

# Java Assessment

Code ▾

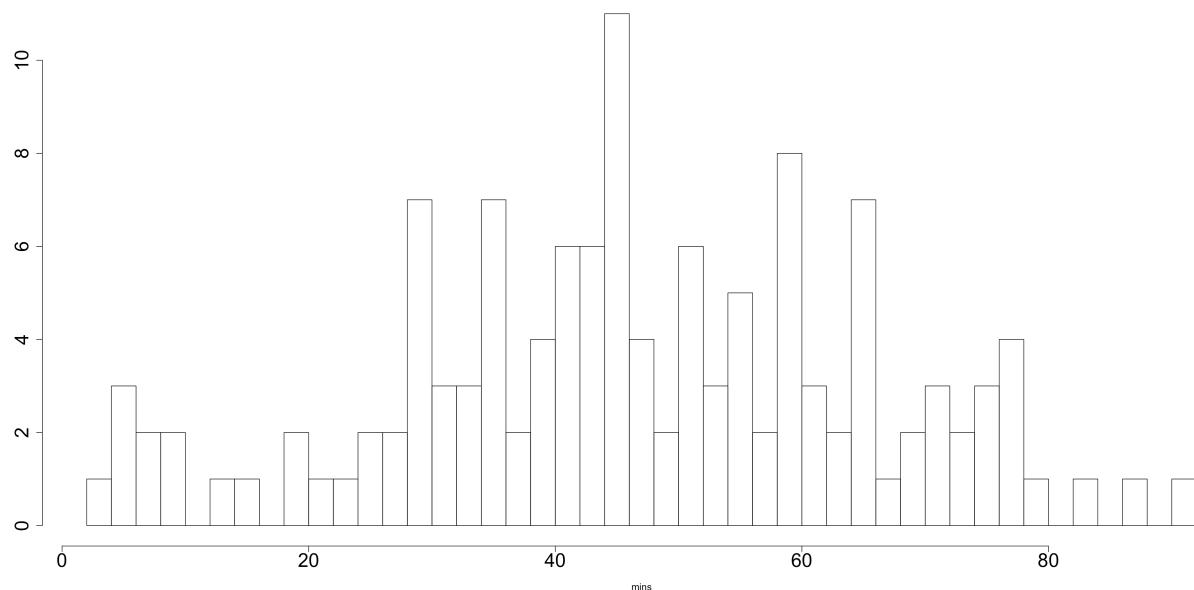
## Calibration, Scoring, & Impact Analyses

### Number of Questions Attempted

1	2	3	4	5	6	7	8	9	10	12	14	19	20	22	29	30	NA
12	9	6	3	6	6	3	4	2	1	3	3	1	1	1	1	128	0

### Time spent on Java Quiz

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



n	min	q1	avg	med	q3	max
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
128	2.02	34.415	46.37687	45.485	59.6775	90.08
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 Language 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 if a particular item is removed. 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:

*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.

Overall JavaScript Reliability is: 0.817

df_blueprint\$question_id	flag	claim	name	target	n_candidates	item_mean	item_total_correlation	alpha_de
1416	<=	Claim 1	lang_java_01	java-syntax-strings	118	0.831	0.259	1
1417		Claim 1	lang_java_02	java-syntax-polymorphism	118	0.780	0.232	1
1418	<=	Claim 1	lang_java_03	java-syntax-modifiers	118	0.780	0.111	1
1419	<=	Claim 1	lang_java_04	java-syntax-inheritance	118	0.890	0.401	1
1420		Claim 1	lang_java_05	java-syntax-classes	118	0.780	0.359	1
1432	<=	Claim 2	lang_java_17	java-containers-wrappers	118	0.915	0.220	1
1433	<=	Claim 2	lang_java_18	java-containers-util-arrays	118	0.915	0.294	1
1434		Claim 2	lang_java_19	java-containers-hashset	118	0.610	0.476	1
1435	<=	Claim 2	lang_java_20	java-containers-hashmap	118	0.873	0.469	1
1436	<=	Claim 2	lang_java_21	java-containers-array	118	0.864	0.026	1
1427		Claim 3	lang_java_12	java-features-reference	118	0.737	0.394	1
1428		Claim 3	lang_java_13	java-features-override	118	0.610	0.408	1
1429		Claim 3	lang_java_14	java-features-lambda	118	0.686	0.434	1
1430		Claim 3	lang_java_15	java-features-interface	118	0.712	0.273	1
1431		Claim 3	lang_java_16	java-features-enum	118	0.703	0.485	1
1421	<=	Claim 4	lang_java_06	java-std-time	118	0.864	0.220	1
1422		Claim 4	lang_java_07	java-std-thread	118	0.610	0.356	1
1423	<=	Claim 4	lang_java_08	java-std-streams	118	0.814	0.479	1
1424		Claim 4	lang_java_09	java-std-regex	118	0.636	0.336	1
1425		Claim 4	lang_java_10	java-standard-file-watcher	118	0.441	0.301	1

df_blueprint\$question_id	flag	claim	name	target	n_candidates	item_mean	item_total_correlation	alpha_de
1426		Claim	lang_java_11	java-libraries-gson 5	118	0.297	0.235	1
1437		Claim	lang_java_22	java-libraries-junit 5	118	0.619	0.560	1
1441		Claim	lang_java_26	java-libraries-guava 5	118	0.678	0.374	1
1444		Claim	lang_java_29	java-libraries-async 5	118	0.568	0.290	1
1445 <=		Claim	lang_java_30	java-libraries-sql 5	118	0.873	0.447	1
1438		Claim	lang_java_23	java-domain-javac 6	118	0.585	0.305	1
1439		Claim	lang_java_24	java-domain-blocking-queue 6	118	0.568	0.395	1
1440 <=		Claim	lang_java_25	java-domain-wait-notify 6	118	0.161	0.020	1
1442		Claim	lang_java_27	java-domain-heap 6	118	0.500	0.388	1
1443 <=		Claim	lang_java_28	java-domain-jvm 6	118	0.924	0.147	1

## IRT Analysis

### Unidimensional Model

#### IRT Item Parameters

	a	b	g	u
1426	0.607	1.694	0	1
1417	0.607	-1.952	0	1
1440	0.204	7.916	0	1
1432	0.575	-3.796	0	1
1428	0.962	-0.342	0	1
1416	0.616	-2.438	0	1
1425	0.660	0.592	0	1
1439	0.924	-0.139	0	1
1422	0.883	-0.376	0	1
1431	1.394	-0.622	0	1
1443	0.459	-4.898	0	1
1424	0.835	-0.545	0	1
1438	0.742	-0.301	0	1
1436	0.265	-6.443	0	1
1433	0.781	-2.880	0	1
1437	1.806	-0.211	0	1
1444	0.631	-0.255	0	1
1419	1.222	-1.798	0	1
1435	1.533	-1.435	0	1
1441	0.883	-0.754	0	1
1420	0.847	-1.443	0	1
1429	1.217	-0.609	0	1

	<b>a</b>	<b>b</b>	<b>g</b>	<b>u</b>
1434	1.230	-0.263	0	1
1423	1.460	-1.128	0	1
1442	0.990	0.204	0	1
1430	0.672	-1.229	0	1
1421	0.557	-3.110	0	1
1445	1.342	-1.563	0	1
1427	1.012	-0.991	0	1
1418	0.299	-3.911	0	1

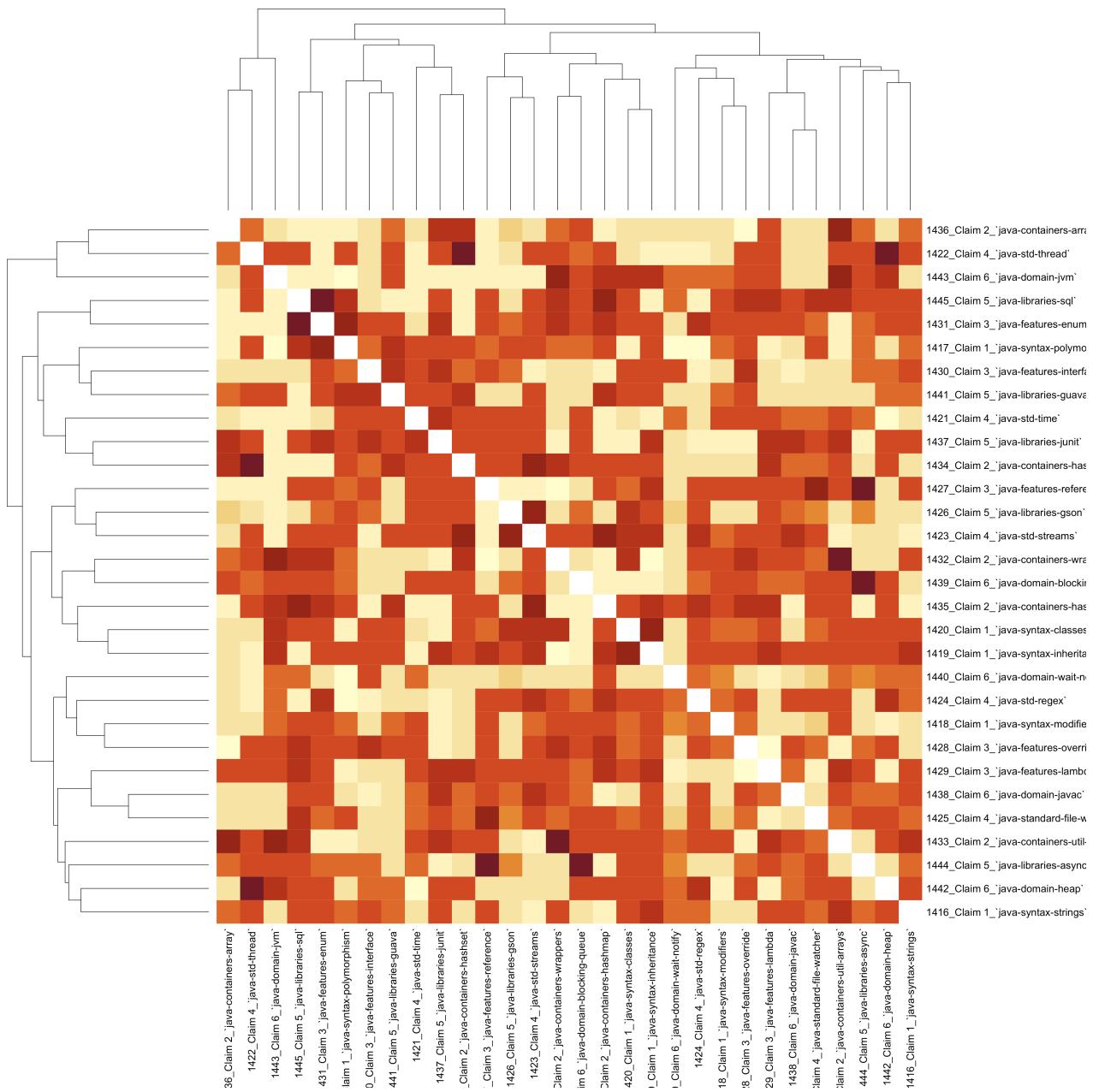
**Item Fit Statistics (S-X2): p-val < 0.05 => generally indicates model misfit**

claim	name	target	item	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2	flag
Claim 5	lang_java_11	java-libraries-gson	1426	4.211	10	0.000	0.937	
Claim 1	lang_java_02	java-syntax-polymorphism	1417	7.667	13	0.000	0.865	
Claim 6	lang_java_25	java-domain-wait-notify	1440	9.843	11	0.000	0.545	
Claim 2	lang_java_17	java-containers-wrappers	1432	3.993	7	0.000	0.781	
Claim 3	lang_java_13	java-features-override	1428	17.965	11	0.074	0.082	
Claim 1	lang_java_01	java-syntax-strings	1416	7.246	10	0.000	0.702	
Claim 4	lang_java_10	java-standard-file-watcher	1425	14.044	11	0.049	0.231	
Claim 6	lang_java_24	java-domain-blocking-queue	1439	10.625	11	0.000	0.475	
Claim 4	lang_java_07	java-std-thread	1422	12.774	12	0.023	0.386	
Claim 3	lang_java_16	java-features-enum	1431	9.967	10	0.000	0.443	
Claim 6	lang_java_28	java-domain-jvm	1443	9.176	6	0.067	0.164	
Claim 4	lang_java_09	java-std-regex	1424	5.550	11	0.000	0.902	
Claim 6	lang_java_23	java-domain-javac	1438	8.229	12	0.000	0.767	
Claim 2	lang_java_21	java-containers-array	1436	5.727	8	0.000	0.678	
Claim 2	lang_java_18	java-containers-util-arrays	1433	4.701	6	0.000	0.583	
Claim 5	lang_java_22	java-libraries-junit	1437	5.689	8	0.000	0.682	
Claim 5	lang_java_29	java-libraries-async	1444	7.796	12	0.000	0.801	
Claim 1	lang_java_04	java-syntax-inheritance	1419	15.725	7	0.103	0.028	<=
Claim 2	lang_java_20	java-containers-hashmap	1435	11.666	7	0.075	0.112	
Claim 5	lang_java_26	java-libraries-guava	1441	7.830	12	0.000	0.798	
Claim 1	lang_java_05	java-syntax-classes	1420	8.930	12	0.000	0.709	
Claim 3	lang_java_14	java-features-lambda	1429	7.053	11	0.000	0.795	
Claim 2	lang_java_19	java-containers-hashset	1434	10.256	10	0.015	0.418	
Claim 4	lang_java_08	java-std-streams	1423	8.017	9	0.000	0.532	
Claim 6	lang_java_27	java-domain-heap	1442	12.673	11	0.036	0.315	
Claim 3	lang_java_15	java-features-interface	1430	6.942	13	0.000	0.905	
Claim 4	lang_java_06	java-std-time	1421	3.983	8	0.000	0.859	
Claim 5	lang_java_30	java-libraries-sql	1445	7.857	7	0.032	0.345	
Claim 3	lang_java_12	java-features-reference	1427	7.729	12	0.000	0.806	
Claim 1	lang_java_03	java-syntax-modifiers	1418	11.784	13	0.000	0.545	

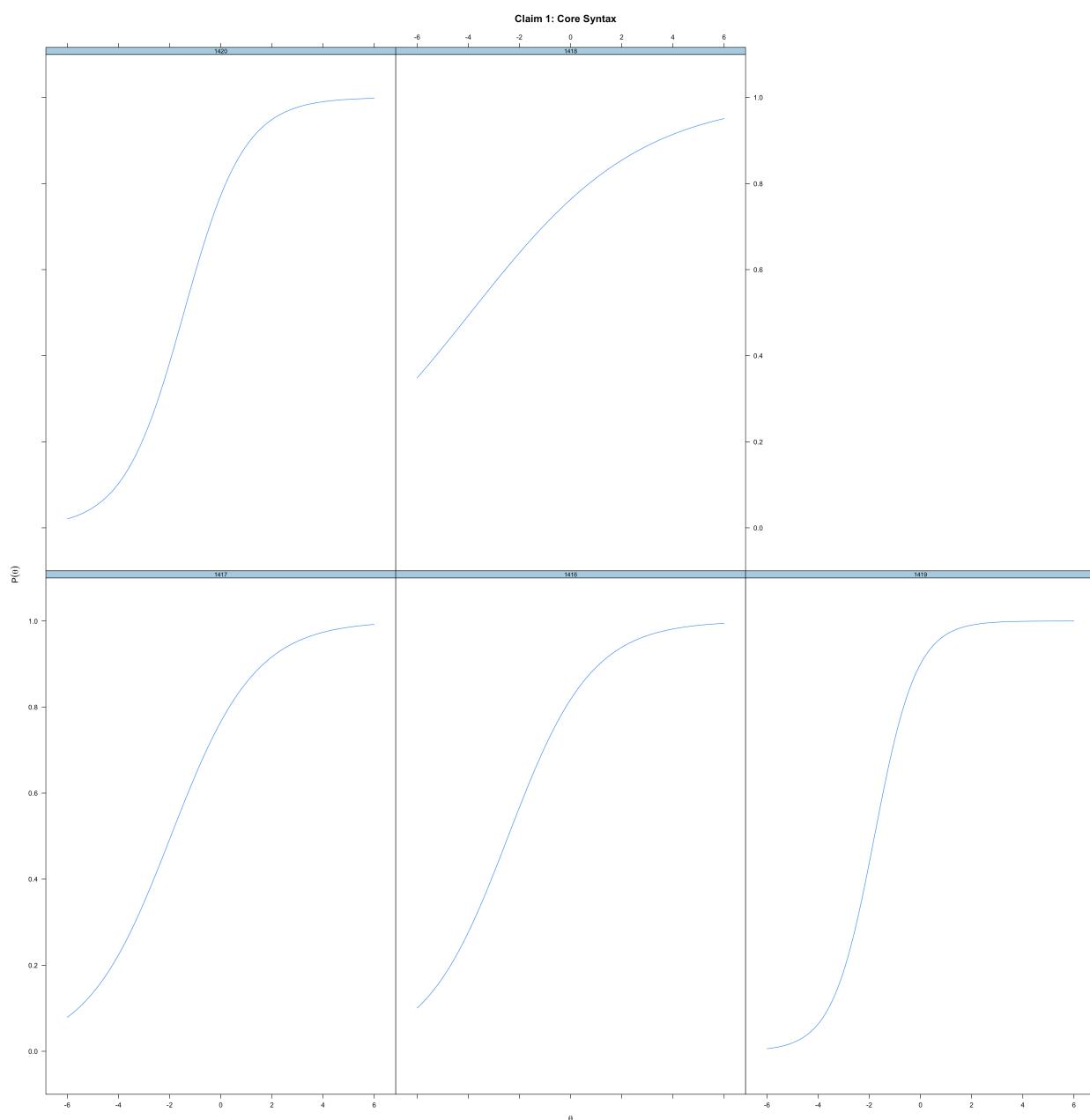
Items are flagged with '<=' to be reviewed or removed if *p-value* < 0.05. This indicates poor fit of 2PL model.

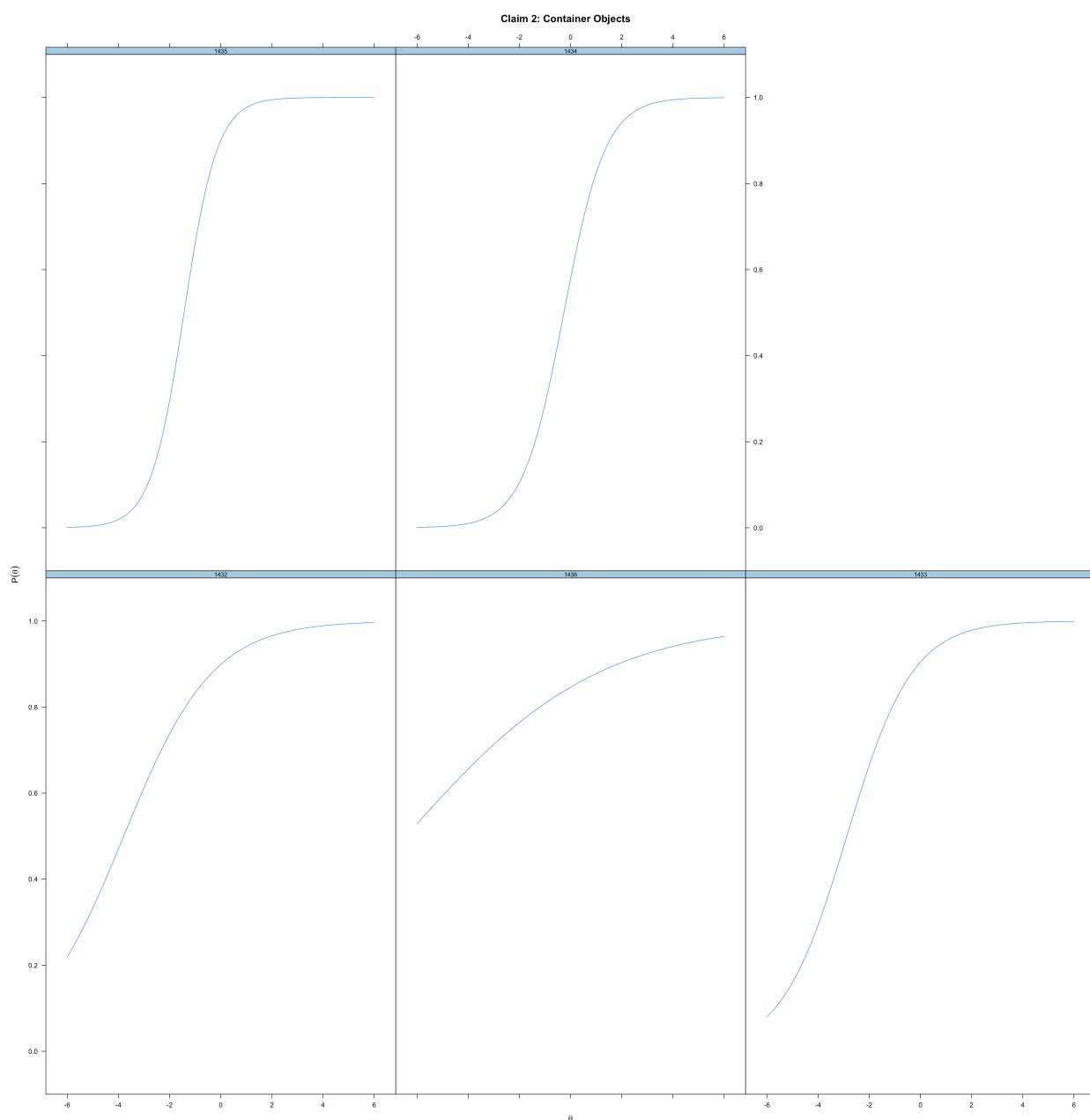
# IRT Plots

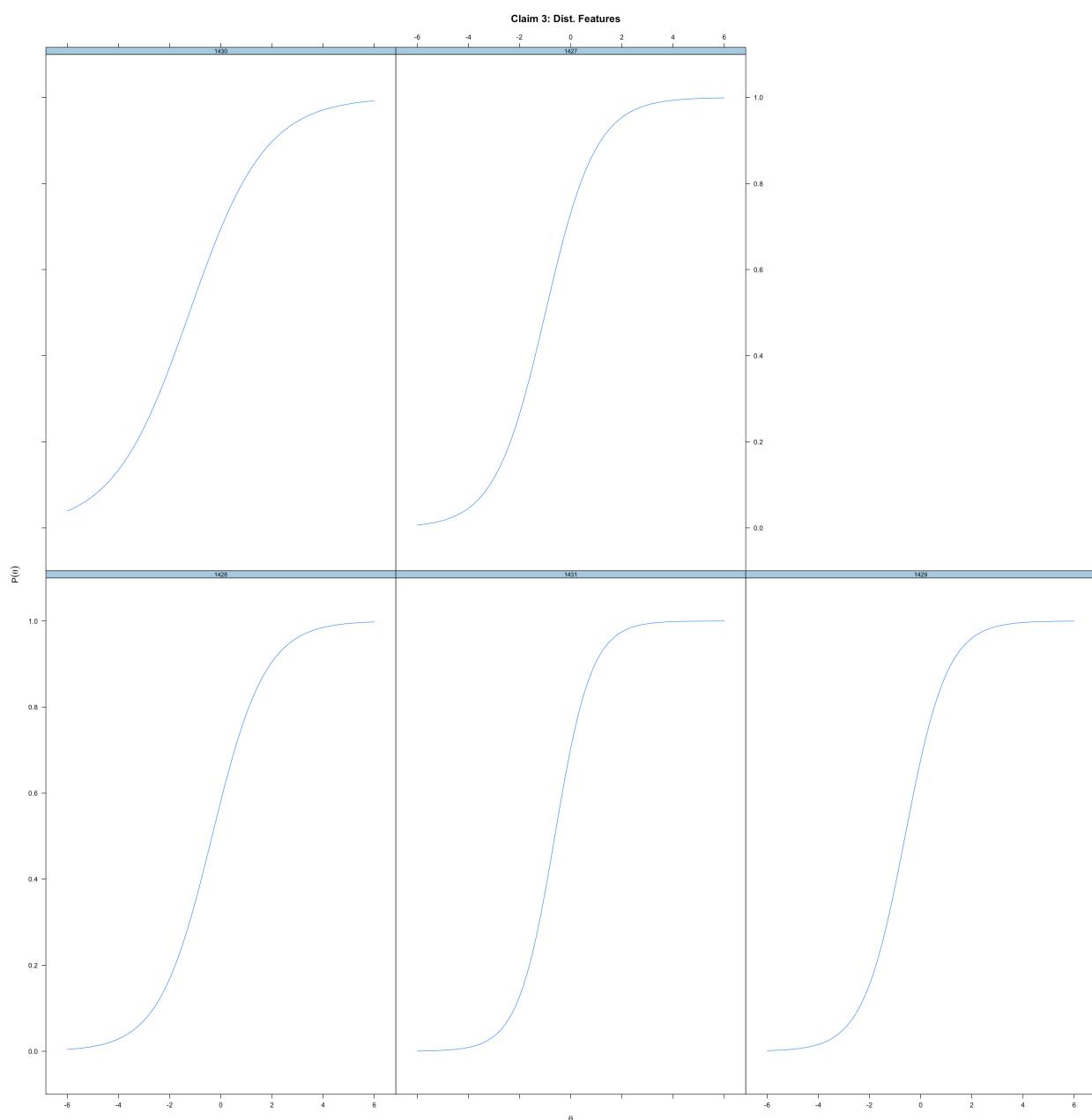
Item Local-Dependence Plots - Residual Dependencies given a unidimensional model:

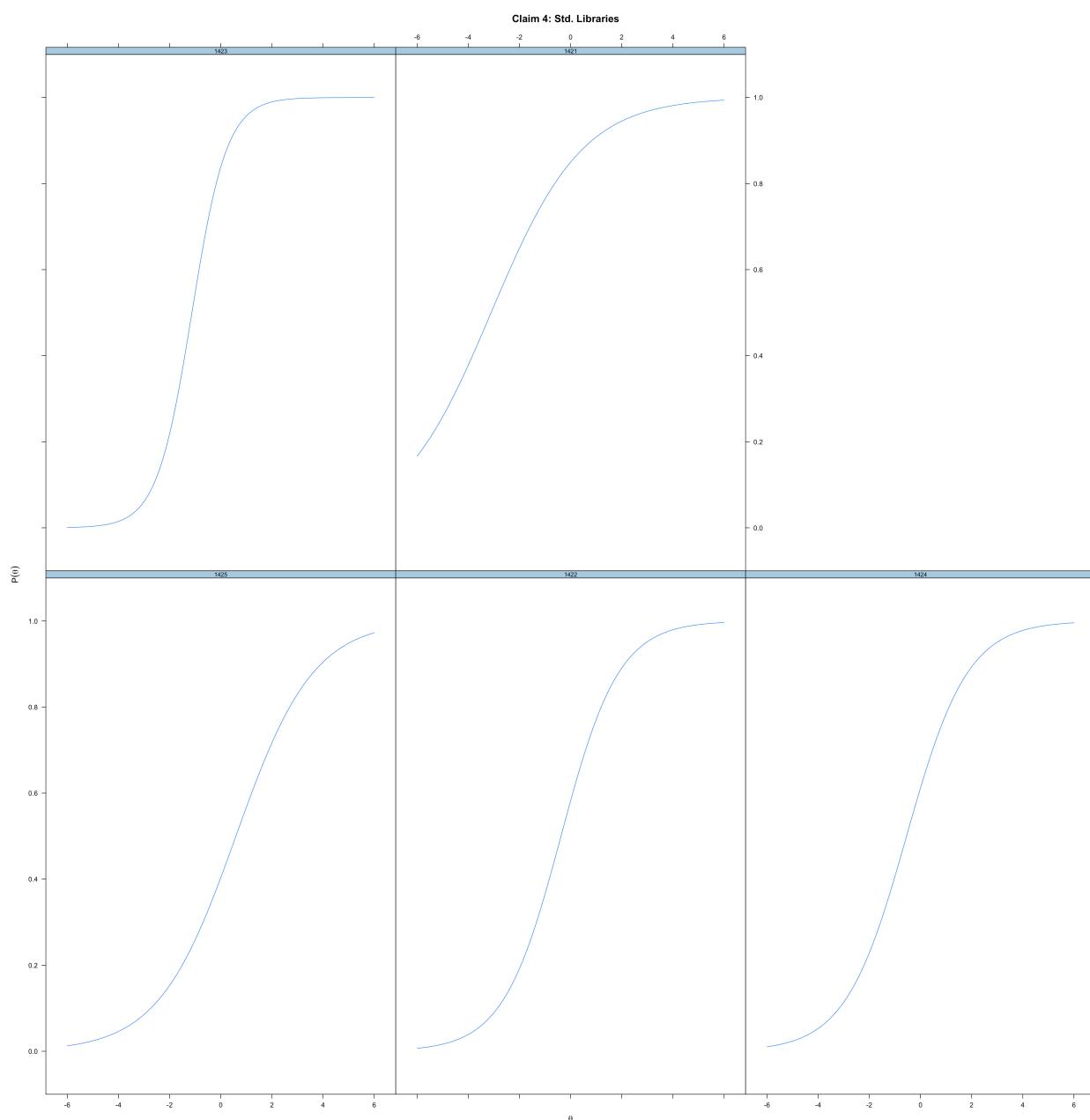


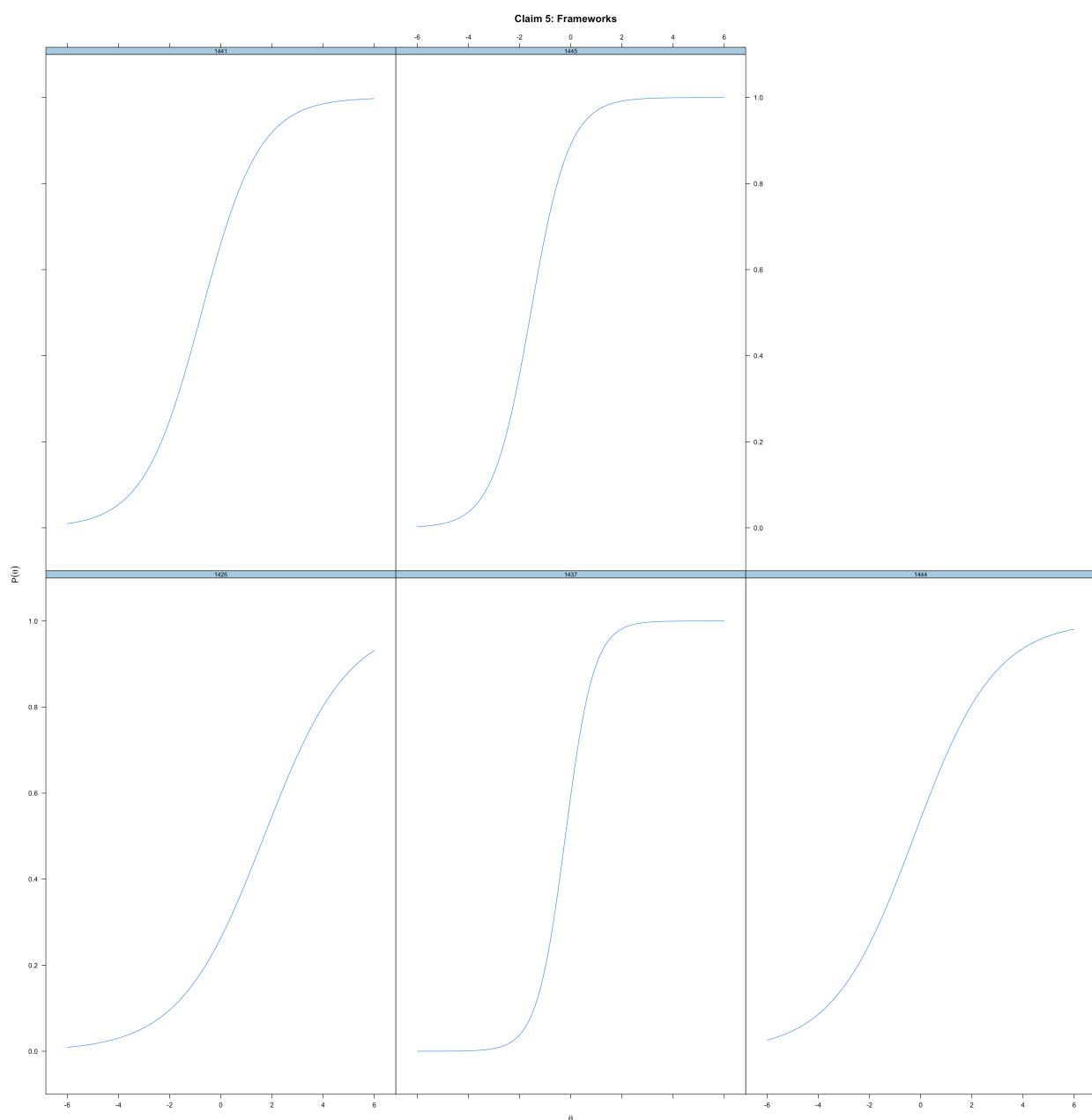
Item Characteristic Curves (Tracelines), by Claims

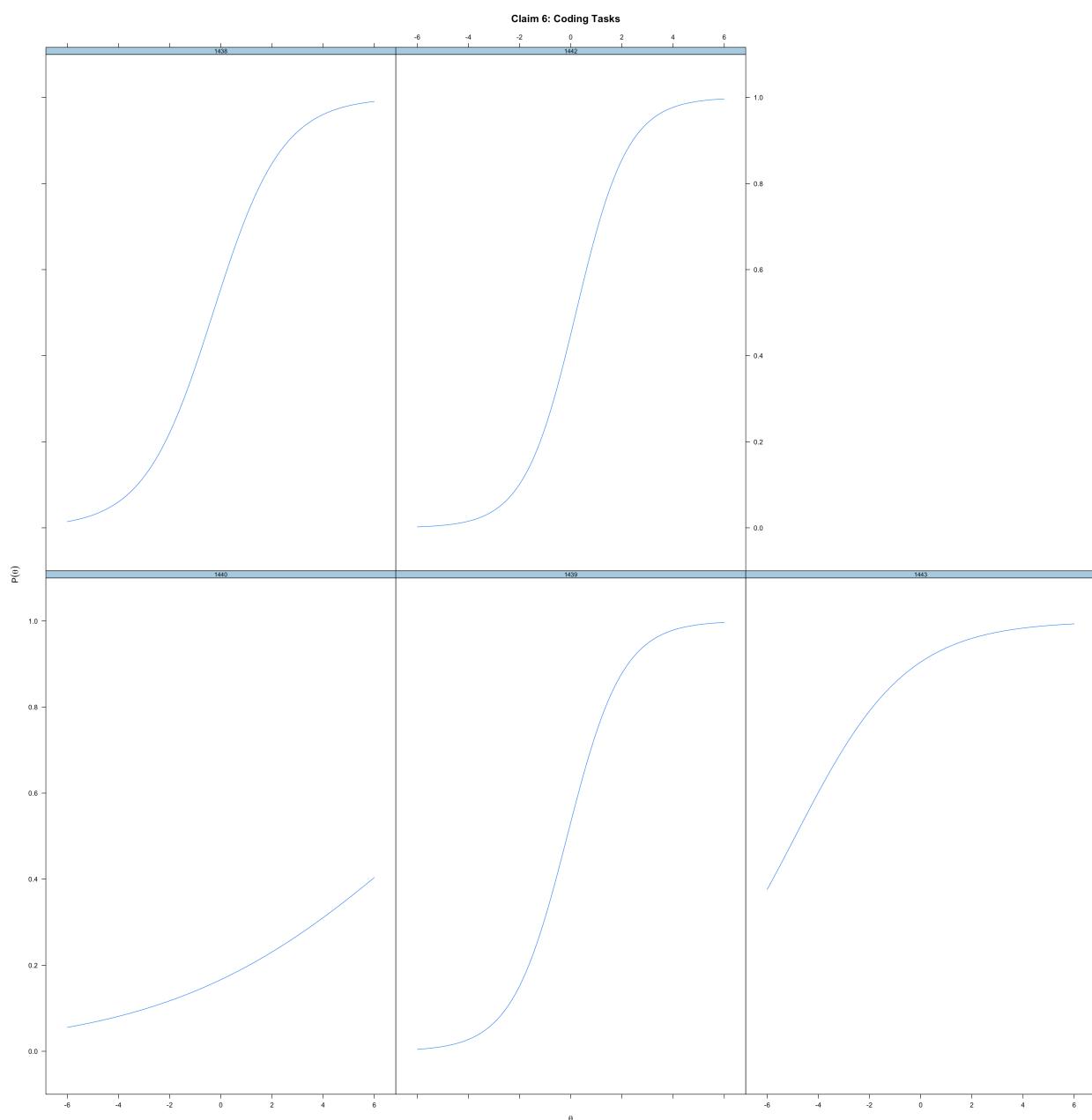




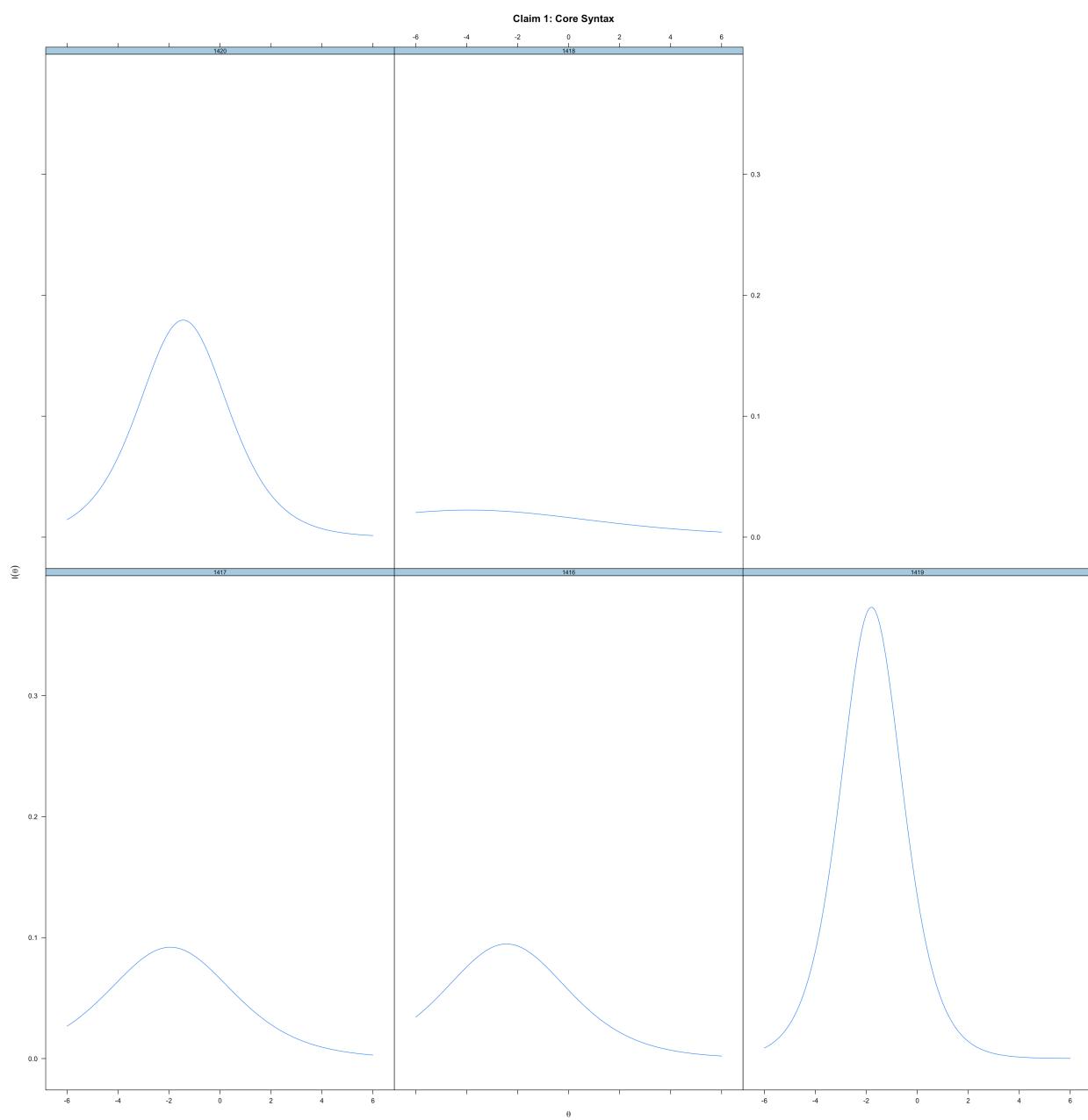


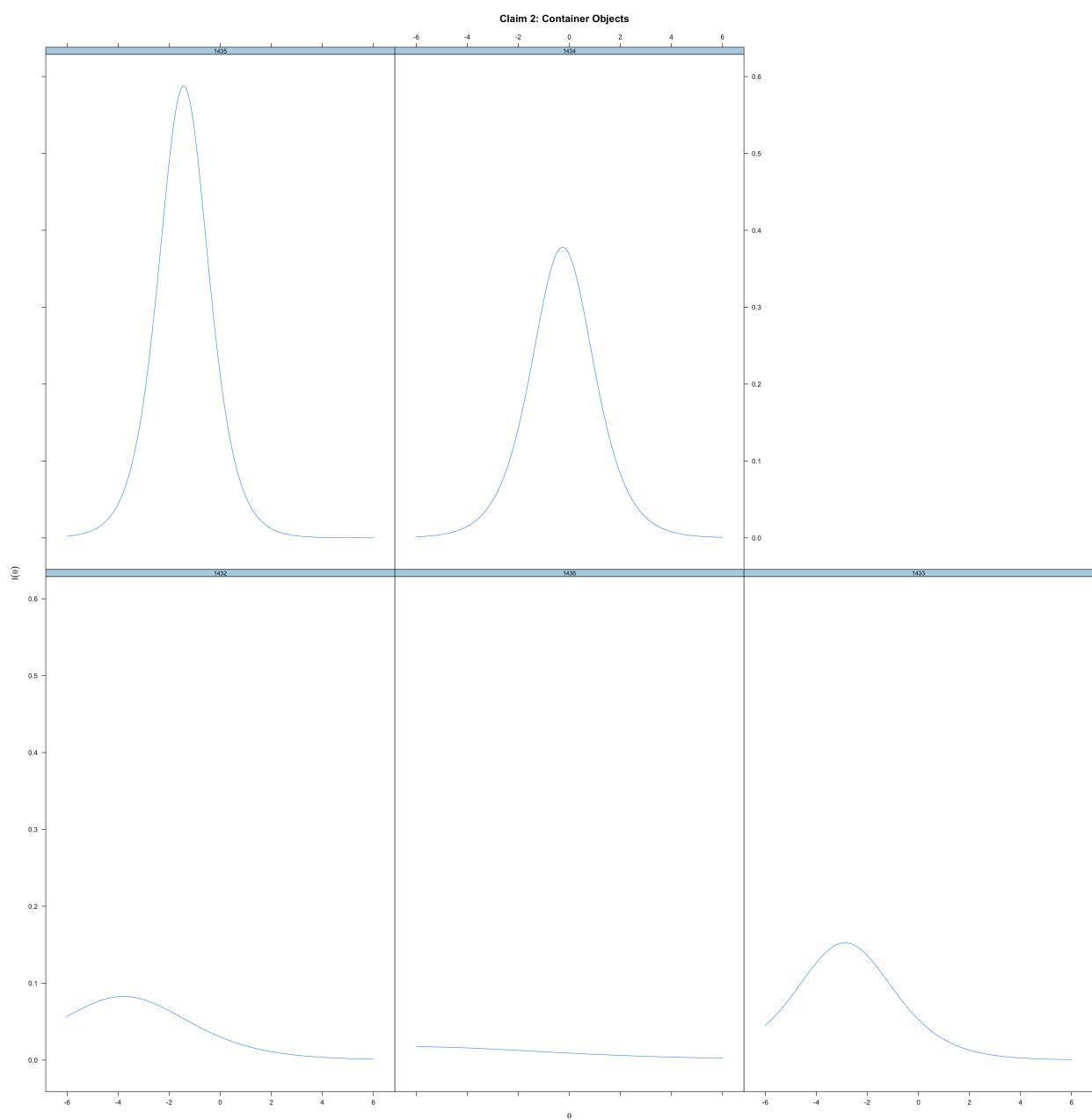


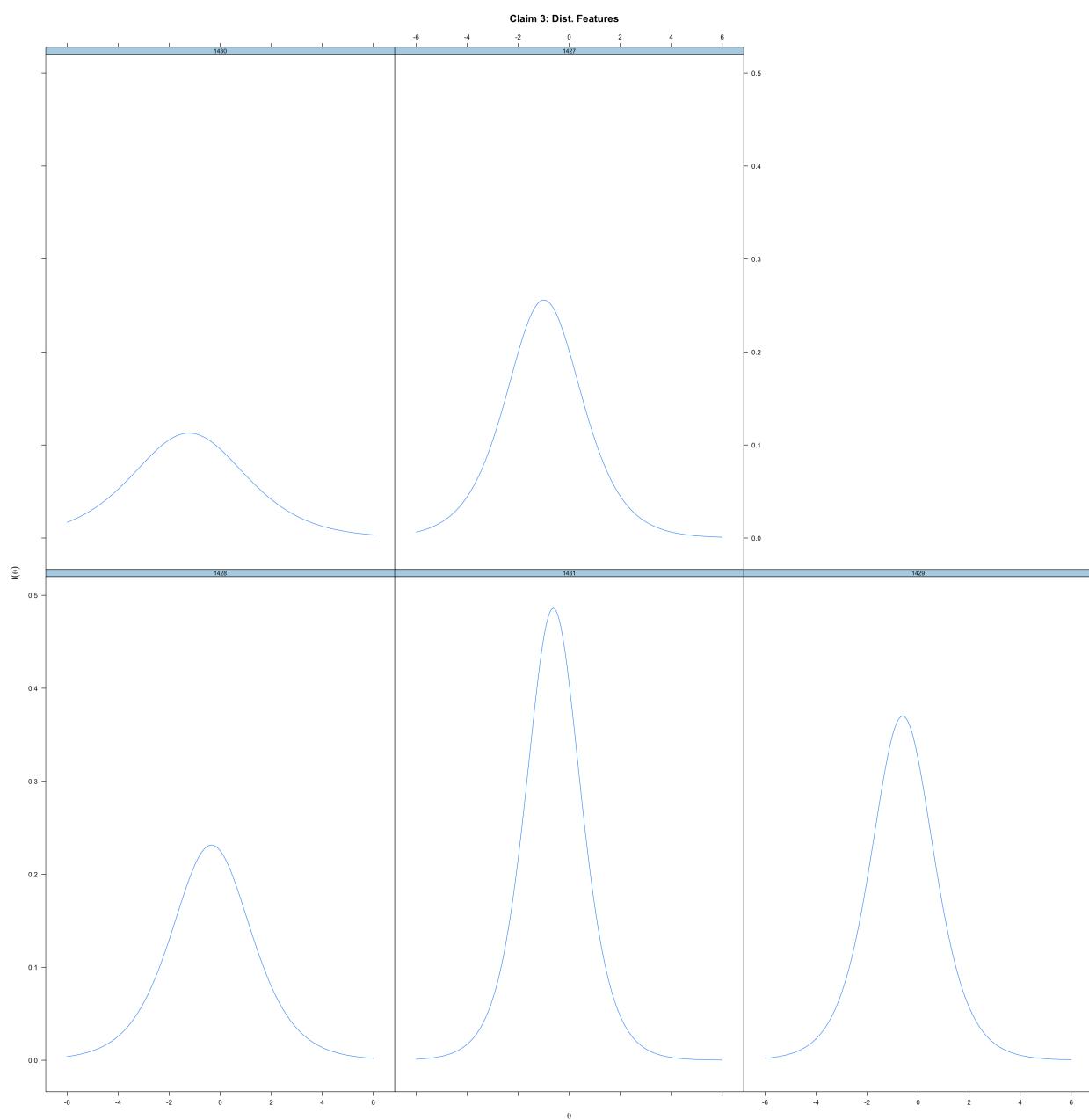


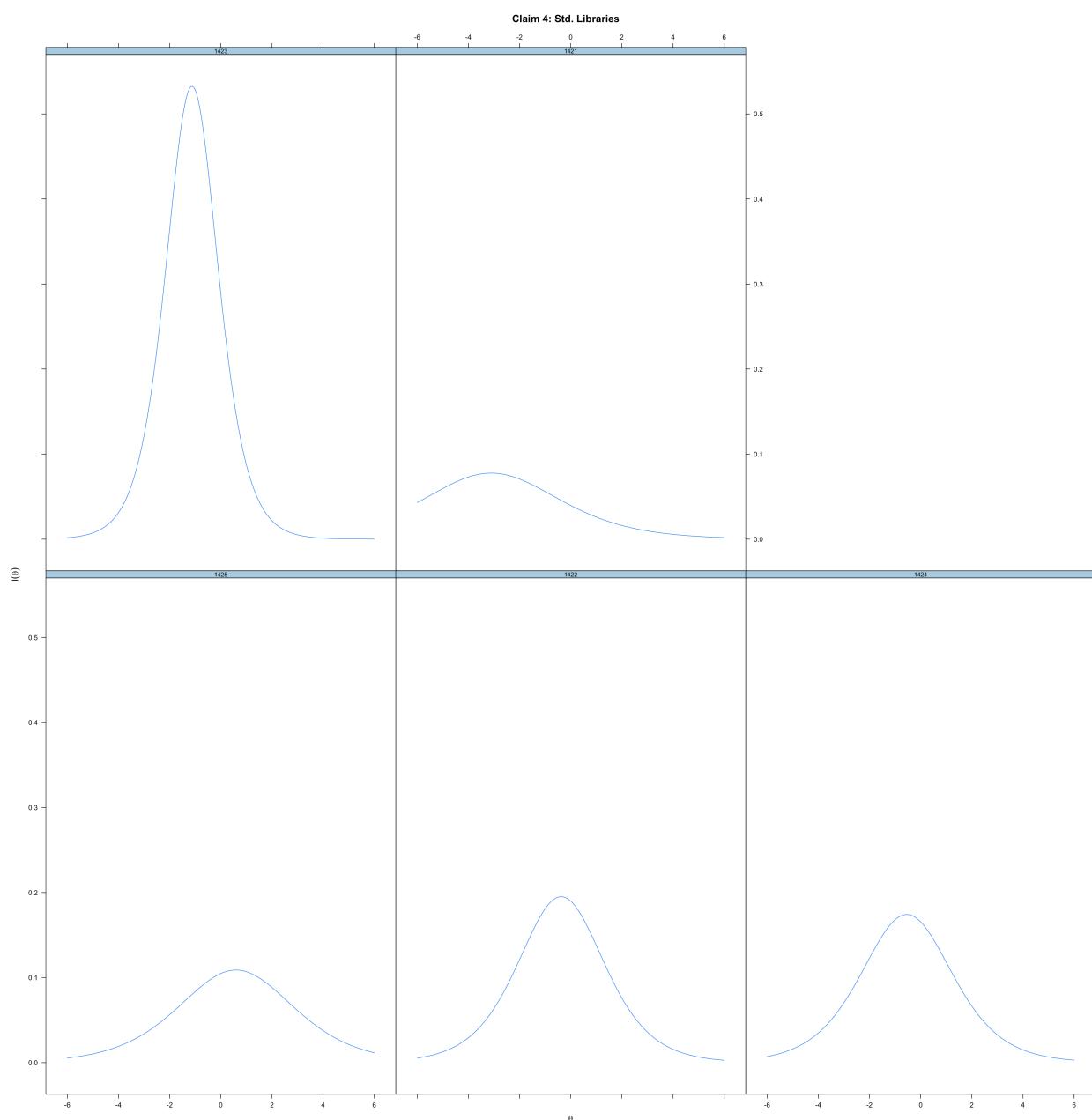


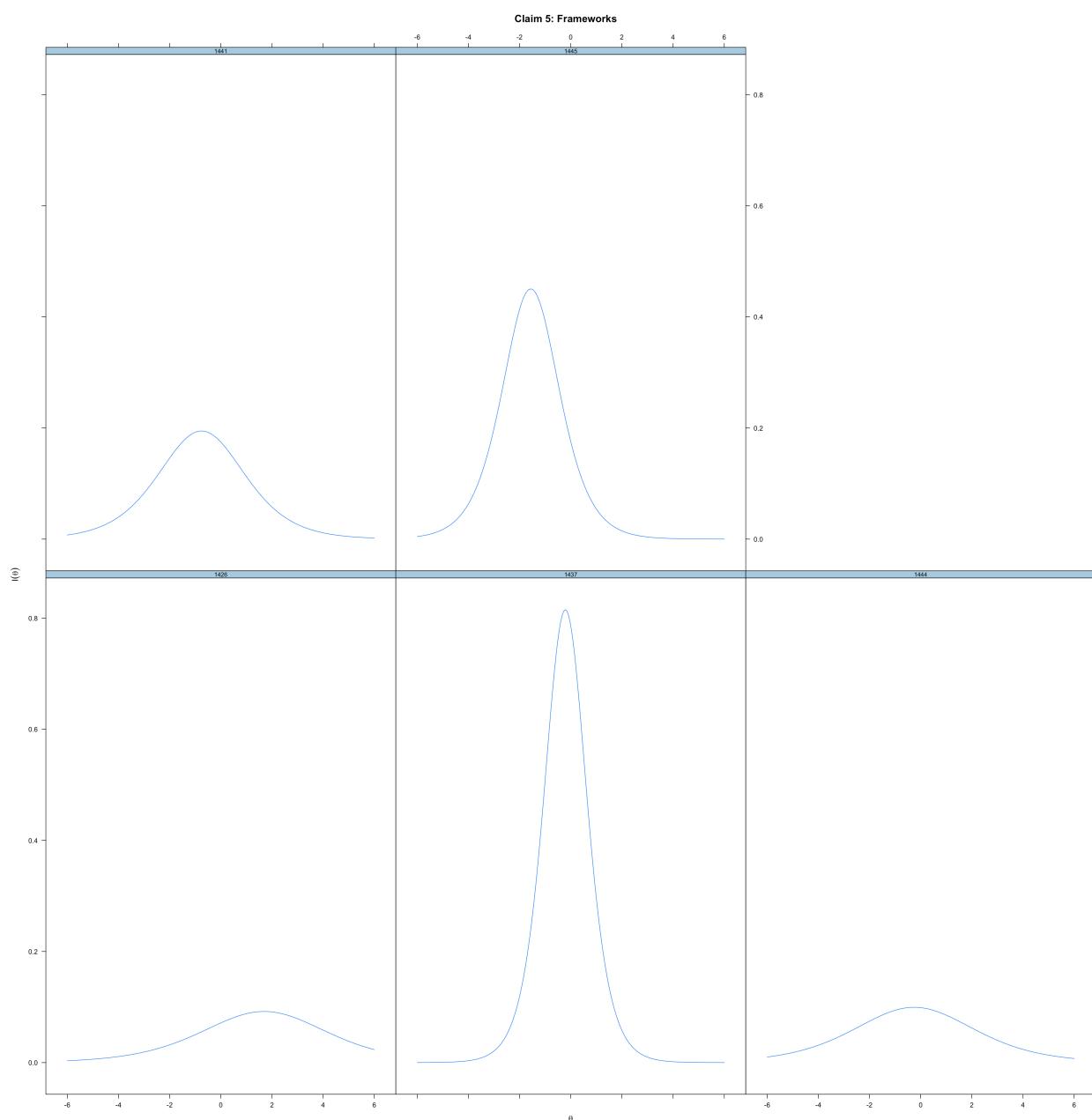
**Item Information Curves by Claims**

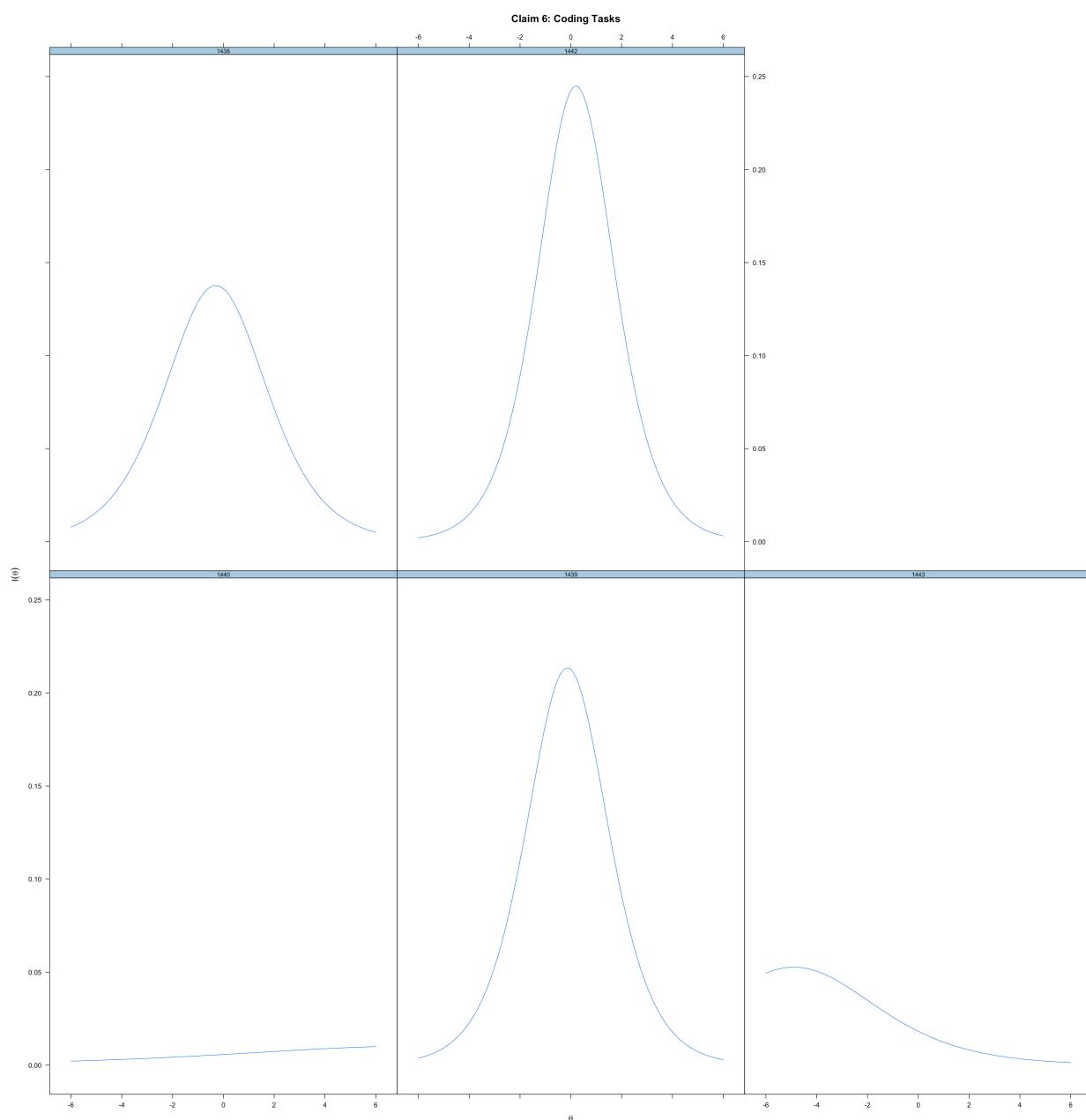




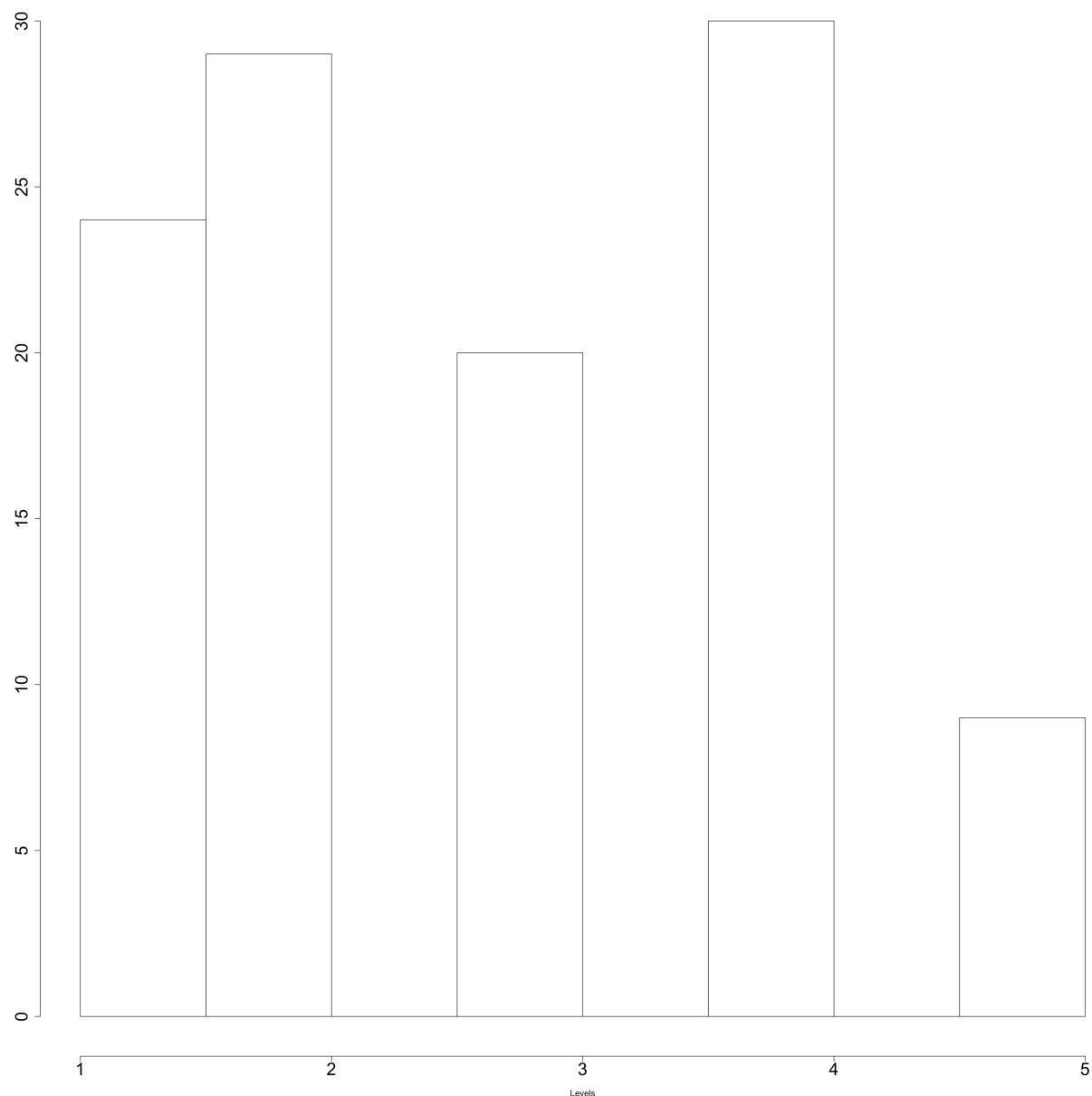


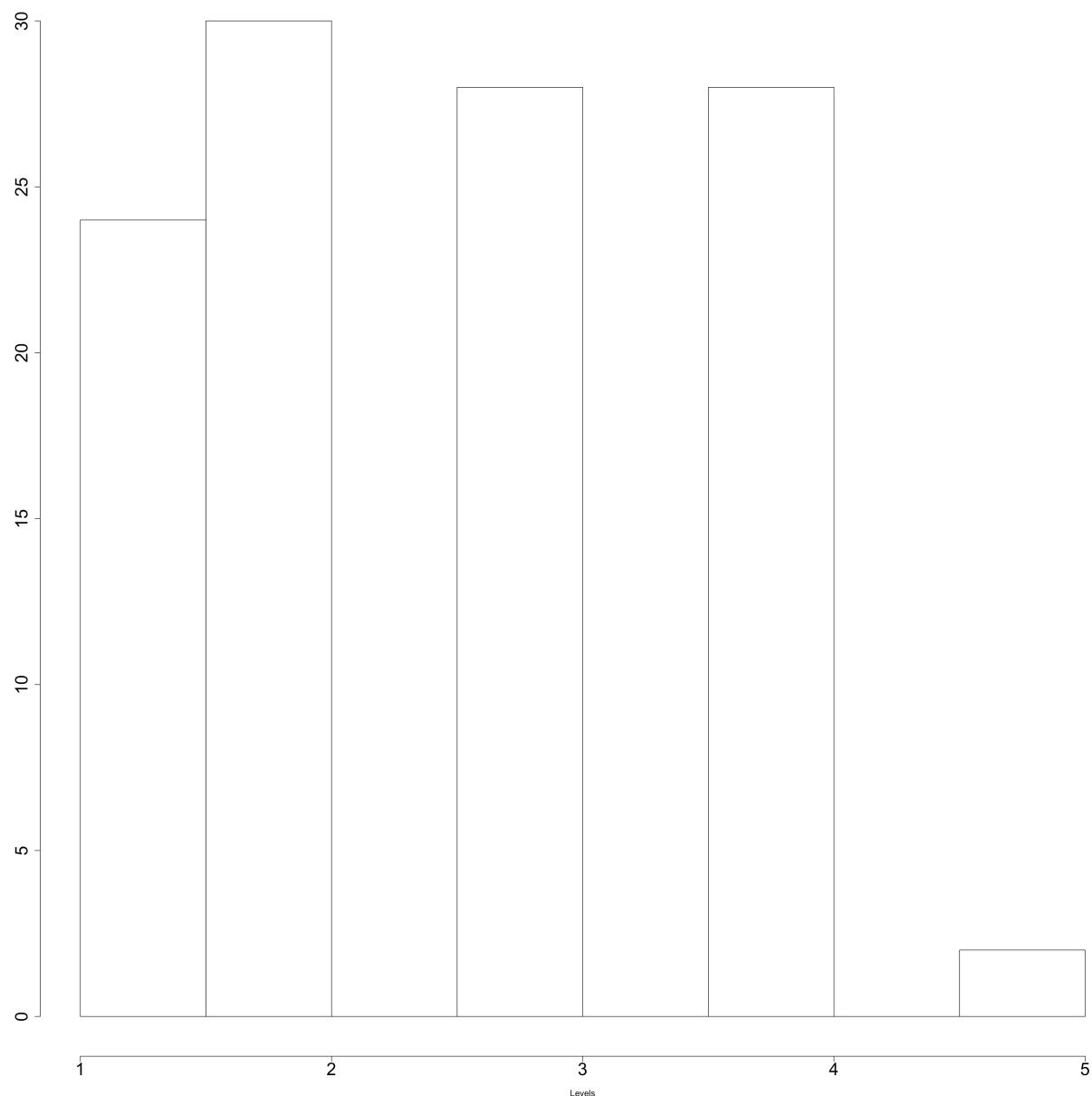






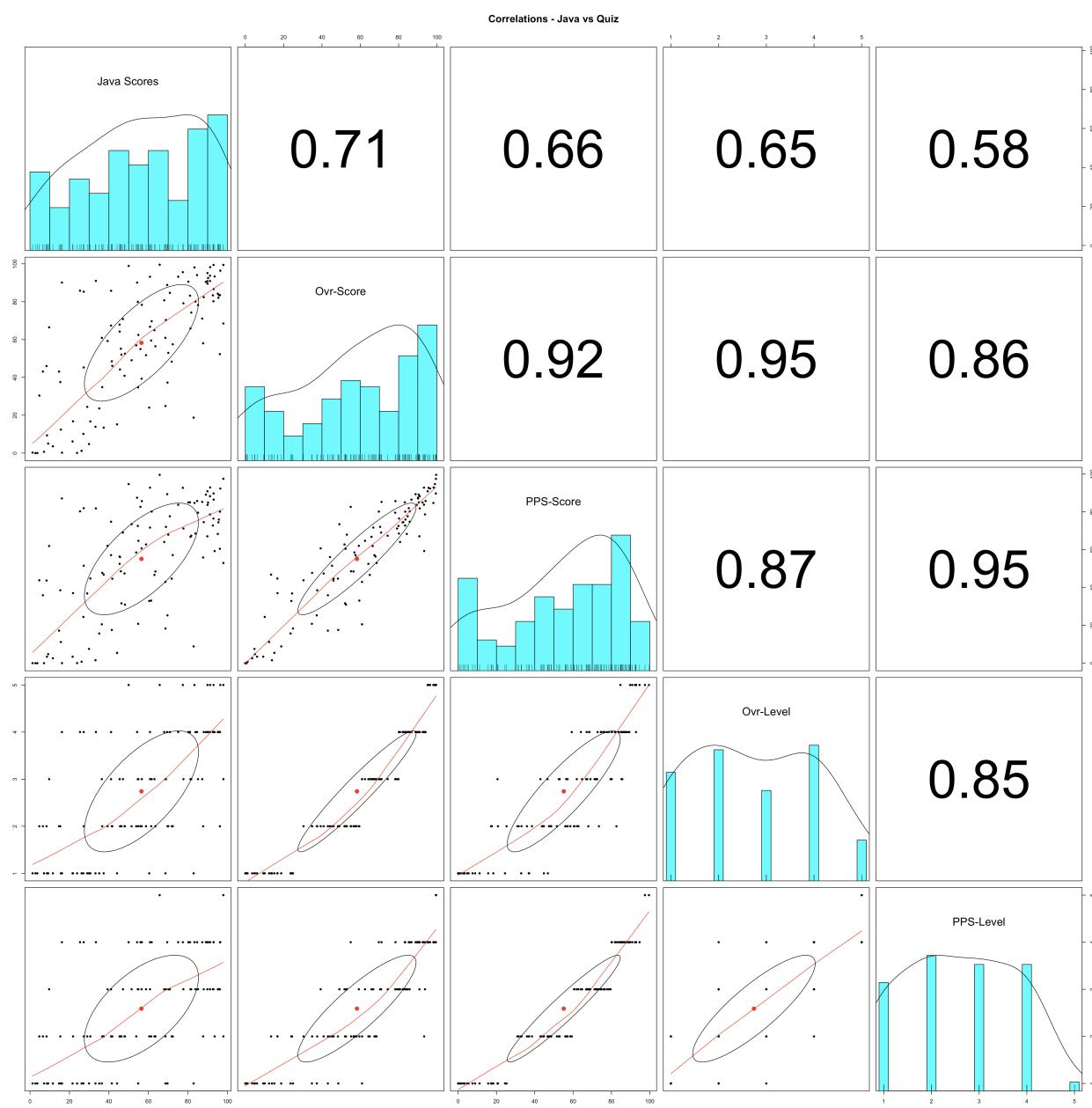
## Calibration Sample Characteristics

**Generalist - Overall Level**

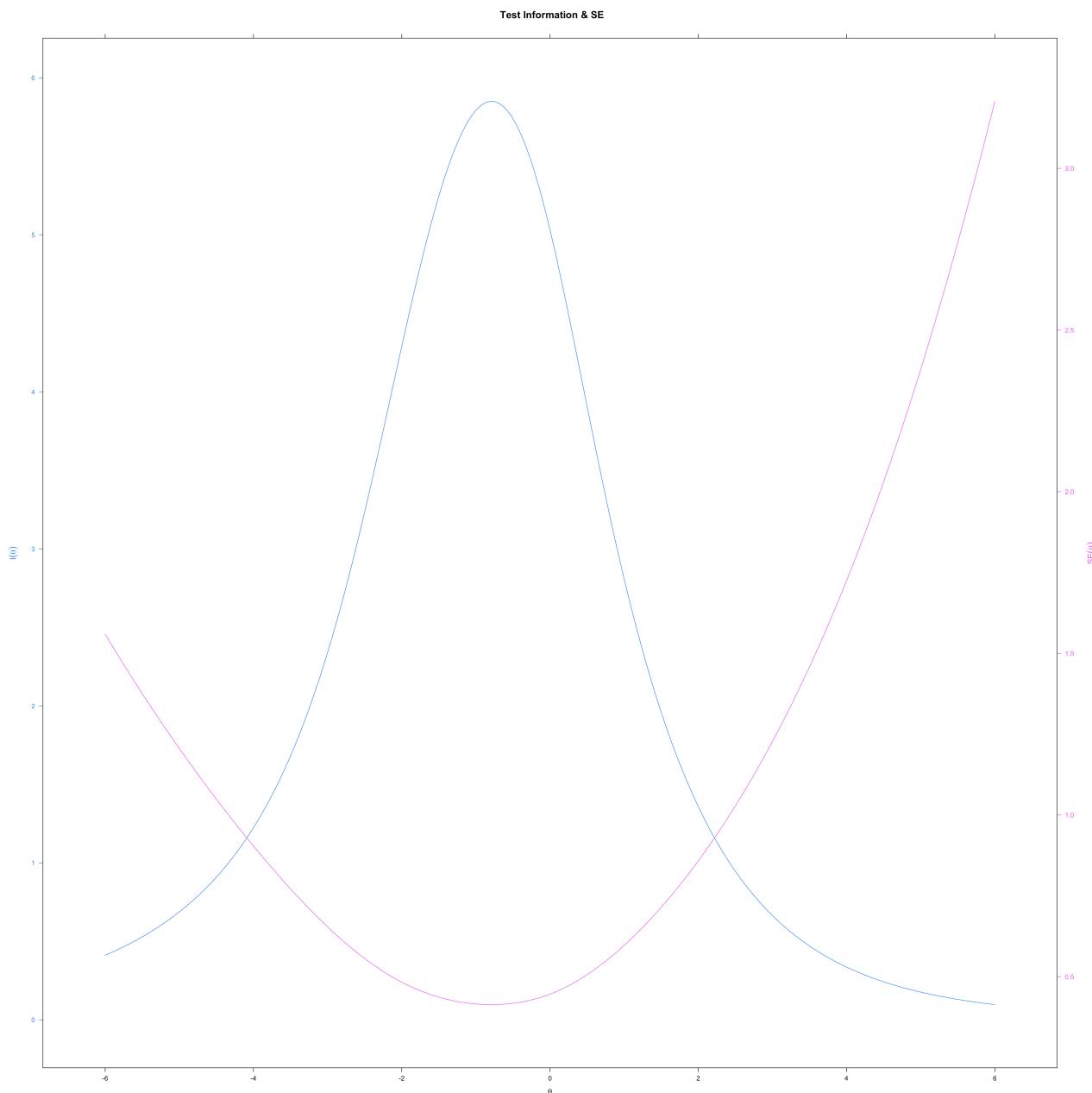
**Generalist - Prog Prob Solving Levels**

## Scoring\*

*Distributions & Correlations*



Avg. SE



## Equating/Linking Scores

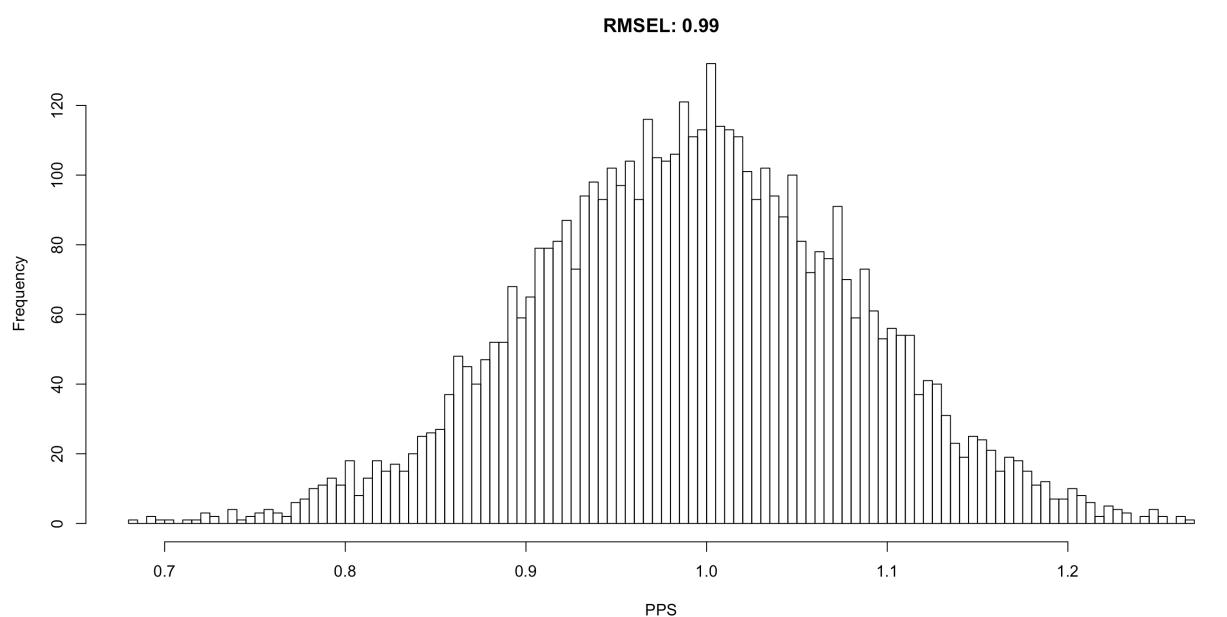
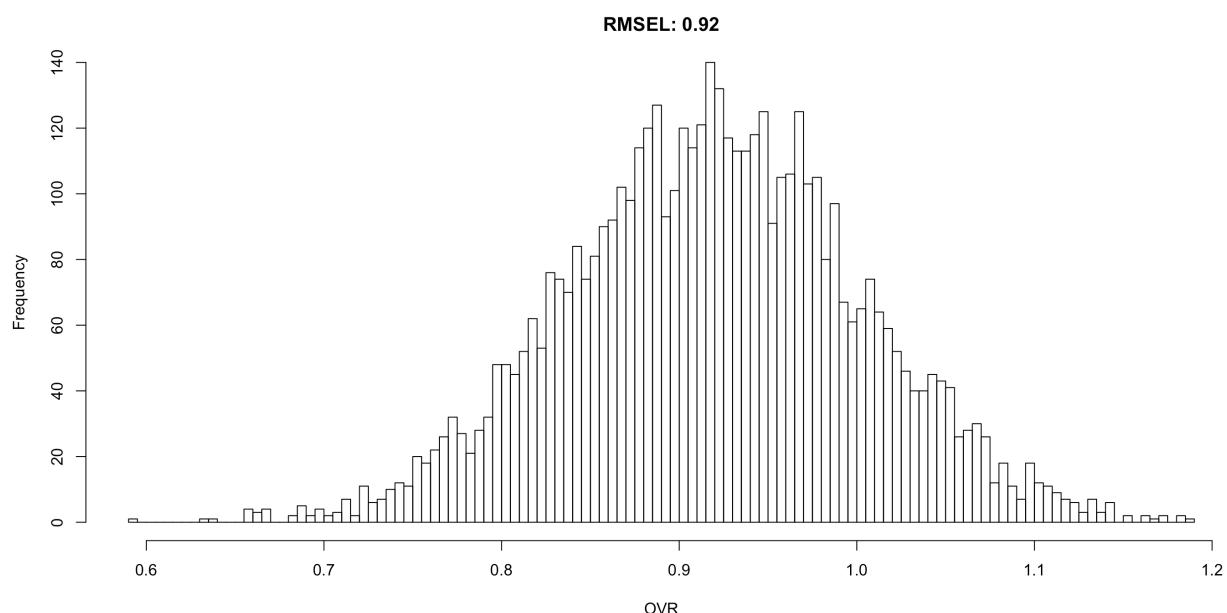
In order to get language assessment SkillEstimates on the same scale as the other quizzes, it requires being able to produce item parameters or IRT-ability estimates on the same scale as the other assessments.

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>

0.9189787

0.08280764

1 row

\$PPS

**avg**  
<dbl>**sd**  
<dbl>

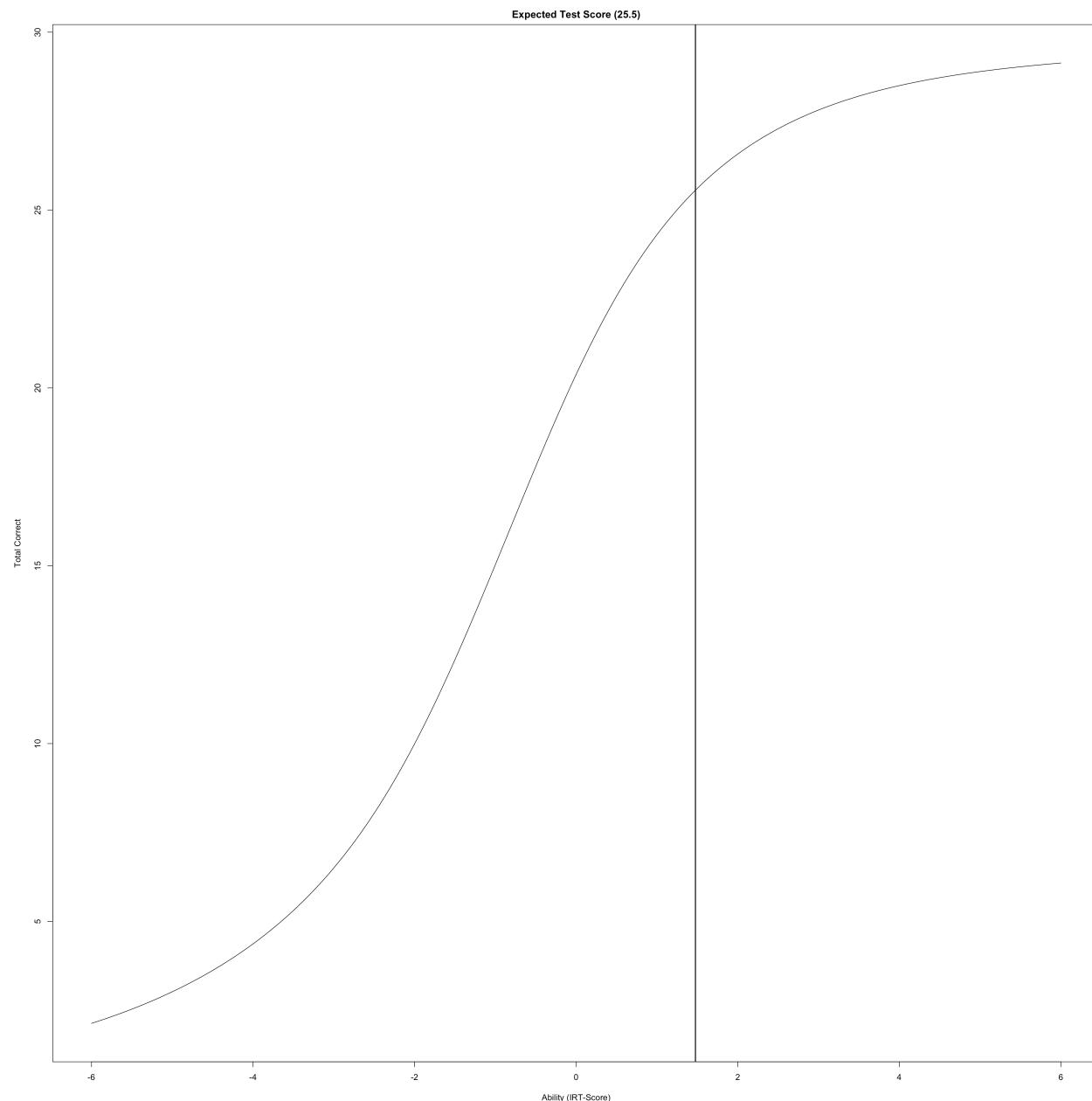
0.9924214

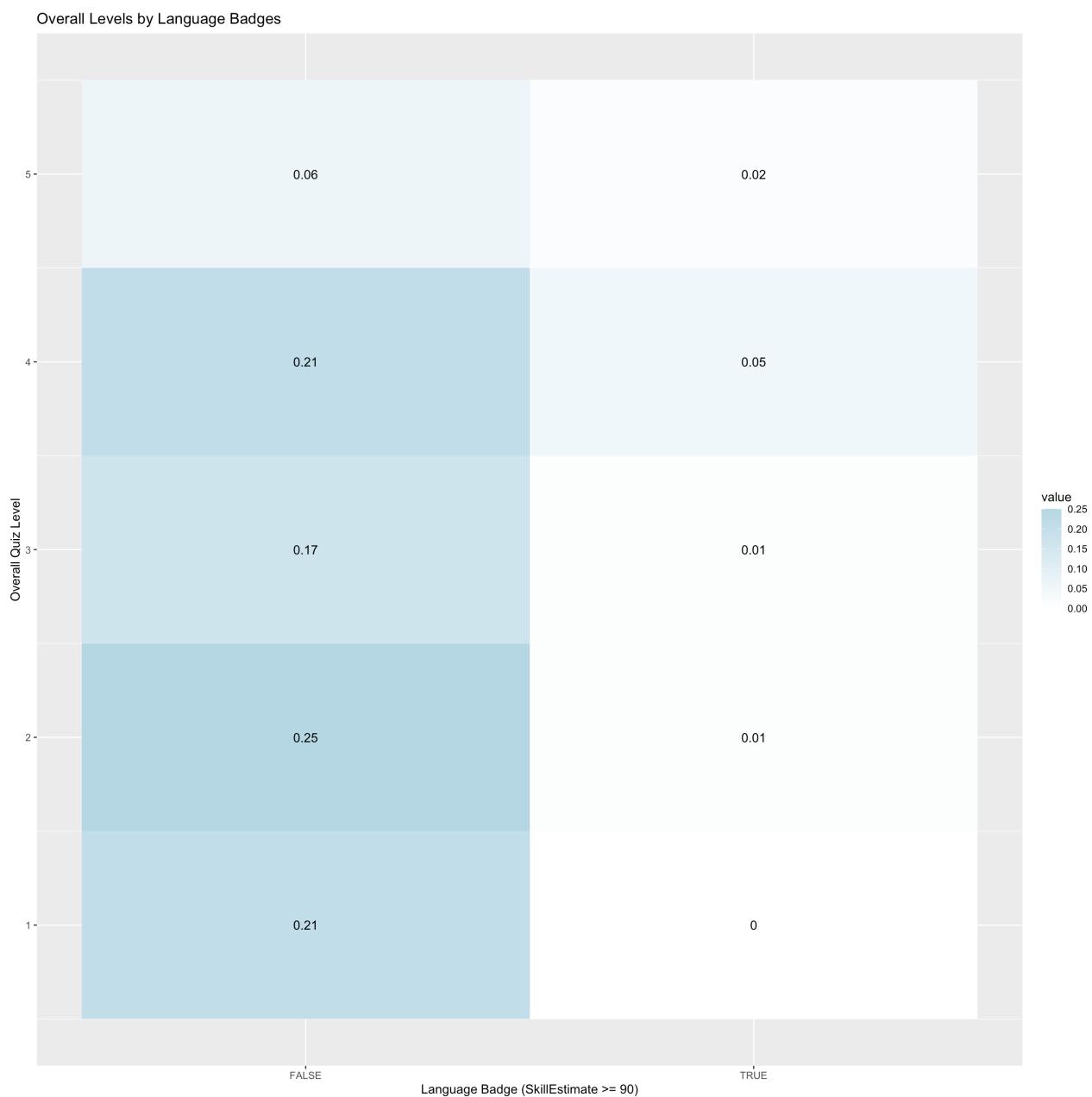
0.08910226

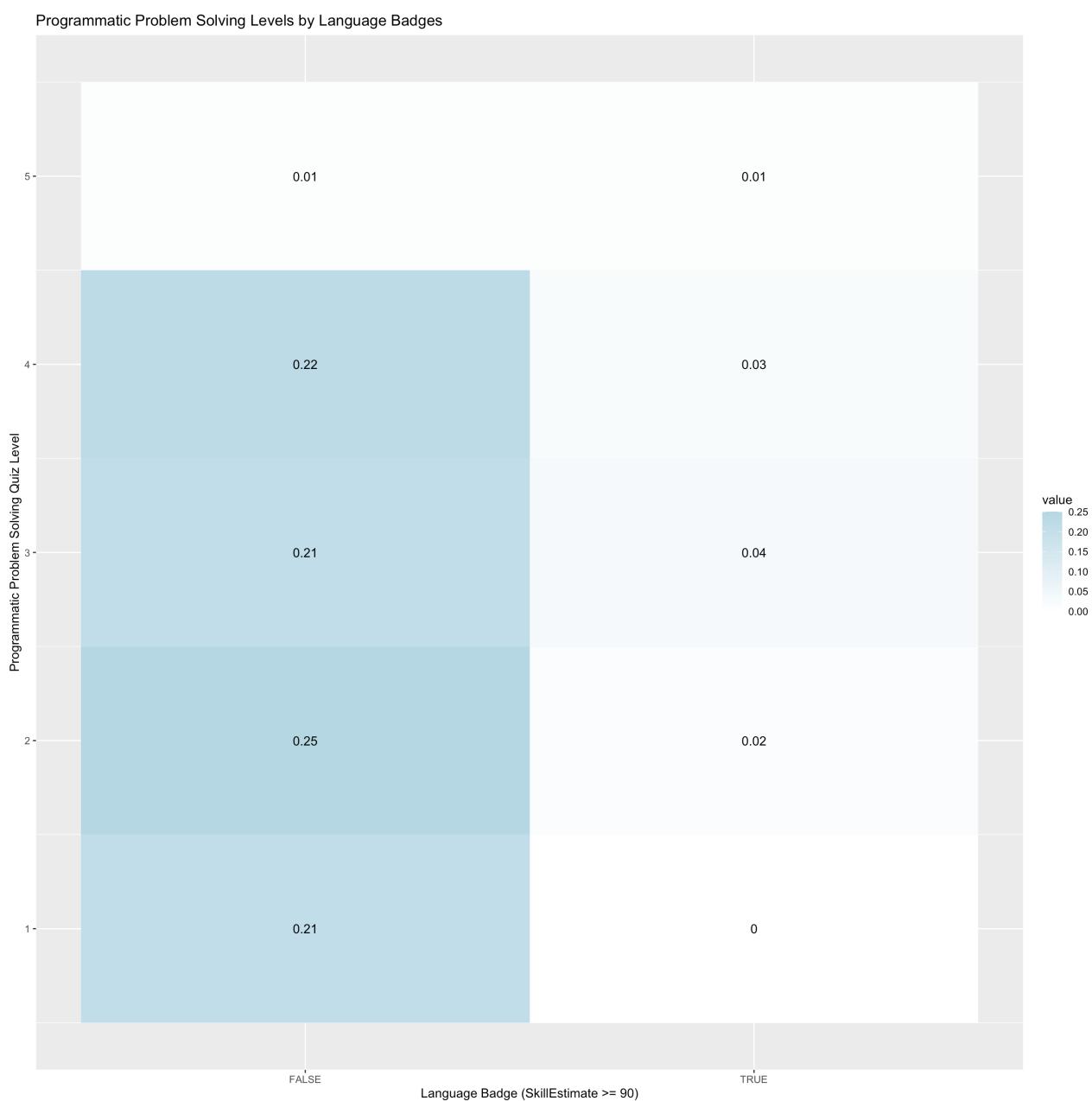
1 row

NA

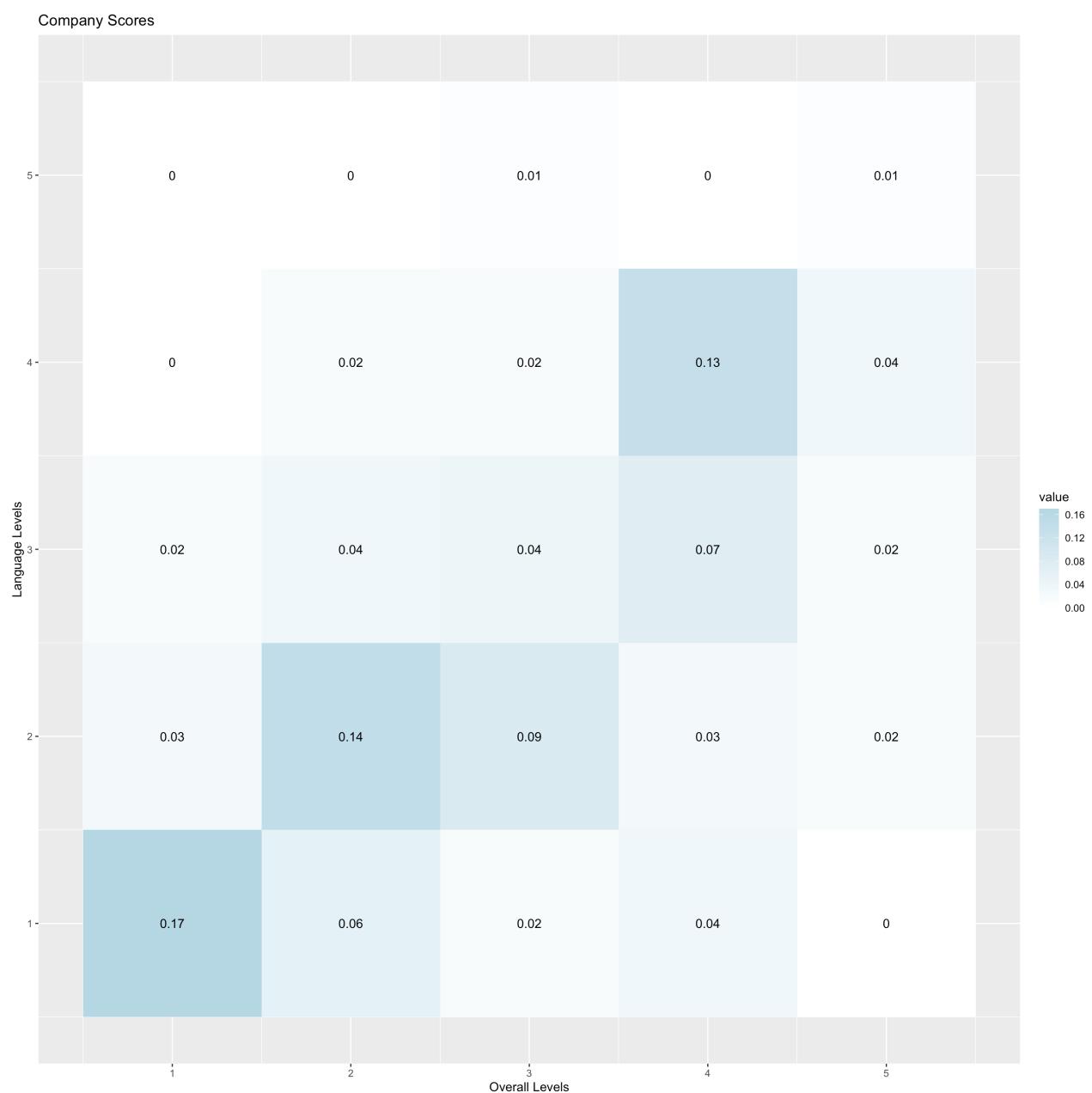
Result: Overall Scales score produce smaller linking error

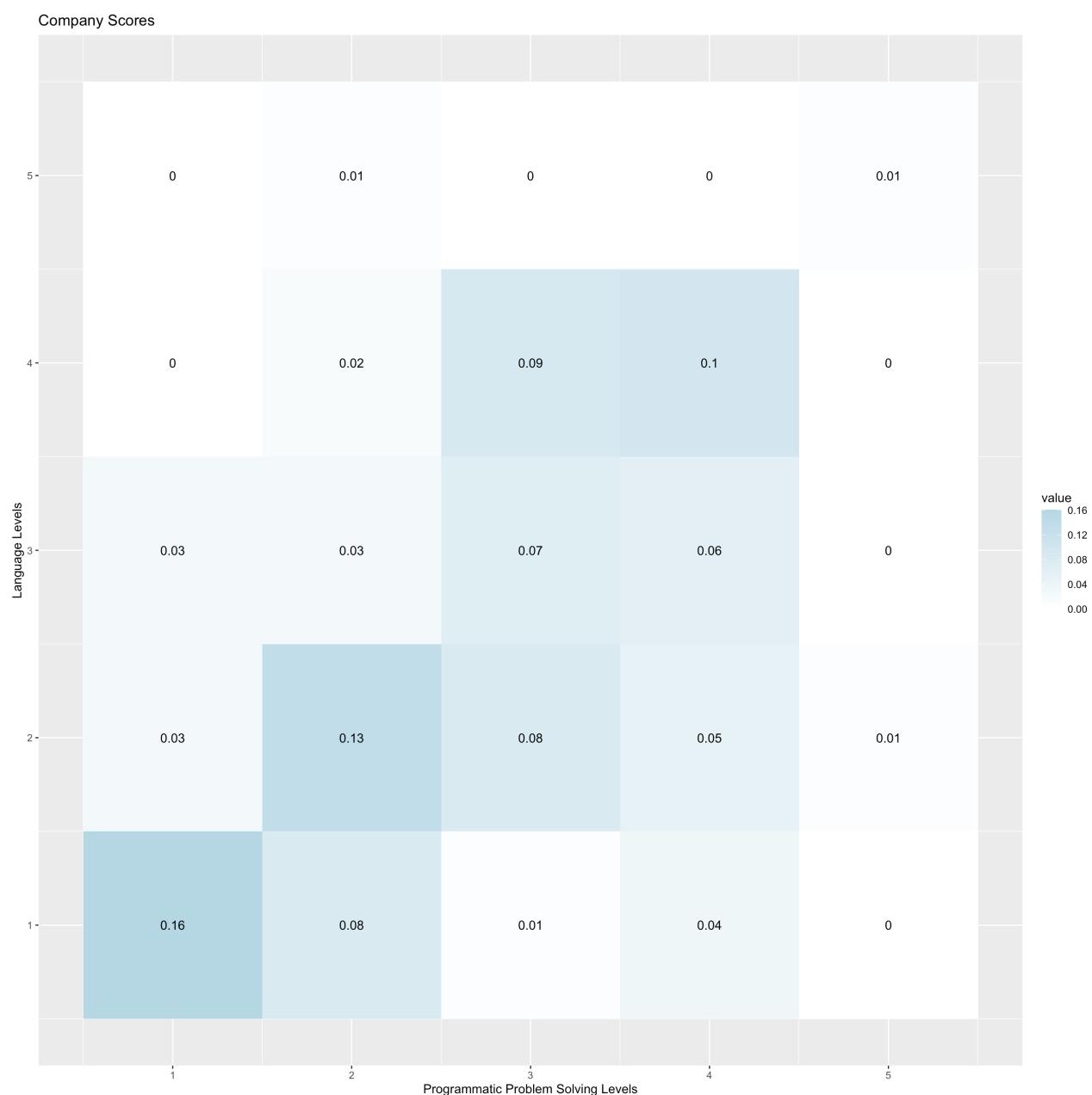
**Badges - Test Characteristic Curve****Badges - Quiz Level, by Badge**



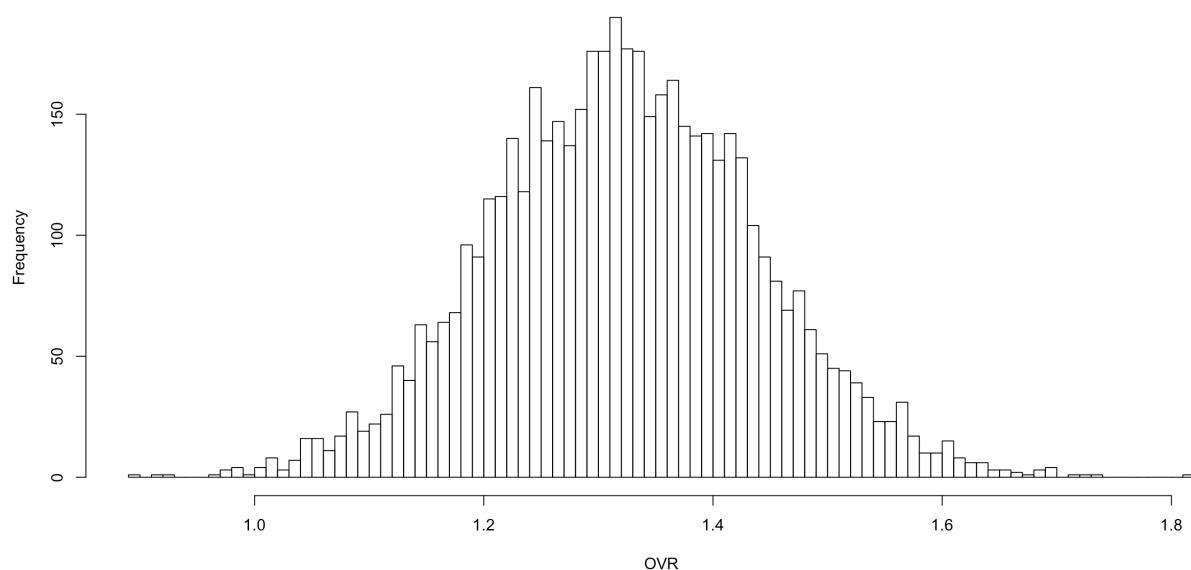
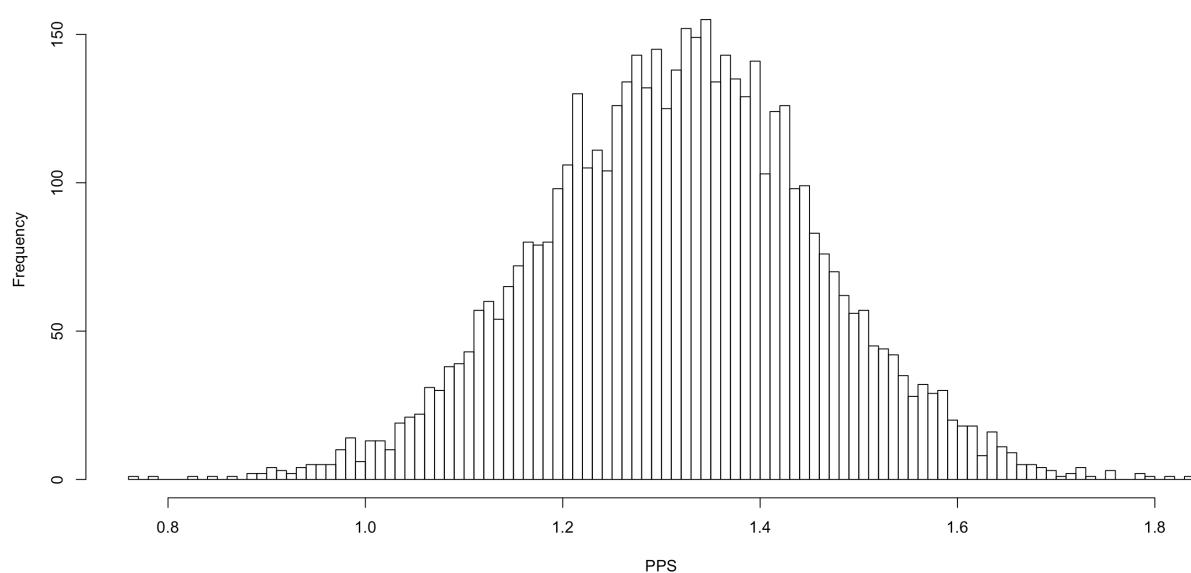


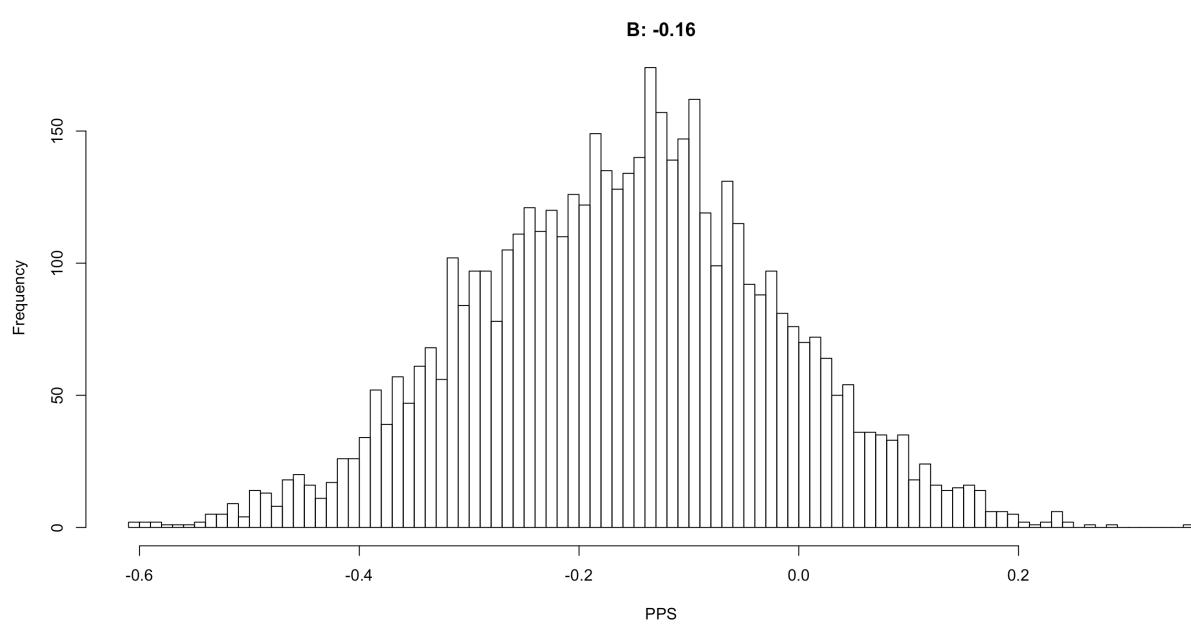
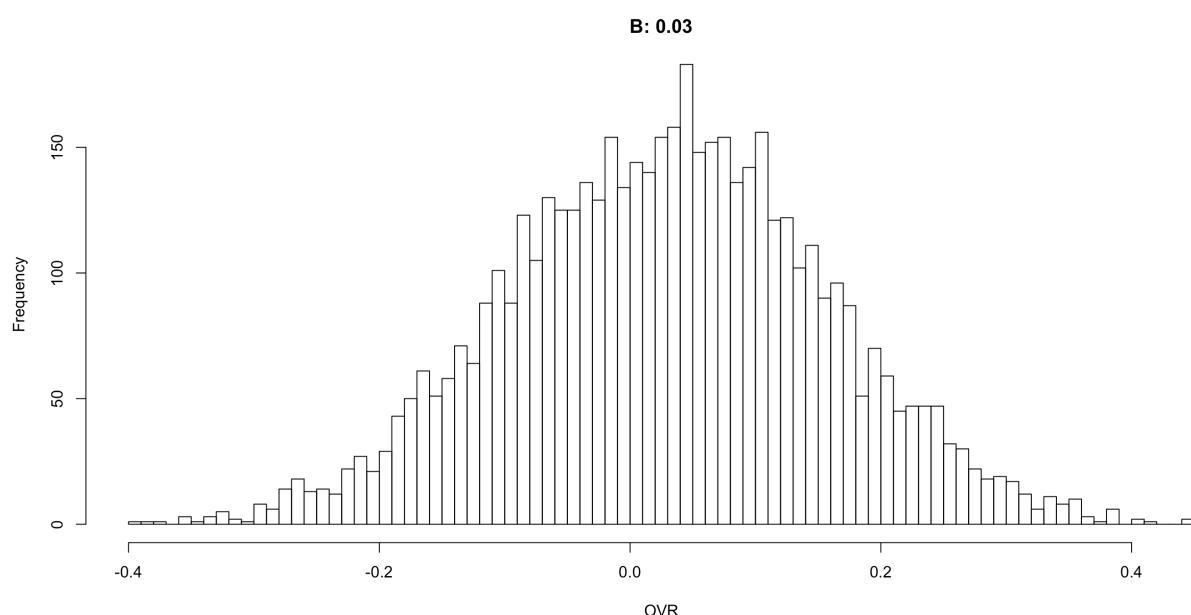
### Language Levels, Quiz Levels

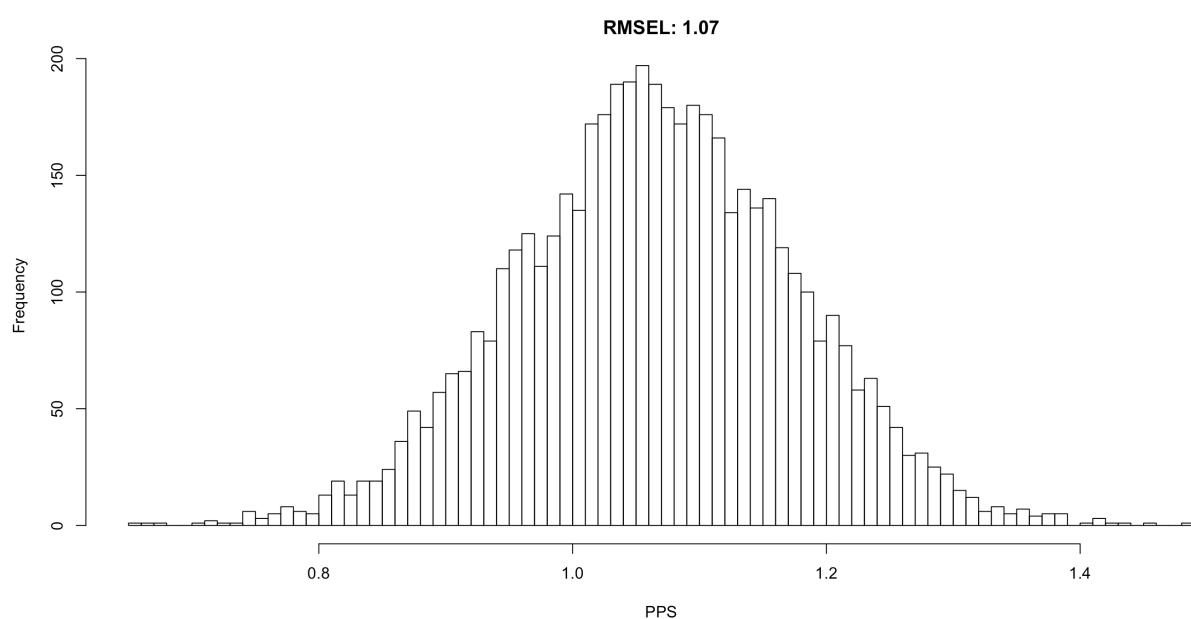
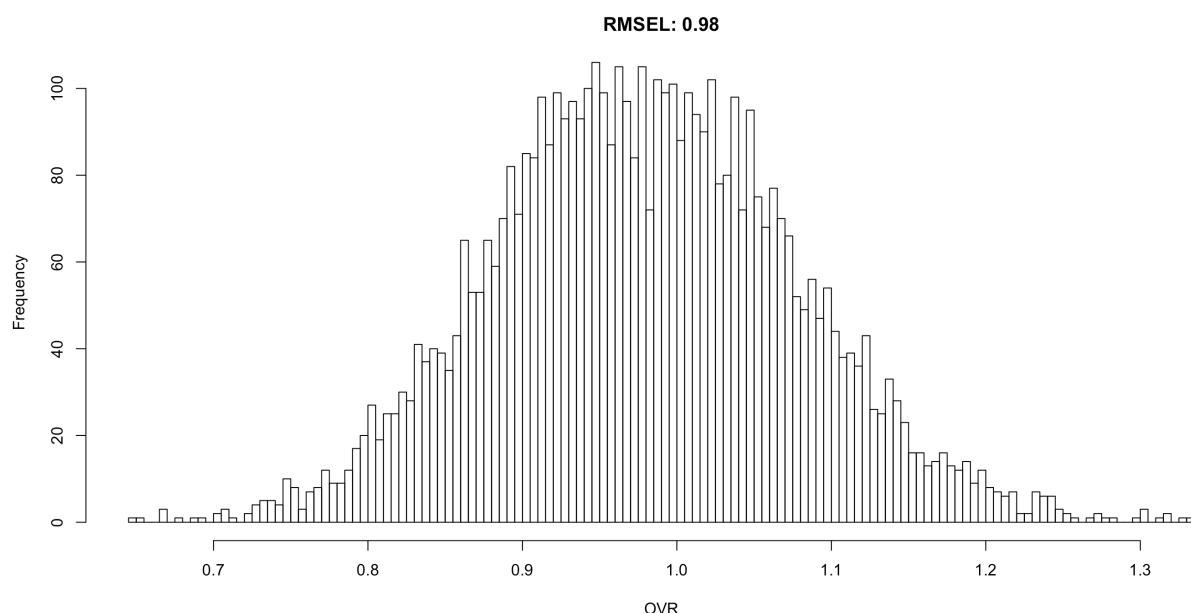




## Linear Method

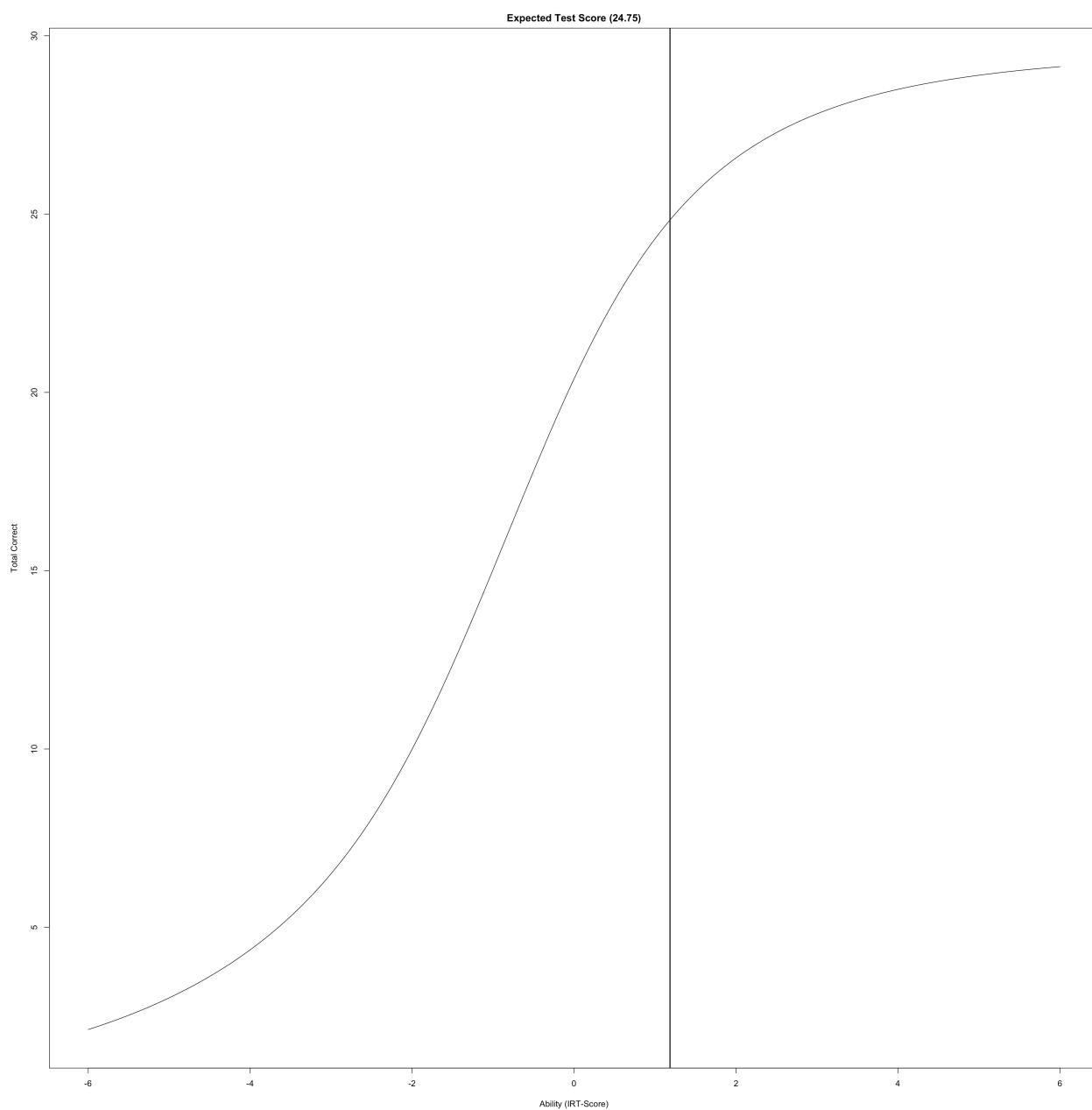
**A: 1.32****A: 1.32**

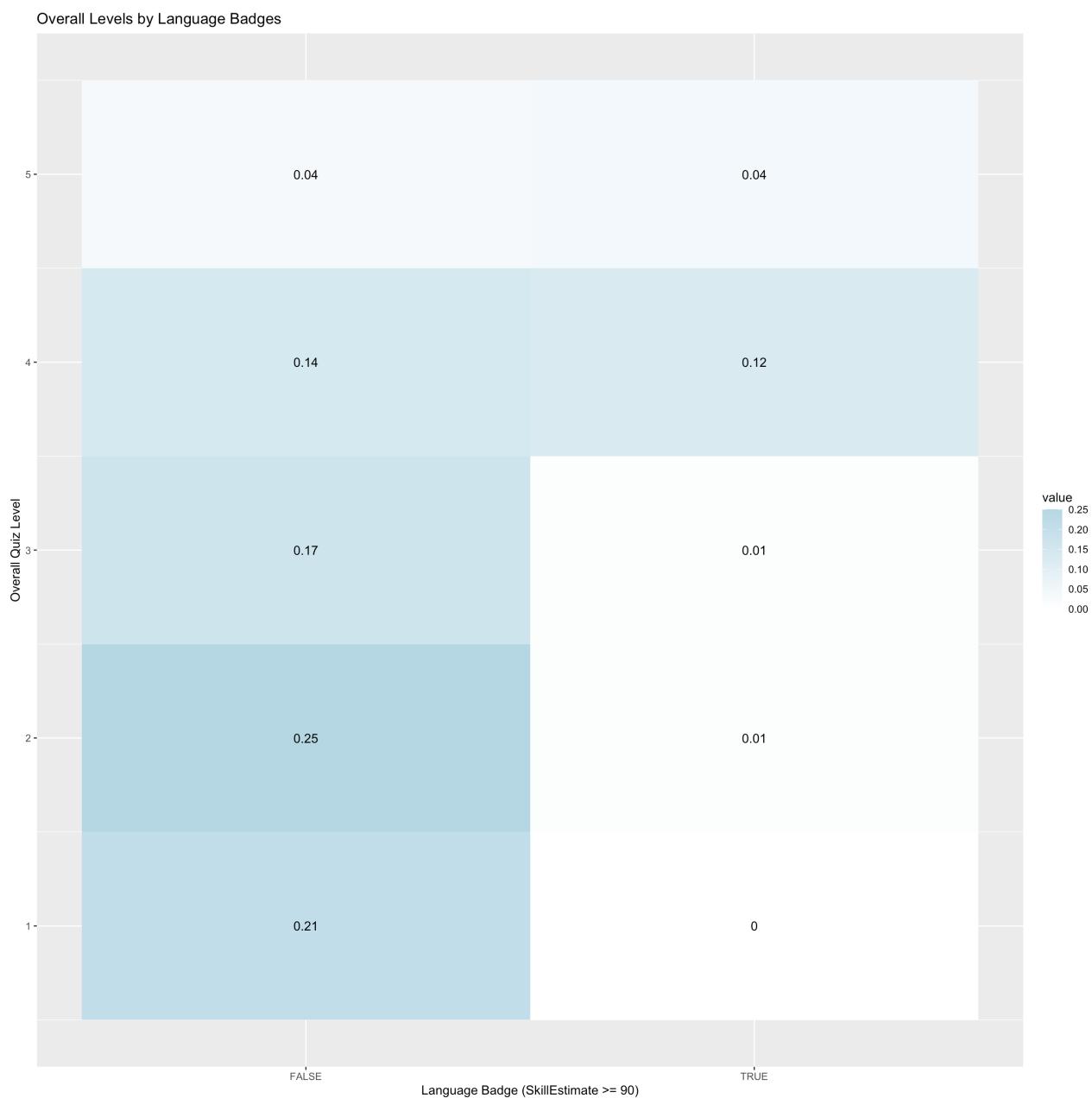


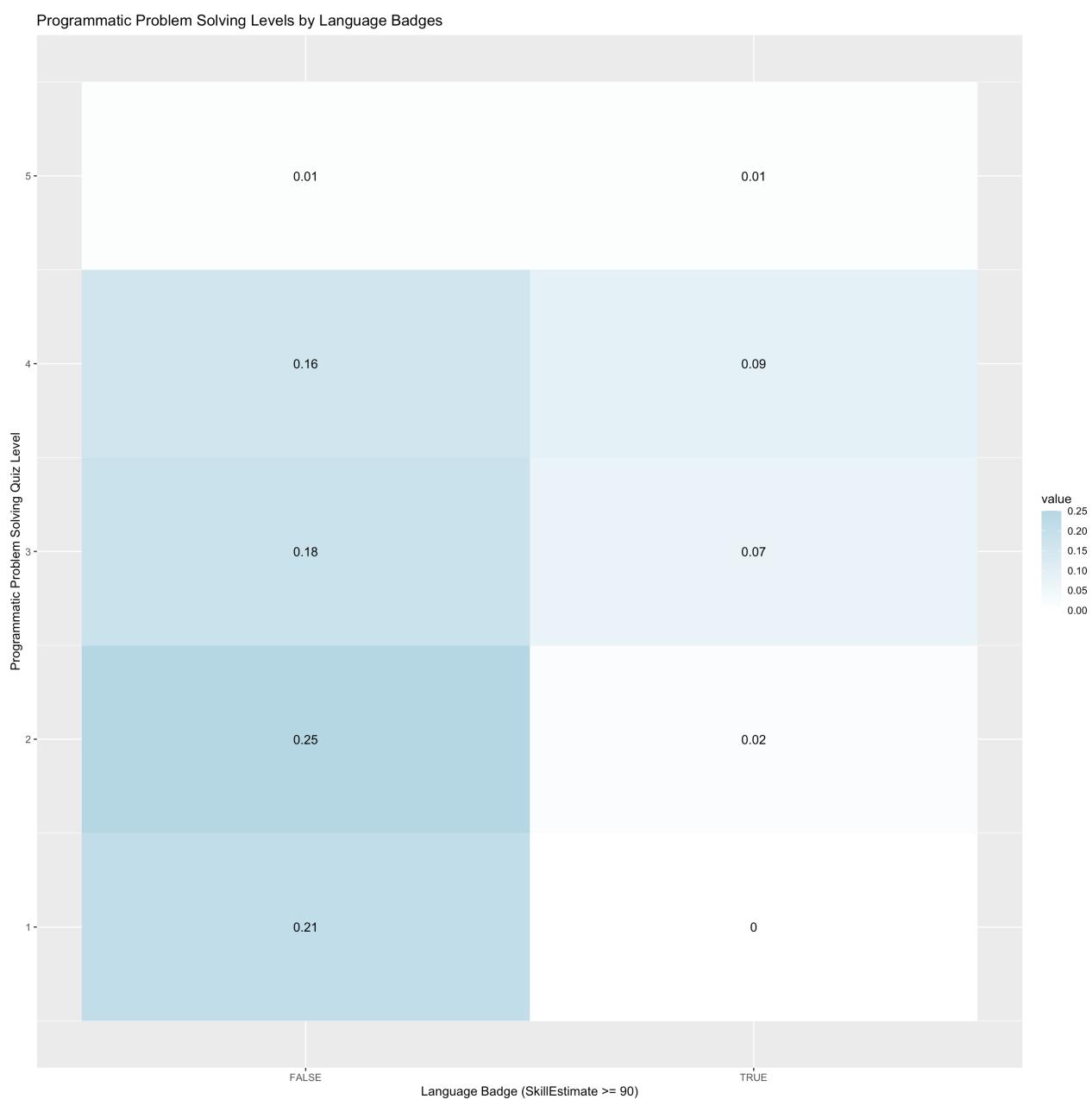


Result: Overall score produce smaller linking error

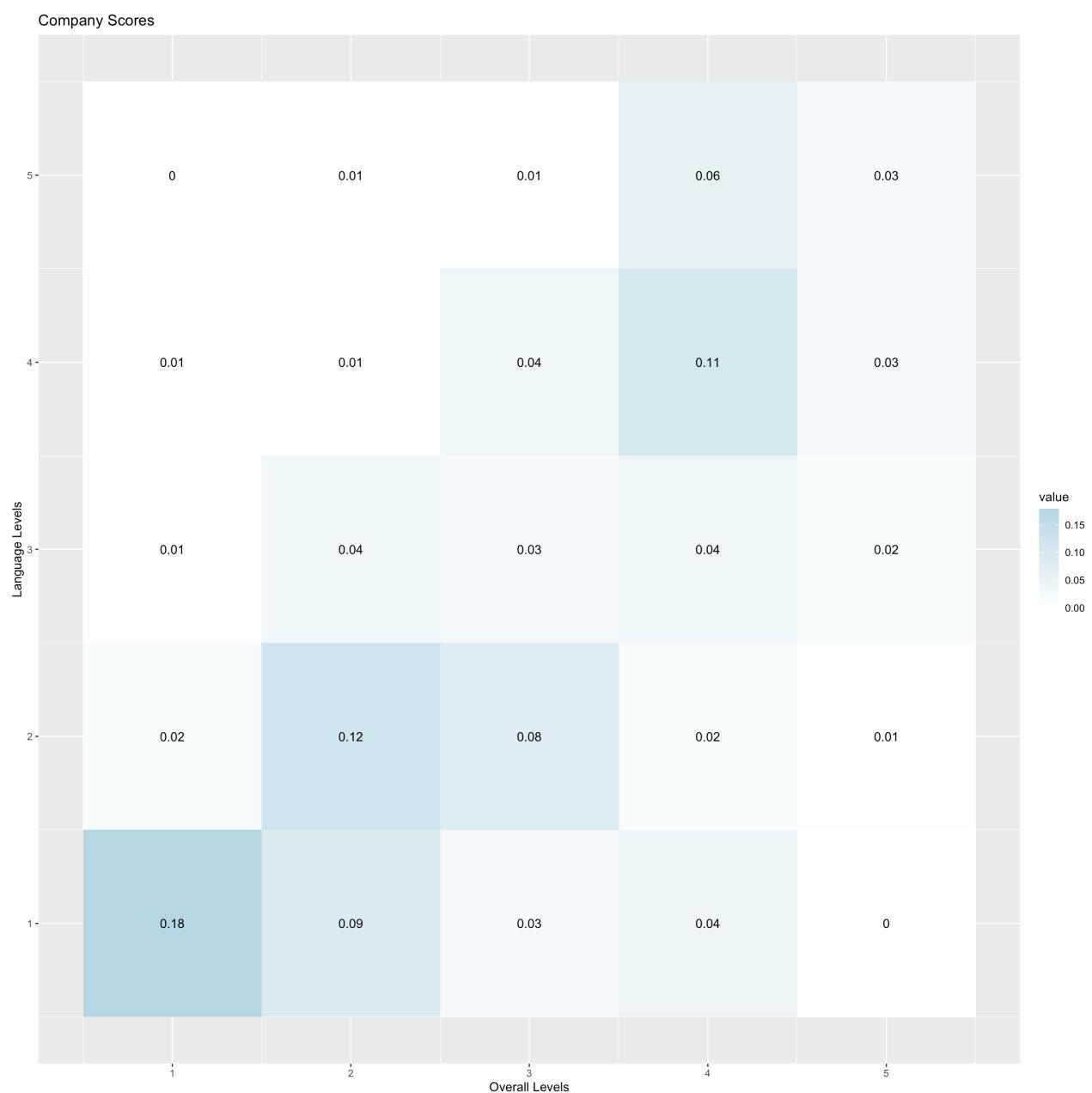
#### Badges - Test Characteristic Curve

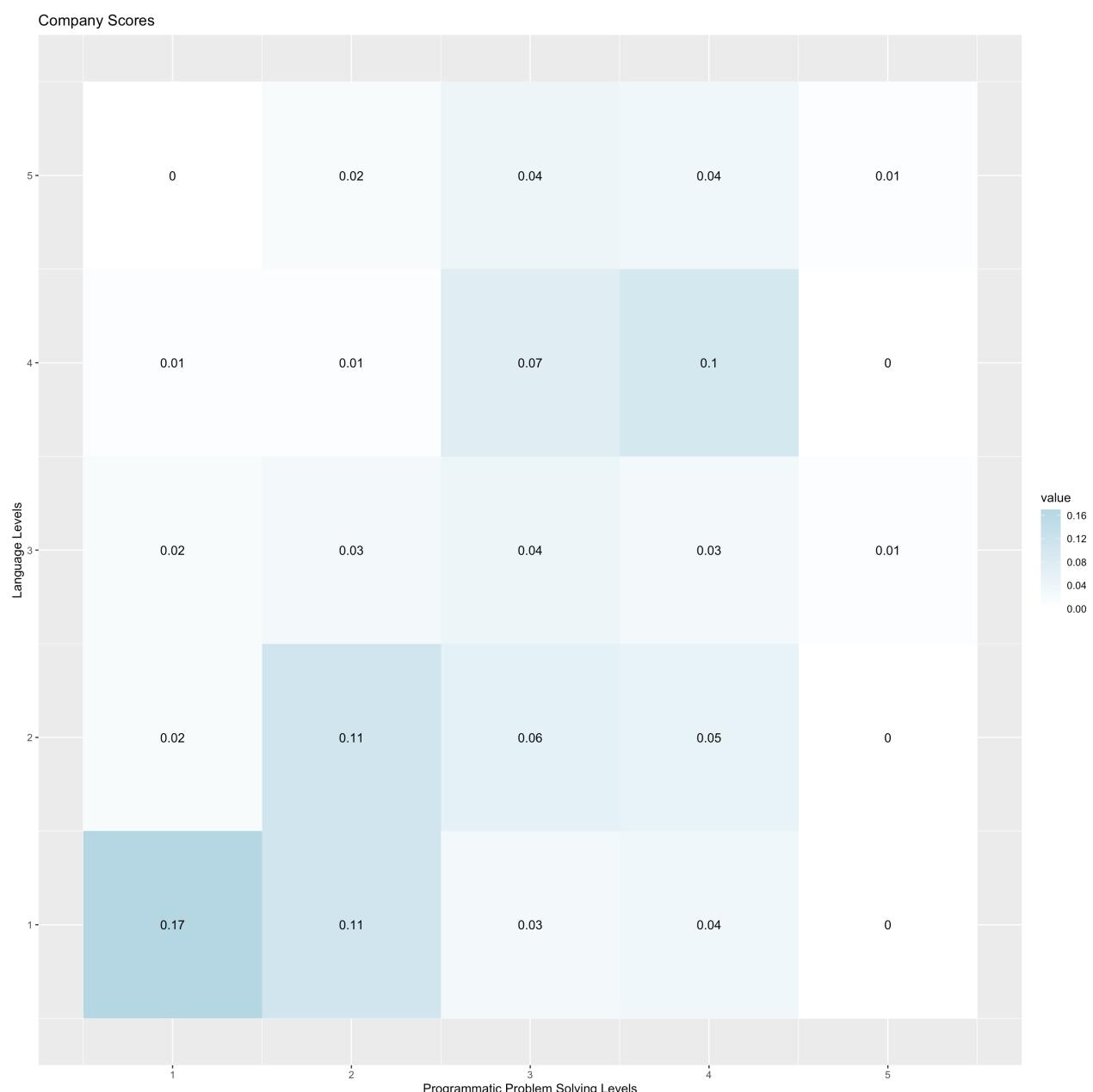
**Badges - Quiz Level, by Badge**





### Language Levels, Quiz Levels





## Percentile Method

### Summary