Model

IRT Item Parameters

| | a | b | g | u |
|------|-------|--------|---|---|
| 1358 | 0.686 | 0.302 | 0 | 1 |
| 1365 | 0.740 | -0.265 | 0 | 1 |
| 1361 | 0.640 | -3.767 | 0 | 1 |
| 1359 | 0.180 | -1.201 | 0 | 1 |
| 1368 | 0.287 | 0.883 | 0 | 1 |
| 1366 | 0.512 | -4.639 | 0 | 1 |
| 1374 | 0.400 | -1.022 | 0 | 1 |
| 1357 | 0.631 | -1.621 | 0 | 1 |
| 1383 | 0.890 | -1.059 | 0 | 1 |
| 1379 | 0.607 | -0.090 | 0 | 1 |
| 1370 | 1.298 | -0.675 | 0 | 1 |
| 1371 | 0.977 | -1.882 | 0 | 1 |
| 1384 | 1.753 | 0.684 | 0 | 1 |
| 1375 | 0.185 | 2.036 | 0 | 1 |
| 1385 | 0.636 | -1.775 | 0 | 1 |
| 1363 | 0.956 | -1.618 | 0 | 1 |
| 1362 | 0.221 | 1.404 | 0 | 1 |
| 1377 | 0.631 | 0.559 | 0 | 1 |
| 1367 | 0.519 | -1.773 | 0 | 1 |

| | a | b | g | u |
|------|----------|----------|----------|----------|
| 1360 | 0.398 | -0.051 | 0 | 1 |
| 1382 | 0.468 | -2.386 | 0 | 1 |
| 1356 | 0.522 | -2.144 | 0 | 1 |
| 1369 | 1.263 | -1.685 | 0 | 1 |
| 1372 | 1.074 | -0.897 | 0 | 1 |
| 1376 | 0.410 | 3.425 | 0 | 1 |
| 1373 | 1.051 | -0.325 | 0 | 1 |
| 1378 | 0.746 | -0.378 | 0 | 1 |
| 1364 | 0.747 | 1.841 | 0 | 1 |
| 1380 | 1.103 | -0.877 | 0 | 1 |
| 1381 | 1.368 | -0.155 | 0 | 1 |

Item Fit Statistics (S-X2): p-val < 0.05 => generally indicates model misfit

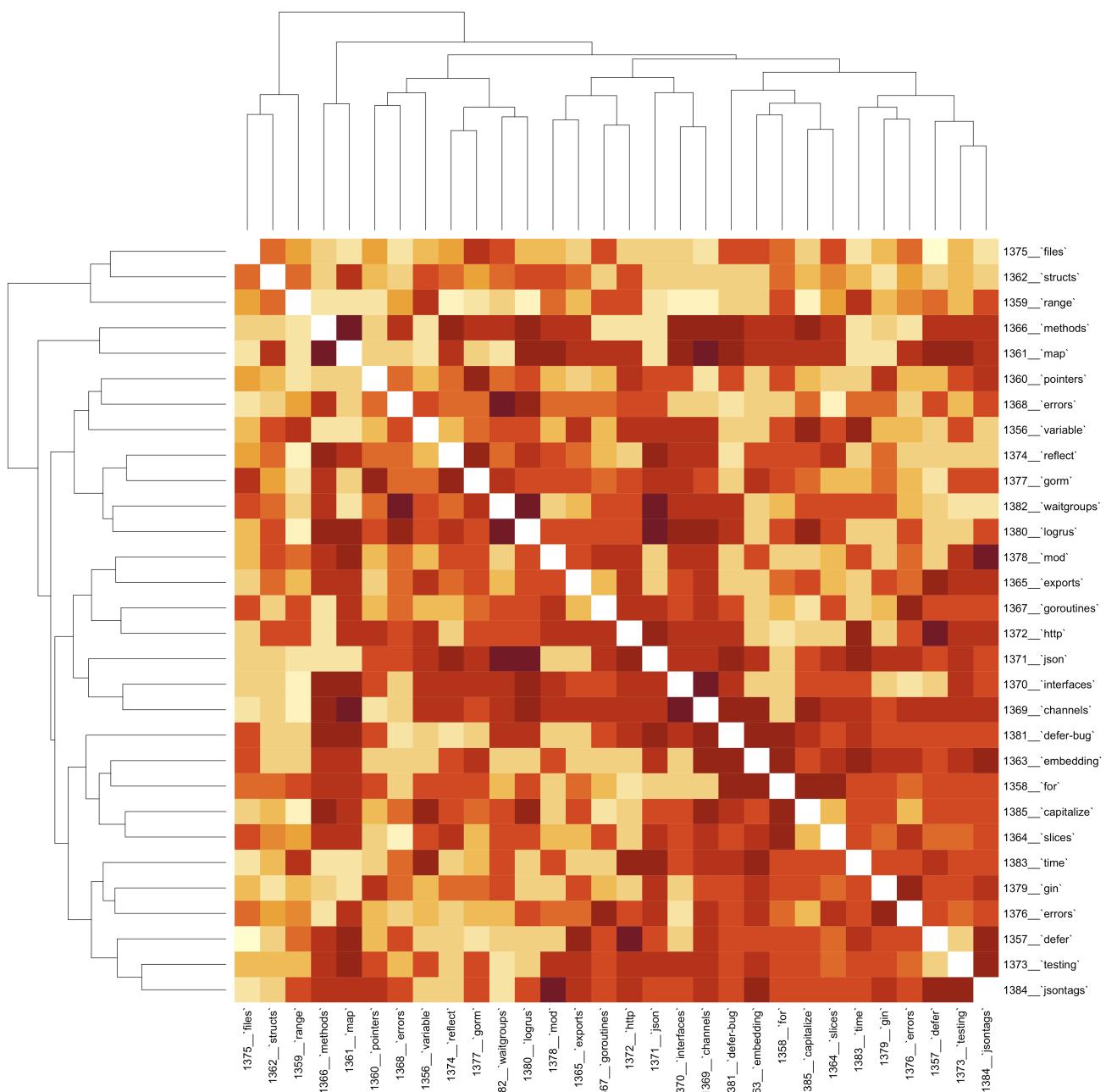
| name | target | item | S_X2 | df.S_X2 | RMSEA.S_X2 | p.S_X2 | flag |
|-------------|---------------|-------------|-------------|----------------|-------------------|---------------|-------------|
| lang_go_03 | for | 1358 | 8.759 | 7 | 0.071 | 0.270 | |
| lang_go_10 | exports | 1365 | 2.685 | 5 | 0.000 | 0.748 | |
| lang_go_06 | map | 1361 | 2.126 | 1 | 0.150 | 0.145 | |
| lang_go_04 | range | 1359 | 12.964 | 9 | 0.094 | 0.164 | |
| lang_go_13 | errors | 1368 | 10.977 | 8 | 0.086 | 0.203 | |
| lang_go_11 | methods | 1366 | NaN | 0 | NaN | NaN NA | |
| lang_go_19 | reflect | 1374 | 2.033 | 5 | 0.000 | 0.845 | |
| lang_go_02 | defer | 1357 | 7.322 | 5 | 0.096 | 0.198 | |
| lang_go_28 | time | 1383 | 1.605 | 4 | 0.000 | 0.808 | |
| lang_go_24 | gin | 1379 | 3.143 | 6 | 0.000 | 0.791 | |
| lang_go_15 | interfaces | 1370 | 1.751 | 4 | 0.000 | 0.781 | |
| lang_go_16 | json | 1371 | 2.211 | 3 | 0.000 | 0.530 | |
| lang_go_29 | jsontags | 1384 | 5.175 | 5 | 0.026 | 0.395 | |
| lang_go_20 | files | 1375 | 6.809 | 8 | 0.000 | 0.557 | |
| lang_go_30 | capitalize | 1385 | 5.911 | 5 | 0.060 | 0.315 | |
| lang_go_08 | embedding | 1363 | 7.042 | 4 | 0.123 | 0.134 | |
| lang_go_07 | structs | 1362 | 8.261 | 8 | 0.026 | 0.408 | |
| lang_go_22 | gorm | 1377 | 15.555 | 8 | 0.137 | 0.049 | <= |
| lang_go_12 | goroutines | 1367 | 8.005 | 6 | 0.082 | 0.238 | |
| lang_go_05 | pointers | 1360 | 9.446 | 8 | 0.060 | 0.306 | |
| lang_go_27 | waitgroups | 1382 | 2.642 | 5 | 0.000 | 0.755 | |

| name | target | item | S_X2 | df.S_X2 | RMSEA.S_X2 | p.S_X2 | flag |
|-------------|---------------|-------------|-------------|----------------|-------------------|---------------|-------------|
| lang_go_01 | variable | 1356 | 3.623 | 5 | 0.000 | 0.605 | |
| lang_go_14 | channels | 1369 | 2.549 | 3 | 0.000 | 0.467 | |
| lang_go_17 | http | 1372 | 1.241 | 4 | 0.000 | 0.871 | |
| lang_go_21 | errors | 1376 | 7.688 | 5 | 0.104 | 0.174 | |
| lang_go_18 | testing | 1373 | 1.288 | 5 | 0.000 | 0.936 | |
| lang_go_23 | mod | 1378 | 3.420 | 5 | 0.000 | 0.636 | |
| lang_go_09 | slices | 1364 | 6.300 | 5 | 0.072 | 0.278 | |
| lang_go_25 | logrus | 1380 | 5.525 | 4 | 0.087 | 0.238 | |
| lang_go_26 | defer-bug | 1381 | 1.651 | 4 | 0.000 | 0.800 | |

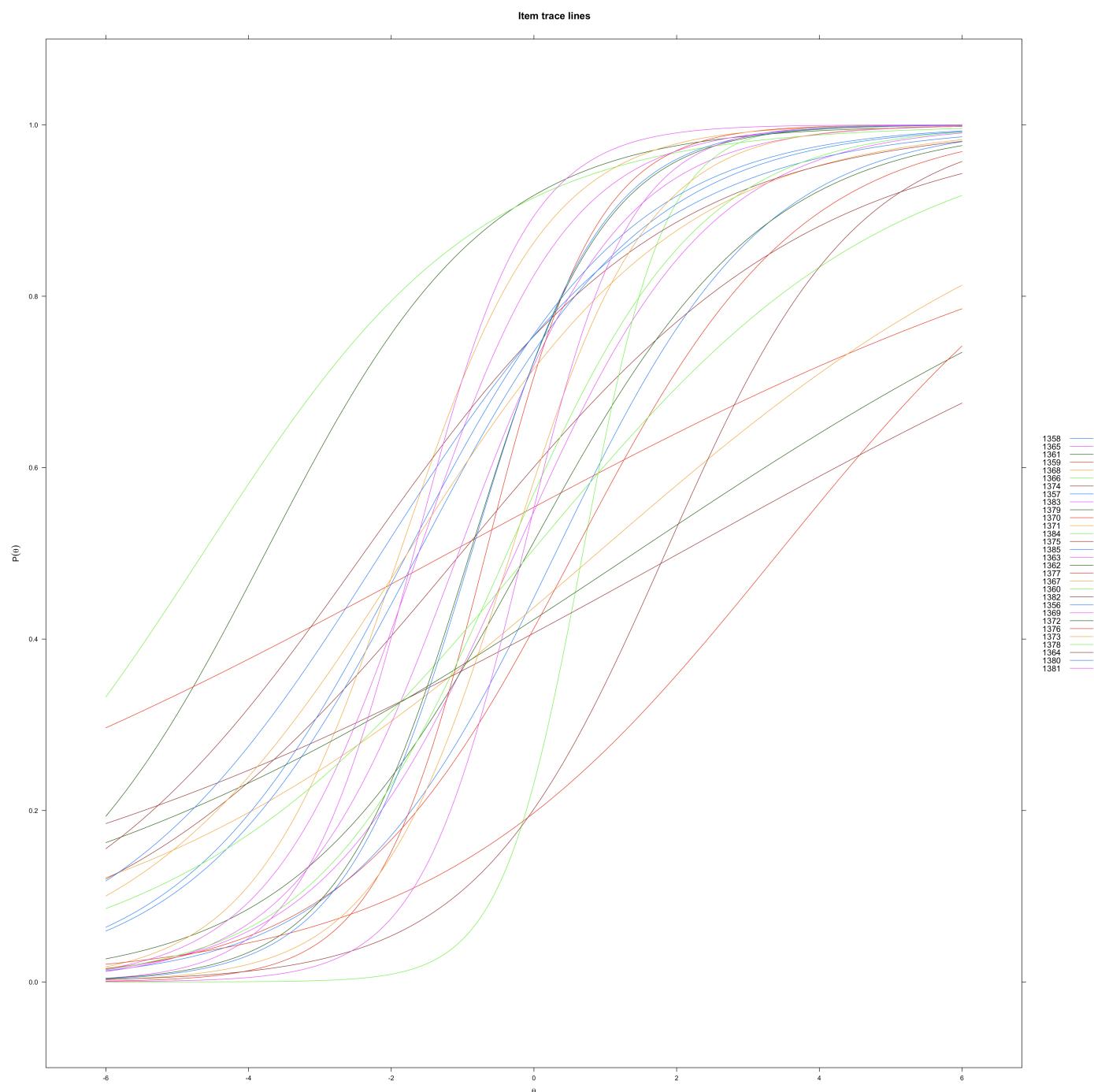
Items are flagged with ‘<=’ to be reviewed or removed if $p\text{-value} < 0.05$. This indicates poor fit of 2PL model.

IRT Plots

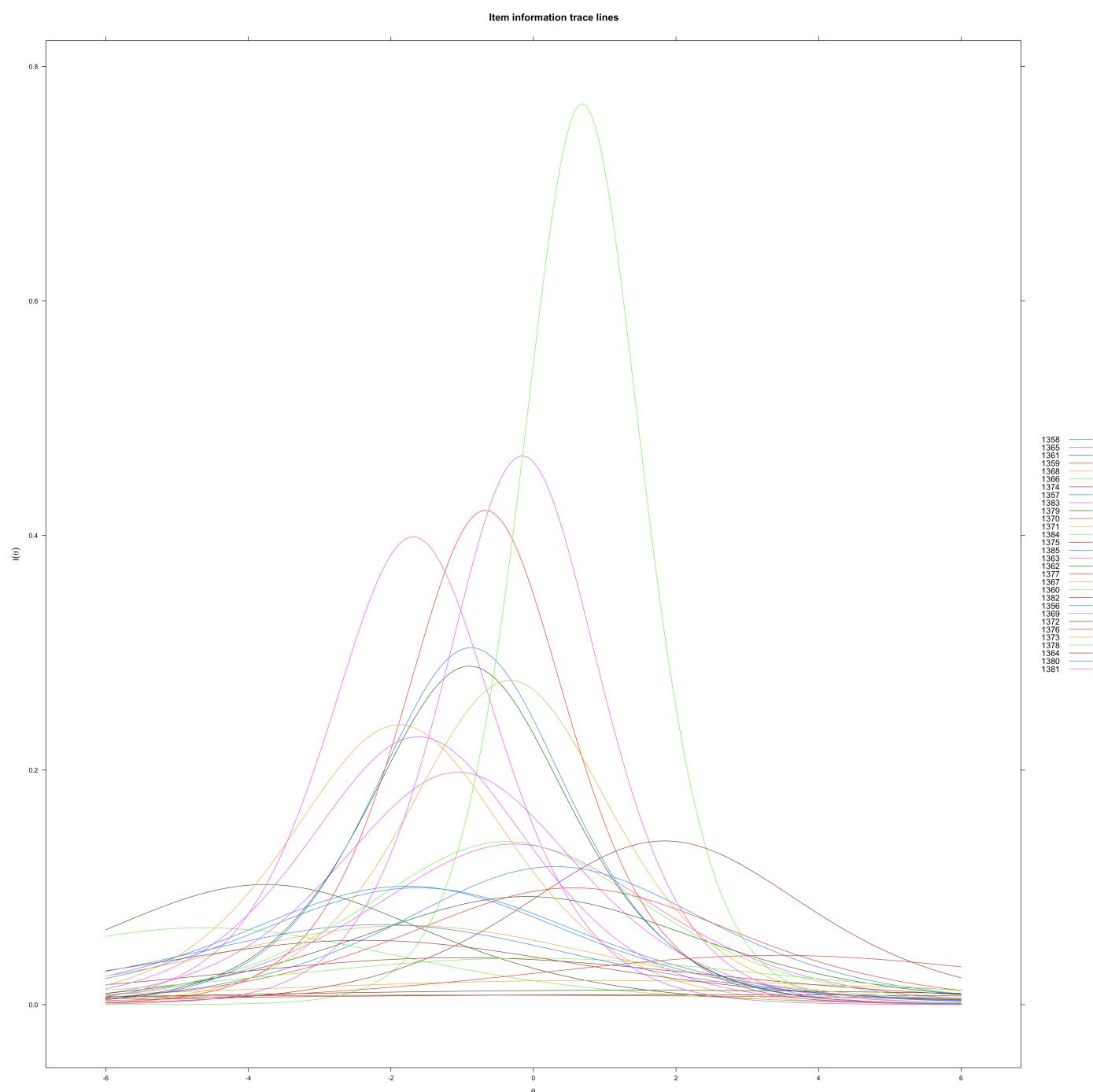
Item Local-Dependence Plots - Residual Dependencies given a unidimensional model:



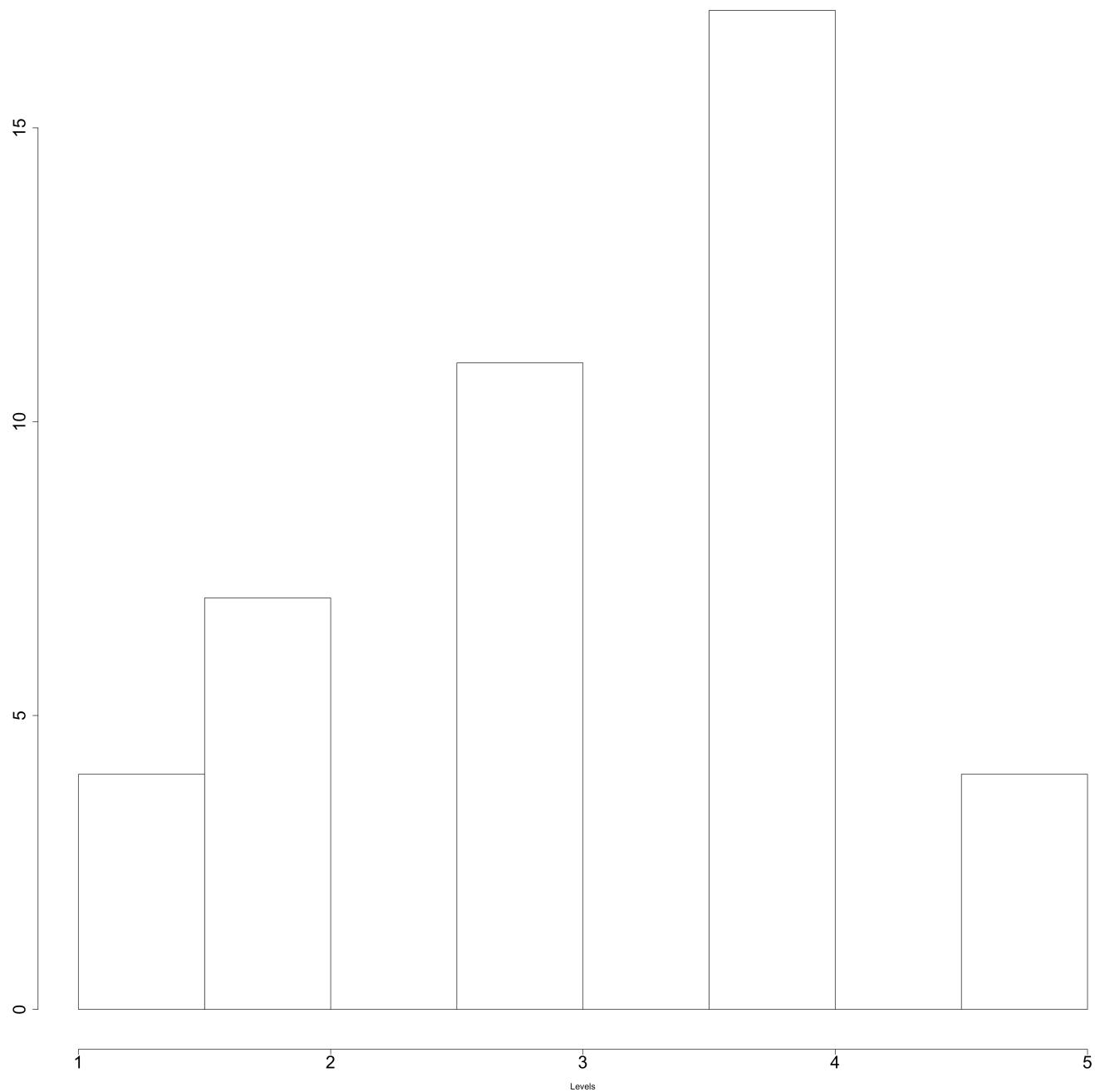
Item Characteristic Curves (Tracelines), by Claims

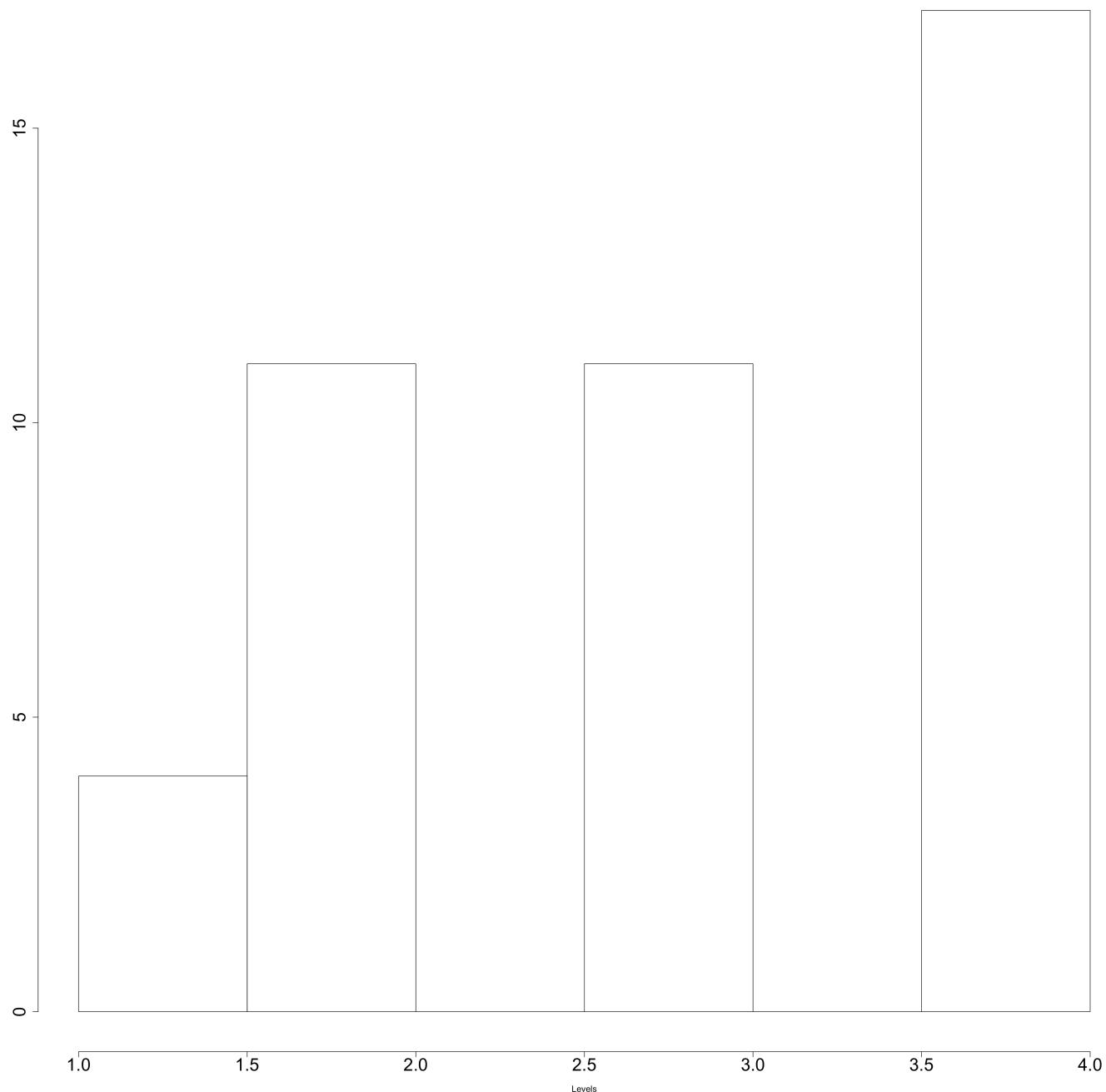


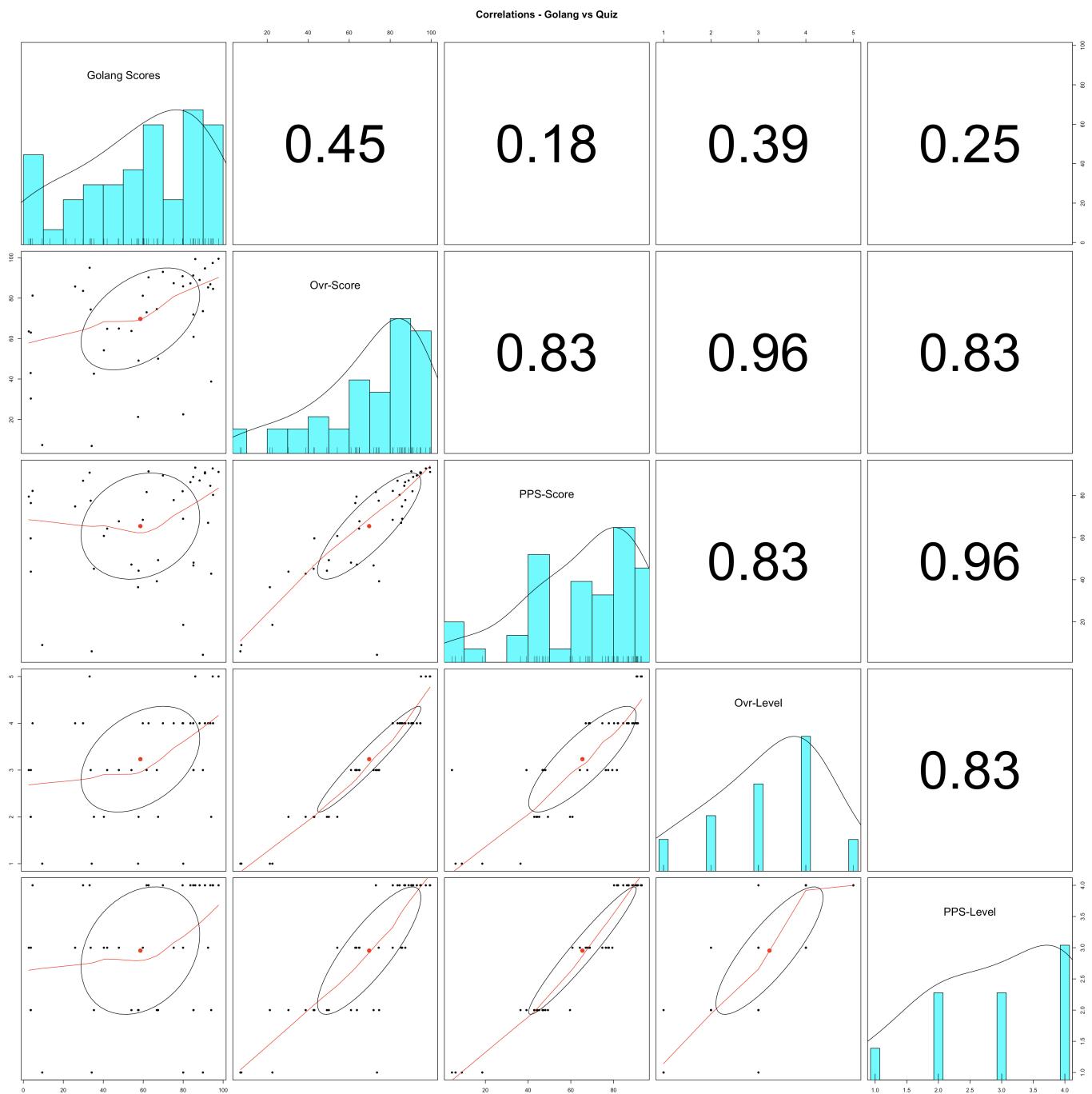
Item Information Curves by Claims



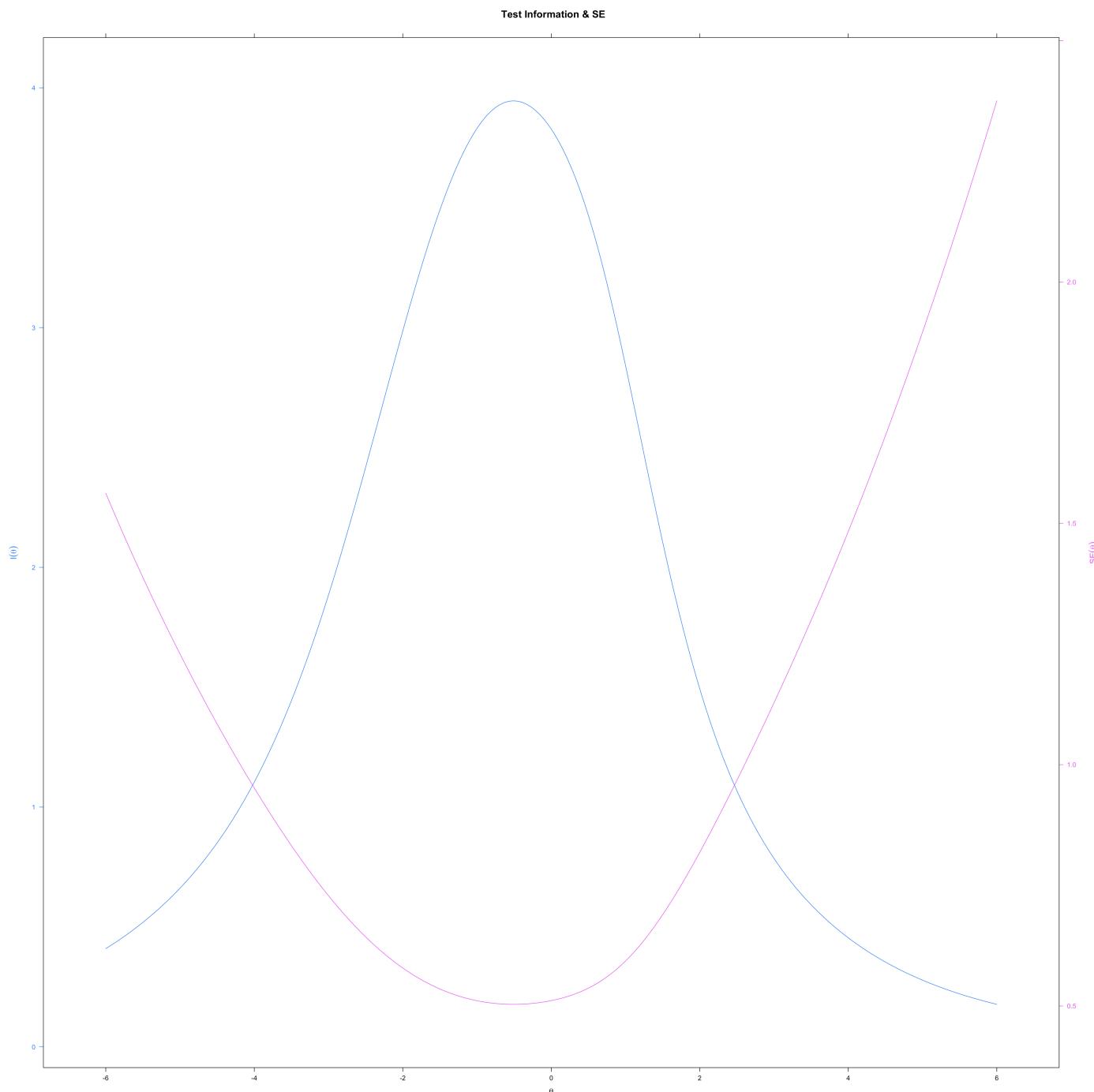
Calibration Sample Characteristics

Generalist - Overall Level

Generalist - Prog Prob Solving Levels**Scoring****Distributions & Correlations*



Avg. SE



Equating/Linking Scores

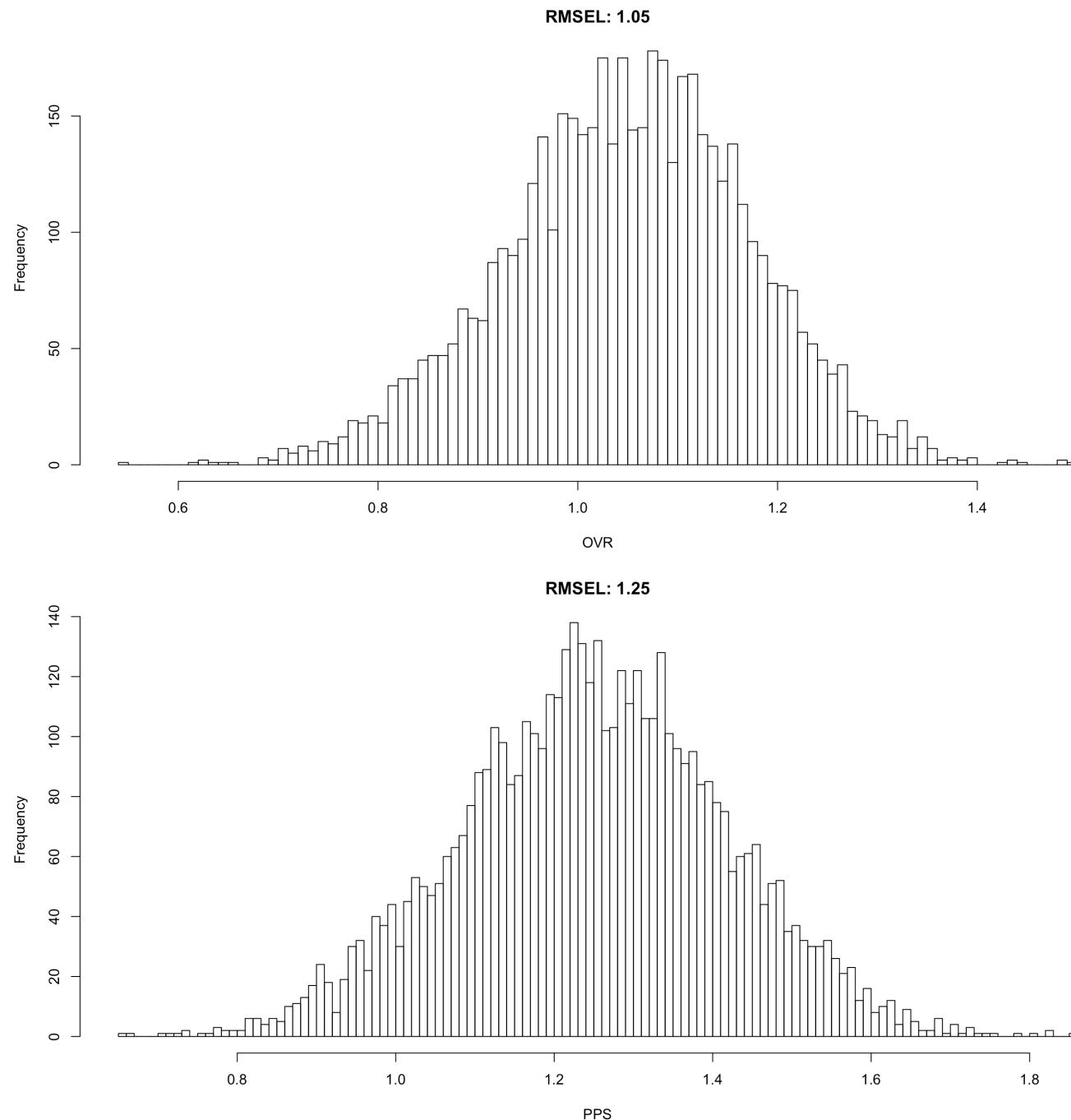
In order to get language assessment SkillEstimates on the same scale as the other quizzes, it requires being able to produce item parameters or IRT-ability estimates on the same scale as the other assessments.

The equating/linking design available is a *Single Group Design* where the same candidates have taken both a previous core quiz and a language assessment. Therefore, there are limited methodologies available to link the two scales together (plus with a limited sample size).

Three approaches explored are:

- Mean Method
- Linear Method
- Equipercentile Method

Mean Method



\$OVR

avg
<dbl>

sd
<dbl>

1.050434

0.1230221

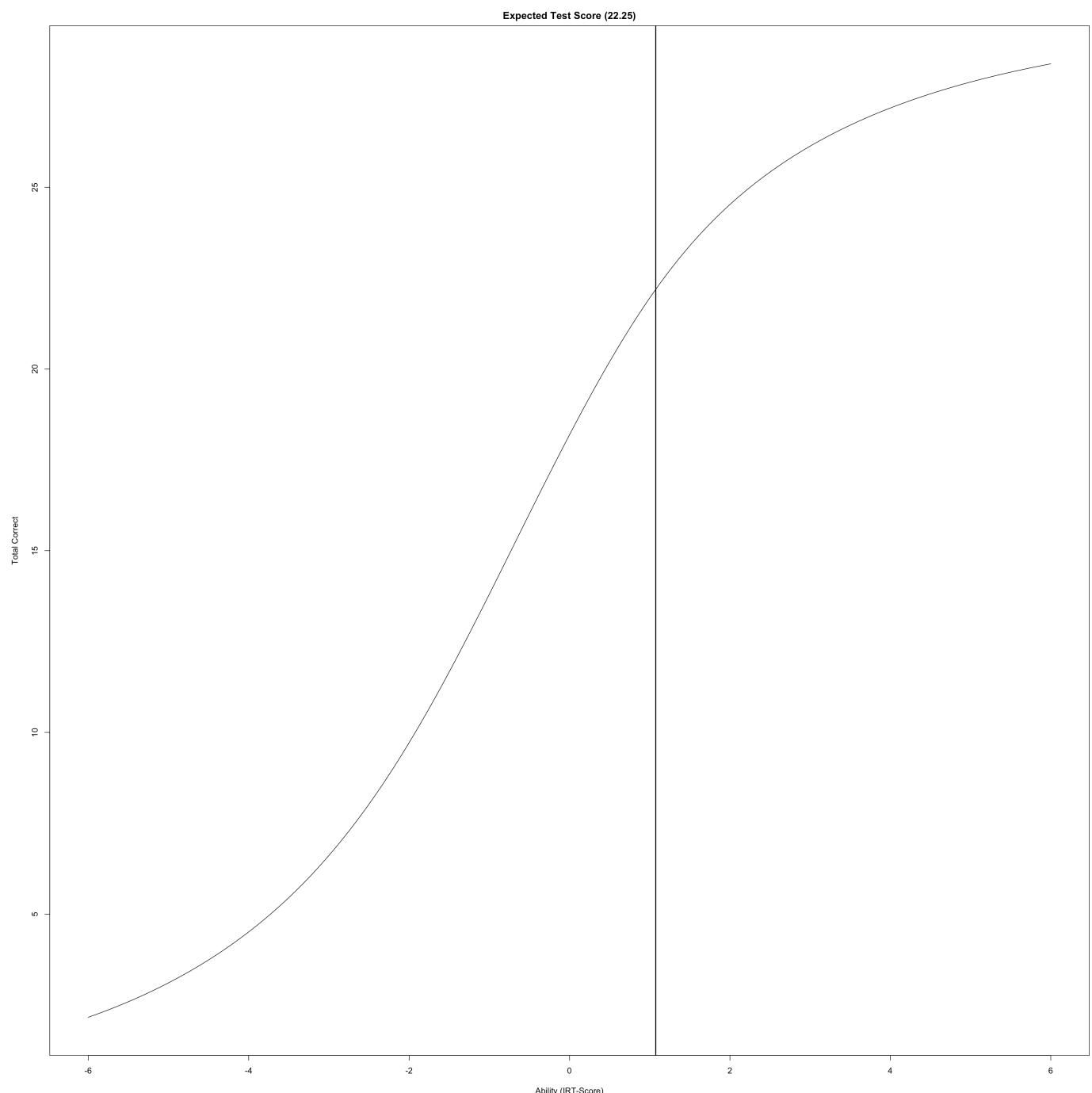
1 row

\$PPS

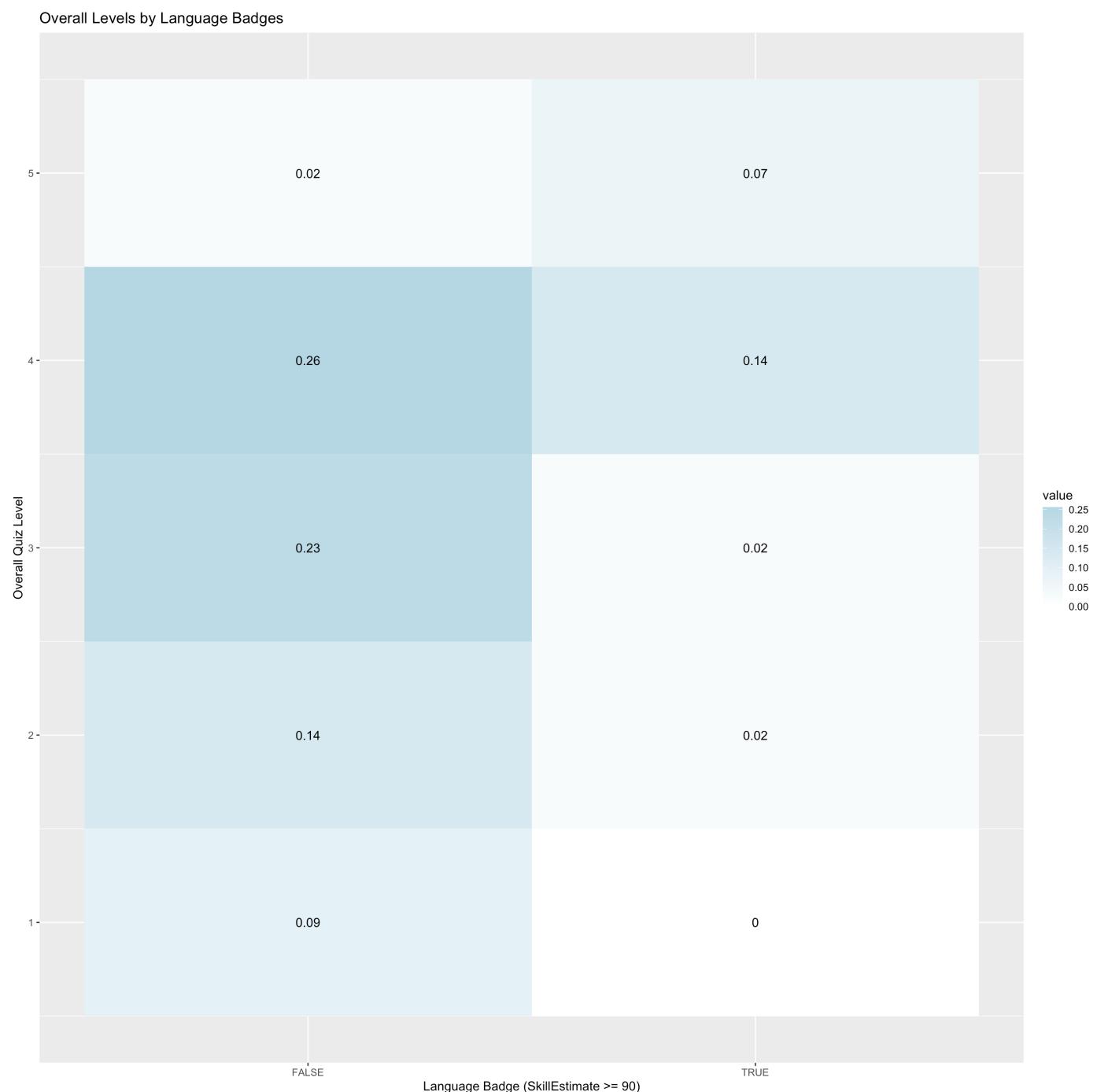
| | avg <dbl> | sd <dbl> |
|-------|--------------|-------------|
| | 1.249153 | 0.1669365 |
| 1 row | | |
| NA | | |

Result: Overall scales scores produce smaller linking error

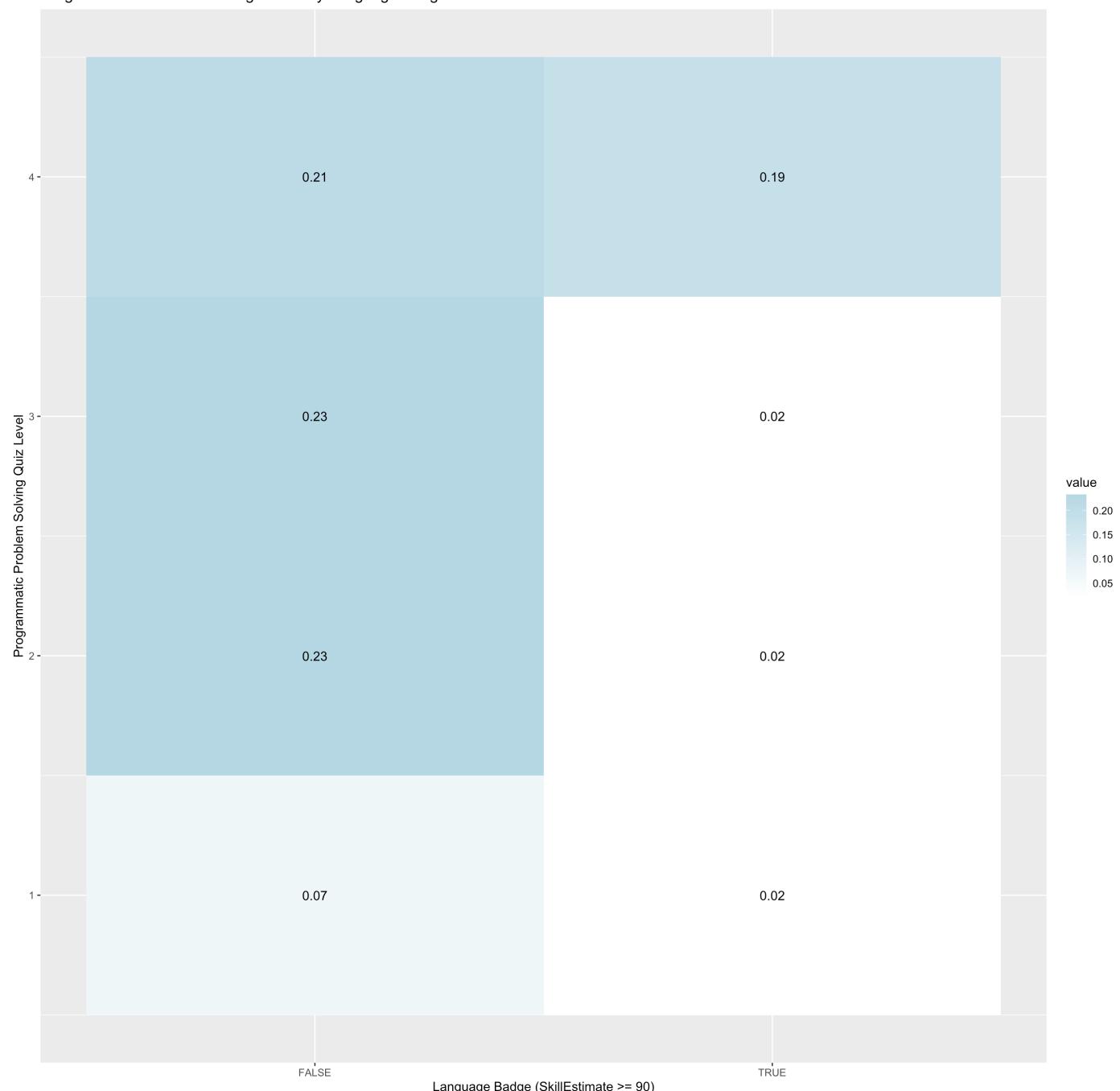
Badges - Test Characteristic Curve



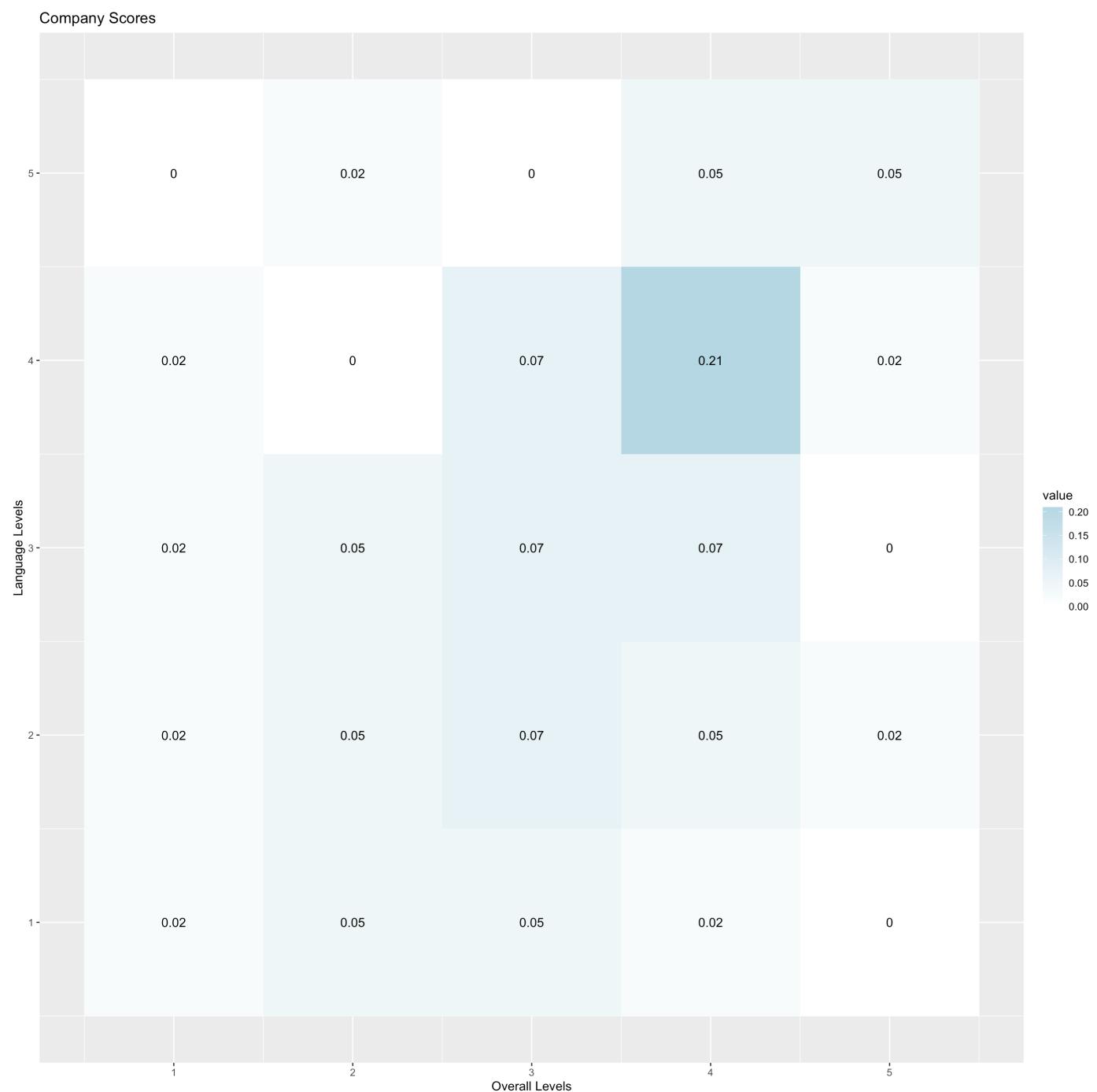
Badges - Quiz Level, by Badge

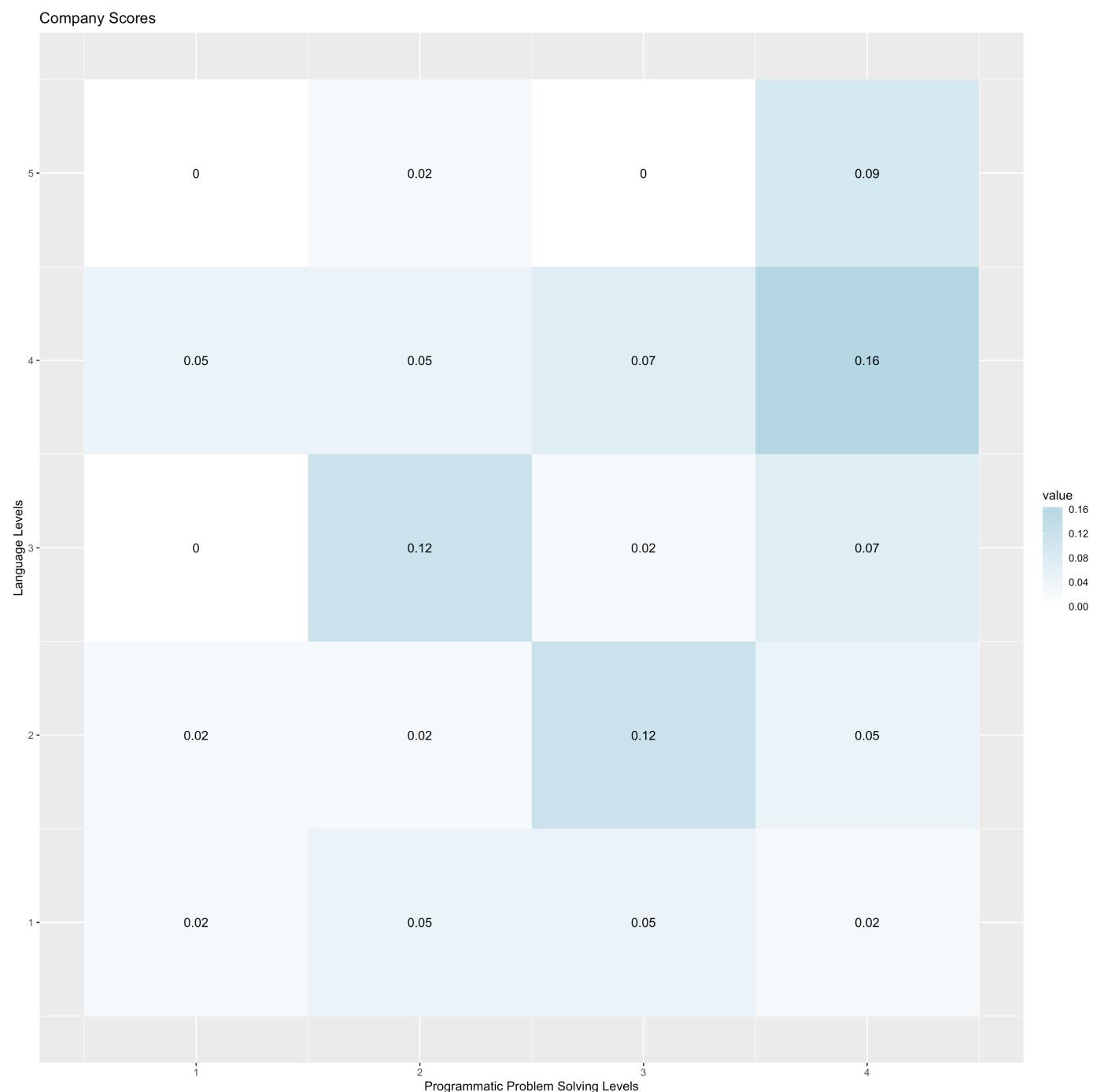


Programmatic Problem Solving Levels by Language Badges

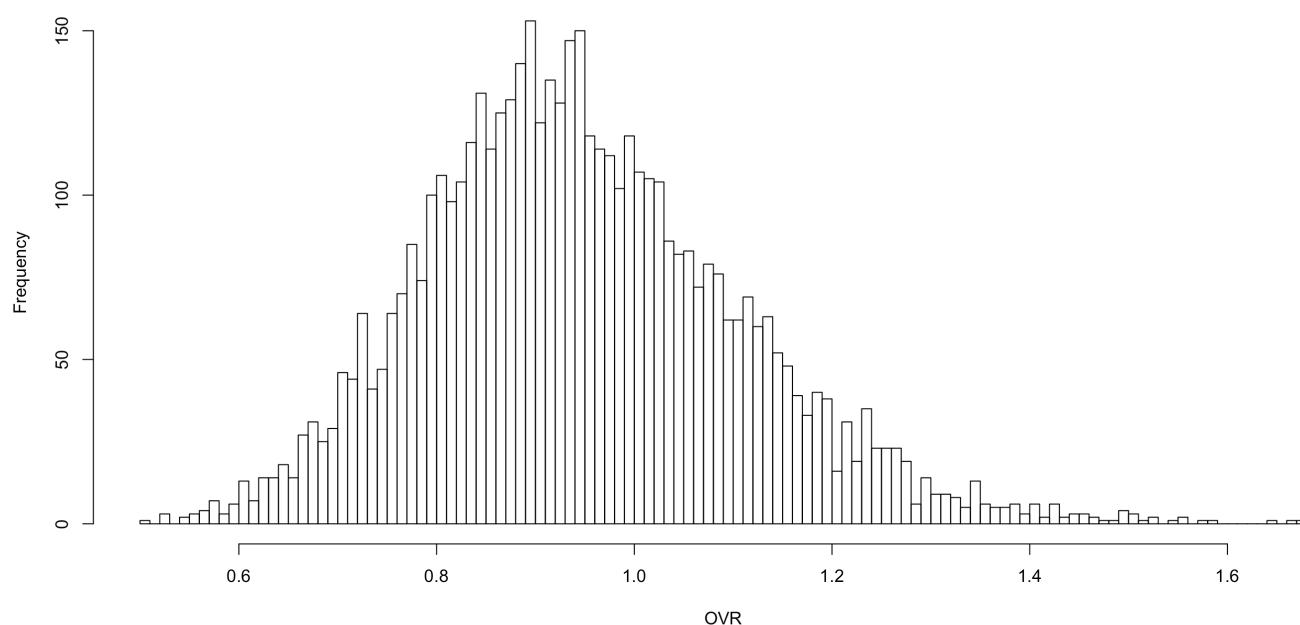
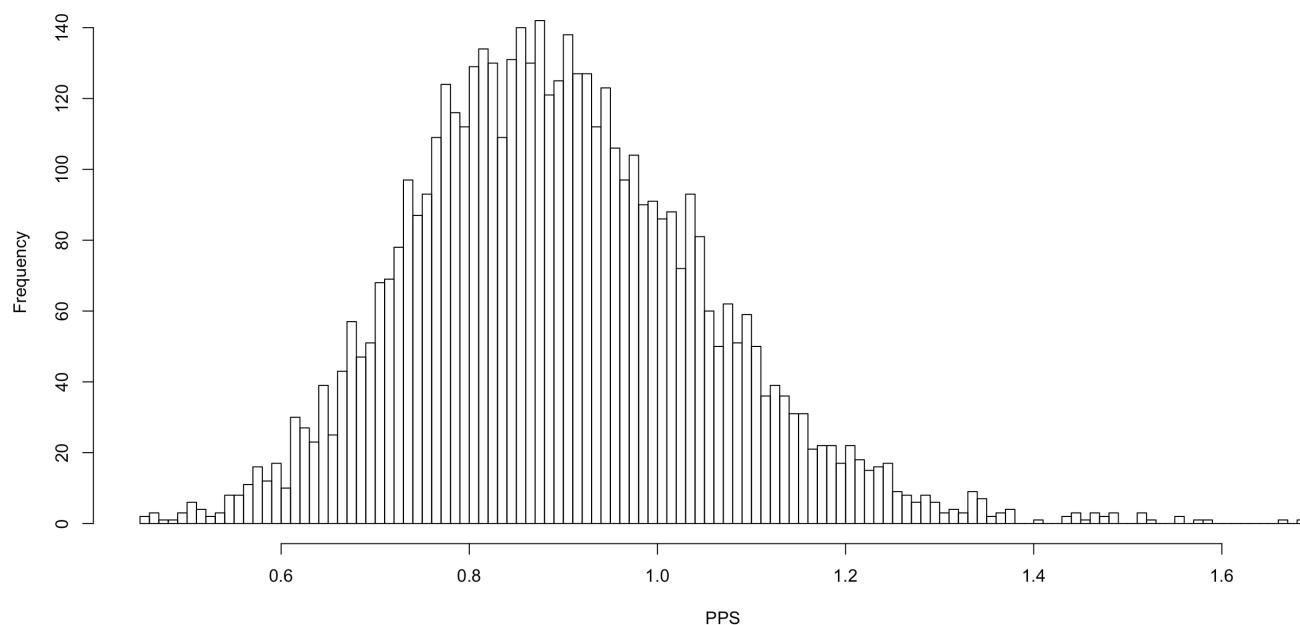


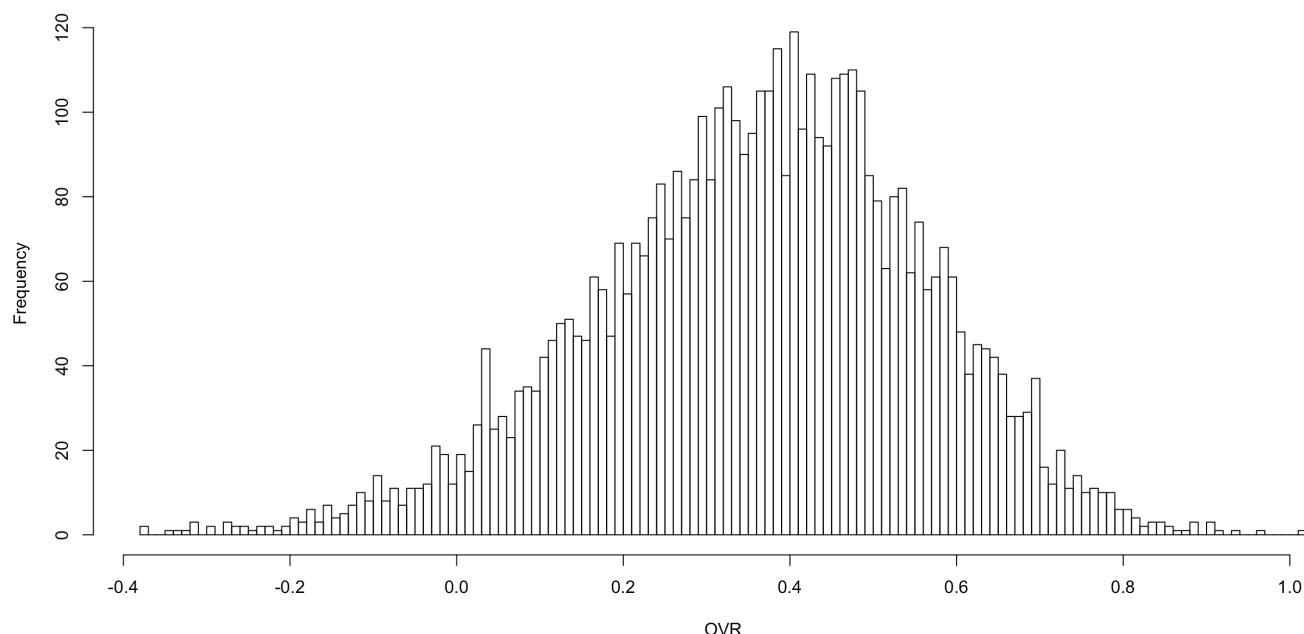
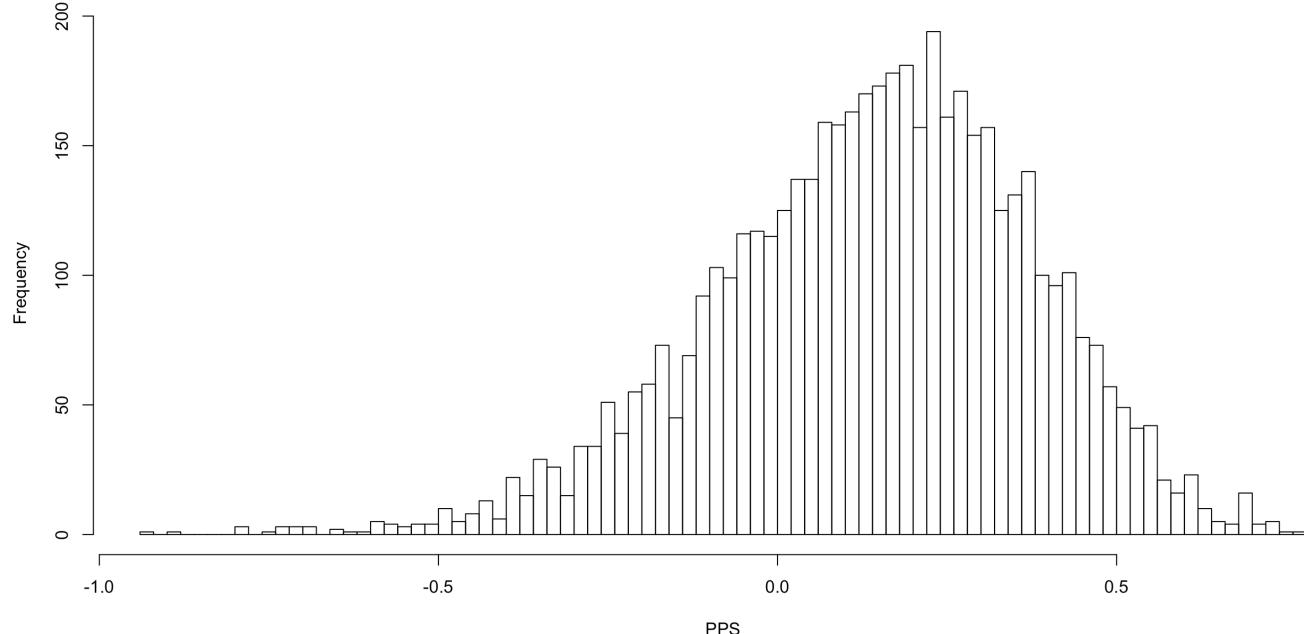
Language Levels, Quiz Levels

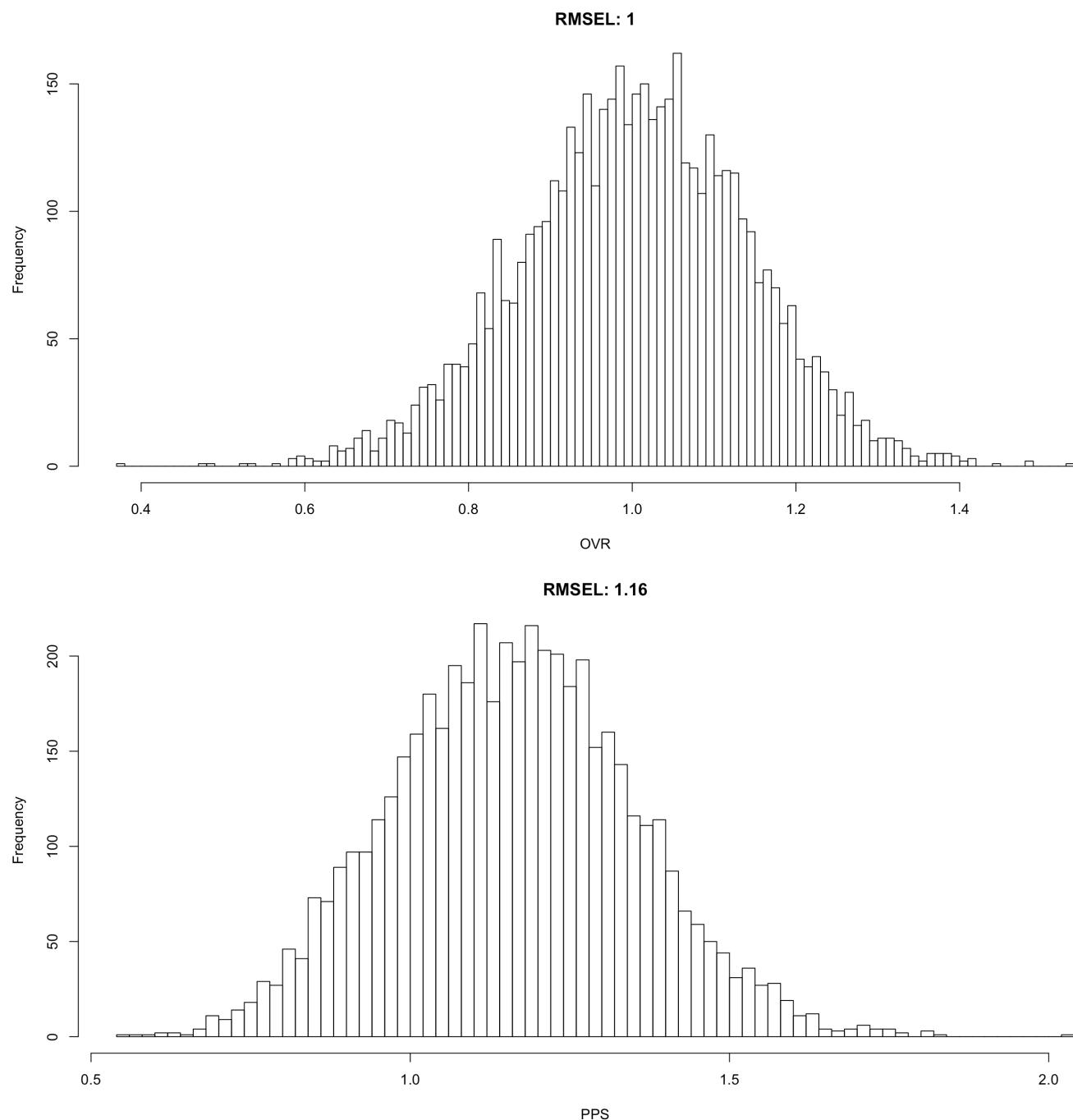




Linear Method

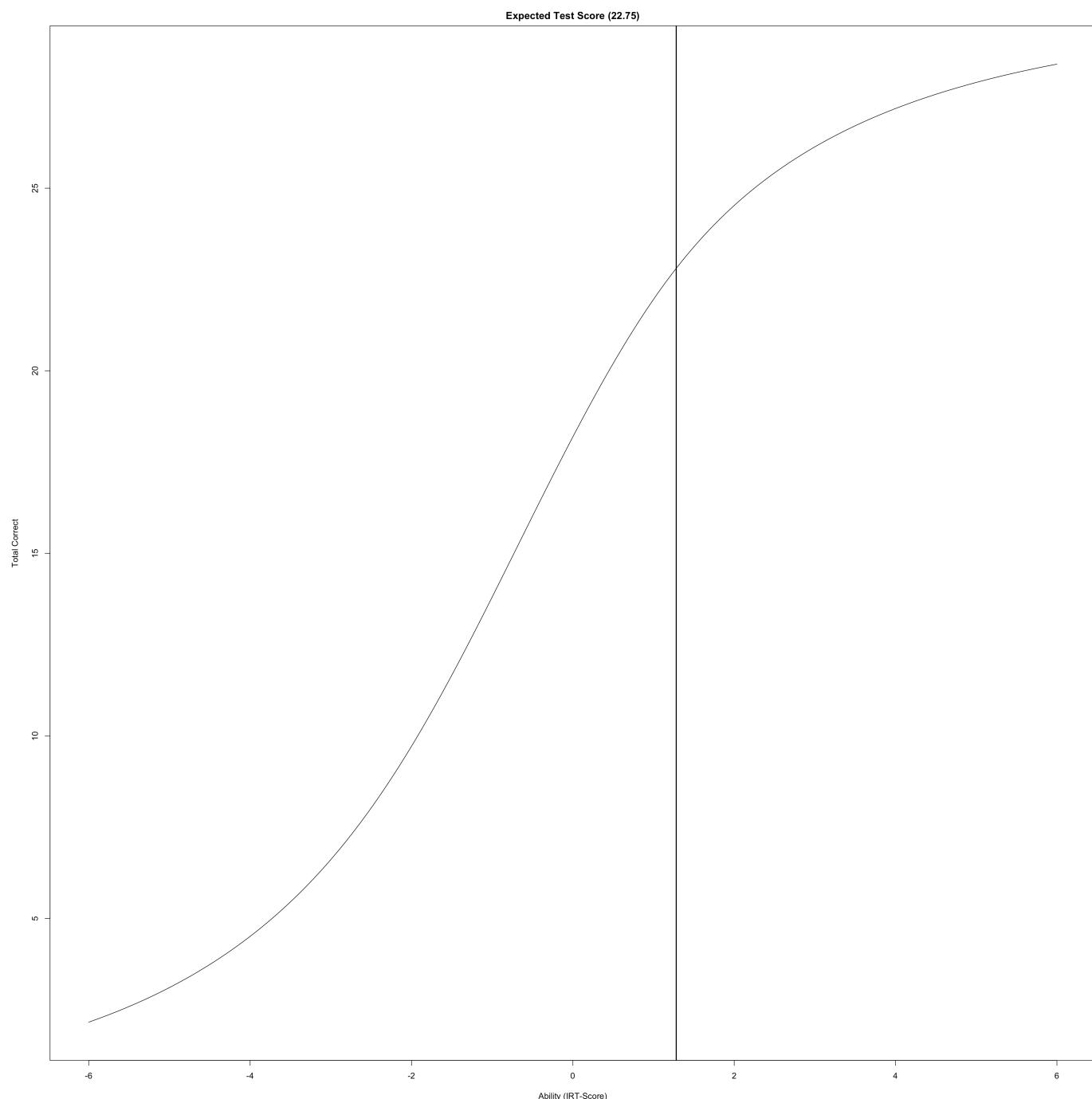
A: 0.95**A: 0.9**

B: 0.36**B: 0.14**

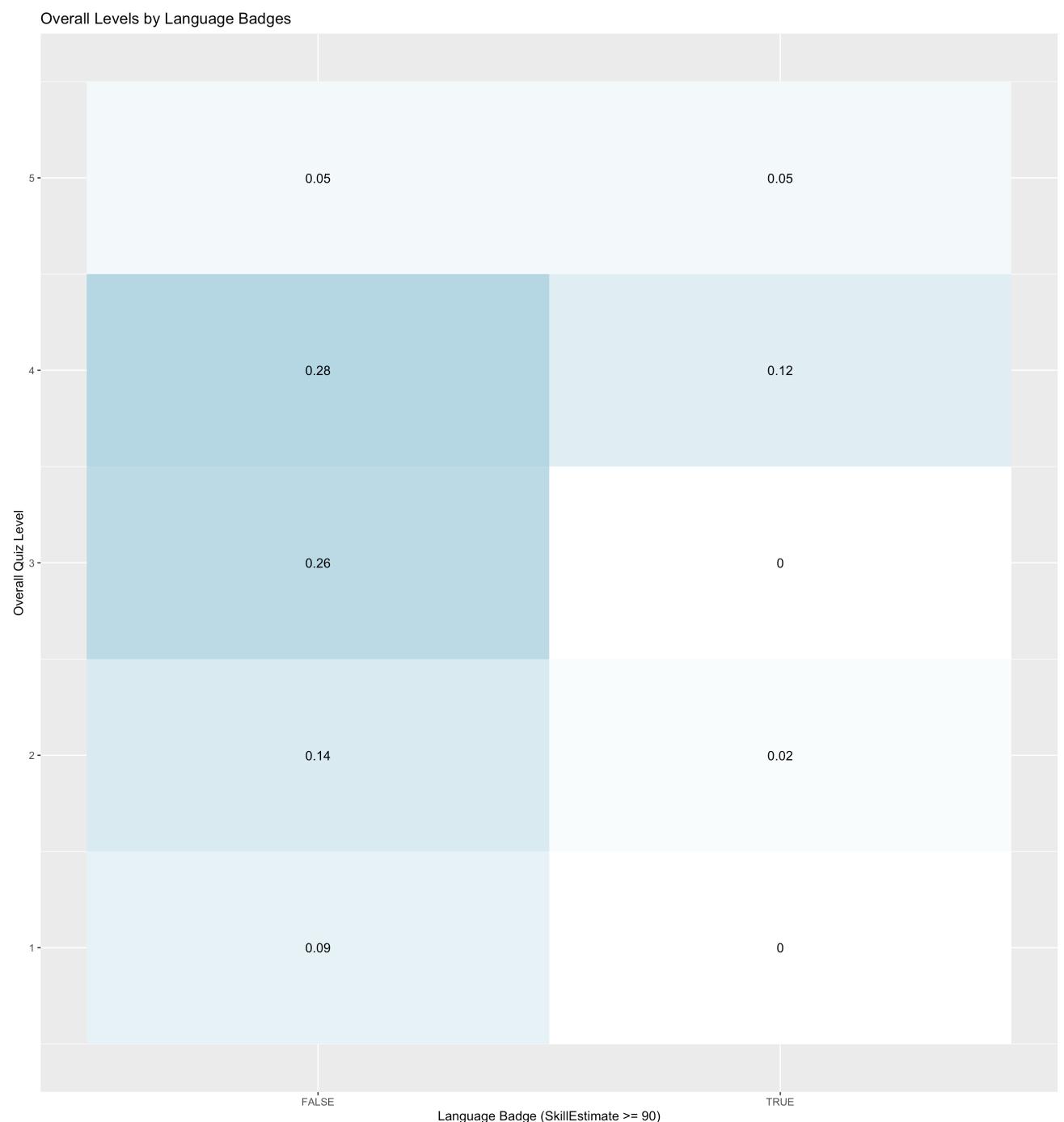


Result: Overall score produce smaller linking error

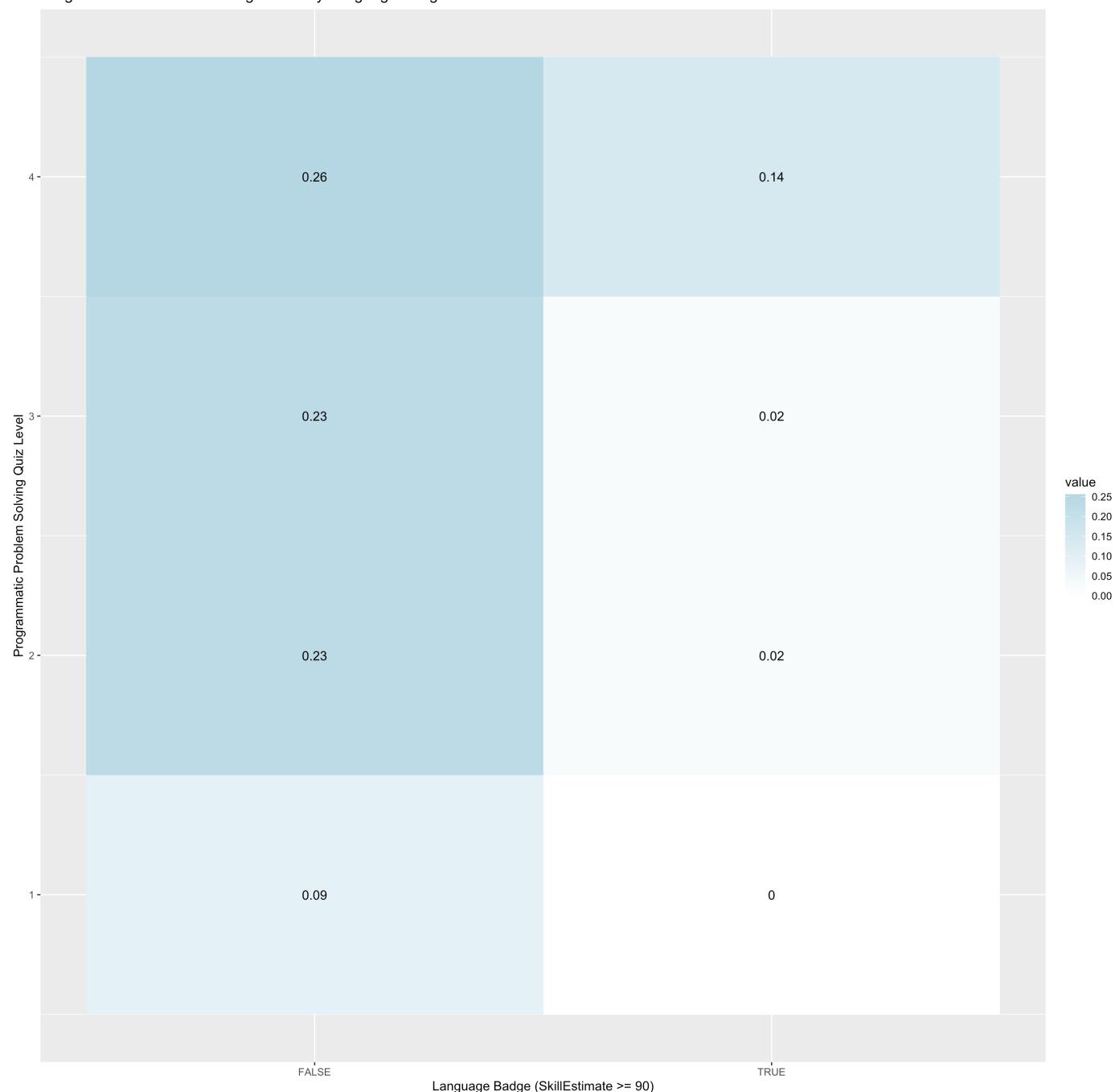
Badges - Test Characteristic Curve



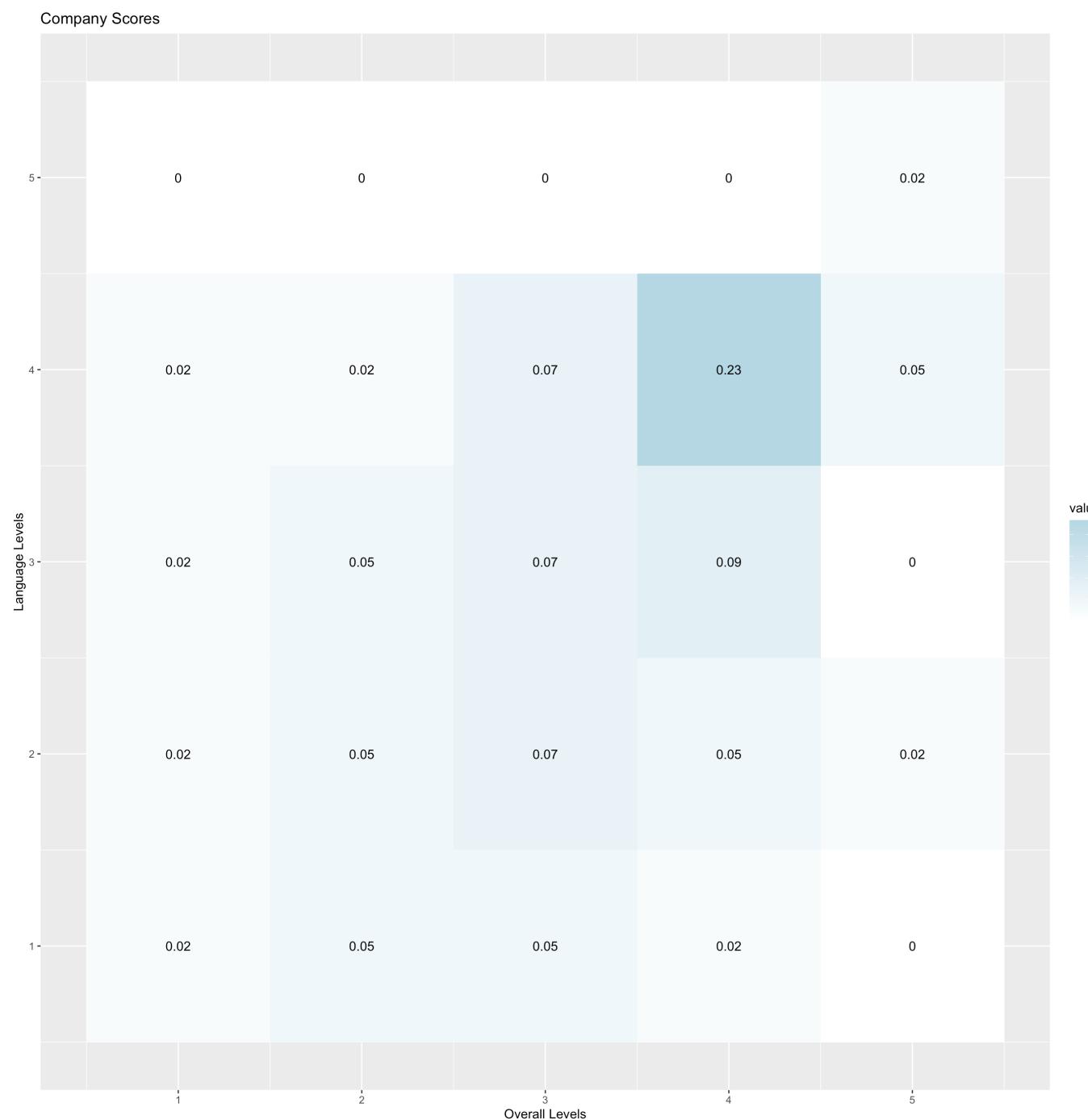
Badges - Quiz Level, by Badge

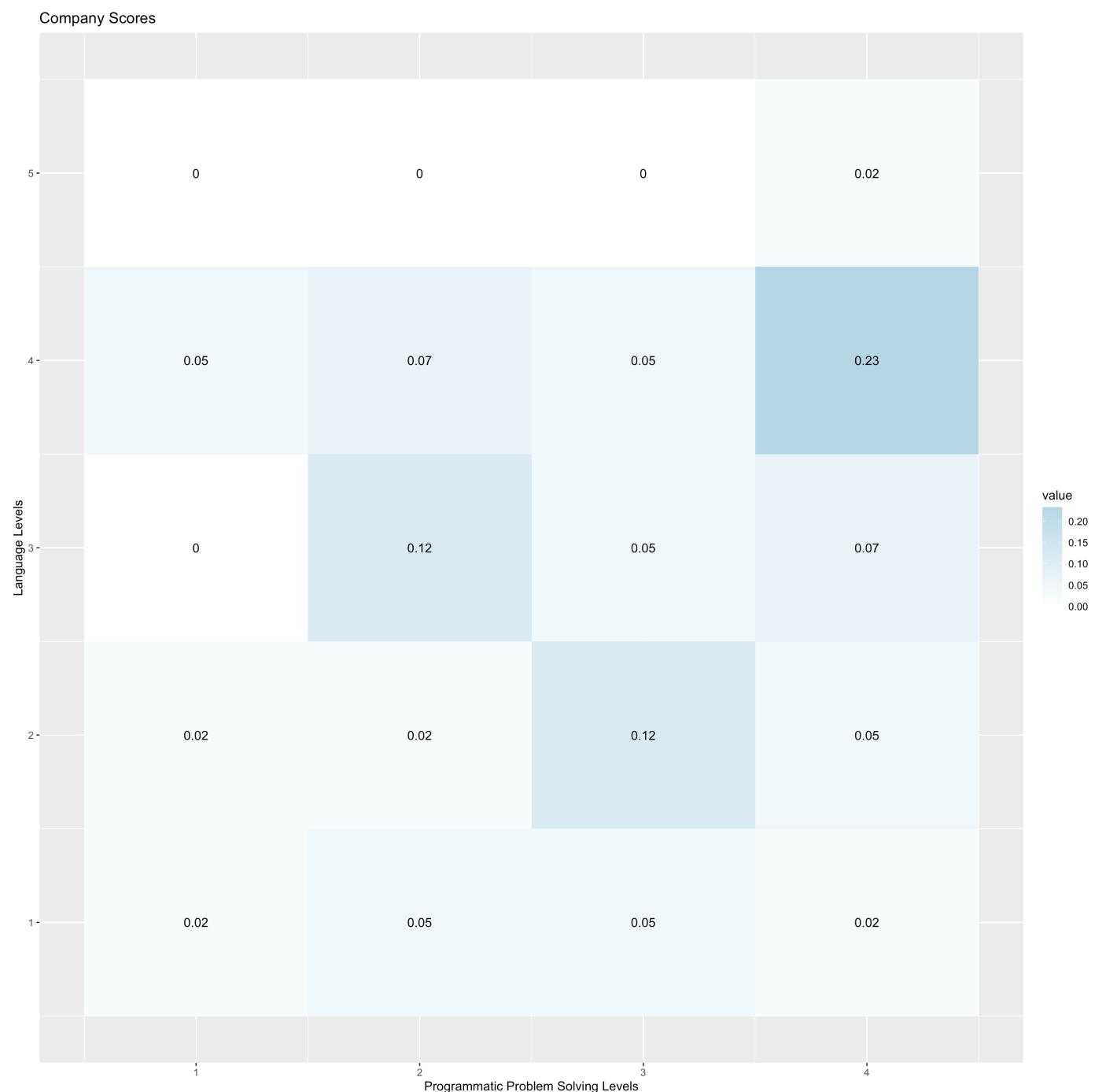


Programmatic Problem Solving Levels by Language Badges



Language Levels, Quiz Levels





Percentile Method

Summary