



Swansea University  
Prifysgol Abertawe

# Making a LLAMA into an ALPACAA: A revised attempt at assessing aptitude

Vivienne Rogers, Paul Meara, Amy Adlam-Hill,  
Megan Davies & Dafydd Llewellyn-Williams



Swansea University  
Prifysgol Abertawe

## Outline

- Background on Aptitude and LLAMA tests
- Methodology: how we've revised the tests
- Results & Discussion
- Conclusion







Swansea University  
Prifysgol Abertawe

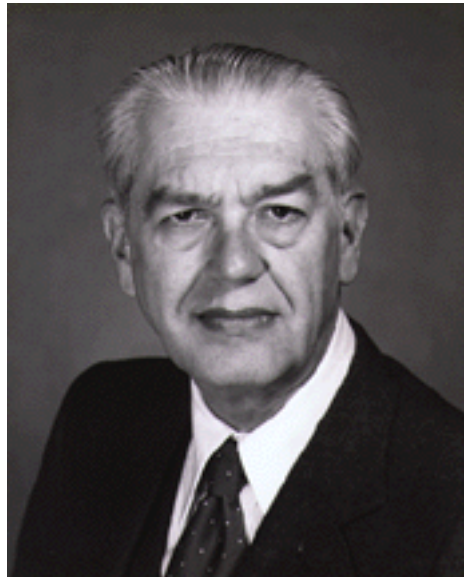
# Background: Aptitude & LLAMA tests



Swansea University  
Prifysgol Abertawe

# What is Language Learning Aptitude

“the amount of time a student needs to learn a given task, unit of instruction, or curriculum to an acceptable criterion of mastery under optimal conditions of instruction and student motivation.”  
(Carroll 1990 p. 26)



- A 'knack for learning languages'. A cognitive variable - something you are born with.
- What does it mean?
- aptitude is different from other cognitive systems, including intelligence
- aptitude is stable (doesn't change)
- aptitude is made up of different components



# WM as L2 aptitude?

- Wen (2016, p. 142)
  - ❖ “to what extent [can] PWM... complement (or even replace) the phonetic coding ability of language aptitude and, similarly, to what extent EWM can outperform the language analytical ability of language aptitude.”
  - ❖ “premature... to claim that WM ‘replaces’ L2 aptitude given our currently limited knowledge of their relationship and relatively scarce empirical evidence”
  - ❖ PWM = language learning device
  - ❖ EWM = language processes
  - ❖ (but see also Baddeley 2003 a & b, 2017; Winke, 2013)

## Li (2015) Construct validity: meta analysis of 66 studies.

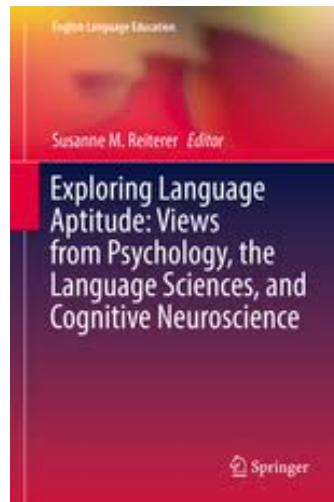
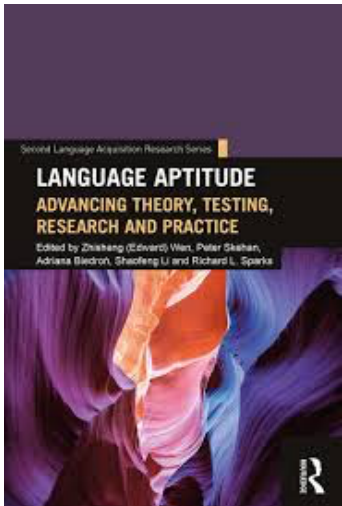


Swansea University  
Prifysgol Abertawe

- Aptitude is independent of other individual differences, e.g. motivation.
- executive working memory (EWM) more strongly associated with aptitude than phonological short-term memory (PSTM).
  - BUT Linck et al (2013): relevance of PSTM to advanced learners.
- strong predictor of general proficiency but not vocabulary learning or L2 writing.
- different components predicted different aspects of learning.
- negative correlation between anxiety and aptitude.
  - Sparks & Patton (2013): anxiety as result not cause of low aptitude
- Granena (2013): LLAMA tests measure 2 different constructs:
  - Implicit (sound recognition task) & explicit (other three tasks)

# • Resurgence in interest

- Over 700 citations on google scholar published since 2015!
- Including 115 in 2019.
- Problem: how to test?
- 82 citations for LLAMA (Meara 2005)



Swansea University  
Prifysgol Abertawe

- “A rather recent and very useful language aptitude test is the LLAMA (Meara, 2005). ... It has certainly gained popularity and, as Granena (2013) points out, only the LLAMA test does not suffer from any limitation or restriction, e.g. being difficult to get, being available only in pencil-and-paper format or only being used for military purposes. ... The LLAMA test is also the test that will appear most frequently in this volume.”
- Ameringer et al (2018, p.27)



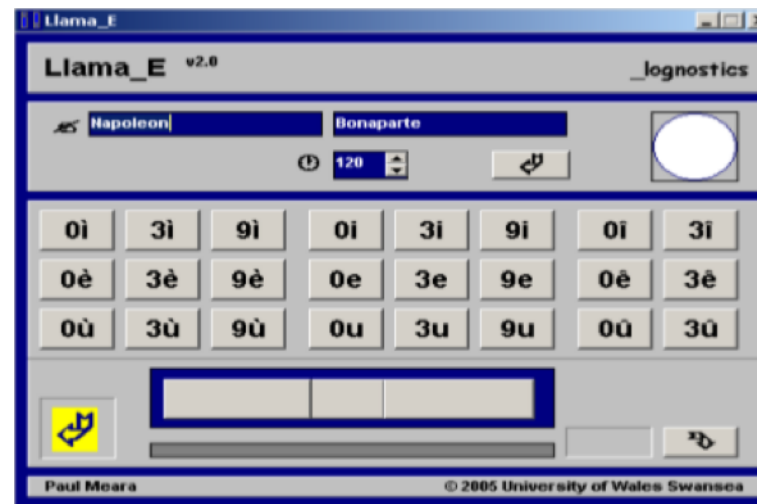
# Swansea LLAMA tests (Meara, 2005)

[www.lognostics.co.uk/tools/llama](http://www.lognostics.co.uk/tools/llama)



Swansea University  
Prifysgol Abertawe

- Free, loosely based on MLAT
- LLAMA B = vocabulary measure
- LLAMA D = sound recognition (implicit learning)
- LLAMA E = sound-symbol correspondence
- LLAMA F = grammatical inferencing
- Has not been fully validated.





Swansea University  
Prifysgol Abertawe

# Purpose/ Research questions

# Purpose/ Research questions



Swansea University  
Prifysgol Abertawe

- This study has three purposes:
- to remedy some of the test flaws.
- to revise the scoring method of the LLAMA test
- to examine if the revised tests overlap with working memory measures

- Research Questions:
- What is the impact of different scoring mechanisms on the distribution of ALPACAA scores?
- Do all the items discriminate between participants?
- What is the relationship between the new scoring method and WM?
- What impact do different background variables have on the ALPACAA scores?

# Methodology



Swansea University  
Prifysgol Abertawe

- Re-programmed the LLAMA tests into OpenSesame – called ALPACAA
- Changed order of administration:
  - D then B, E, F
  - Kept: 2 mins learning B & E, 5 mins learning F
- Fixed errors in original.
- No feedback to participant during test.
- End: given average RT and total correct.
- Clearer instructions (English)
- Can start test early



- **ALPACAA\_1 (sound recognition): Pre-listening instructions**

- Welcome to Part 2 of the experiment.
- There are four sections within this part. This is the first section.
- Please put on your headphones.
- You will hear a series of words in another language.
- All you have to do is listen to the words.
- Press any key when you are ready.



Swansea University  
Prifysgol Abertawe



- **ALPACAA\_1 (sound recognition): Post listening instructions**

- Thank you.
- You will now hear another series of words.
- Some of them are the same as the words you have just heard. Others are not.
- After each word, you will be asked if you have heard the word before.
- If you have heard the word before, press Y.
- If you have not heard the word before, press N.
- Only respond when you see the question.
- Press spacebar to continue.



Swansea University  
Prifysgol Abertawe

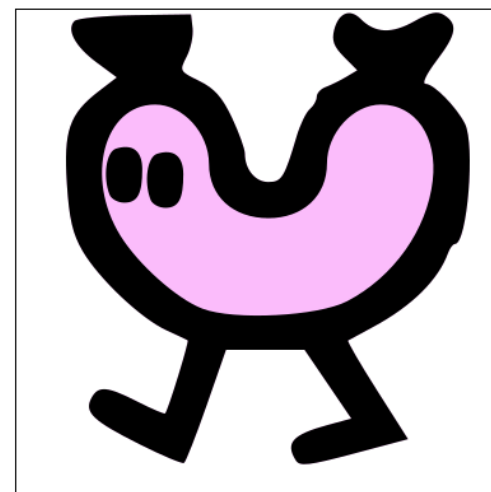
# ALPACAA\_2 (vocabulary) learning



Swansea University  
Prifysgol Abertawe

1 min  
and  
48 sec left

CHUEN



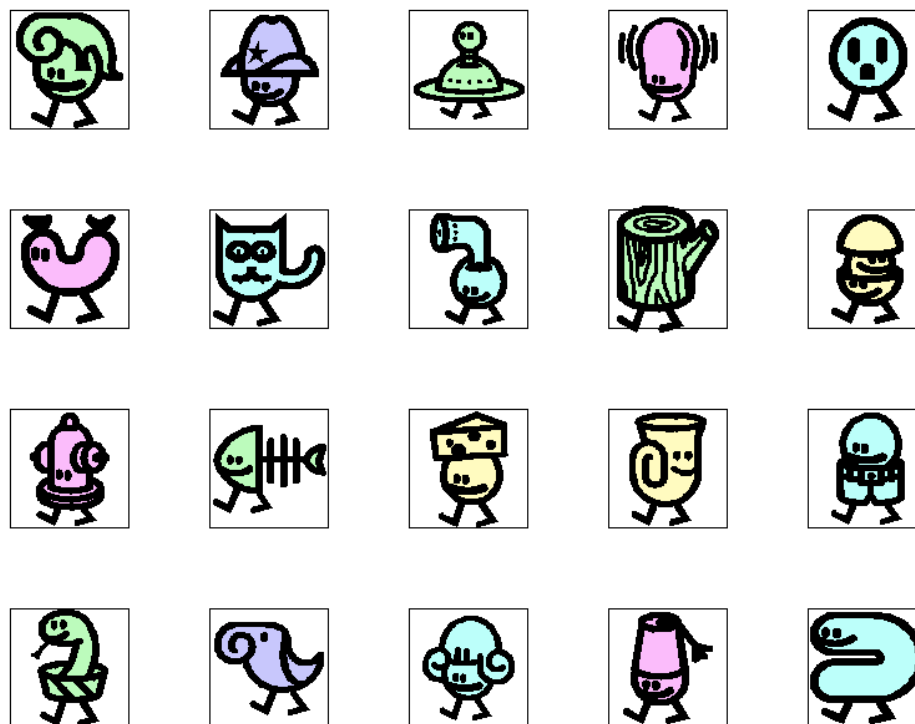
Start test

# ALPACAA\_2 (vocabulary) test phase



Swansea University  
Prifysgol Abertawe

Find the OC



# ALPACAA\_3 (sound-symbol) learning



Swansea University  
Prifysgol Abertawe

|    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|
| 0i | 3i | 9i | 0i | 3i | 9i | 0î | 3î |
| 0ë | 3ë | 9ë | 0e | 3e | 9e | 0ê | 3ê |
| 0ü | 3ü | 9ü | 0u | 3u | 9u | 0û | 3û |

1 min  
and  
59 sec left

Start test

# ALPACAA\_3 (sound-symbol) test instructions



Swansea University  
Prifysgol Abertawe

In this section, you will hear a new sound and be asked which symbol it matches.

These new sounds are a combination of two sounds that you have just learnt.

You should use the mouse to click on the symbol that matches the sound.

Press spacebar to continue.



# ALPACAA\_3 (sound-symbol) test layout



Swansea University  
Prifysgol Abertawe

Click on the word in the grid

|      |      |      |      |      |
|------|------|------|------|------|
| 0ë3ë | 9ü3ë | 3ë3ë | 0i3ë | 9i3e |
| 9ü0ë | 9ë0ë | 0i0ë | 3i3ë | 0ë3û |
| 0i3û | 3i3û | 0ë9û | 3ë3u | 0u0u |
| 3u3i | 0û3i | 3û3i | 3u0i | 9ë3u |

# ALPACAA\_4 (grammatical inferencing) learning

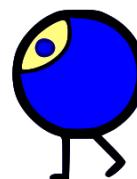


Swansea University  
Prifysgol Abertawe

4 min  
and  
56 sec left

|  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

atak-arap-sa



Start test

# ALPACAA\_4 (grammatical inferencing) test



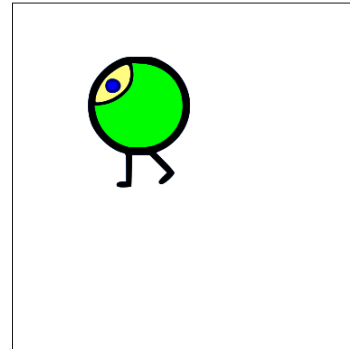
Swansea University  
Prifysgol Abertawe

1

eket-arap-sa

2

eket-arap



# Participants



Swansea University  
Prifysgol Abertawe



- Administered to 123 participants
- Age 17-55, (M=23.5, S.D.=5.576)
- Male = 56, Female = 67
- L1 English speakers = 77
  - (63 with L2, 14 L1 English only)
- Bilingual L1 English speakers = 7
- L2 English speakers = 39
- Also administered Stroop, Flanker and auditory Digits backwards.
- Collected by BA dissertation students (L-R, Dafydd, Megan, Amy)



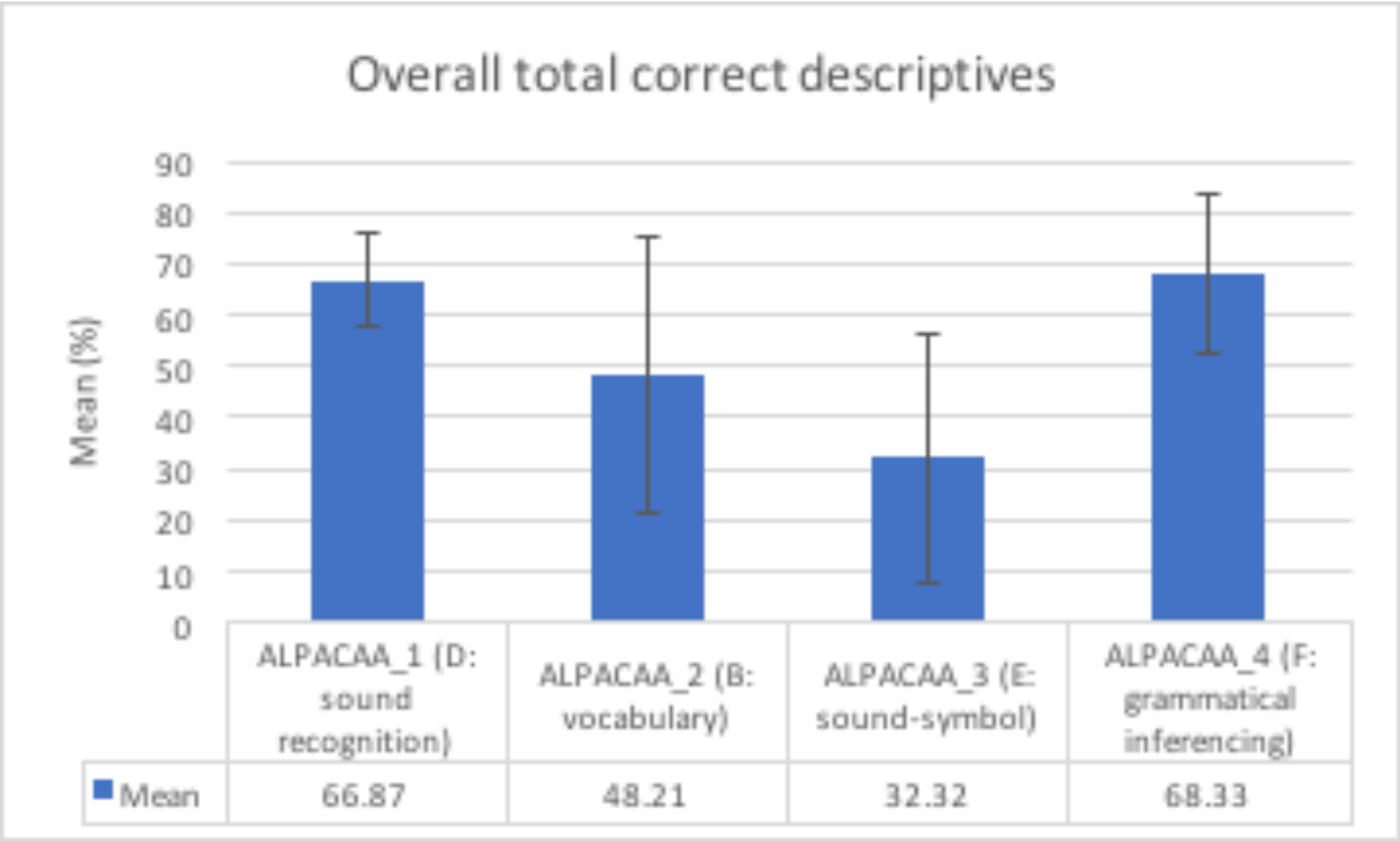
Swansea University  
Prifysgol Abertawe

RQ1:

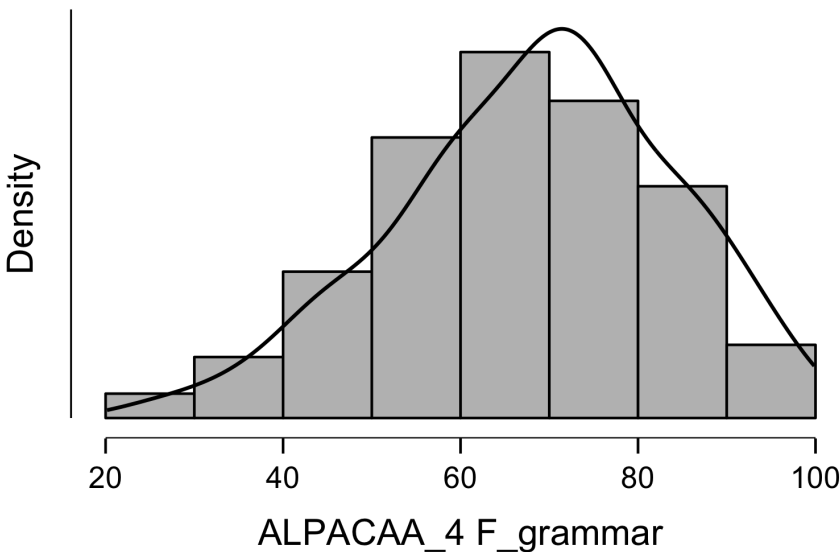
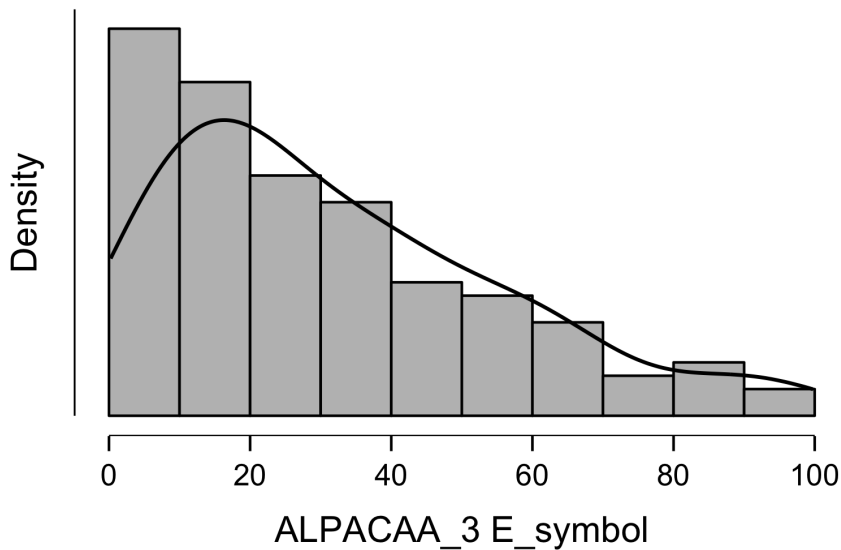
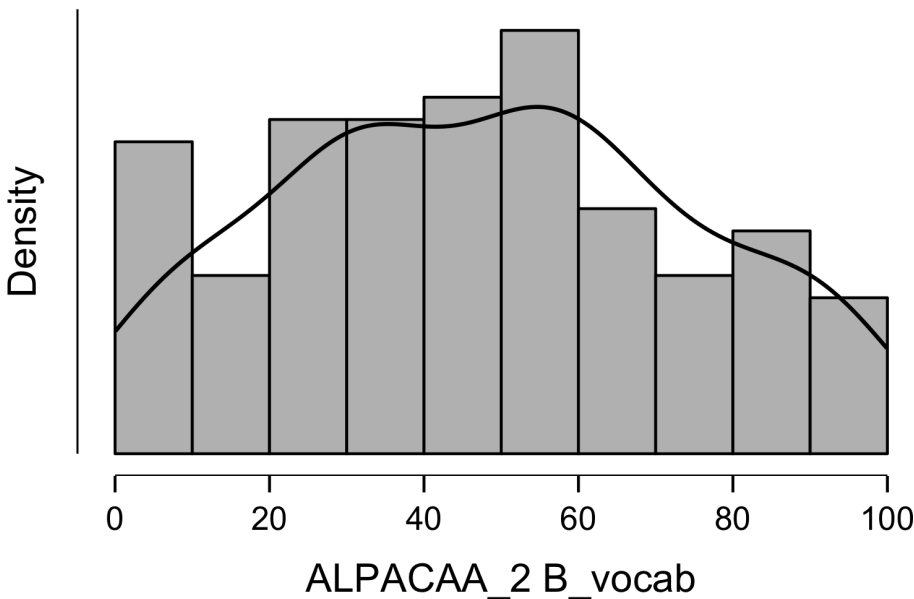
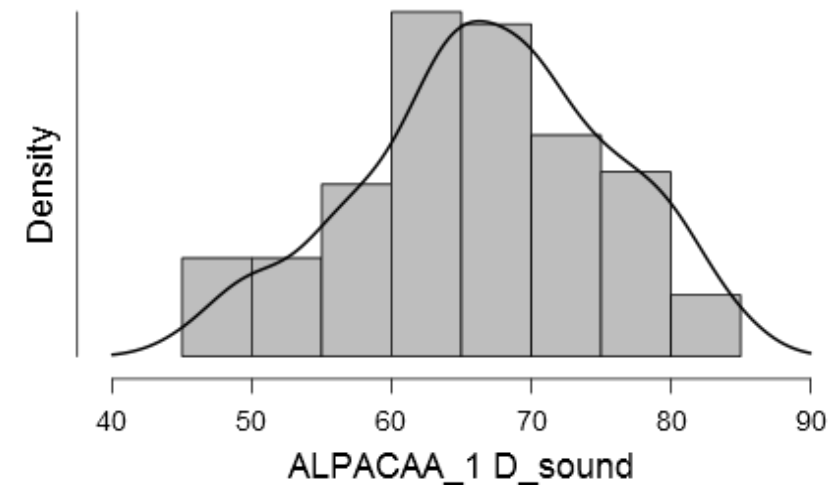
What is the impact of different scoring mechanisms  
on the distribution of ALPACAA scores?



# Overall descriptive: total correct – n=123



- Distribution of scores in tests (total correct)



# Adjusting for guessing



Swansea University  
Prifysgol Abertawe

- **Step 1**

- Adjusted for not doing learning phase (20 items)
- Criteria – must click on each item at least once.
- ALPACAA\_2 (vocab): 6 removed n=117
- ALPACAA\_3 (sound/symbol): 3 removed, n=120
- ALPACAA\_4 (grammatical inferencing): 3 removed, n=120

- **Step 2: Applied LLAMA penalties**

- LLAMA D, E, F – lose 1 mark (5%) for incorrect answer (binary choice)
- ALPACAA\_1 (D)
  - M=33.74, S.D=17.86
  - Mean was 68.67
  - Range: -10 – 70
- ALPACAA\_4 (F)
  - M=36.50, S.D=31.35
  - Mean was 68.33
  - Range: -50 - 100



# What about adjusting for guessing? Using RTs

- Have RTs for all test items for all participants..
- Excluded any RTs faster than 200ms.
- ALPACAA\_1\_D: In test phase, heard sound then question prompt then click.
  - More than 200ms after sound so no exclusions.
- ALPACAA\_2\_B: Three items identified (out of  $117 \times 20 = 2340$ )
  - Two were correct: removed.
- ALPACAA\_3\_E: No items
- ALPACAA\_4\_F: Four items identified (out of  $120 \times 20 = 2400$ )
  - Two were correct: removed.

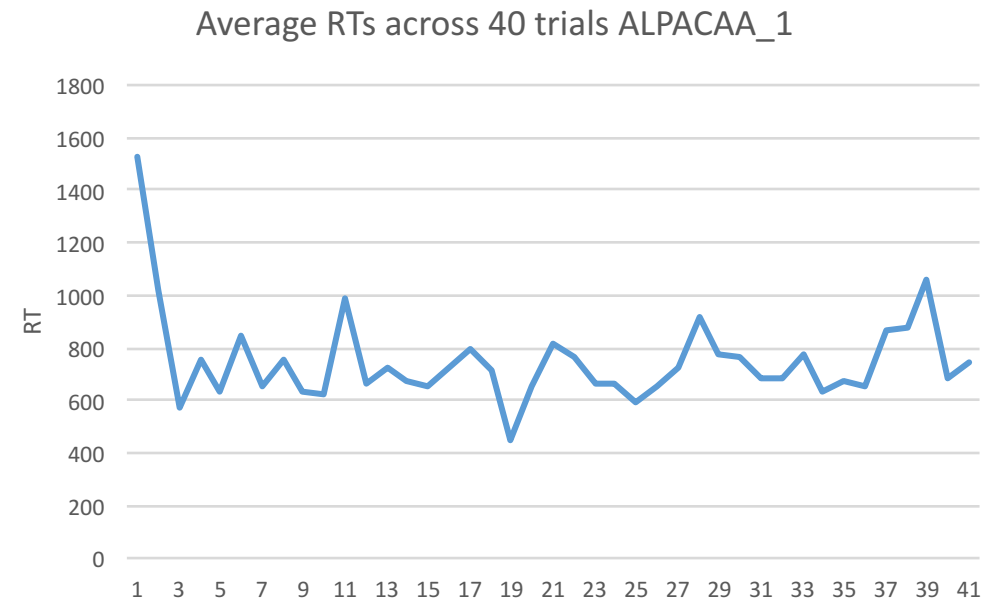
As they have to navigate with mouse then 200ms not an appropriate cut off?

# Discussion



Prifysgol Abertawe  
Swansea University

- Participants use the learning phase (12/369)
- Few react within 200ms (7/7140)
- Penalising doesn't change distribution but lowers mean (a lot).
- Lots more could be done with RT data.
- Very slow on first question but then flatten.
- Exclude items based on  $\pm 2$  S.D.'s.





Swansea University  
Prifysgol Abertawe

RQ2:

Do all the items discriminate between participants?

# Internal reliability (Cronbach's alpha)



Swansea University  
Prifysgol Abertawe

|                                       | n   | Cronbach's $\alpha$ | Average inter item correlation | 95% CI Lower | 95% CI Higher |
|---------------------------------------|-----|---------------------|--------------------------------|--------------|---------------|
| ALPACAA_1 (sound recognition) all     | 123 | 0.385               | 0.017                          | 0.329        | 0.438         |
| ALPACAA_1 (sound recognition) correct | 123 | 0.544               | 0.502                          | 0.502        | 0.584         |
| ALPACAA_2 (vocabulary)                | 117 | 0.850               | 0.227                          | 0.836        | 0.863         |
| ALPACAA_3 (sound/symbol)              | 120 | 0.883               | 0.272                          | 0.872        | 0.893         |
| ALPACAA_4 (grammatical inferencing)   | 120 | 0.617               | 0.079                          | 0.581        | 0.650         |

# ALPACAA\_1 Item Reliability Statistics (all items)



Swansea University  
Prifysgol Abertawe

|           | mean  | sd    | item-rest<br>correlation | If item dropped<br>Cronbach's $\alpha$ |
|-----------|-------|-------|--------------------------|--|
| latd11-n  | 0.463 | 0.501 | -0.137                   | 0.417                                  |
| latd12-n  | 0.545 | 0.500 | -0.076                   | 0.406                                  |
| latd03-y1 | 0.894 | 0.309 | 0.183                    | 0.368                                  |
| latd13-n  | 0.236 | 0.426 | -0.048                   | 0.398                                  |
| latd08-y1 | 0.301 | 0.460 | 0.161                    | 0.365                                  |
| latd14-n  | 0.683 | 0.467 | 0.160                    | 0.365                                  |
| latd15-n  | 0.642 | 0.481 | 0.003                    | 0.392                                  |
| latd05-y1 | 0.813 | 0.391 | 0.015                    | 0.388                                  |
| latd04-y1 | 0.691 | 0.464 | 0.243                    | 0.350                                  |
| latd06-y1 | 0.780 | 0.416 | 0.040                    | 0.385                                  |
| latd16v-n | 0.740 | 0.441 | 0.092                    | 0.377                                  |
| latd09-y1 | 0.585 | 0.495 | -0.168                   | 0.422                                  |
| latd17-n  | 0.740 | 0.441 | -0.030                   | 0.396                                  |
| latd10-y1 | 0.602 | 0.492 | 0.048                    | 0.384                                  |
| latd07-y1 | 0.732 | 0.445 | 0.146                    | 0.368                                  |
| latd18-n  | 0.496 | 0.502 | 0.012                    | 0.391                                  |
| latd19-n  | 0.732 | 0.445 | 0.082                    | 0.378                                  |
| latd20-n  | 0.366 | 0.484 | -0.006                   | 0.393                                  |
| latd01-y1 | 0.951 | 0.216 | 0.115                    | 0.378                                  |
| latd02-y1 | 0.553 | 0.499 | 0.160                    | 0.364                                  |
| latd21-n  | 0.821 | 0.385 | 0.051                    | 0.383                                  |
| latd22-n  | 0.740 | 0.441 | 0.157                    | 0.366                                  |

|           | mean  | sd    | item-rest<br>correlation | If item dropped<br>Cronbach's $\alpha$ |
|-----------|-------|-------|--------------------------|--|
| latd04-y2 | 0.756 | 0.431 | 0.124                    | 0.371                                  |
| latd06-y2 | 0.732 | 0.445 | 0.195                    | 0.359                                  |
| latd07-y2 | 0.748 | 0.436 | 0.129                    | 0.370                                  |
| latd23-n  | 0.829 | 0.378 | 0.189                    | 0.363                                  |
| latd08-y2 | 0.382 | 0.488 | 0.042                    | 0.385                                  |
| latd10-y2 | 0.659 | 0.476 | 0.058                    | 0.382                                  |
| latd24-n  | 0.667 | 0.473 | 0.177                    | 0.361                                  |
| latd25-n  | 0.740 | 0.441 | 0.206                    | 0.358                                  |
| latd26-n  | 0.699 | 0.460 | 0.163                    | 0.364                                  |
| latd03-y2 | 0.764 | 0.426 | 0.057                    | 0.382                                  |
| latd27-n  | 0.780 | 0.416 | -0.010                   | 0.392                                  |
| latd05-y2 | 0.675 | 0.470 | 0.099                    | 0.375                                  |
| latd02-y2 | 0.561 | 0.498 | 0.128                    | 0.370                                  |
| latd01-y2 | 0.846 | 0.363 | 0.222                    | 0.360                                  |
| latd28-n  | 0.675 | 0.470 | 0.054                    | 0.383                                  |
| latd09-y2 | 0.618 | 0.488 | -0.033                   | 0.398                                  |
| latd29-n  | 0.748 | 0.436 | 0.179                    | 0.362                                  |
| latd30-n  | 0.764 | 0.426 | 0.096                    | 0.376                                  |



# ALPACAA\_1 Inter-Item Reliability (all items)

|  | n   | Cronbach's $\alpha$ | Average inter<br>item<br>correlation | 95% CI<br>Lower | 95% CI<br>Higher |
|--|-----|---------------------|--------------------------------------|-----------------|------------------|
| ALPACAA_1<br>(sound<br>recognition) all                  | 123 | 0.385               | 0.017                                | 0.329           | 0.438            |
| <b>ALPACAA_1<br/>(sound<br/>recognition)<br/>revised</b> | 123 | <b>0.535</b>        | 0.036                                | 0.492           | 0.575            |

# ALPACAA\_1 Item Reliability Statistics (yes only)

|           | mean  | sd    | item-rest correlation | If item dropped Cronbach's $\alpha$ |
|-----------|-------|-------|-----------------------|-------------------------------------|
| latd03-y1 | 0.894 | 0.309 | 0.180                 | 0.532                               |
| latd08-y1 | 0.301 | 0.460 | 0.106                 | 0.543                               |
| latd05-y1 | 0.813 | 0.391 | 0.091                 | 0.544                               |
| latd04-y1 | 0.691 | 0.464 | 0.258                 | 0.517                               |
| latd06-y1 | 0.780 | 0.416 | -0.129                | 0.576                               |
| latd09-y1 | 0.585 | 0.495 | -0.007                | 0.564                               |
| latd10-y1 | 0.602 | 0.492 | 0.217                 | 0.524                               |
| latd07-y1 | 0.732 | 0.445 | 0.162                 | 0.533                               |
| latd01-y1 | 0.951 | 0.216 | 0.124                 | 0.539                               |
| latd02-y1 | 0.553 | 0.499 | 0.284                 | 0.511                               |
| latd04-y2 | 0.756 | 0.431 | 0.185                 | 0.530                               |
| latd06-y2 | 0.732 | 0.445 | 0.169                 | 0.532                               |
| latd07-y2 | 0.748 | 0.436 | 0.300                 | 0.510                               |
| latd08-y2 | 0.382 | 0.488 | 0.103                 | 0.544                               |
| latd10-y2 | 0.659 | 0.476 | 0.219                 | 0.523                               |
| latd03-y2 | 0.764 | 0.426 | 0.216                 | 0.525                               |
| latd05-y2 | 0.675 | 0.470 | 0.178                 | 0.531                               |
| latd02-y2 | 0.561 | 0.498 | 0.268                 | 0.514                               |
| latd01-y2 | 0.846 | 0.363 | 0.336                 | 0.509                               |
| latd09-y2 | 0.618 | 0.488 | 0.252                 | 0.517                               |



Swansea University  
Prifysgol Abertawe

# Inter-item reliability (Cronbach's alpha)

|  | n   | Cronbach's $\alpha$ | Average inter item correlation | 95% CI Lower | 95% CI Higher |
|--|-----|---------------------|--------------------------------|--------------|---------------|
| ALPACAA_1 (sound recognition) all                    | 123 | 0.385               | 0.017                          | 0.329        | 0.438         |
| <b>ALPACAA_1 (sound recognition) revised</b>         | 123 | 0.535               | 0.036                          | 0.492        | 0.575         |
| ALPACAA_1 (sound recognition) correct                | 123 | 0.544               | 0.502                          | 0.502        | 0.584         |
| <b>ALPACAA_1 (sound recognition) correct revised</b> | 123 | 0.593               | 0.075                          | 0.555        | 0.629         |

# ALPACAA\_4 Item Reliability Statistics

|                    | mean  | sd    | item-rest correlation | If item dropped Cronbach's $\alpha$ |
|--------------------|-------|-------|-----------------------|-------------------------------------|
| eket-arap-sa       | 0.825 | 0.382 | 0.173                 | 0.609                               |
| ipod-ilad-za       | 0.850 | 0.359 | 0.247                 | 0.601                               |
| eket-arap          | 0.733 | 0.444 | 0.316                 | 0.591                               |
| atak-arap-sa       | 0.767 | 0.425 | 0.299                 | 0.594                               |
| ipot-arap          | 0.592 | 0.494 | 0.095                 | 0.621                               |
| atag-ilad          | 0.583 | 0.495 | 0.349                 | 0.584                               |
| unak atak-arap-sa  | 0.875 | 0.332 | 0.321                 | 0.595                               |
| umush-ek ipot-arap | 0.783 | 0.414 | 0.306                 | 0.593                               |
| unak-ek ipot-arap  | 0.642 | 0.482 | 0.255                 | 0.599                               |
| inut-ek eket-arap  | 0.708 | 0.456 | 0.222                 | 0.603                               |
| unak-em eked-ilad  | 0.592 | 0.494 | 0.245                 | 0.600                               |
| umush-em ipod-ilad | 0.675 | 0.470 | 0.253                 | 0.599                               |
| unak ipot-arap-sa  | 0.692 | 0.464 | 0.253                 | 0.599                               |
| umush ipot-arap-sa | 0.633 | 0.484 | 0.137                 | 0.615                               |
| ipod-orad-za       | 0.817 | 0.389 | 0.220                 | 0.604                               |
| atag-orad-za       | 0.508 | 0.502 | -0.003                | 0.635                               |
| eked-orad-za       | 0.650 | 0.479 | 0.315                 | 0.590                               |
| umush-ek atag-orad | 0.658 | 0.476 | 0.334                 | 0.587                               |
| unak-em atag-orad  | 0.650 | 0.479 | 0.291                 | 0.593                               |
| ipod-orad          | 0.400 | 0.492 | -0.219                | 0.662                               |



Swansea University  
Prifysgol Abertawe

# Internal reliability (Cronbach's alpha)



Swansea University  
Prifysgol Abertawe

|  | n   | Cronbach's $\alpha$ | Average inter item correlation | 95% CI Lower | 95% CI Higher |
|--|-----|---------------------|--------------------------------|--------------|---------------|
| ALPACAA_4<br>(grammatical inferencing)                 | 120 | 0.617               | 0.079                          | 0.581        | 0.650         |
| <b>ALPACAA_4<br/>(grammatical inferencing) revised</b> | 120 | <b>0.682</b>        | 0.108                          | 0.652        | 0.710         |

# Discussion



Swansea University  
Prifysgol Abertawe

- ALPACAA\_2 & 3 (vocab and sound/symbol) discriminate well.
  - Participants chose from 20 pictures.
- ALPACAA\_1 & 4 (sound recognition and grammatical inferencing) do not discriminate well.
  - Participants given binary choice.
- Need more participants.
- More detailed analysis of items.
- Follow Bokander & Bylund (2019)



Swansea University  
Prifysgol Abertawe

RQ3: What is the relationship  
between the new scoring method  
and WM



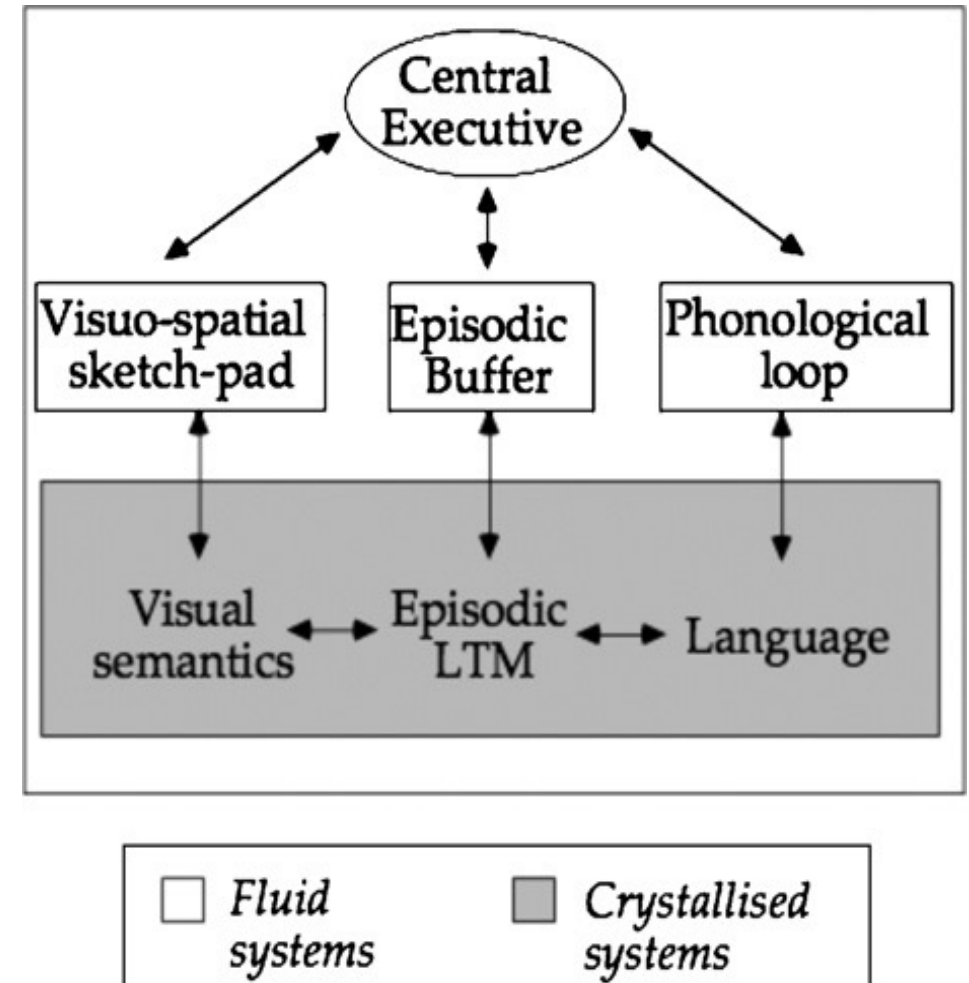
# What is working memory?

“Working memory refers to the system or systems that are assumed to be necessary in order to keep things in mind while performing complex tasks such as reasoning, comprehension and learning.”

Baddeley (2010, p. 136)

STM: maintenance of information

WM: maintenance and manipulation

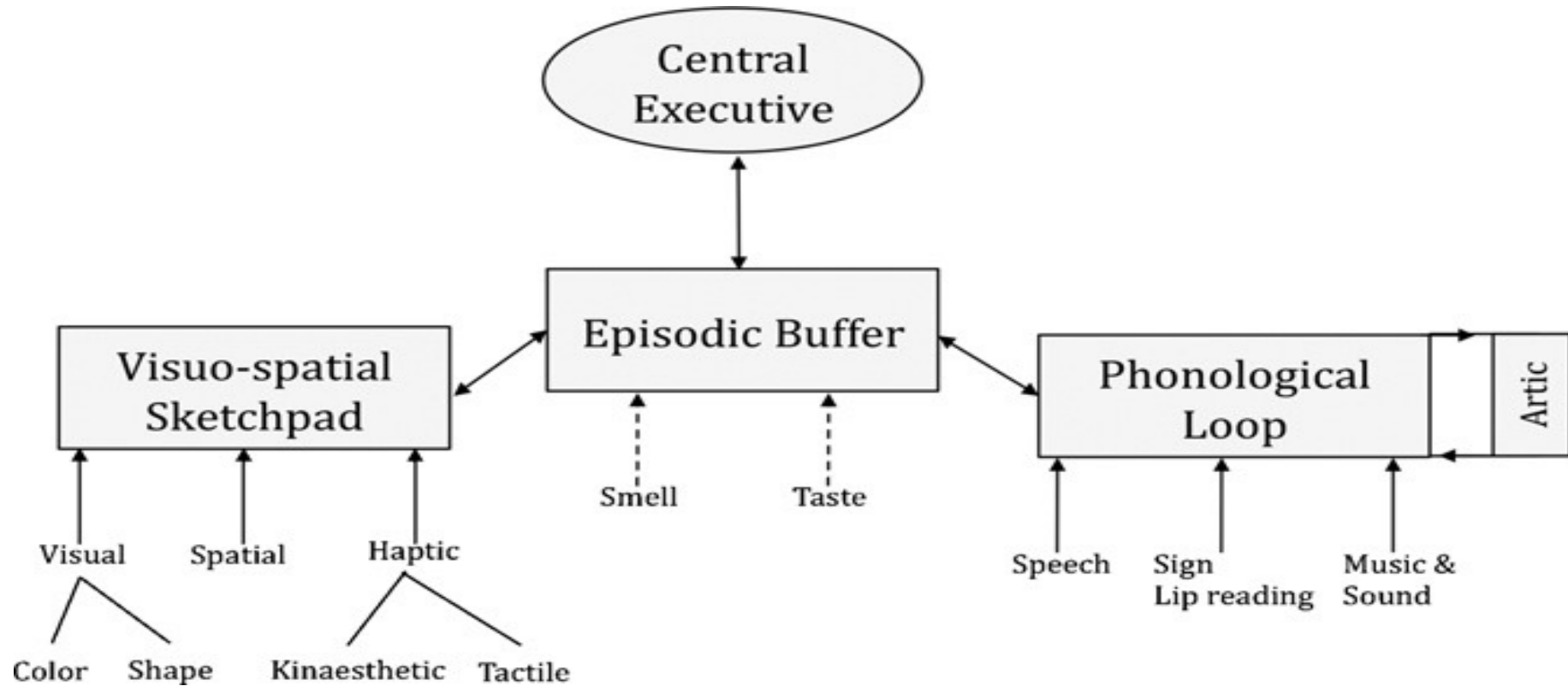




# Revised WM model (Baddeley et al 2011)



Swansea University  
Prifysgol Abertawe



# Previous work with LLAMA (presented at EUROSILA 2017)



Swansea University  
Prifysgol Abertawe

- Data collected by BA dissertation students:
  - Tesni Galvin, Amelia Cobner, Martha Chisholm, Jake Clothier & Issy Greenfield
- 127 participants
  - predominantly students
- Typically L1 English speakers

**Table I – Participant Data**

|             |       |
|-------------|-------|
| No. Females | 60    |
| No. Males   | 67    |
| Age Range   | 16-78 |
| Average Age | 33.5  |



# Results: PCA

- No LLAMA test loads on the same factor as any of the working memory and attention tests.

| Pattern Matrix <sup>a</sup>                                |           |       |
|--|-----------|-------|
|  | Component |       |
|  | 1         | 2     |
| LLAMA E  | .807      |       |
| LLAMA F  | .799      |       |
| LLAMA B  | .670      |       |
| LLAMA D  | .546      |       |
| WM3 (A)  |           | .906  |
| WM3 (B)  |           | .877  |
| WM1 (Visual)   |           | -.498 |
| WM2 (Digits)   |           | -.392 |
| Extraction Method: Principal Component Analysis.           |           |       |
| Rotation Method: <u>Oblimin</u> with Kaiser Normalization. |           |       |
| a. Rotation converged in 6 iterations.                     |           |       |

# Results: PCA part 2

- Even if forced to four factors, LLAMA tests load differently to the WM/attention tests.
- LLAMA B, E & F measure something different to LLAMA D (similar to Grañena 2013).
- TMT parts A & B measure different aspect of WM to the digits backwards (PSTM) and visuo-spatial/ storage measures.

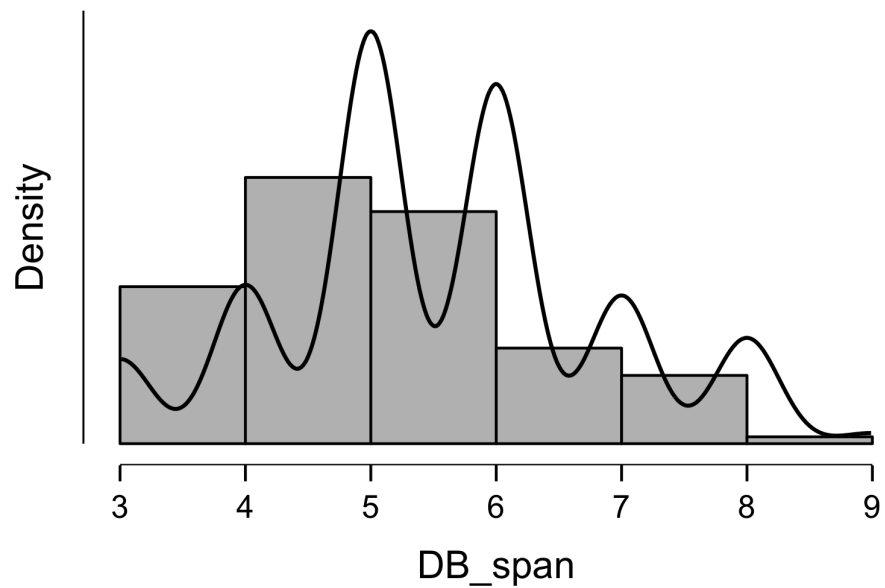
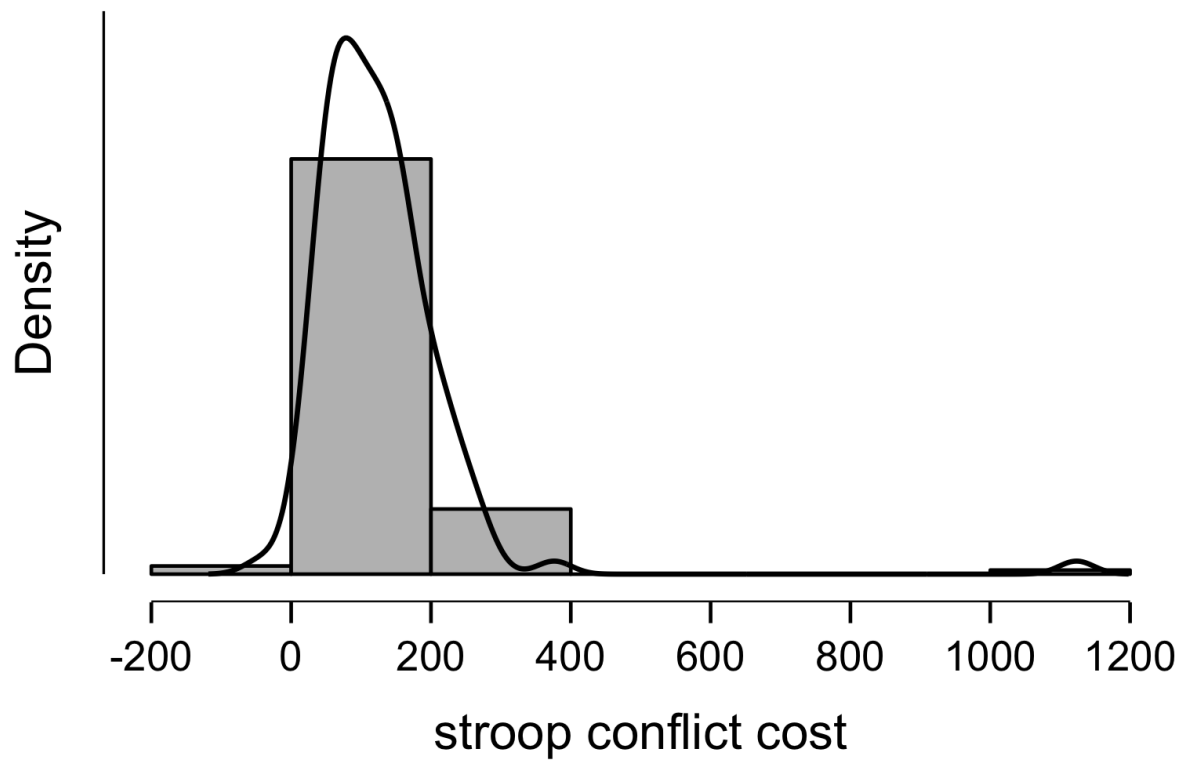
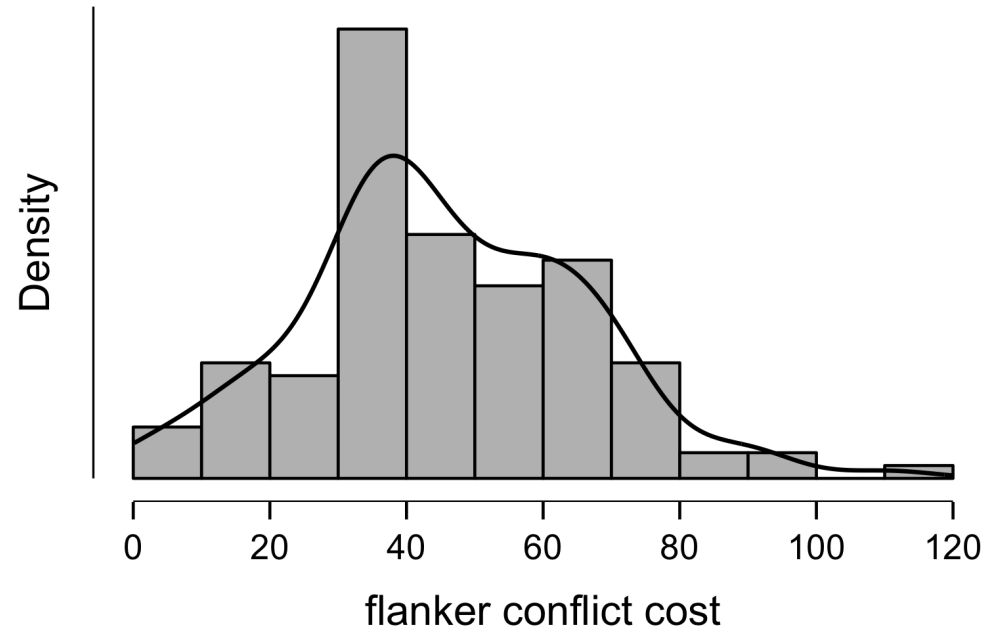
| Pattern Matrix <sup>a</sup>                                |           |      |      |      |
|--|-----------|------|------|------|
|  | Component |      |      |      |
|  | 1         | 2    | 3    | 4    |
| LLAMA F  | .831      |      |      |      |
| LLAMA E  | .828      |      |      |      |
| LLAMA B  | .672      |      |      |      |
| WM3 (A)  |           | .914 |      |      |
| WM3 (B)  |           | .867 |      |      |
| WM2 (Digits)   |           |      | .897 |      |
| WM1 (Visual)   |           |      | .586 |      |
| LLAMA D  |           |      |      | .947 |
| Extraction Method: Principal Component Analysis.           |           |      |      |      |
| Rotation Method: <u>Oblimin</u> with Kaiser Normalization. |           |      |      |      |
| a. Rotation converged in 6 iterations.                     |           |      |      |      |



# Working memory results (n=123)

|                | flanker conflict<br>cost | stroop conflict<br>cost | DB_span |
|----------------|--------------------------|-------------------------|---------|
| Mean           | 45.440                   | 127.297                 | 5.537   |
| Std. Deviation | 20.379                   | 114.391                 | 1.317   |
| Minimum        | 0.025                    | -43.880                 | 3.000   |
| Maximum        | 110.550                  | 1123.434                | 9.000   |

Only significant correlation (Spearman's) between  
Stroop and Digits Backwards ( $r = -0.252$ ,  $p = .005$ )



## Spearman Correlations

|                      |                | flanker cost | stroop cost | DB_span   | A1     | A2        | A3        | A4 |
|----------------------|----------------|--------------|-------------|-----------|--------|-----------|-----------|----|
| flanker              | Spearman's rho | —            |             |           |        |           |           |    |
| conflict cost        | p-value        | —            |             |           |        |           |           |    |
| stroop               | Spearman's rho | 0.065        | —           |           |        |           |           |    |
| conflict cost        | p-value        | 0.478        | —           |           |        |           |           |    |
| DB_span              | Spearman's rho | 0.049        | -0.252**    | —         |        |           |           |    |
|                      | p-value        | 0.590        | 0.005       | —         |        |           |           |    |
| A1_total_corr<br>ect | Spearman's rho | 0.006        | -0.046      | 0.073     | —      |           |           |    |
|                      | p-value        | 0.948        | 0.616       | 0.428     | —      |           |           |    |
| A2_total_corr<br>ect | Spearman's rho | -0.012       | -0.186*     | 0.432***  | 0.200* | —         |           |    |
|                      | p-value        | 0.901        | 0.045       | 1.161e -6 | 0.031  | —         |           |    |
| A3_otal_corr<br>ect  | Spearman's rho | 0.019        | -0.178      | 0.252**   | 0.178  | 0.467***  | —         |    |
|                      | p-value        | 0.834        | 0.052       | 0.005     | 0.052  | 1.253e -7 | —         |    |
| A4_total_corr<br>ect | Spearman's rho | -0.021       | -0.193      | 0.200*    | 0.191* | 0.523***  | 0.455***  | —  |
|                      | p-value        | 0.821        | 0.036       | 0.029     | 0.037  | 1.988e -9 | 2.254e -7 | —  |

\* p < .05, \*\* p < .01, \*\*\* p < .001

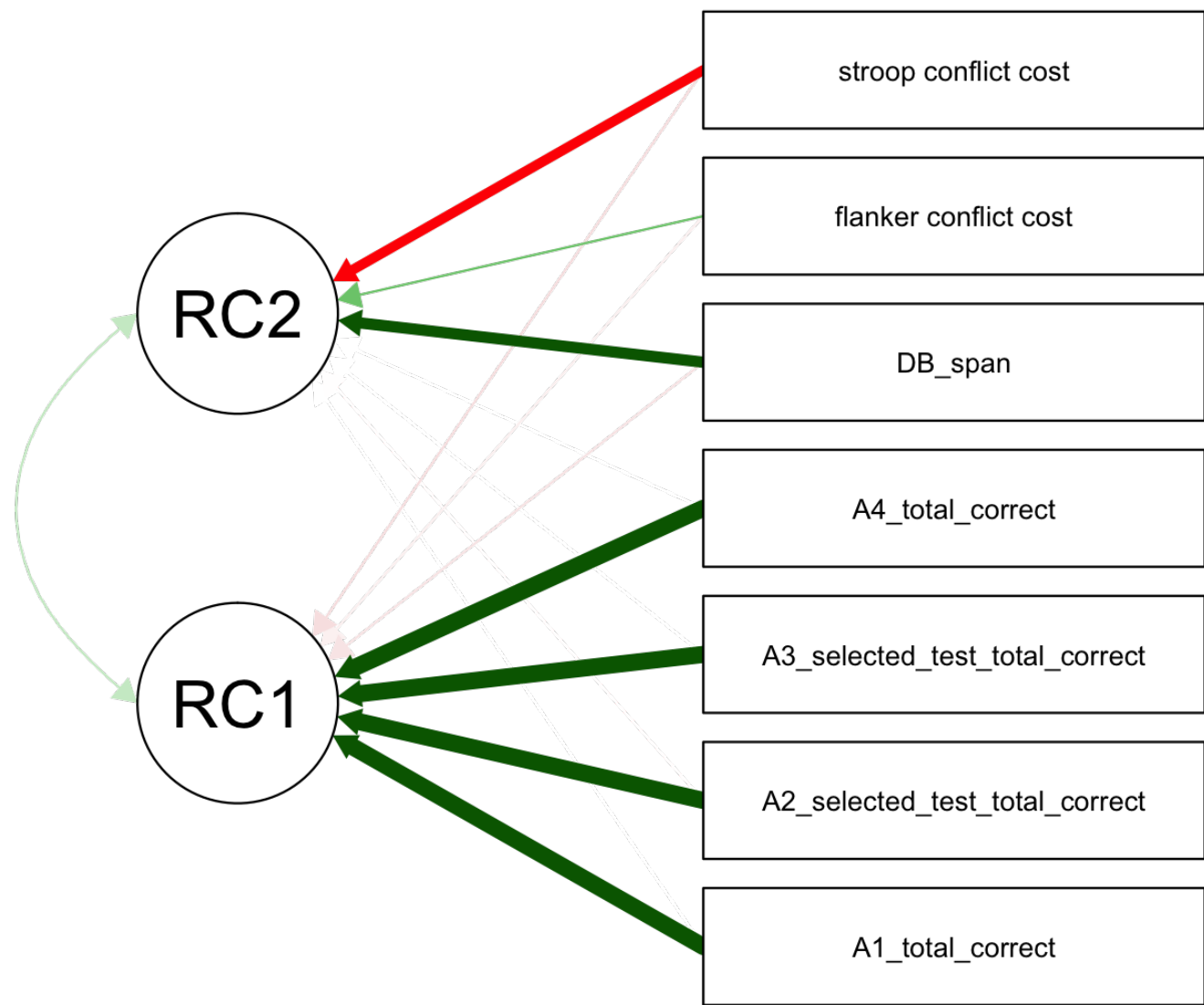
# PCA analysis: WM and ALPACAA component total correct scores

Component Loadings

|                                | RC 1  | RC 2   | Uniqueness |
|--------------------------------|-------|--------|------------|
| A1_total_correct               | 1.000 | .      | 1.821e -4  |
| A2_selected_test_total_correct | 1.000 | .      | 1.895e -4  |
| A3_selected_test_total_correct | 1.000 | .      | 2.005e -4  |
| A4_total_correct               | 1.000 | .      | 1.792e -4  |
| DB_span                        | .     | 0.765  | 0.417      |
| flanker conflict cost          | .     | .      | 0.958      |
| stroop conflict cost           | .     | -0.714 | 0.484      |

*Note.* Applied rotation method is promax.





# Discussion



Swansea University  
Prifysgol Abertawe

- WM tests and ALPACAA aptitude tests (total correct) are measuring different things.
- WM may be part of aptitude but doesn't replace it
  - (cf Wen, 2016)
- Comparable to previous findings on LLAMA and WM.
  - Different WM tests (Corsi block, TMT A&B & Digits backwards)
- Didn't find difference with sound recognition and other tests.
  - Scores to 100 and no penalties?



Swansea University  
Prifysgol Abertawe

RQ4: What impact do different background variables have on the ALPACAA scores?

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017).  
Examining the LLAMA aptitude tests.. *Journal of the European  
Second Language Association*, 1(1), 49–60.  
DOI: <http://doi.org/10.22599/jesla.24>



Swansea University  
Prifysgol Abertawe

- **How much of the LLAMA test score variance do the individual factors measures account for?**
- Factors included age, L1, L2 status, education level, gender, playing of logic puzzles.
- 404 participants in total.
- 346 took all 4 parts of the LLAMA tests and background questionnaires.
- Multiple regression analysis for 6 factors.  
Overall variance for:
  - LLAMA B:  $R^2 = 9.1\%$
  - LLAMA D:  $R^2 = 4.8\%$
  - LLAMA E:  $R^2 = 3.4\%$
  - LLAMA F:  $R^2 = 6.6\%$
- Only L2 status consistently was significant  $p < .05$  (not for E).
  - LLAMA B:  $\beta = -.250$ , contribution to variance = 6.0
  - LLAMA D:  $\beta = .136$ , contribution to variance = 1.8
  - LLAMA F:  $\beta = -.165$ , contribution to variance = 2.6



Swansea University  
Prifysgol Abertawe

# Background variables (n=123)

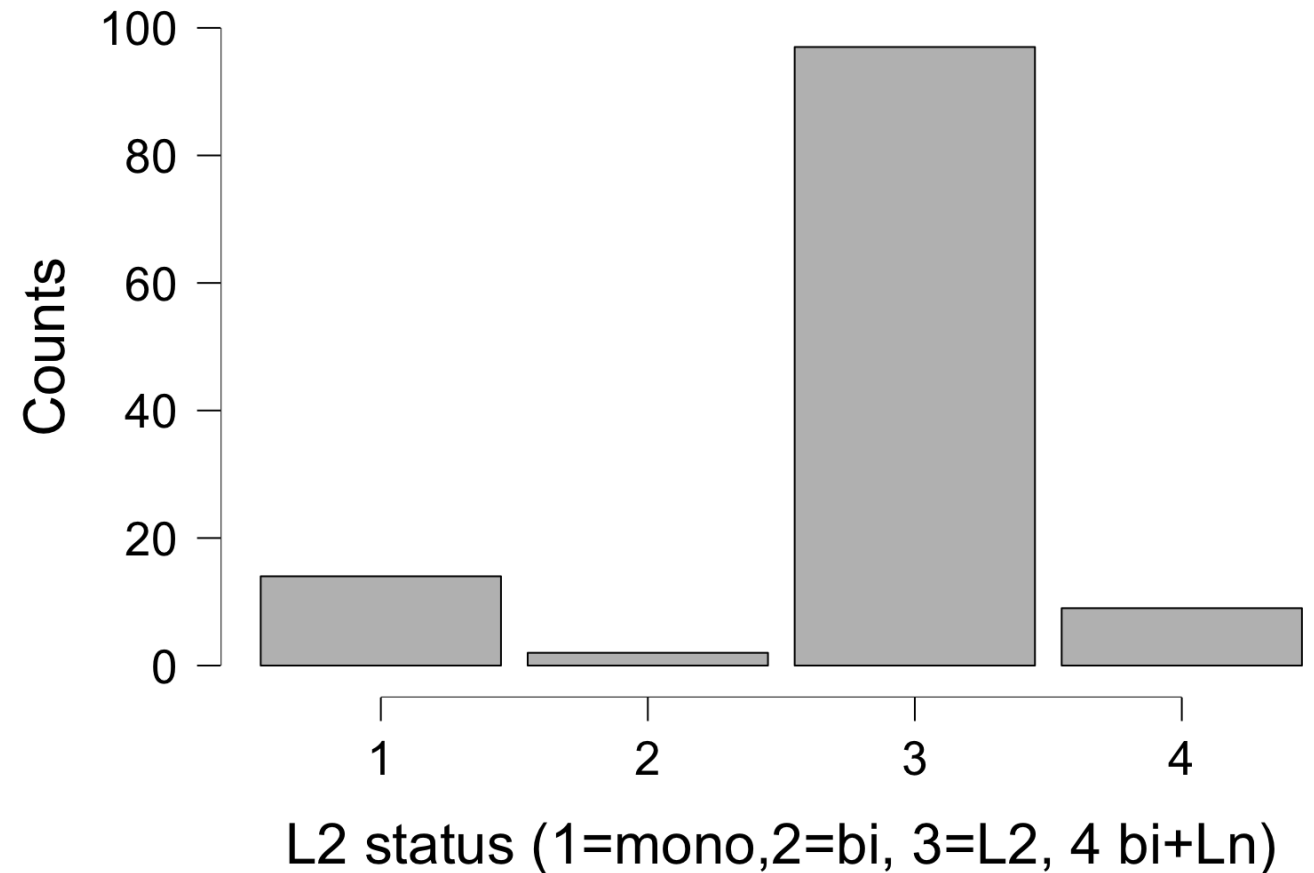
- Working memory scores: (Stroop, Flanker, Digits Backwards)
- L2 status:
- Age:
- Test anxiety scores
  - (self-report and from Horwitz et al 1986, FLCAS)
- Sleep
  - Hours (average and last night), tiredness rating
- Total response time

# L2 status



Swansea University  
Prifysgol Abertawe

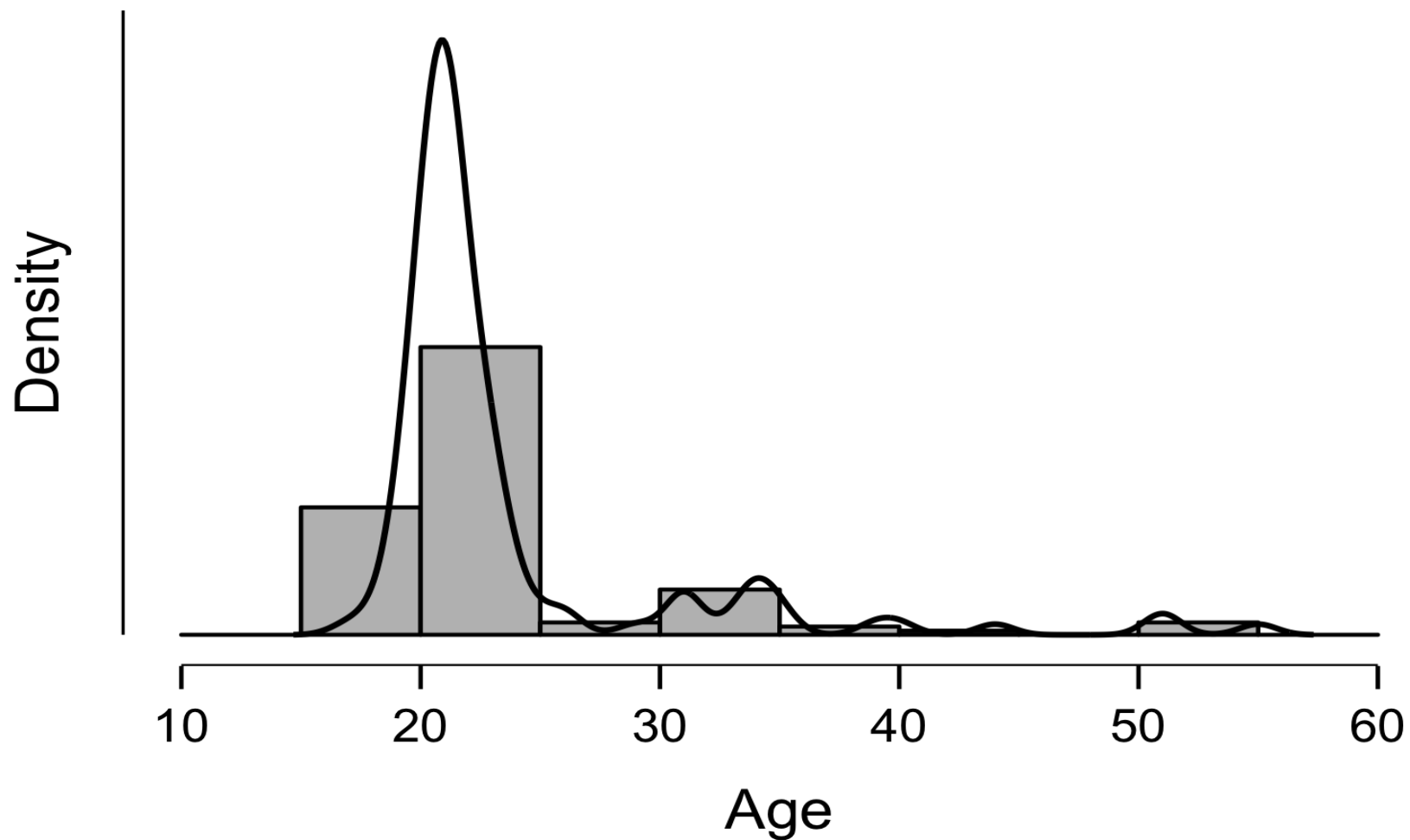
|                            | Number |
|----------------------------|--------|
| Monolingual                | 14     |
| Bilingual                  | 2      |
| Instructed L2              | 97     |
| Bilingual + Ln instruction | 9      |



# Age



Swansea University  
Prifysgol Abertawe



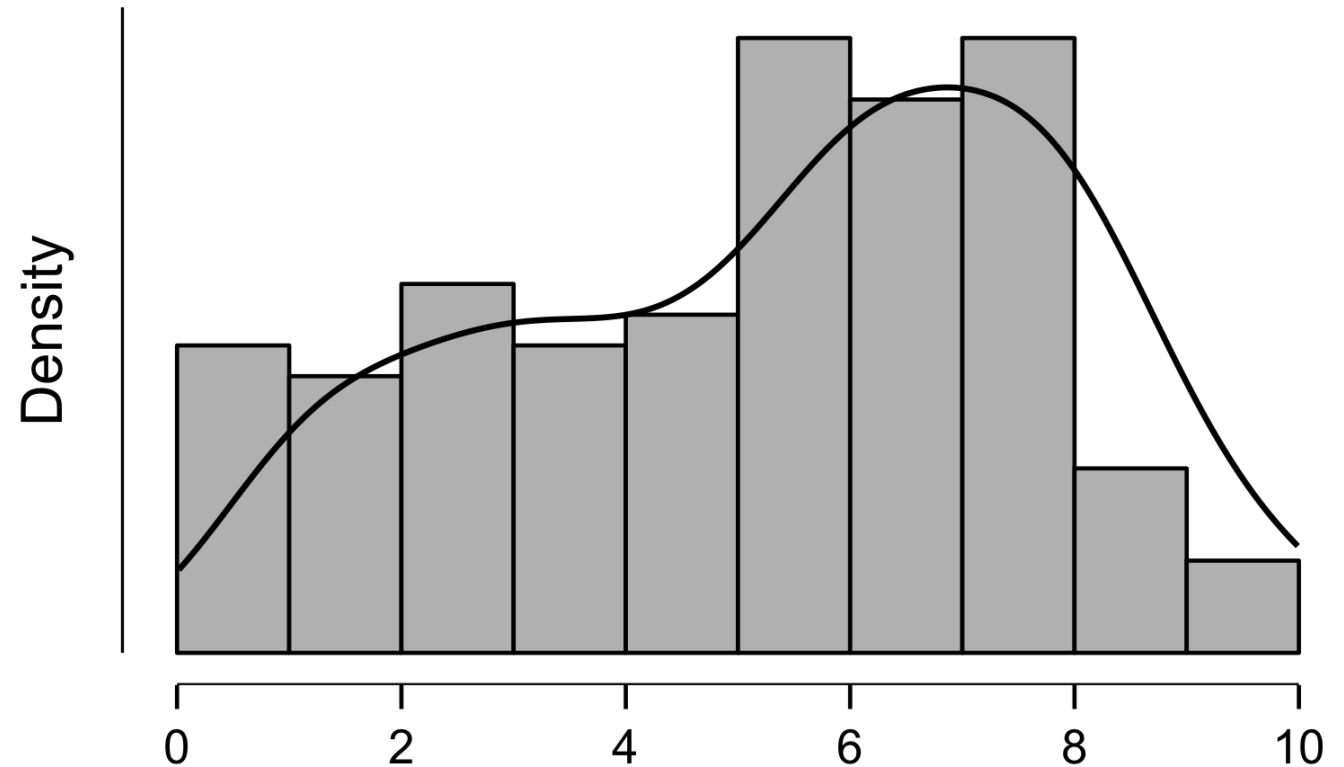
- Mean= 23.6
- SD = 6.619
- Min = 17
- Max=55

# Self-reported anxiety when taking tests



Swansea University  
Prifysgol Abertawe

- Mean = 5.45
- SD = 2.466
- Range = 1-10



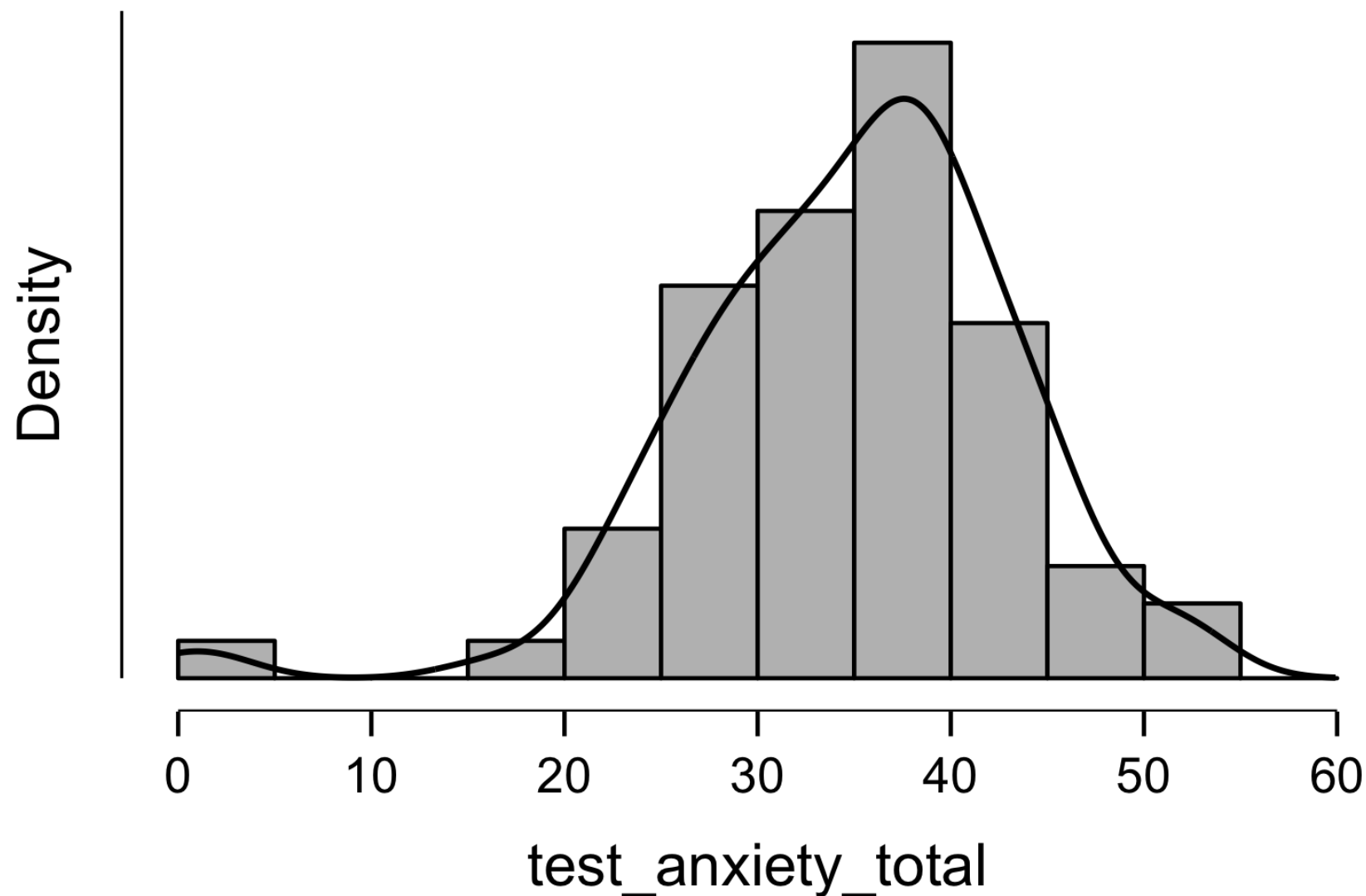
) On a scale of 1-10, how anxious do you feel while ta



# Test anxiety (FLCAS: 11 items)



Swansea University  
Prifysgol Abertawe

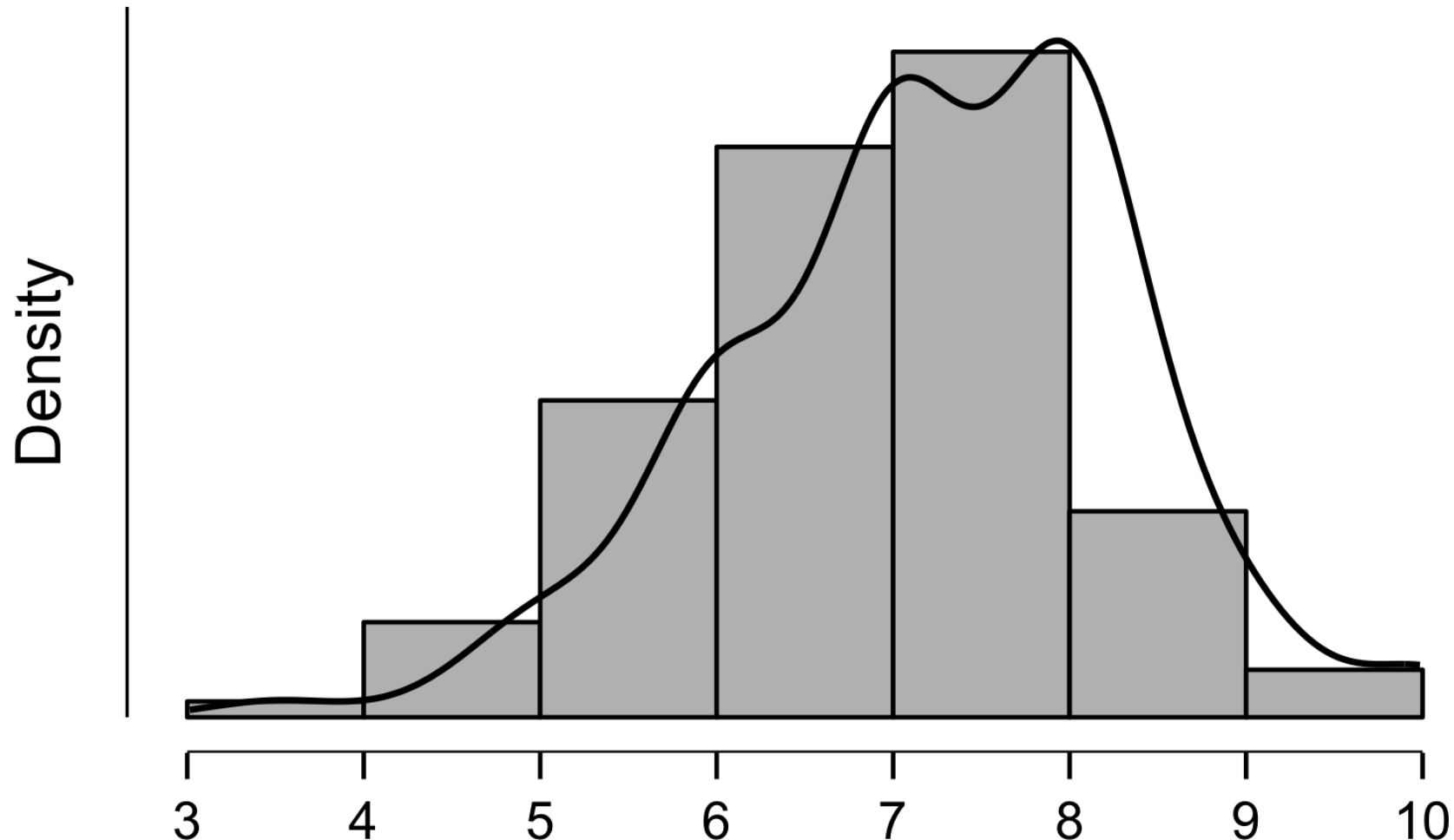


- Mean = 34.88
- SD = 8.485
- Range = 0-53



Swansea University  
Prifysgol Abertawe

# Average amount of sleep (generally)



- Mean = 7.2
- SD = 1.114
- Range = 3.5-10

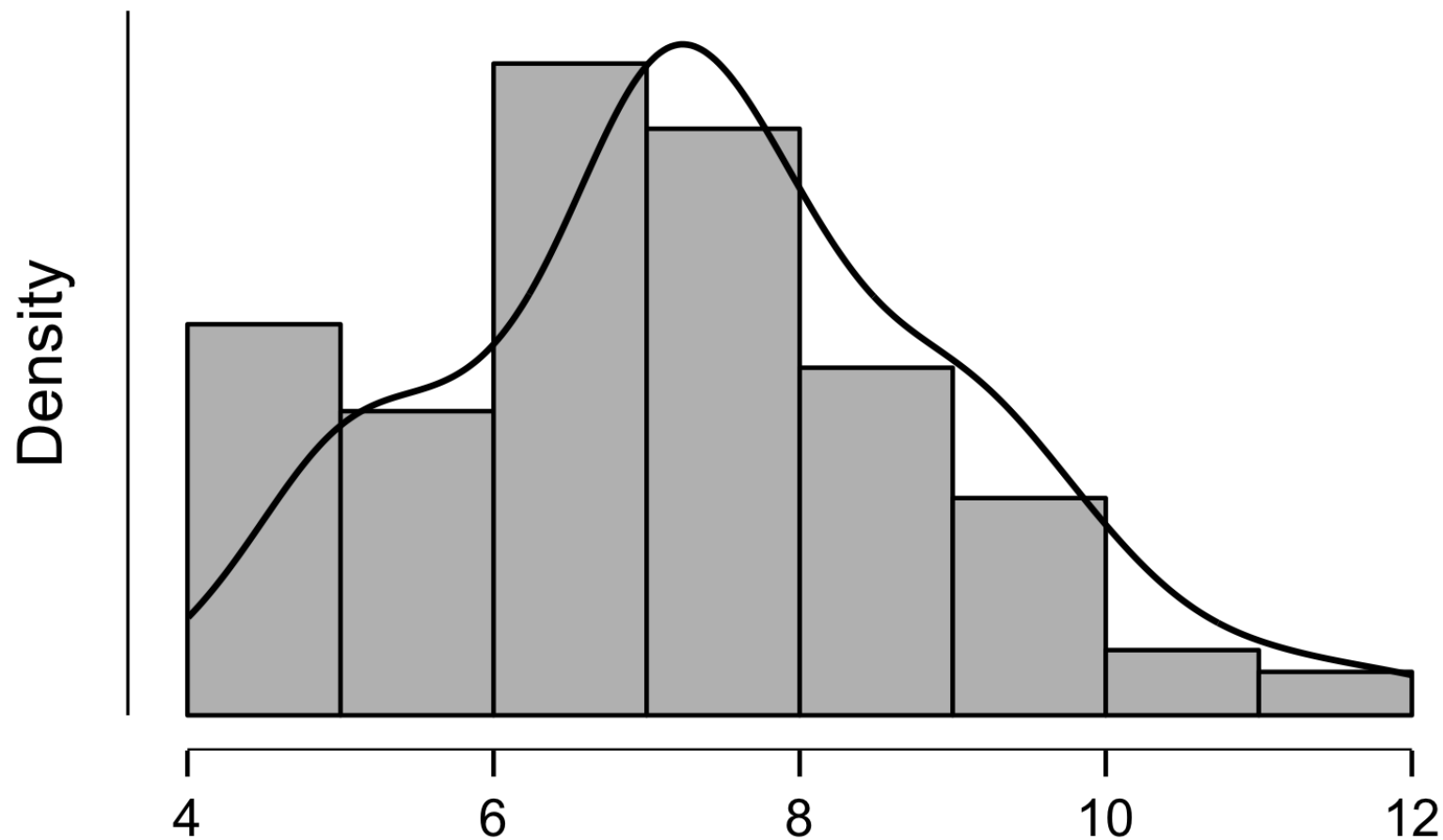
How many hours of sleep do you get per night on average?

# Amount of sleep previous night



Swansea University  
Prifysgol Abertawe

- Mean= 7.4
- SD=1.689
- Range =4-12



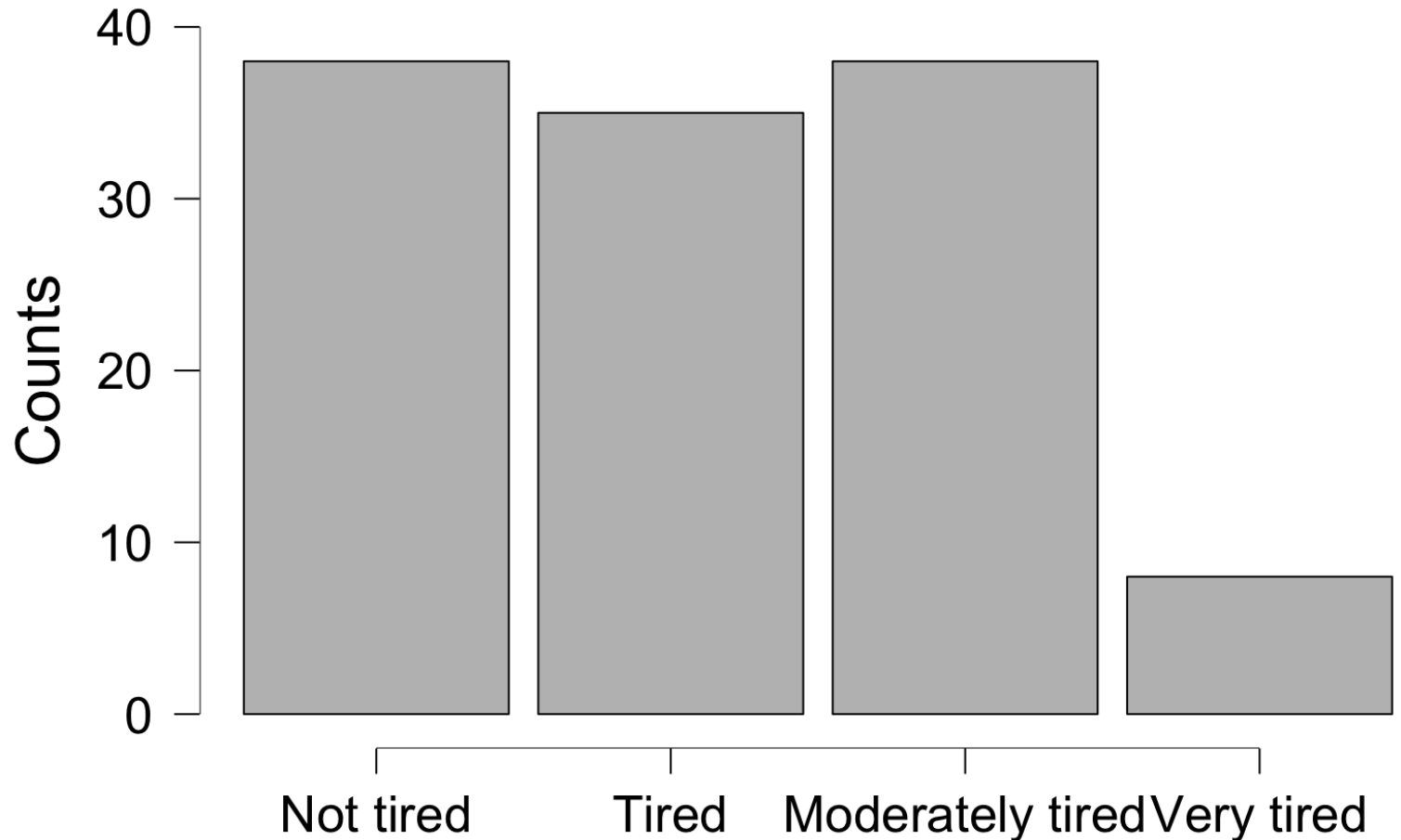
Q1) How many hours of sleep did you get last night? (

# Current tiredness rating



Swansea University  
Prifysgol Abertawe

- Very tired = 8
- Moderately tired = 38
- Tired = 35
- Not tired = 38



S3Q4) How tired do you feel at this moment? (VT, M1

# Linear Regression model (stepwise)



Swansea University  
Prifysgol Abertawe

- Dependent variable = ALPACAA total correct score
- Co-variates:
  - Age
  - DB
  - Flanker
  - Stroop
  - L2 status
  - Hours sleep (previous night)
  - Test anxiety FLCAS score
  - Total time on test

# Overall results (significant only) inc: time



Swansea University  
Prifysgol Abertawe

|           | F                 | p        | r <sup>2</sup> | Effect size | Significant variable | β         |
|-----------|-------------------|----------|----------------|-------------|----------------------|-----------|
| ALPACAA_2 | (1,113) = 20.306  | 1.618e-5 | 0.152          | medium      | Digits Backwards     | 1.486     |
| ALPACAA_3 | (1,116) = 36.606  | 1.817e-8 | 0.240          | medium      | A3 total RT time     | 1.319 e-5 |
| ALPACAA_4 | (1, 115) = 25.853 | 1.445e-6 | 0.184          | medium      | A4 total RT time     | 3.753e-5  |

# Overall results (significant only) exc: time



Swansea University  
Prifysgol Abertawe

|           | F                | p        | r <sup>2</sup> | Effect size | Significant variable | β     |
|-----------|------------------|----------|----------------|-------------|----------------------|-------|
| ALPACAA_2 | (1,113) = 10.159 | 1.618e-5 | 0.152          | medium      | Digits Backwards     | 1.486 |
| ALPACAA_3 | (1,116) = 6.947  | 0.010    | 0.057          | small       | Digits backwards     | 0.857 |
| ALPACAA_4 | (1, 115)         | 0.048    | 0.033          | small       | Digits backwards     | 0.432 |

# Discussion



Swansea University  
Prifysgol Abertawe

- Regression suggests that DB is significant predictor for ‘explicit’ tests only.
- No predictor for ALPACAA\_1 (LLAMA D/ implicit test)
- Counter to Rogers et al (2017) finding for L2 status.
- However,
  - L2 status coded differently as no control for proficiency.
  - Total correct not LLAMA adjusted scores.
- More detailed modelling needed.





# Overall conclusion

- ALPACAA are an (initial) attempt to refine the LLAMA tests.
- Further work needed on:
  - ALPACAA\_4 (grammatical inferencing) and ALPACAA\_1 (sound recognition) in terms of reliability.
  - ALPACAA\_3 (sound/symbol) in terms of negative skew.
  - Are layout revisions enough?
- More detailed analysis of RT and items needed.
- More detailed analysis of predictor variables.
- New versions of LLAMA are also in development – see Paul Meara's website ([www.lognostics.co.uk](http://www.lognostics.co.uk)) for updates.



Thank you!

Diolch yn fawr!

Tusen takk!

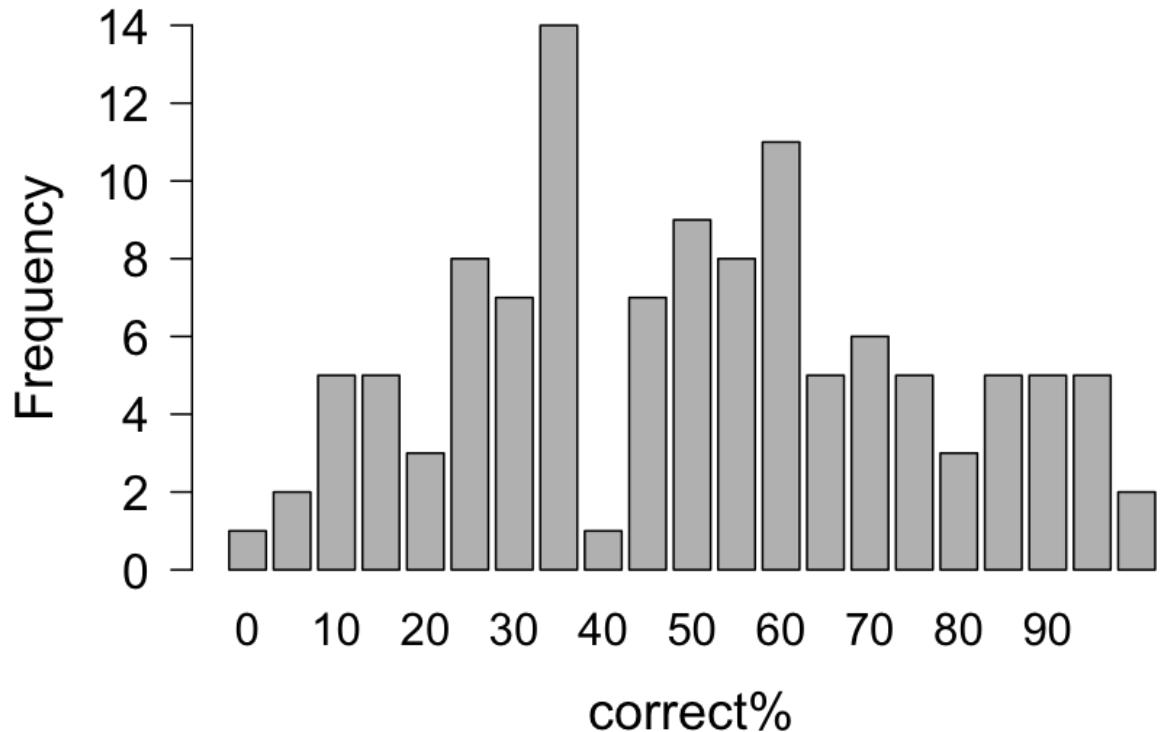
Vivienne Rogers: [v.e.rogers@swansea.ac.uk](mailto:v.e.rogers@swansea.ac.uk)

Paul Meara: [p.m.meara@gmail.com](mailto:p.m.meara@gmail.com)



- **ALPACAA 2 (vocabulary)**
- Adjusted for not doing learning phase (20 items)
- Criteria – must click on each item at least once

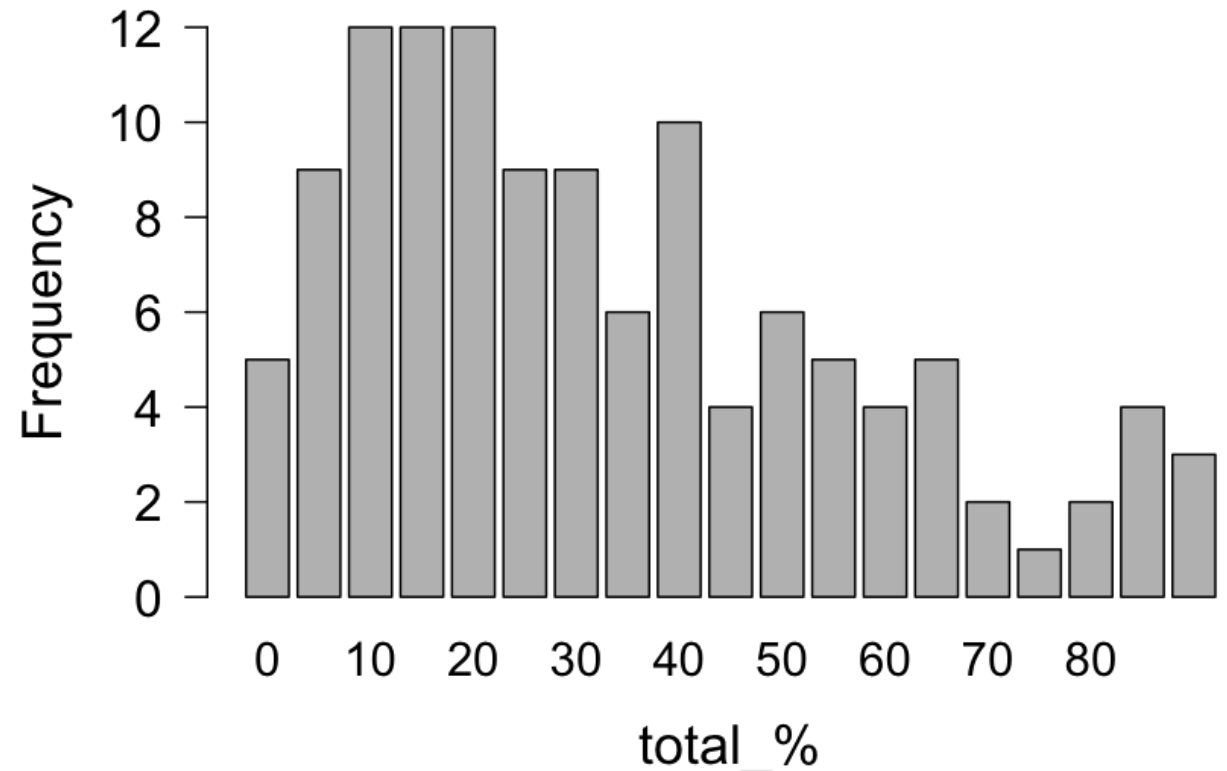
- Removal of 6 participants
- $n=117$
- Average clicks in learning phase = 49.89, S.D = 15.643
- Range = 21-95
- Adjusted mean=50.5%
- Adjusted S.D = 25.337





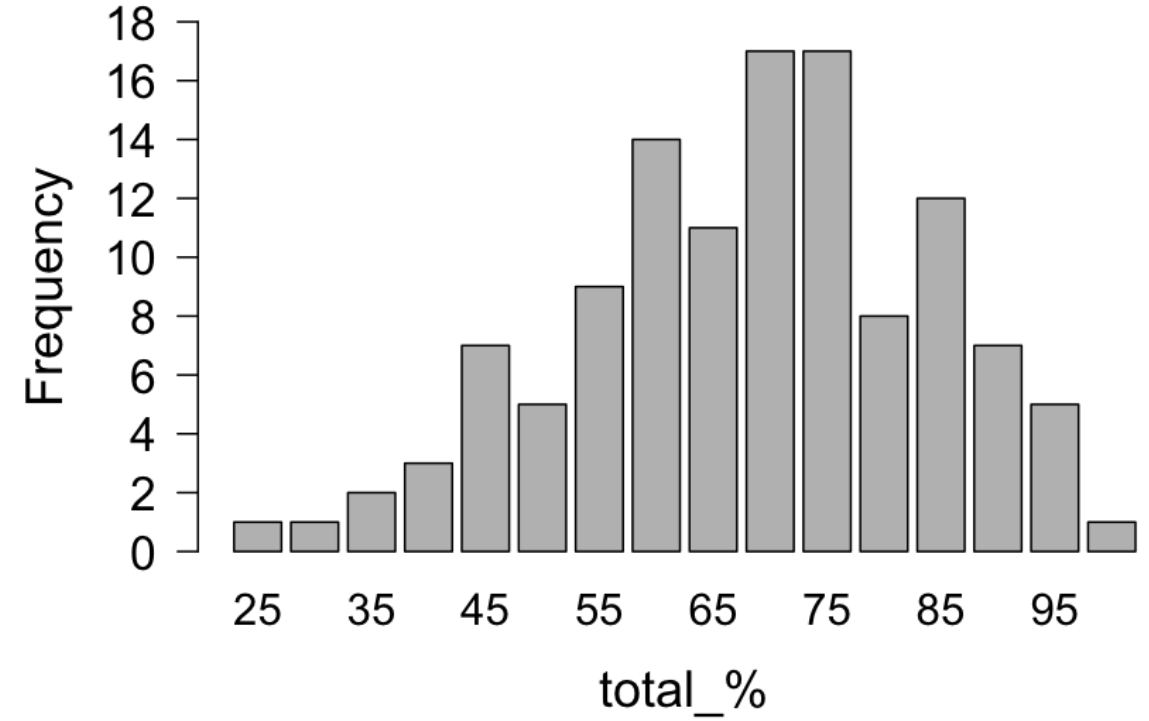
- **ALPACAA 3 (sound-symbol)**
- Adjusted for not doing learning phase (20 items)
- Criteria – must click on each item at least once

- Removal of 3 participants
- $n=120$
- Average clicks in learning phase = 62.66, S.D = 17.191
- Range = 21-109
- Adjusted mean=33.71%
- Adjusted S.D = 24.94



- **ALPACAA 4 (grammatical inferencing)**
- Adjusted for not doing learning phase (20 items)
- Criteria – must click on each item at least once

- Removal of 3 participants
- $n=120$
- Average clicks in learning phase = 102.5, S.D = 44.56
- Range = 21-259
- Adjusted mean=68.25%
- Adjusted S.D = 15.67





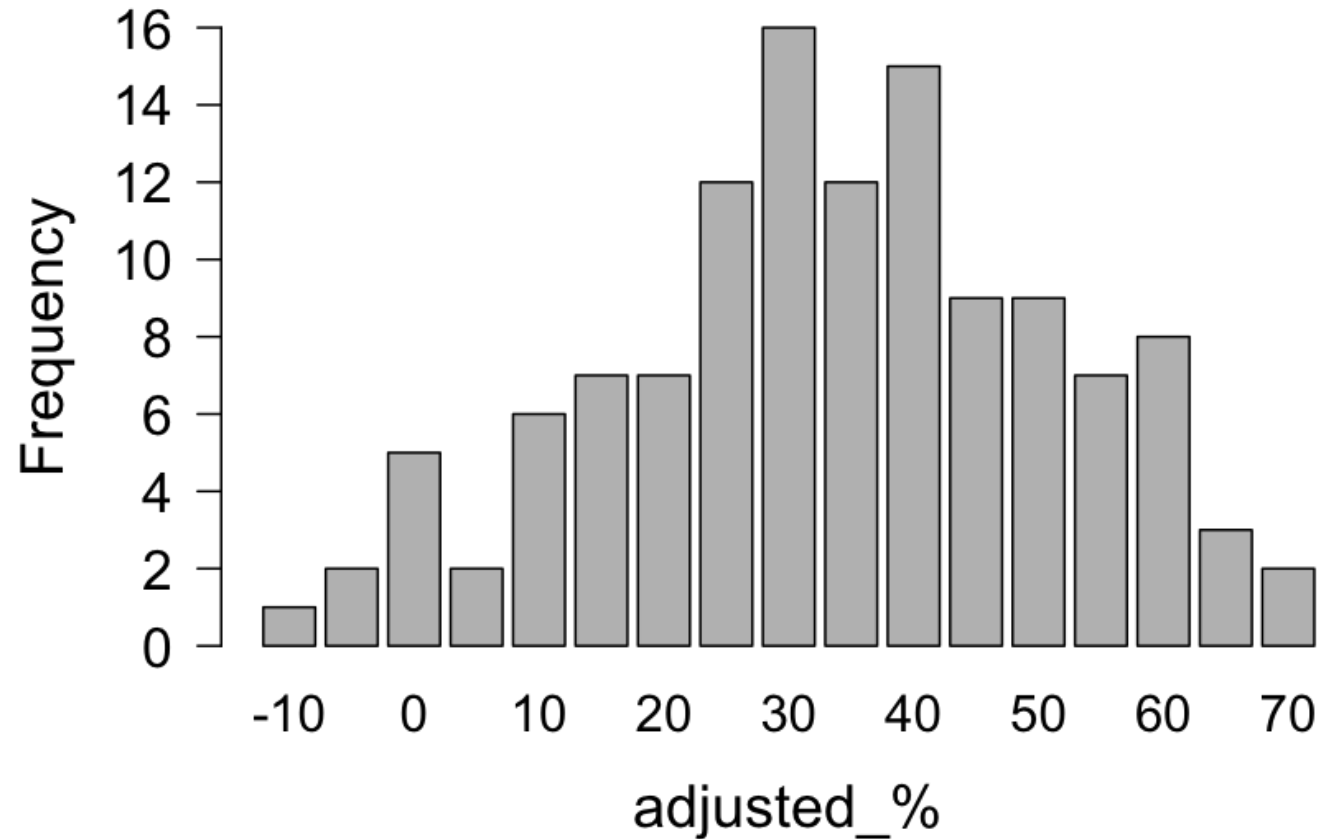
# What about adjusting for guessing?

- Previous figures were total correct only.
- ALPACAA\_2 (vocabulary): chose one from 20 options – guessing at 5%
- ALPACAA\_3 (sound/symbol): chose one from 20 options – guessing at 5%
- BUT:
- ALPACAA\_1 (sound recognition): chose one from 2 options – guessing at 50%
- ALPACAA\_4 (grammatical inferencing): chose one from 2 options
  - first answer: 12/20 (or 60%)
  - second answer: 8/20 (or 40%)



# What about adjusting for guessing?

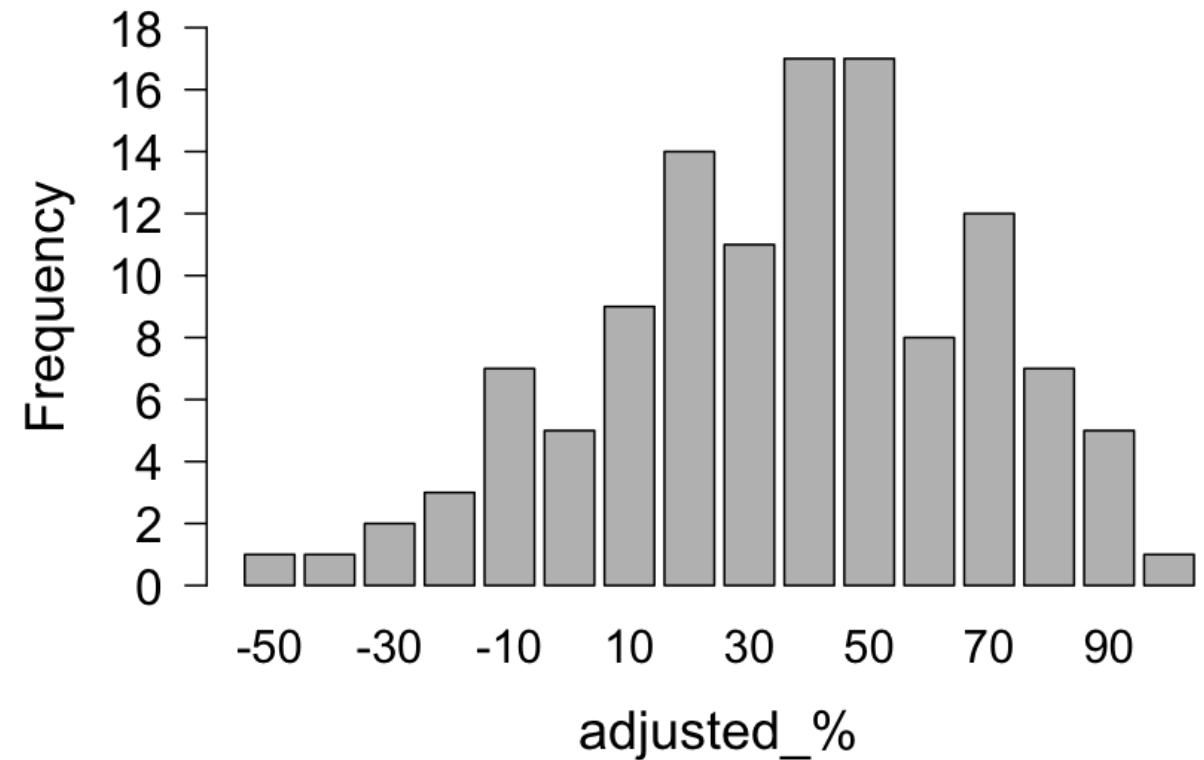
- ALPACAA\_1 (sound recognition):  
chose one from 2 options
- LLAMA: lose 1 mark (5%) for  
every incorrect choice
- Adjusted for incorrect scores:
- $M=33.74$ ,  $S.D=17.86$ 
  - Mean was 68.67
- Range: -10 - 70





# What about adjusting for guessing?

- ALPACAA\_4 (grammatical inferencing):  
chose one from 2 options
- LLAMA: lose 1 mark (5%) for every  
incorrect choice
- Adjusted for incorrect scores:
- $M=36.50$ ,  $S.D=31.35$ 
  - Mean was 68.33
- Range: -50 - 100





## Previous validation work: Grañena



Swansea University  
Prifysgol Abertawe

- Grañena (2013):
  - Internal consistency, Gender and Language neutrality
  - n=187 aged 18-39
  - L1s: Spanish, Chinese and English
  - internal consistency but two forms of aptitude
  - LLAMA D measures implicit and others explicit?
- Grañena (2018):
  - Compared 4 LLAMA tests with 4 Hi-LAB (n=135)
  - Found 3 underlying constructs across the tests.
  - Only the factor with LLAMA D and ALTM Synonym
  - (Hi-LAB) significantly predicted L2 fluency (pruned speech rate per min).

# Further evidence: age and bilingualism



Swansea University  
Prifysgol Abertawe

- WM and aptitude are affected by age and bilingualism in different ways.
- Bilingual advantage in older group across 3 of the LLAMA aptitude tests.
- Age advantage on one of the WM tests (TMT A).

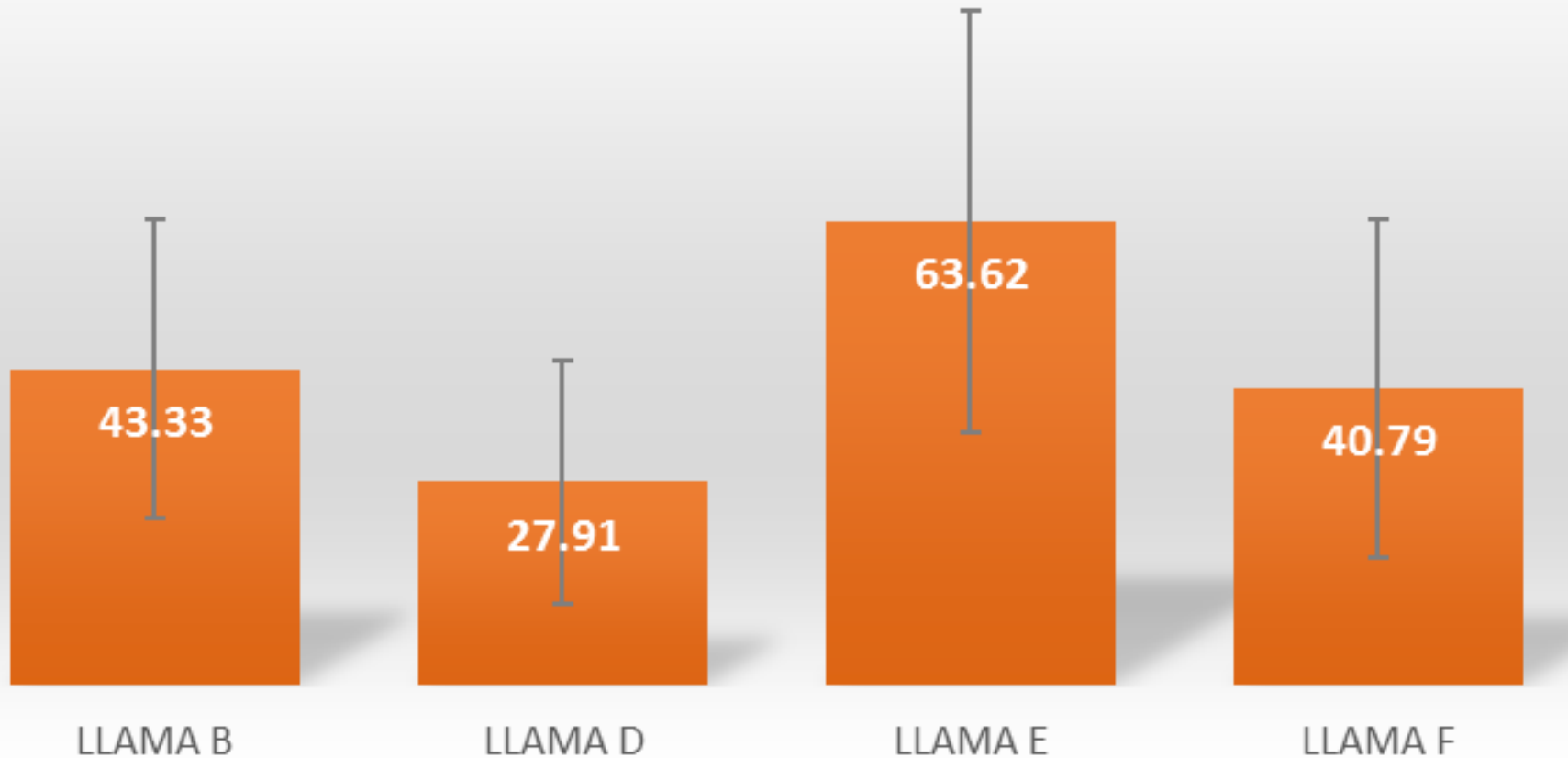
|                  | Group 1      | Group 2      |
|------------------|--------------|--------------|
| Mean age (range) | 21 (18-23)   | 61.5 (50-78) |
| Bilingual        | 14 (7 F, 7M) | 14 (7 F, 7M) |
| Monolingual      | 14 (7 F, 7M) | 14 (7 F, 7M) |
| n                | 28           | 28           |

# Results: Aptitude



Swansea University  
Prifysgol Abertawe

## LLAMA test scores

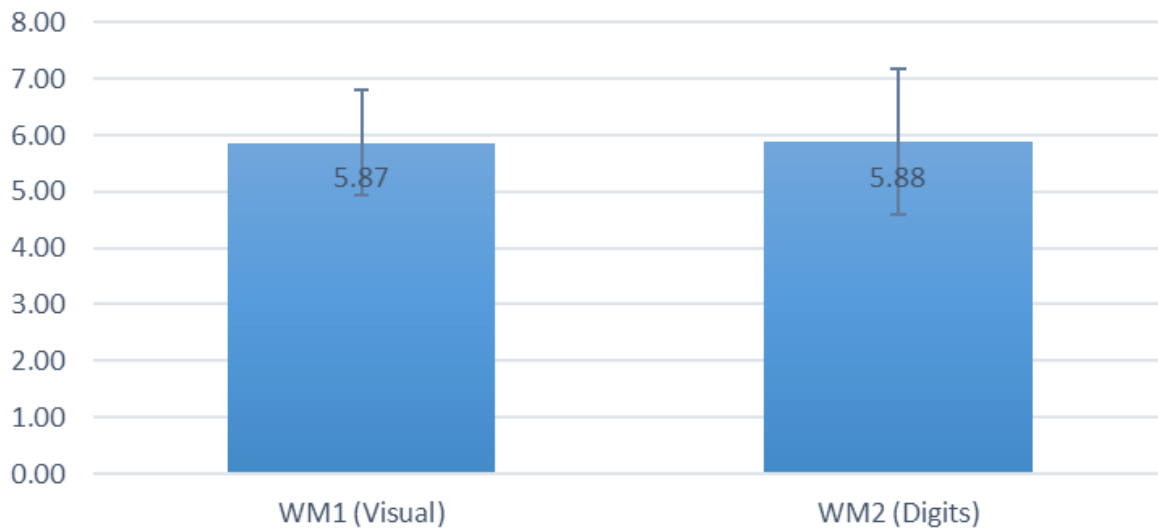


# Results: WM

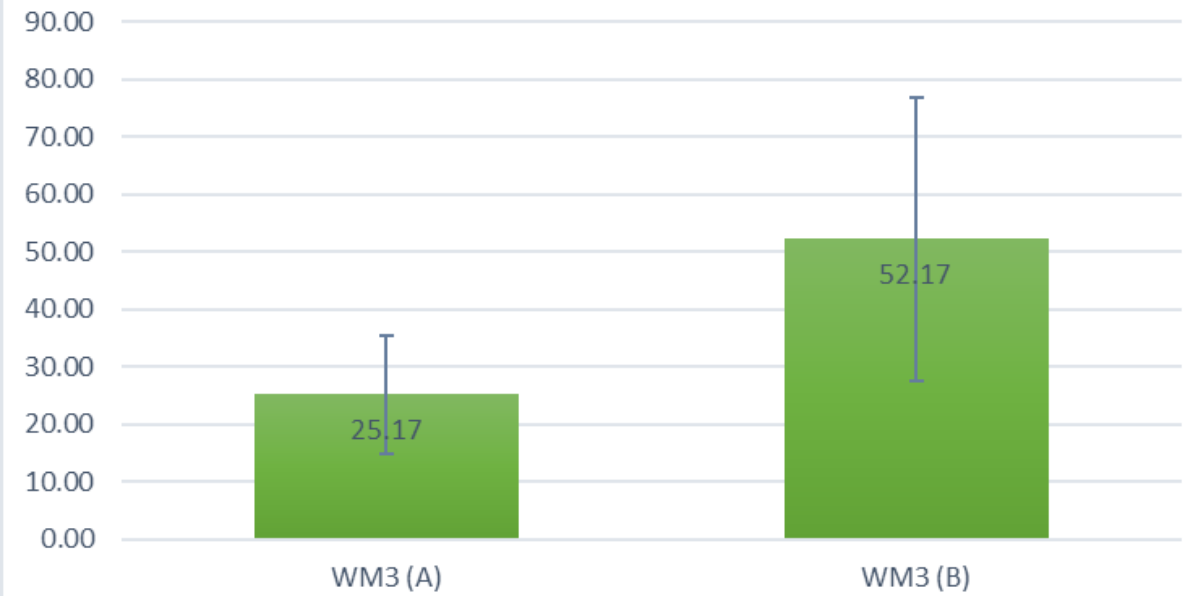


Swansea University  
Prifysgol Abertawe

**Working Memory tests  
(visual and backwards digit span)**



**TMT A & B (attentional control)**



|                 |             | LLAMA B   | LLAMA D  | LLAMA E   | LLAMA F   | WM1 (Visual) | WM2 (Digits) | WM3 (A)   |
|-----------------|-------------|-----------|----------|-----------|-----------|--------------|--------------|-----------|
| LLAMA B         | Pearson's r | —         |          |           |           |              |              |           |
|                 | p-value     | —         |          |           |           |              |              |           |
| LLAMA D         | Pearson's r | 0.299 *** | —        |           |           |              |              |           |
|                 | p-value     | < .001    | —        |           |           |              |              |           |
| LLAMA E         | Pearson's r | 0.387 *** | 0.240 ** | —         |           |              |              |           |
|                 | p-value     | < .001    | 0.006    | —         |           |              |              |           |
| LLAMA F         | Pearson's r | 0.500 *** | 0.263 ** | 0.524 *** | —         |              |              |           |
|                 | p-value     | < .001    | 0.003    | < .001    | —         |              |              |           |
| WM1<br>(Visual) | Pearson's r | 0.242 **  | 0.143    | 0.345 *** | 0.340 *** | —            |              |           |
|                 | p-value     | 0.006     | 0.107    | < .001    | < .001    | —            |              |           |
| WM2 (Digits)    | Pearson's r | 0.201 *   | 0.149    | 0.233 **  | 0.258 **  | 0.440 ***    | —            |           |
|                 | p-value     | 0.023     | 0.092    | 0.008     | 0.003     | < .001       | —            |           |
| WM3 (A)         | Pearson's r | -0.263 ** | -0.153   | -0.089    | -0.152    | -0.234 **    | -0.169       | —         |
|                 | p-value     | 0.003     | 0.083    | 0.318     | 0.086     | 0.008        | 0.056        | —         |
| WM3 (B)         | Pearson's r | -0.253 ** | -0.107   | -0.166    | -0.281 ** | -0.274 **    | -0.195 *     | 0.639 *** |
|                 | p-value     | 0.004     | 0.226    | 0.060     | 0.001     | 0.002        | 0.027        | < .001    |

# Correlational results



Swansea University  
Prifysgol Abertawe

- Significant weak correlations found with LLAMA B, E & F with Visual and Digits WM scores.
  - LLAMA B, E & F = explicit measures
  - WM = visuo-spatial (reading) & phonological loop
  - Lack of correlation with LLAMA D
- Significant weak correlations between TMT B and LLAMA B & F.
- Significant weak correlations between TMT A and LLAMA B
  - TMT A & B = central executive / attentional control
  - LLAMA B = vocabulary, LLAMA F = grammatical inferencing