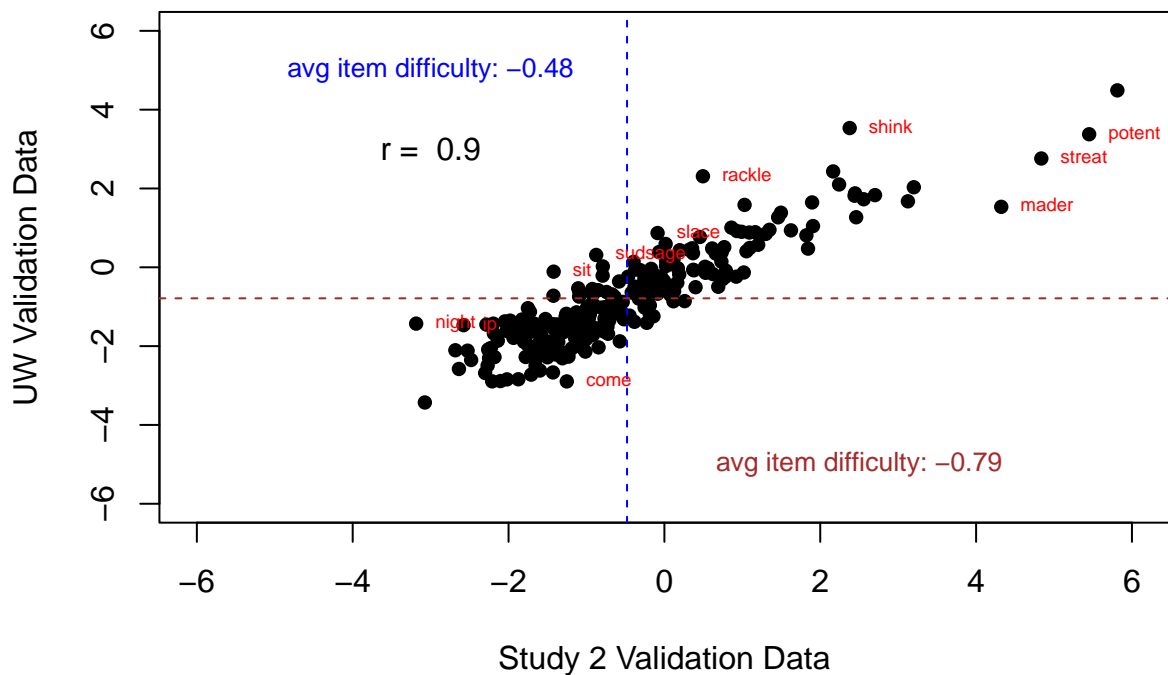# memo-item-difficulty

Item Difficulty Estimates

Previously, difficulty estimates were calibrated using response data from 185 respondents from Study 2. Here, I combine these 185 response patterns with those of the 124 participants from the UW study. The purpose of this work is to provide item parameter estimates that inform item selection in the adaptive testing version of the ROAR.

Right now (I think) adaptive testing is focused on items within Block C of the ROAR. As the staircase method is currently implemented, it relies on these items being classified into several different groups of items.

### Item Difficulty Estimates (Study 2 vs. UW)

When we compare the item difficulty estimates for these two models, we find that they are strongly correlated (r = .90), and there are only a few items that behave differently.

Combined Validation Data Item Estimates

The plots below present the distribution of item difficulty when the model was calibrated with all validation data (n = 309); first with all item difficulties combined, and then by block.

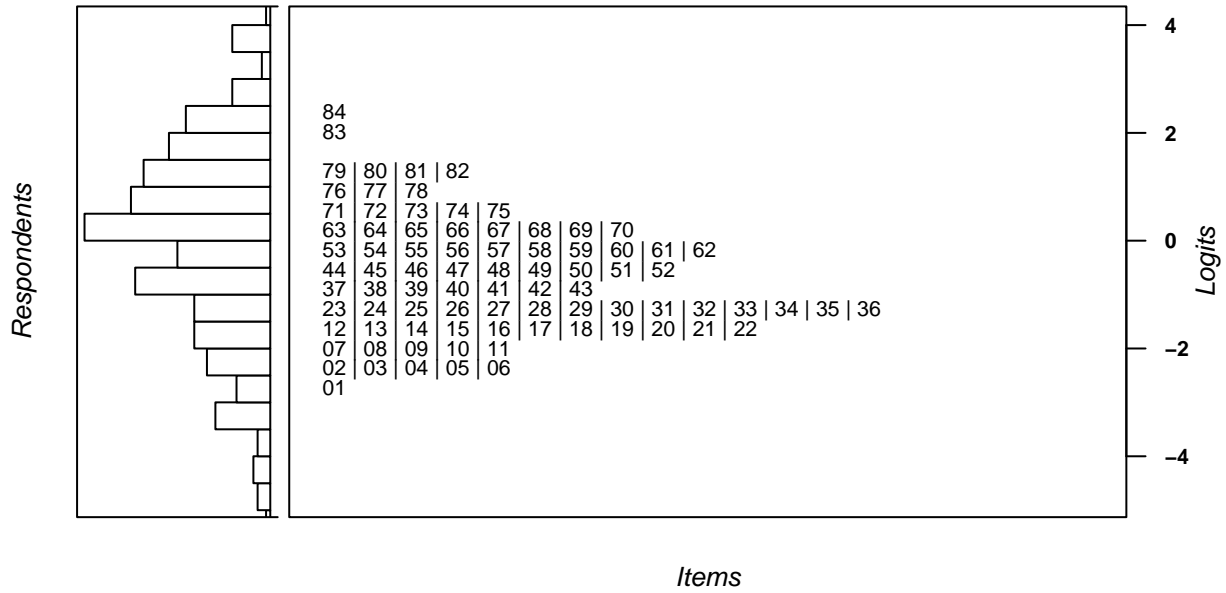**distribution of item difficulty estimates**
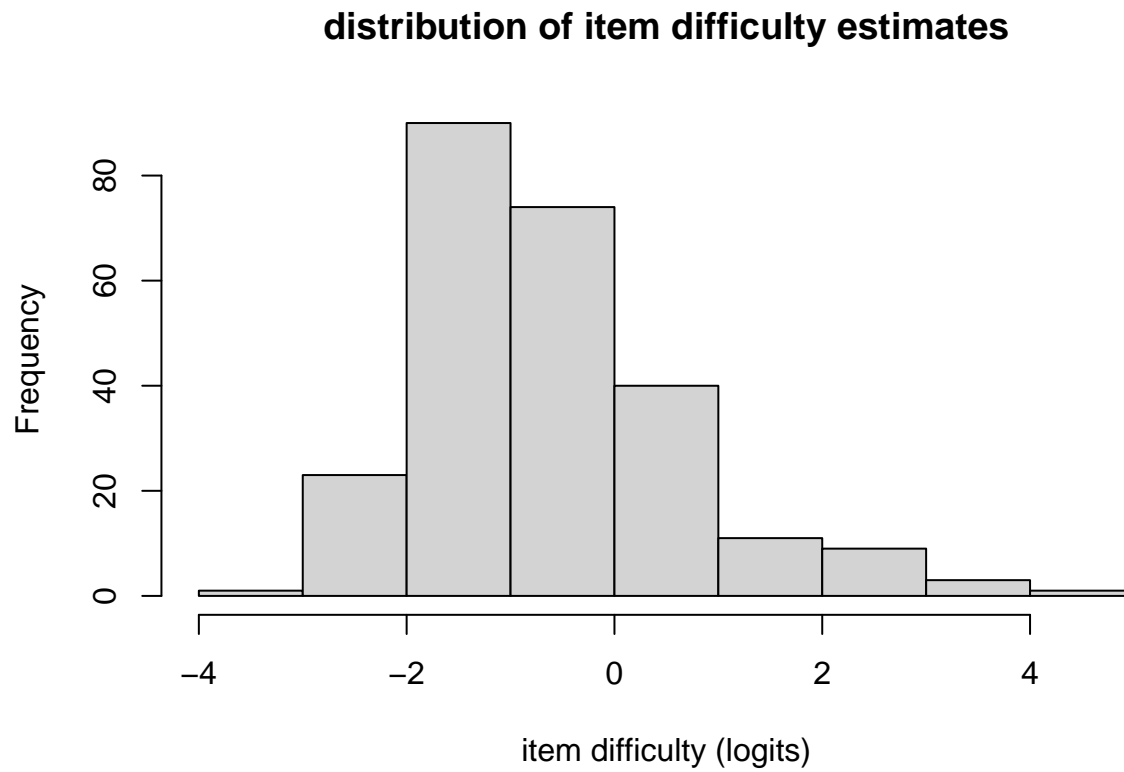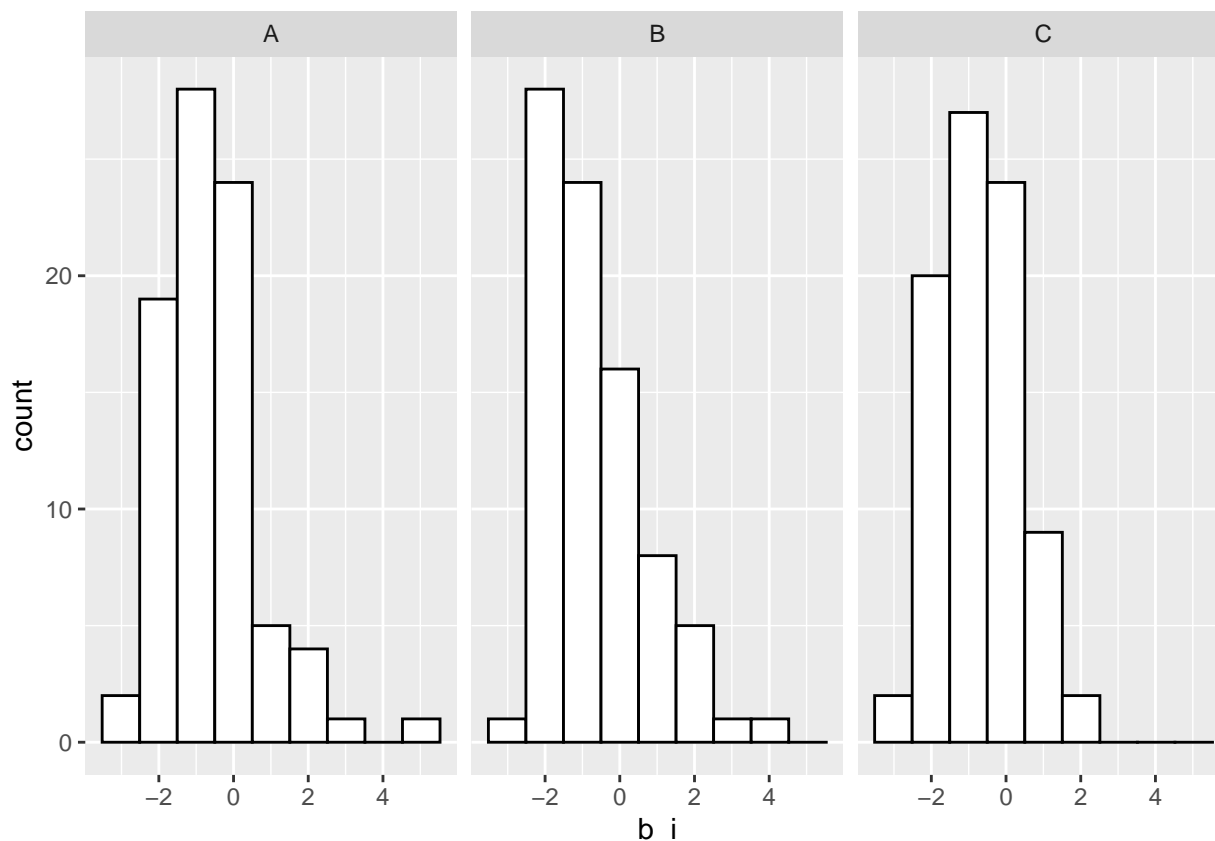
Item-Person Map for Block C

One approach for considering how to classify these words into different groups is to explore the distribution of item difficulties in the context of the distribution of participants' reading abilities. That is – since both the estimates of ability and difficulty are on the same scale (logits) – we can identify which items seem to be clustered together in different ranges of ability.

To examine the data in this way, we use an item-person map. On the left-hand side is the distribution of the ability estimates of the participants. At the top of the chart are those who have more reading ability, and towards the bottom of the chart are those with less ability. The average ability is 0 logits. The right-hand side of the chart represents the distribution of item difficulties. Note that we do not have items that are at the higher – or lower – levels of reading ability (something that we might want to work on)

Item–Person Map


```
                                                                          ─ 4

                        84
                        83                                                ─ 2

                        79 | 80 | 81 | 82
                        76 | 77 | 78
                        71 | 72 | 73 | 74 | 75
                        63 | 64 | 65 | 66 | 67 | 68 | 69 | 70
                        53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62    ─ 0
                        44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52
                        37 | 38 | 39 | 40 | 41 | 42 | 43
                        23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36
                        12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22
                        07 | 08 | 09 | 10 | 11                            ─ −2
                        02 | 03 | 04 | 05 | 06
                        01


                                                                          ─ −4
```

Respondents (left axis) · *Logits* (right axis) · *Items* (bottom)

Below are some proposed classifications of students (and justifications) **but these should be reviewed with the group.**

These word clusters are based on items that are likely to be endorsed by the following sets of respondents:

- Those with the lowest ability for which we have items (22 words)

- Those with ability below average (40 words)

- Those with ability at or slightly above average (13 words)

- Those with the highest reading ability (9 words)

Item Group 1

- Respondents lower on the ability scale (-2.6 to -1.86) are likely to get the following items correct:

  - people, hello, good, hot, open, look, watch, money, round, more, and name.

  Since all of these are real words, we will need to add in the next most difficult set of words (-1.85 to -1.51):

  - hungry, short, him, cok, gurynh, candy, morning, khtna, orn, almost, and che

Items Group 2

- Respondents that are closer to average ability (but still below average) are likely to get the following item correct (-1.45 to -.0.03)

  - begin, through, garden, complete, daople, alk, eamn, guoys, loleh, sick, oban, twew, utashing, pretty, phlere, ap, history, imagine, rebosks, cas, question, collosks, brubs, cotosks, distance, agreed, dogo, albust, chance, runey, hane, compliment, investigate, liffs, fafever, pame, duttery, scays, molent

Items Group 3

- Respondents that are close to average ability are likely to get the following items correct (.05 - .63 logits)

  - swane, banny, fistory, pealt, contain, grasshopper, lame, woft, drustion, gravel, slace, overwhelm, naw

Items Group 4

- Respondents that the highest abilities are likely to get the following items correct (.94 - 2.36)

  - an, snose, freat, chanding, configuration, rackle, qualify, appropriates, garment

Legend

```
## [1] "1 : people"
## [1] "2 : hello"
## [1] "3 : good"
## [1] "4 : hot"
## [1] "5 : open"
## [1] "6 : look"
## [1] "7 : watch"
## [1] "8 : money"
## [1] "9 : round"
## [1] "10 : more"
## [1] "11 : name"
## [1] "12 : hungry"
## [1] "13 : short"
## [1] "14 : him"
## [1] "15 : cok"
## [1] "16 : gurynh"
## [1] "17 : candy"
## [1] "18 : morning"
## [1] "19 : khtna"
## [1] "20 : orn"
## [1] "21 : almost"
## [1] "22 : che"
## [1] "23 : begin"
## [1] "24 : through"
## [1] "25 : garden"
## [1] "26 : complete"
## [1] "27 : daople"
## [1] "28 : alk"
## [1] "29 : eamn"
## [1] "30 : guoys"
## [1] "31 : loleh"
## [1] "32 : sick"
## [1] "33 : oban"
## [1] "34 : twew"
## [1] "35 : utashing"
## [1] "36 : pretty"
## [1] "37 : phlere"
## [1] "38 : ap"
## [1] "39 : history"
## [1] "40 : imagine"
## [1] "41 : rebosks"
## [1] "42 : cas"
## [1] "43 : question"
## [1] "44 : spy"
## [1] "45 : collosks"
## [1] "46 : brubs"
## [1] "47 : cotosks"
## [1] "48 : distance"
## [1] "49 : agreed"
## [1] "50 : dogo"
## [1] "51 : albust"
## [1] "52 : chance"
```

```
## [1] "53 : runey"
## [1] "54 : hane"
## [1] "55 : compliment"
## [1] "56 : investigate"
## [1] "57 : liffs"
## [1] "58 : fafever"
## [1] "59 : pame"
## [1] "60 : duttery"
## [1] "61 : scays"
## [1] "62 : molent"
## [1] "63 : swane"
## [1] "64 : banny"
## [1] "65 : fistory"
## [1] "66 : pealt"
## [1] "67 : contain"
## [1] "68 : grasshopper"
## [1] "69 : lame"
## [1] "70 : woft"
## [1] "71 : drustion"
## [1] "72 : gravel"
## [1] "73 : slace"
## [1] "74 : overwhelm"
## [1] "75 : naw"
## [1] "76 : an"
## [1] "77 : snose"
## [1] "78 : freat"
## [1] "79 : chanding"
## [1] "80 : configuration"
## [1] "81 : rackle"
## [1] "82 : qualify"
## [1] "83 : appropriates"
## [1] "84 : garment"
```

Note: the phrase "are likely to get the items correct" instead of have a 50% chance of getting these items correct" is because of the fixed guessing parameter of .5. As such the probability of getting the item correct is the average of the max (1) and the min (.05) = .75.