Swansea University
Prifysgol Abertawe

LLAMAs, ALPACAAs and testing aptitude

Vivienne Rogers & Paul Meara

Outline

- What is aptitude?
- Background on LLAMA tests
- Methodology: how we've revised the tests
- Results & Discussion
- And back to LLAMA

What is Language Learning Aptitude

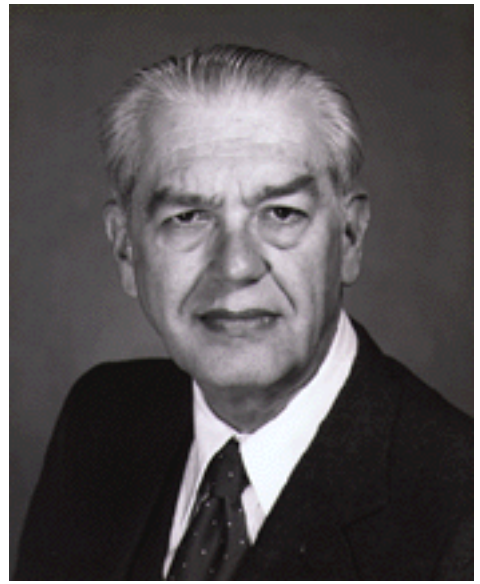
Swansea University
Prifysgol Abertawe

- A ’knack for learning languages’.
  A cognitive variable - something you are born with.
- What does it mean?
  - Learn faster?
  - Better capacity for learning? Ultimate attainment?
  - More intelligent?

What is Language Learning Aptitude

Swansea University
Prifysgol Abertawe

"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)



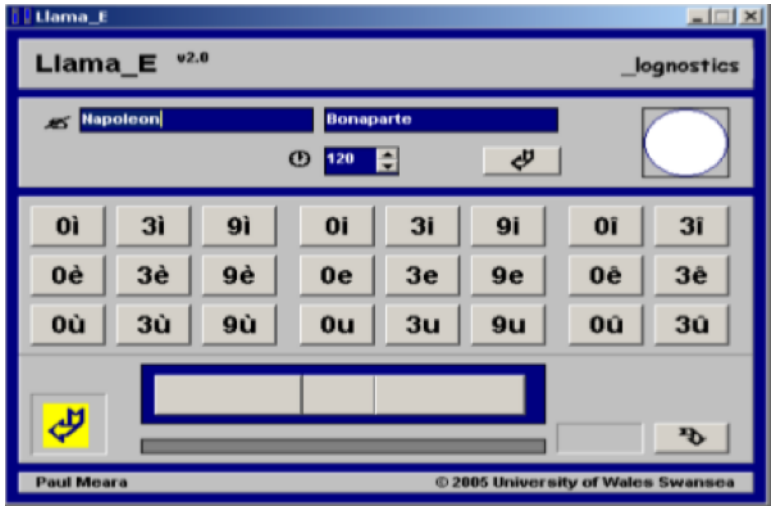
- aptitude is different from other cognitive systems, including intelligence
- aptitude is stable (doesn't change)
- aptitude is made up of different components

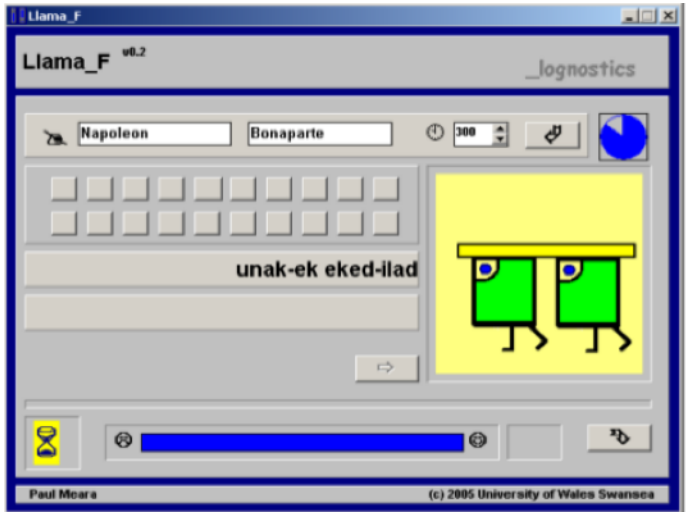
Swansea LLAMA tests (Meara, 2005)

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.





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).

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: β = -.250, contribution to variance = 6.0
  - LLAMA D: β = .136, contribution to variance = 1.8
  - LLAMA F: β = -.165, contribution to variance = 2.6

Previous validation work: Bokander & Bylund (2019)

Swansea University
Prifysgol Abertawe

- Scoring
  - LLAMA B performed well.
  - Others did not (particularly D)
- Generalization
  - Internal consistency
  - LLAMA B & E met .70 criterion
  - LLAMA E: analytic/ strategic use of vowels only rather than sound/symbol

- Explanation
  - Construct & content validity
  - Possibly doesn't reflect Skehan's (1998) three components of aptitude
  - Two component: LLAMA D is different to the others.

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?

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

Participants

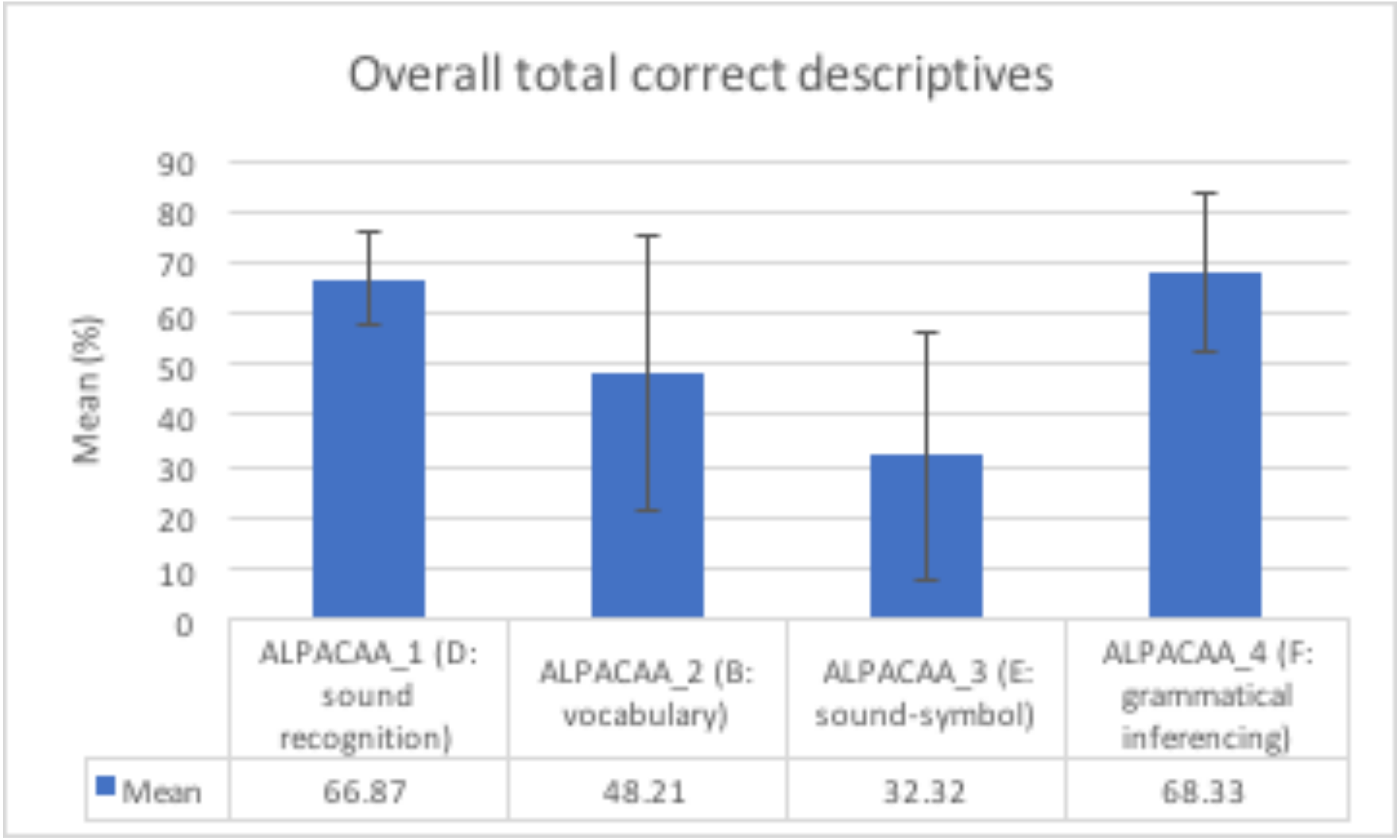
Swansea University
Prifysgol Abertawe

- 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

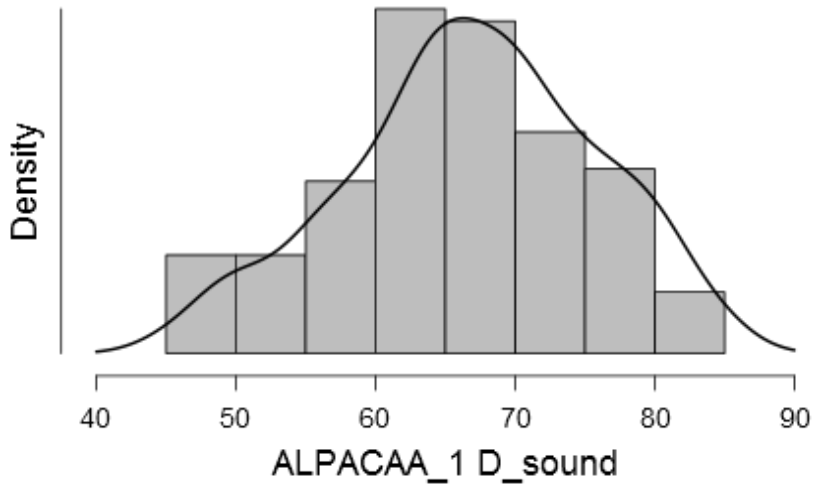
RQ1:
What is the impact of different scoring mechanisms on the distribution of ALPACAA scores?

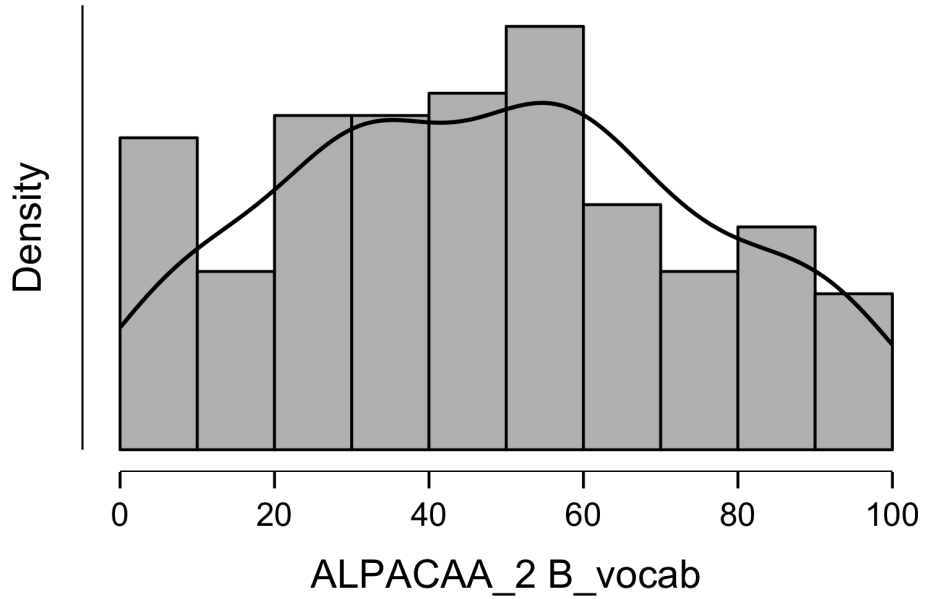
Overall descriptive: total correct – n=123

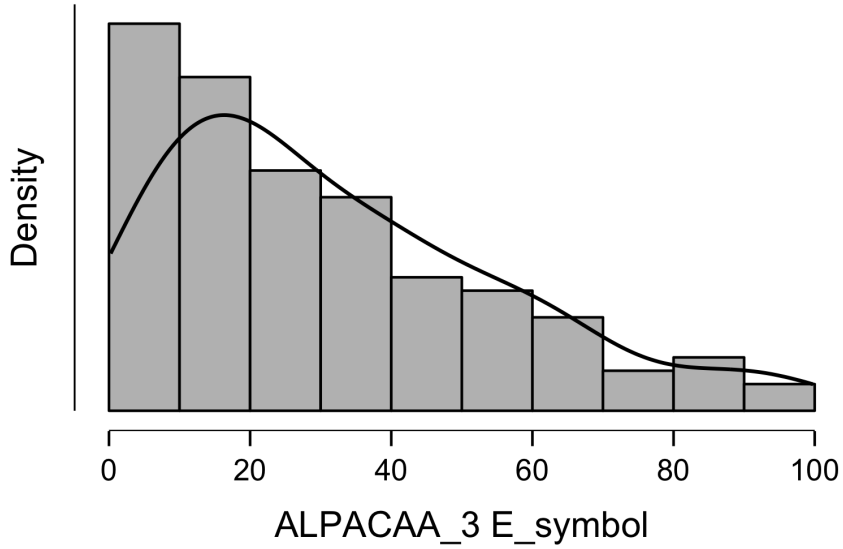


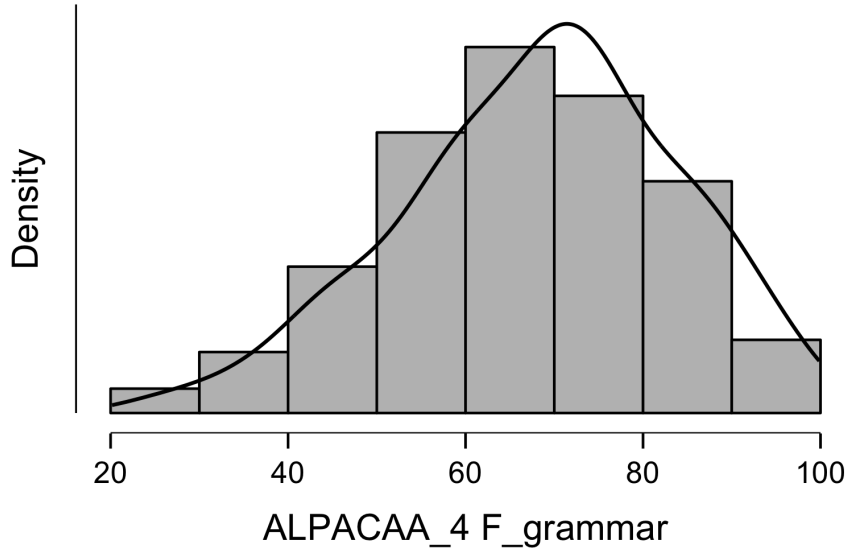
Swansea University
Prifysgol Abertawe

- Distribution of scores in tests (total correct)









Swansea University
Prifysgol Abertawe

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

Prifysgol Abertawe
Swansea University

- Have RTs for all test items for all participants..
- Excluded any RTs faster than 200ms.

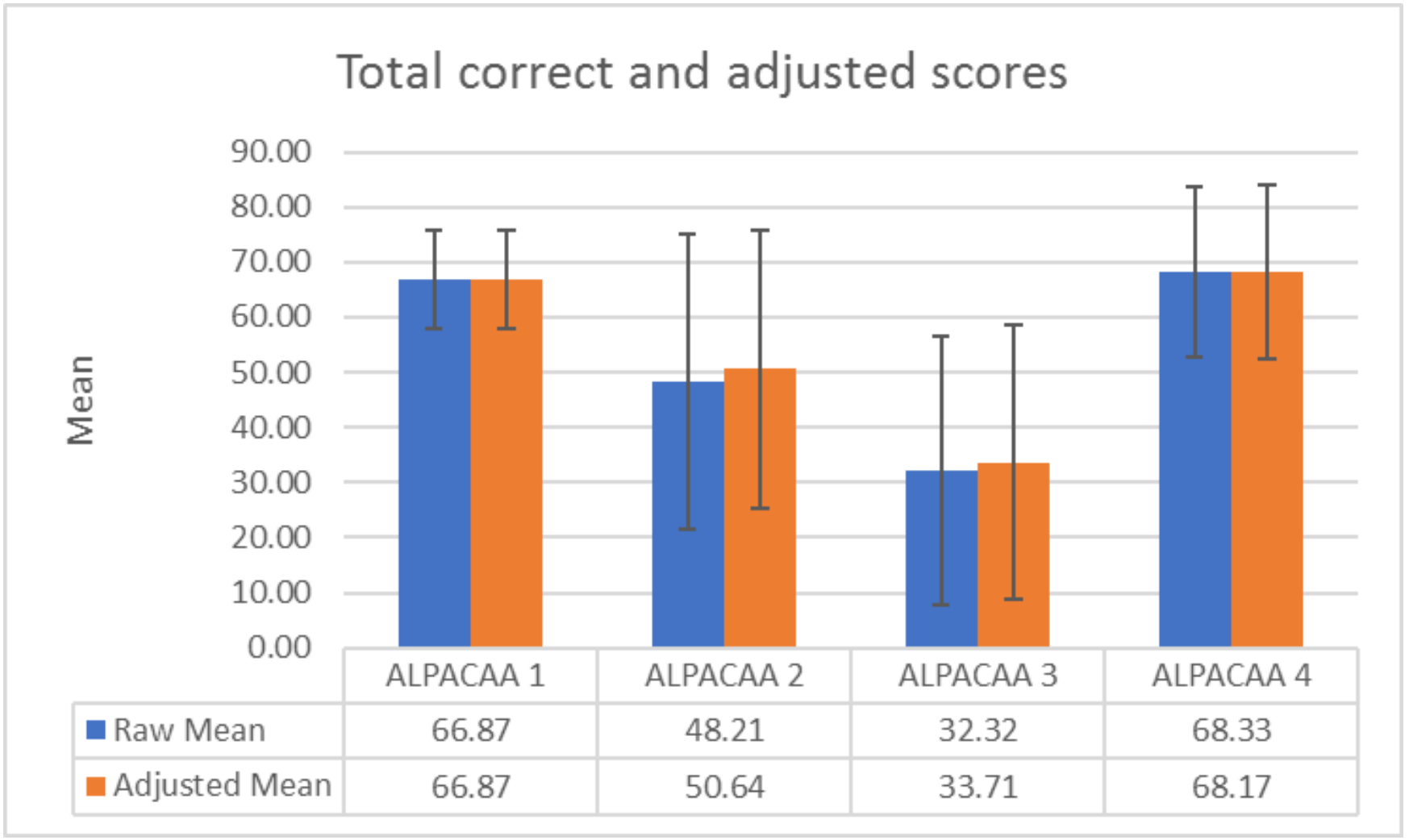
- ALPACAA_1_D: In test phase, heard sound then question prompt then click.
  - More than 200ms after sound so no exclusions.

- ALPACAA_2_B: Three items identified (out of 117*20 = 2340)
  - Two were correct: removed.

- ALPACAA_3_E: No items

- ALPACAA_4_F: Four items identified (out of 120*20=2400)
  - Two were correct: removed.

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



	ALPACAA 1	ALPACAA 2	ALPACAA 3	ALPACAA 4
Raw Mean	66.87	48.21	32.32	68.33
Adjusted Mean	66.87	50.64	33.71	68.17

Spearman's correlations: average RT and total correct (adjusted for learning phase)

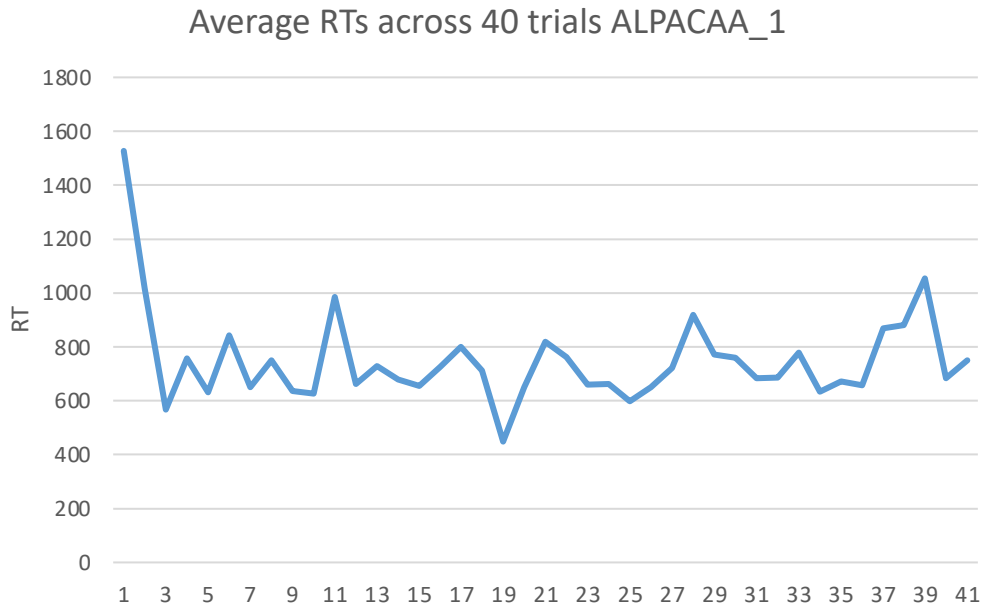
Swansea University
Prifysgol Abertawe

	rho	*p*	Lower 95% CI	Upper 95% CI
ALPACAA 1 (sound recognition)	-0.009	0.918	-0.186	0.168
ALPACAA 2 (vocabulary)	-0.078	0.406	-0.255	0.106
ALPACAA 3 (sound-symbol)	0.550***	<.001	0.664	0.411
ALPACAA 4 (grammatical inferencing)	0.401***	<0.001	0.239	0.541

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 +/- 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 α	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 correlation	If item dropped Cronbach's α
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 correlation	If item dropped Cronbach's α
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)

Swansea University
Prifysgol Abertawe

	n	Cronbach's α	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) revised	123	0.535	0.036	0.492	0.575

ALPACAA_1 Item Reliability Statistics (yes only)

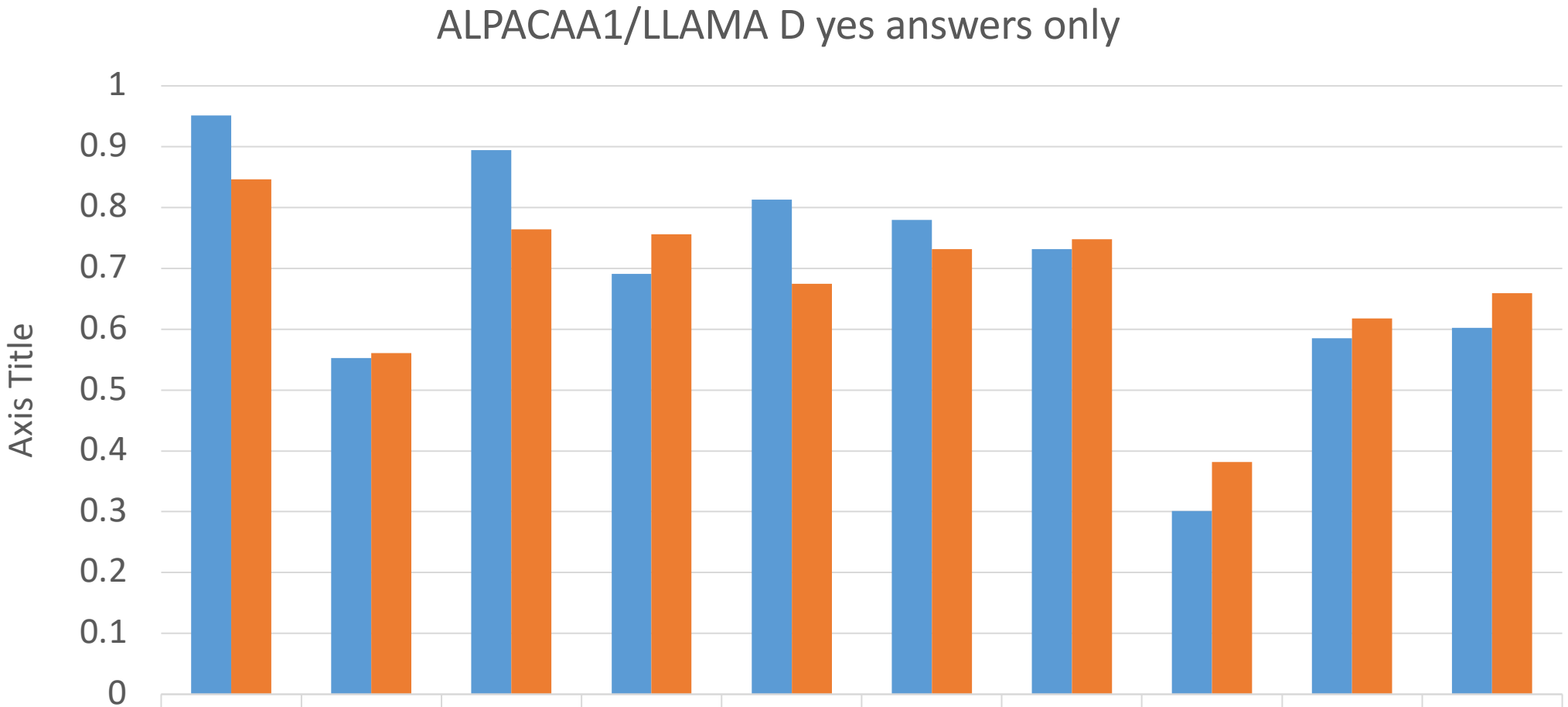
	mean	sd	item-rest correlation	If item dropped Cronbach's α
latd03-y1	0.894	0.309	0.180	0.532
latd08-y1	0.301	0.460	0.106	0.543
latd05-y1	0.813	0.391	0.091	0.544
latd04-y1	0.691	0.464	0.258	0.517
latd06-y1	0.780	0.416	-0.129	0.576
latd09-y1	0.585	0.495	-0.007	0.564
latd10-y1	0.602	0.492	0.217	0.524
latd07-y1	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)

Swansea University
Prifysgol Abertawe

	n	Cronbach's α	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) revised	123	0.535	0.036	0.492	0.575
ALPACAA_1 (sound recognition) correct	123	0.544	0.502	0.502	0.584
ALPACAA_1 (sound recognition) correct revised	123	0.593	0.075	0.555	0.629



	latd01	latd02	latd03	latd04	latd05	latd06	latd07	latd08	latd09	latd10
Listen1	0.951	0.553	0.894	0.691	0.813	0.78	0.732	0.301	0.585	0.602
Listen2	0.846	0.561	0.764	0.756	0.675	0.732	0.748	0.382	0.618	0.659

Paired t-tests first and second listen

Swansea University
Prifysgol Abertawe

Paired Samples T-Test

Measure 1	Measure 2	t	df	p
latd01-y1	latd01-y2	3.275	122	0.001
latd02-y1	latd02-y2	-0.164	122	0.870
latd03-y1	latd03-y2	3.130	122	0.002
latd04-y1	latd04-y2	-1.301	122	0.196
latd05-y1	latd05-y2	3.058	122	0.003
latd06-y1	latd06-y2	0.948	122	0.345
latd07-y1	latd07-y2	-0.364	122	0.717
latd08-y1	latd08-y2	-1.515	122	0.132
latd09-y1	latd09-y2	-0.616	122	0.539
latd10-y1	latd10-y2	-1.068	122	0.288

Note. Student's t-test.

ALPACAA_4 Item Reliability Statistics

	mean	sd	item-rest correlation	If item dropped Cronbach's α
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 α	Average inter item correlation	95% CI Lower	95% CI Higher
ALPACAA_4 (grammatical inferencing)	120	0.617	0.079	0.581	0.650
ALPACAA_4 (grammatical inferencing) revised	120	0.682	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)

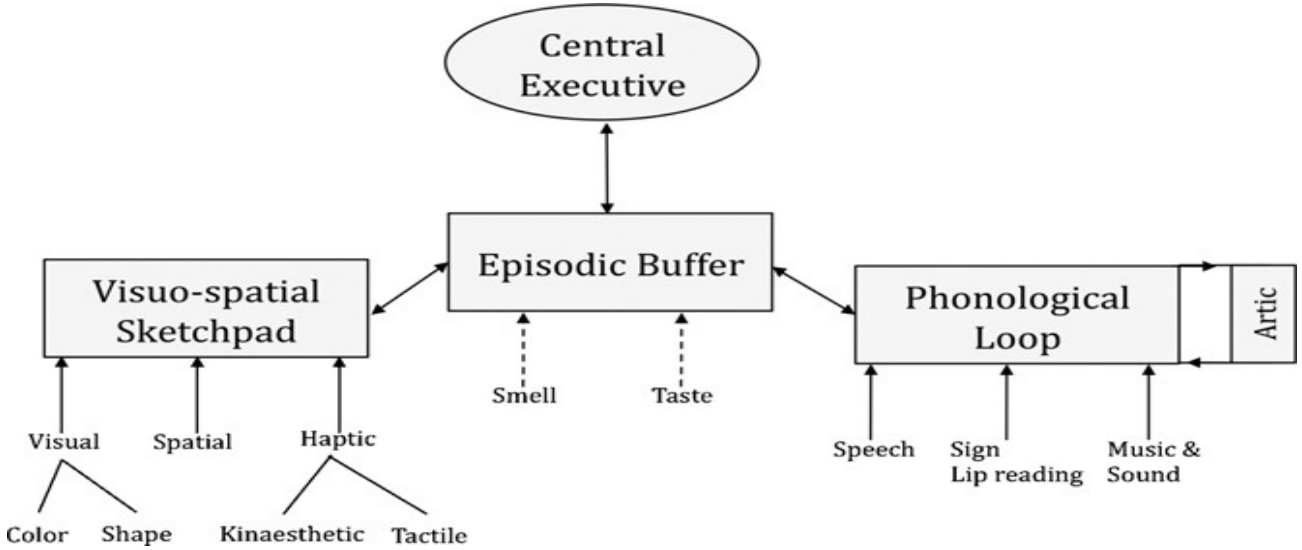
Swansea University
Prifysgol Abertawe

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

What is working memory?

Swansea University
Prifysgol Abertawe

"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

Previous work with LLAMA
(presented at EUROSLA 2017)

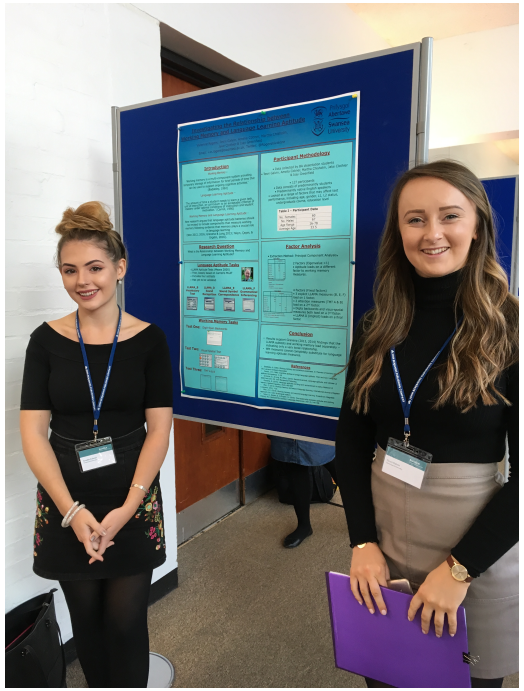
Swansea University
Prifysgol Abertawe

- Data collected by BA dissertation students:
  - Tesni Galvin, Amelia Cobner, Martha Chisholm, Jake Clothier & Issy Greenfield

Table I – Participant Data

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

- 127 participants
  - predominantly students
- Typically L1 English speakers



Results: PCA

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

Pattern Matrix[a]

	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: Oblimin 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[a]

	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: Oblimin with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Working memory results (n=123)

Swansea University
Prifysgol Abertawe

	flanker conflict cost	stroop conflict 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 conflict cost	Spearman's rho	—						
	p-value	—						
stroop conflict cost	Spearman's rho	0.065	—					
	p-value	0.478	—					
DB_span	Spearman's rho	0.049	-0.252**	—				
	p-value	0.590	0.005	—				
A1_total_correct	Spearman's rho	0.006	-0.046	0.073	—			
	p-value	0.948	0.616	0.428	—			
A2_total_correct	Spearman's rho	-0.012	-0.186*	0.432***	0.200*	—		
	p-value	0.901	0.045	1.161e -6	0.031	—		
A3_otal_correct	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_correct	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

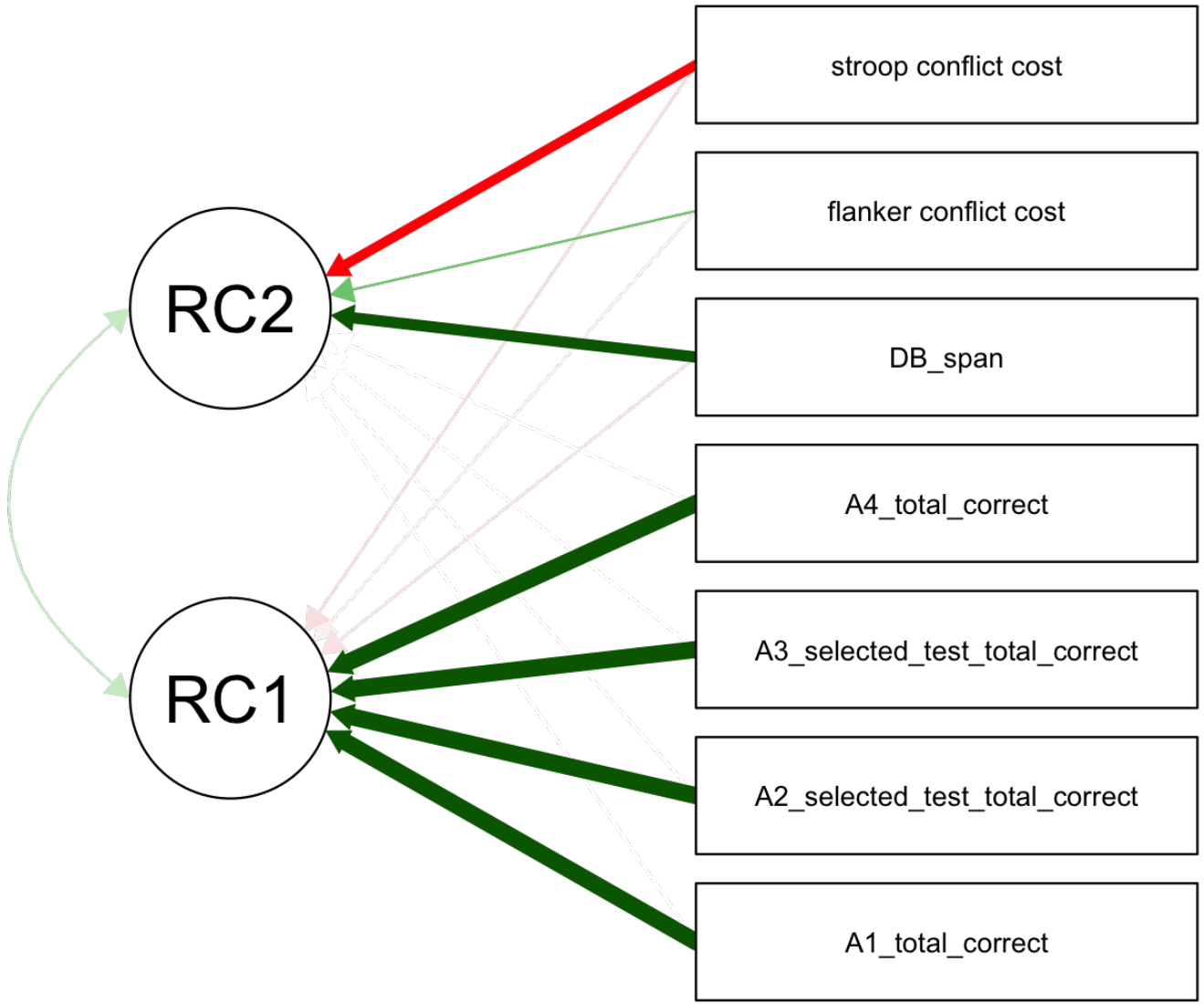
PCA analysis: WM and ALPACAA component total correct scores

Swansea University
Prifysgol Abertawe

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.


RC2
RC1
stroop conflict cost
flanker conflict cost
DB_span
A4_total_correct
A3_selected_test_total_correct
A2_selected_test_total_correct
A1_total_correct

Exploratory factor analysis

Swansea University
Prifysgol Abertawe

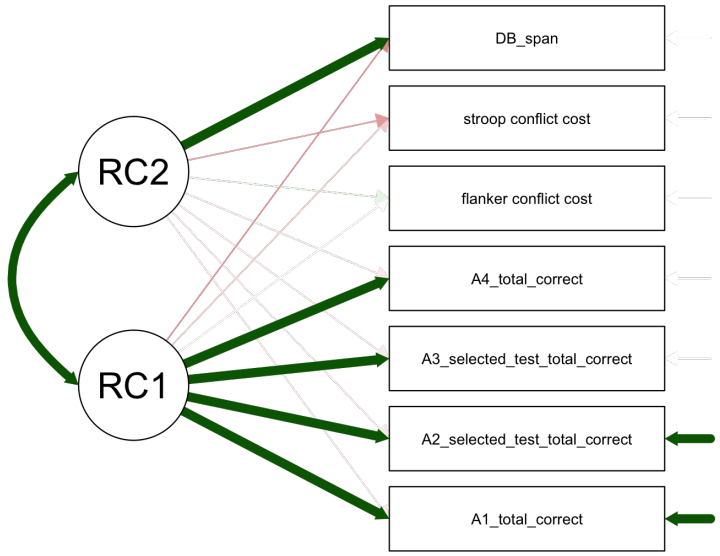
Factor Loadings

	Factor 1	Factor 2	Uniqueness
flanker conflict cost			0.999
stroop conflict cost			0.975
DB_span		1.019	0.005
A1_total_correct	1.005		0.001
A2_selected_test_total_correct	1.005		0.001
A3_selected_test_total_correct	1.004		0.001
A4_total_correct	1.005		0.001

Note. Applied rotation method is promax.

Chi-squared Test

	Value	df	p
Model	36364 .184	8	< .001



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 penalities?

Overall ALPACAA conclusion

Swansea University
Prifysgol Abertawe

- 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.

Back to LLAMA

- Updated test versions on Paul Meara's website (www.lognostics.co.uk)
- Web-based, work on different browsers.
- LLAMA B (vocabulary) = same
- LLAMA D (sound recognition) = now out of 100
- LLAMA E (sound/symbol) = revised test layout
- LLAMA F (grammatical inferencing) = revised test layout

Download versions will not be supported after Jan 2020

The LLAMA tests

_lognostics

This page contains beta versions of the LLAMA_3 language aptitude tests.
If you need access to earlier versions of the LLAMA tests, contact p.m.meara@gmail.com

READ

The Manual
last update: 05.04.2020

RUN MAN

LLAMA_B3: *learning vocabulary*
v3.00 beta last update 05.04.2020

RUN MAN

LLAMA_D3: *listening for new words*
v3.00 beta last update 05.04.2020

RUN MAN

LLAMA_E3: *sounds and symbols*
v3.00 beta last update 05.04.2020

RUN MAN

LLAMA_F3: *grammar rules*
v3.00 beta last update 05.04.2020

READ

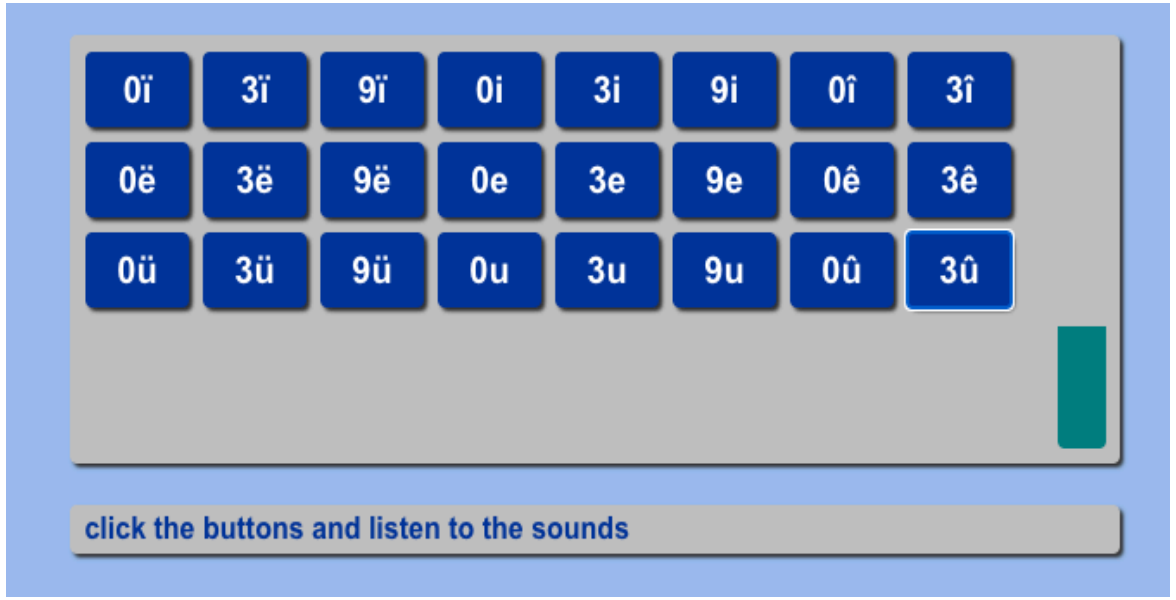
FAQs and Updates
last update: 05.04.2020

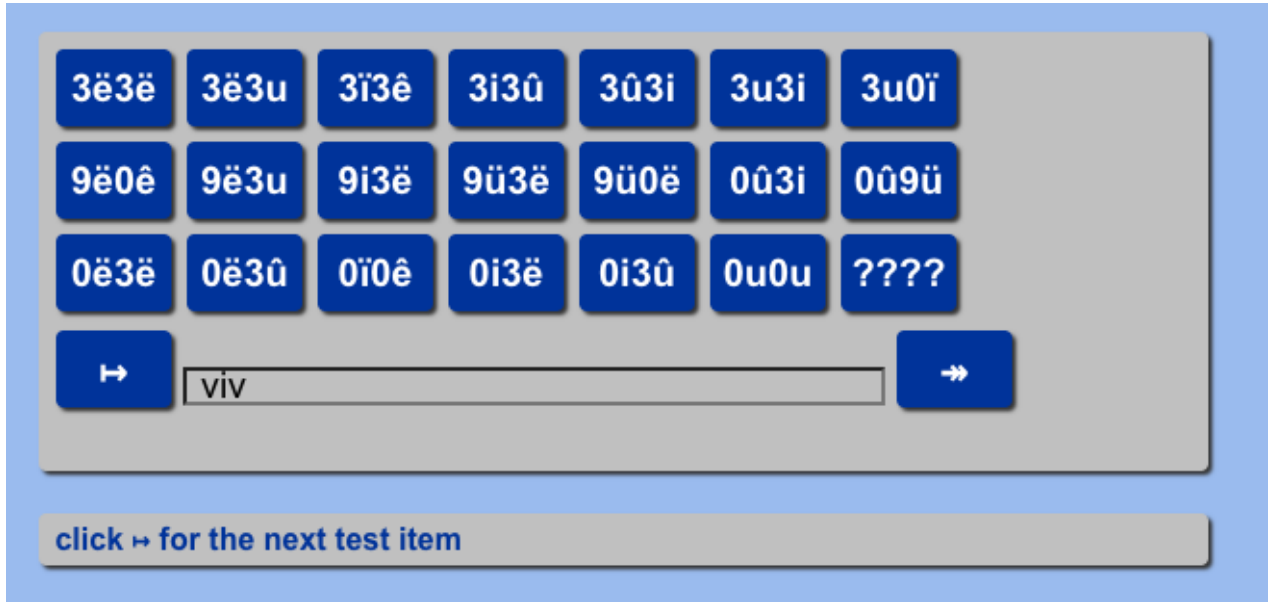
READ

Research Notes
last update: 23.04.2020

Revised layouts: LLAMA E

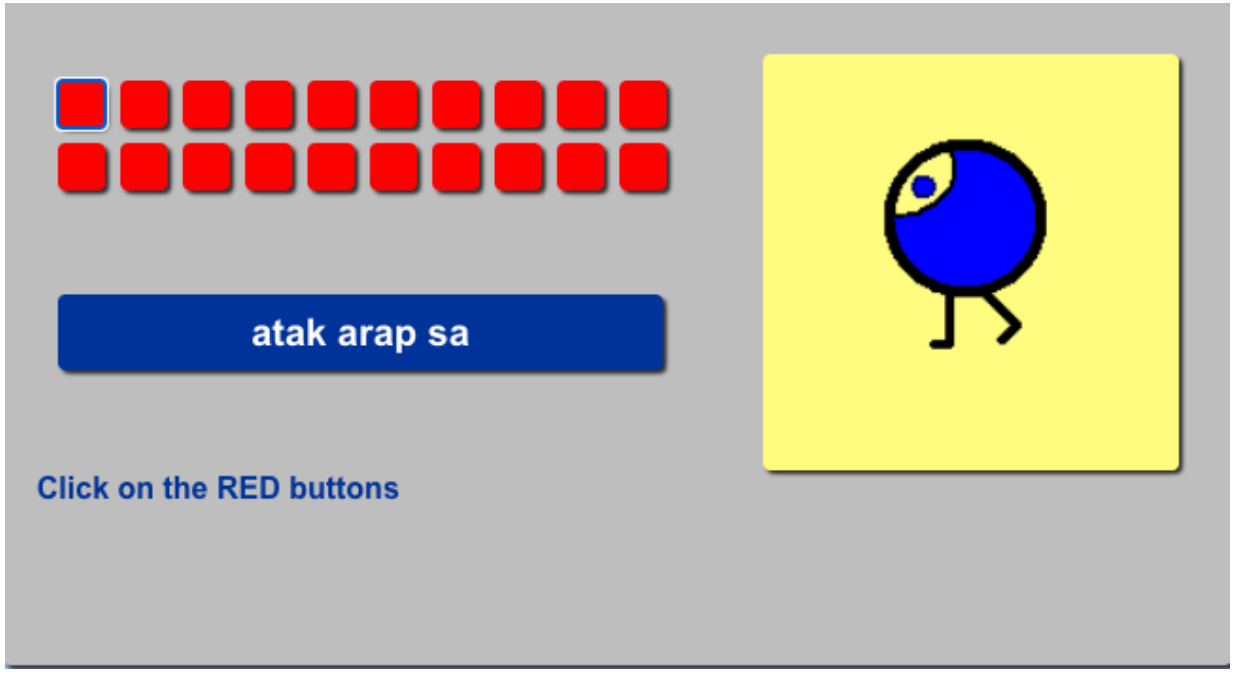
Swansea University
Prifysgol Abertawe





Revised layout LLAMA F

Swansea University
Prifysgol Abertawe



Numbers taking new LLAMAs

LLAMA B	LLAMA D	LLAMA E	LLAMA F
812	2261 (-1500)	448	287

Swansea University
Prifysgol Abertawe

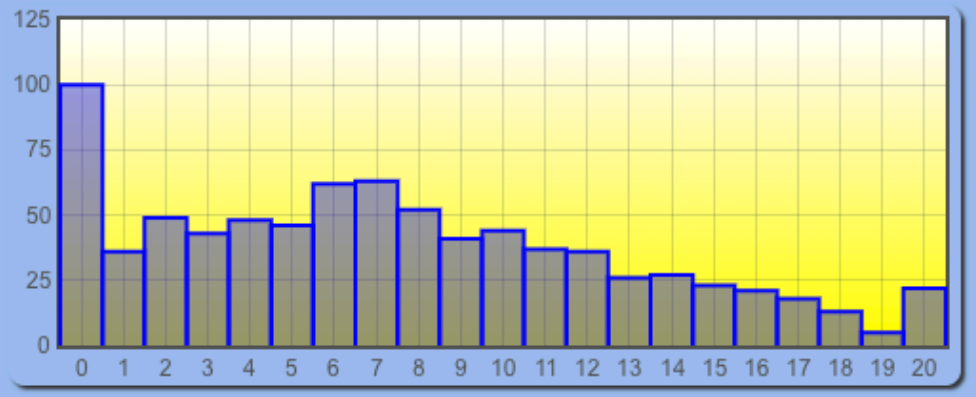
LLAMA_B v3.0
Learning new words

_lognostics

Report for

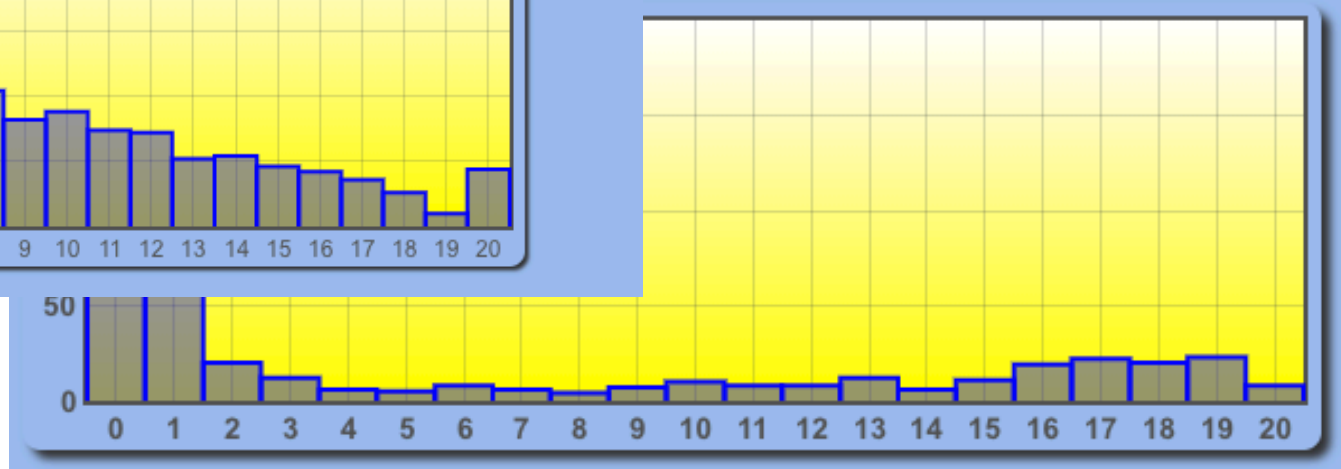
The maximum score was 20. Your score was 0

812 people have taken the test so far. Here are their scores:

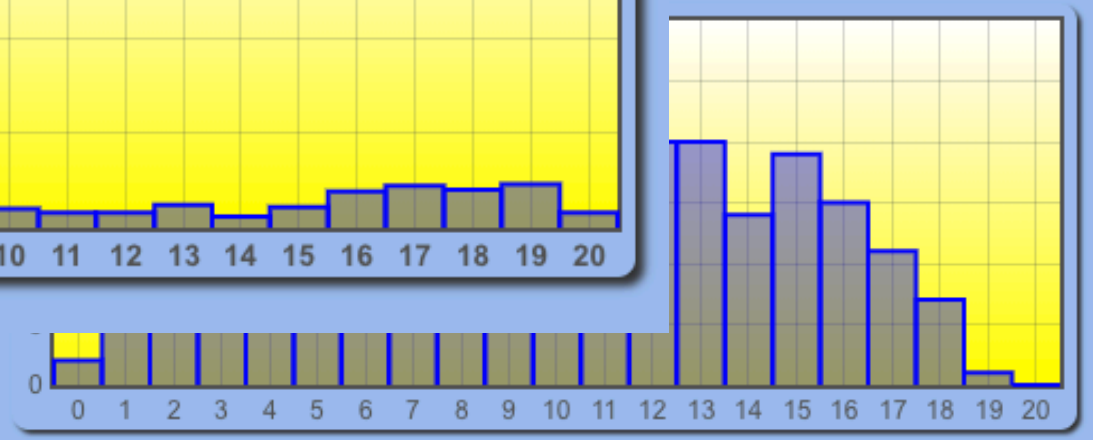


_logno

. Here are their scores:



e their scores:



Next steps….

Swansea University
Prifysgol Abertawe

- Item analysis on new LLAMAs
- Re-programme into Gorilla (or similar)
  - Background variables
- Memory effects – not just WM but declarative and procedural

Thank you! Diolch yn fawr!

Vivienne Rogers: v.e.rogers@swansea.ac.uk
twitter: @RogersVivienne

Paul Meara: p.m.meara@gmail.com