



Swansea University
Prifysgol Abertawe

LLAMA test development

Vivienne Rogers, Paul Meara and Brian Rogers



Outline

- What is aptitude?
- Background on LLAMA v.1 tests
- How we've revised the tests (with ALPACAAs)
 - Introducing LLAMA v. 3
- Has it made any difference?





What is Language Learning Aptitude

- A 'knack for learning languages'.
A cognitive variable - something you are born with.
- What does it mean?
 - Learn faster? (Carroll's approach)
 - Better capacity for learning? Ultimate attainment? (Robinson, 2005)
 - Same as working memory? (Wen, 2019)
 - Same as long term memory? (Buffington & Morgan-Short, 2019)
 - More intelligent?
 - Overlap between highest MLAT scores and IQ.

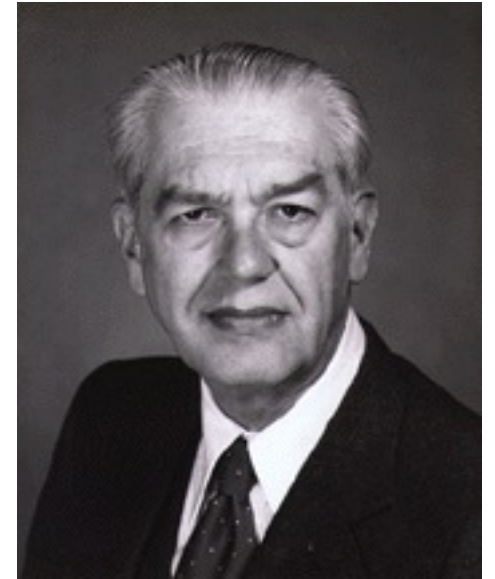


Swansea University
Prifysgol Abertawe

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)

- aptitude is different from other cognitive systems, including intelligence
- aptitude is stable (doesn't change)
 - This is also debated (training effects, prior instruction)
 - Is it aptitude or performance on the test?
- aptitude is made up of different components
- MLAT test (Carroll & Sapon, 1959)





Examples of previous work on aptitude

- types of feedback (Kourtali & Révész, 2020; Yilmaz, 2013; Yilmaz & Grañena, 2019);
- language attrition (Bylund et al., 2010; Bylund & Ramírez-Galan, 2016);
- explicit and implicit knowledge (Suzuki & DeKeyser, 2017b);
- development of proficiency (Artieda & Muñoz, 2016; Saito, 2017; Saito et al., 2019; Suzuki & DeKeyser, 2017a);
- age effects (Saito, 2015, Roehr-Brackin et al, 2023);
- study abroad (Serrano & Llanes, 2015);
- near-native language attainment (Abrahamsson & Hyltenstam, 2008);
- working memory (Wen, 2016);
- Long term memory (Buffington & Morgan-Short, 2021)

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



Swansea University
Prifysgol Abertawe

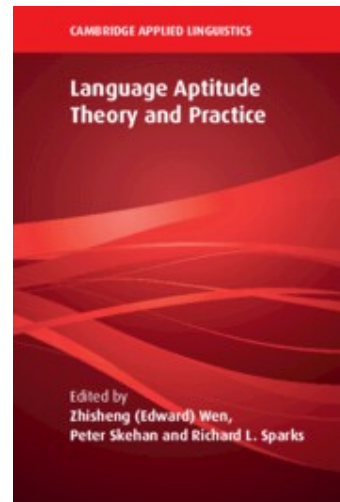
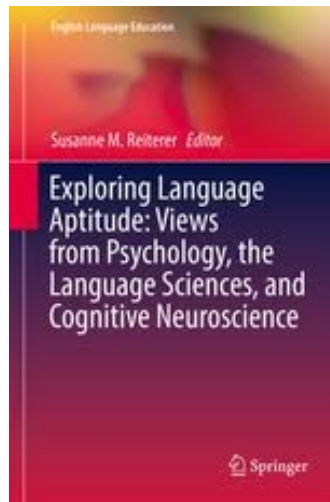
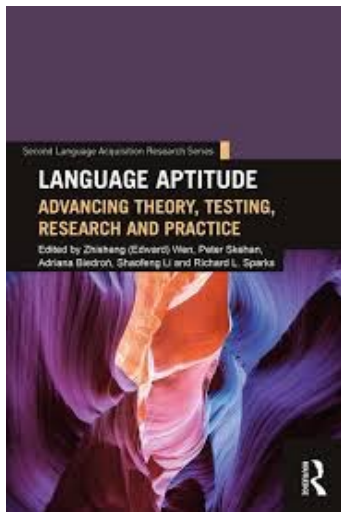
- Aptitude is independent of other individual differences, e.g. motivation.
- executive working memory (EWM) more strongly associated with aptitude than phonological short-term memory (PSTM).
 - BUT Linck et al (2013): relevance of PSTM to advanced learners.
- strong predictor of general proficiency 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 (LLAMA?)



Swansea University
Prifysgol Abertawe

- Over 1500 citations on google scholar published since 2019!
- BUT 252 citations for LLAMA manual (Meara 2005)
- “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 University
Prifysgol Abertawe

Background on LLAMA tests

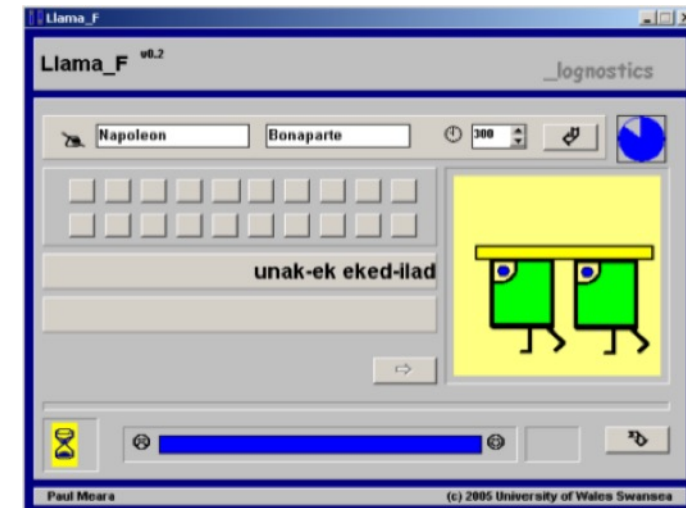
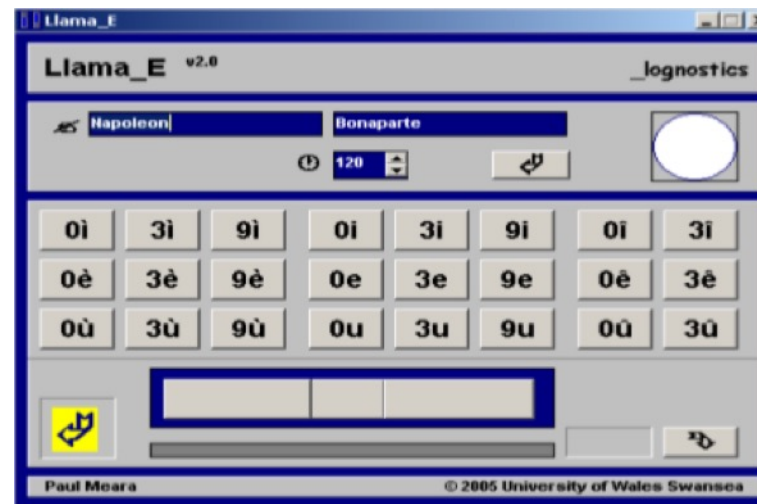
Swansea LLAMA tests v.1 (Meara, 2005)

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



Previous validation work: Bokander & Bylund (2020)

- 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

Revising the tests!

Test development



Swansea University
Prifysgol Abertawe

- Since 2013/14, two parallel strands of development
 1. Creating a web-based, cross-platform version of the tests.
 - a) 2016: LLAMA B (vocabulary test) made available online (v.2)
 - b) 2018/19: other tests available online
 - c) Beta version (ALPACAA) created (presented at EUROSIA 2019)
 - d) 2019: version 3 online – major changes to various subcomponents
 2. Making the tests more reliable.
 - a) Rogers, Meara et al (2016 & 2017) on factors that can influence the test scores.
 - b) Bokander & Bylund (2020)
 - c) Upcoming paper with Rogers, Bokander, Meara & Rogers
 - a) (sneak preview: reliability is much better)

General changes

- Online: compatible with Chrome, Firefox and Safari
- Manual for each programme
- New ID screen
 - Have to input in order to proceed
 - Removes upper and lower case problems
 - Removes use of non Roman alphabet
- Instructions in English
- Key presses now recorded

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**
v3.2 beta last update 25.04.2021

RUN **MAN** **LLAMA_D3: listening for new words**
v3.2 beta last update 28.04.2021

RUN **MAN** **LLAMA_E3: sounds and symbols**
v3.00 beta last update 26.05.2021

RUN **MAN** **LLAMA_F3: grammar rules**
v3.22 beta last update 26.05.2021

 **LLAMA_B v3.0 Learning Words** _lognostics

LLAMA_B tests how good you are at learning words in a foreign language.
The test takes about 10 minutes.
The program will give you two minutes to learn the names of 20 unfamiliar objects.
Then it will give you a short test.

CLEAR **BEGIN >>**

A B C D E F G H I J K L M 1 2 3 4 5
N O P Q R S T U V W X Y Z 6 7 8 9 0

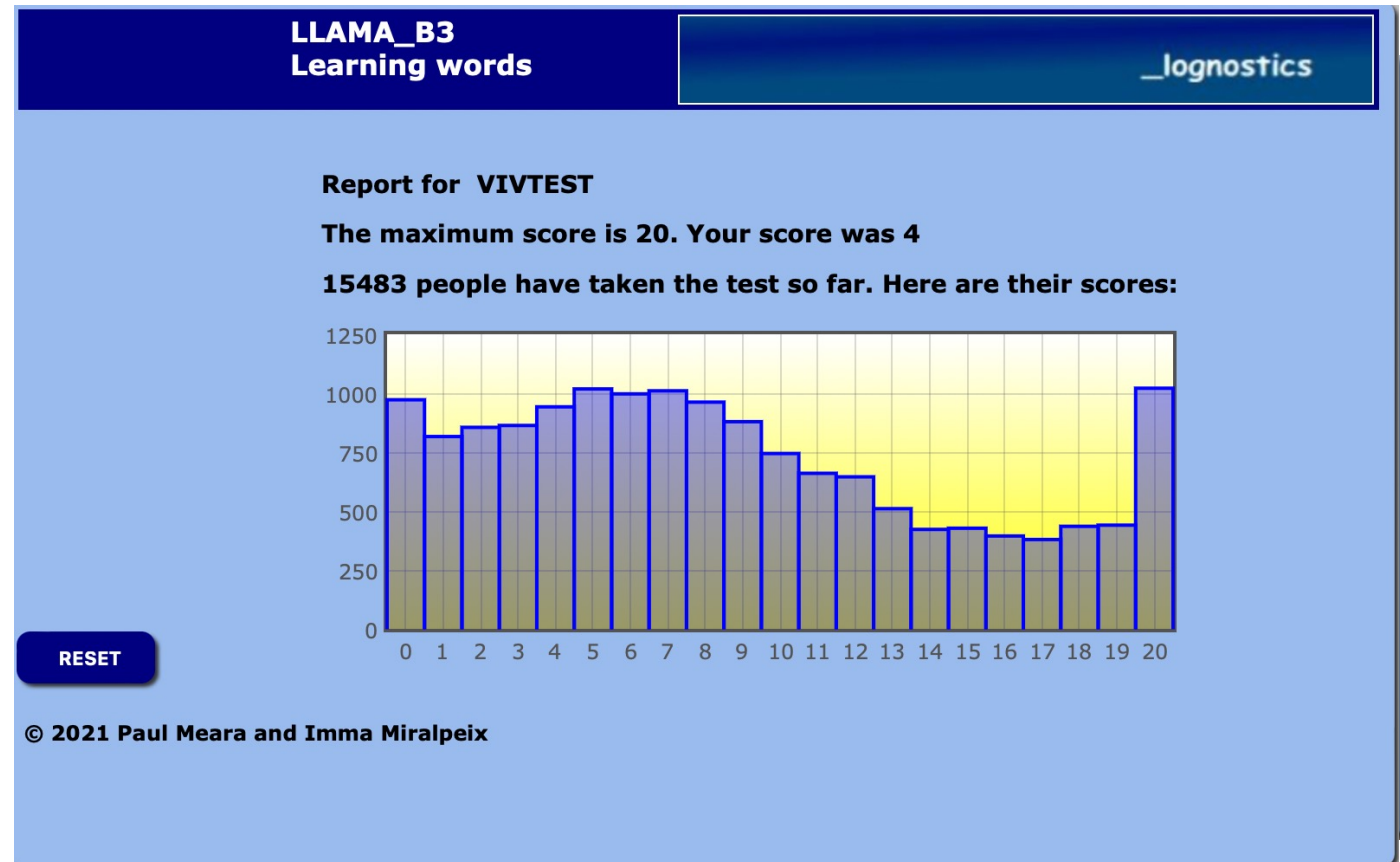
Use the orange buttons to enter your ID Code, then click **BEGIN >>**

© 2021 Paul Meara and Imma Miralpeix
version 3.2.2 2021-03-17



LLAMA B (Learning words) changes

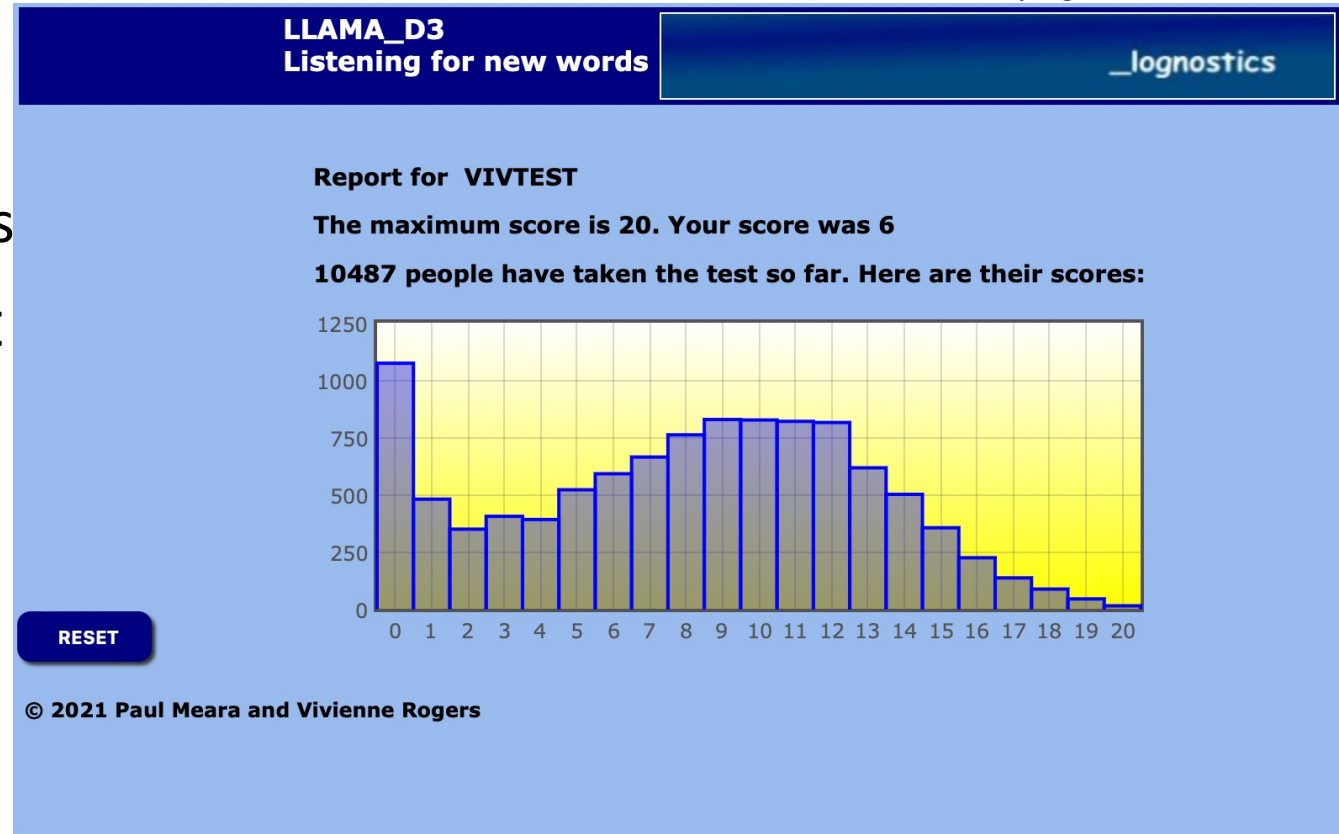
- No major changes
- 2 mins to learn
- Some pictures and words have changed
- Test screen: items moved
- 1 point for each correct answer
- Displayed in chart





LLAMA D (sound recognition) changes

- Two main changes from v.1:
 - Now 40 test items (+10 learning)
 - No separate learning and test phases
- 10 “learning items” presented first
- Each repeated twice in test 40.
 - 20 distractors
- Current website scoring: all 40 items divided by 2.
- Displayed in graph
- ALPACAA scoring only 20 “yes” items.

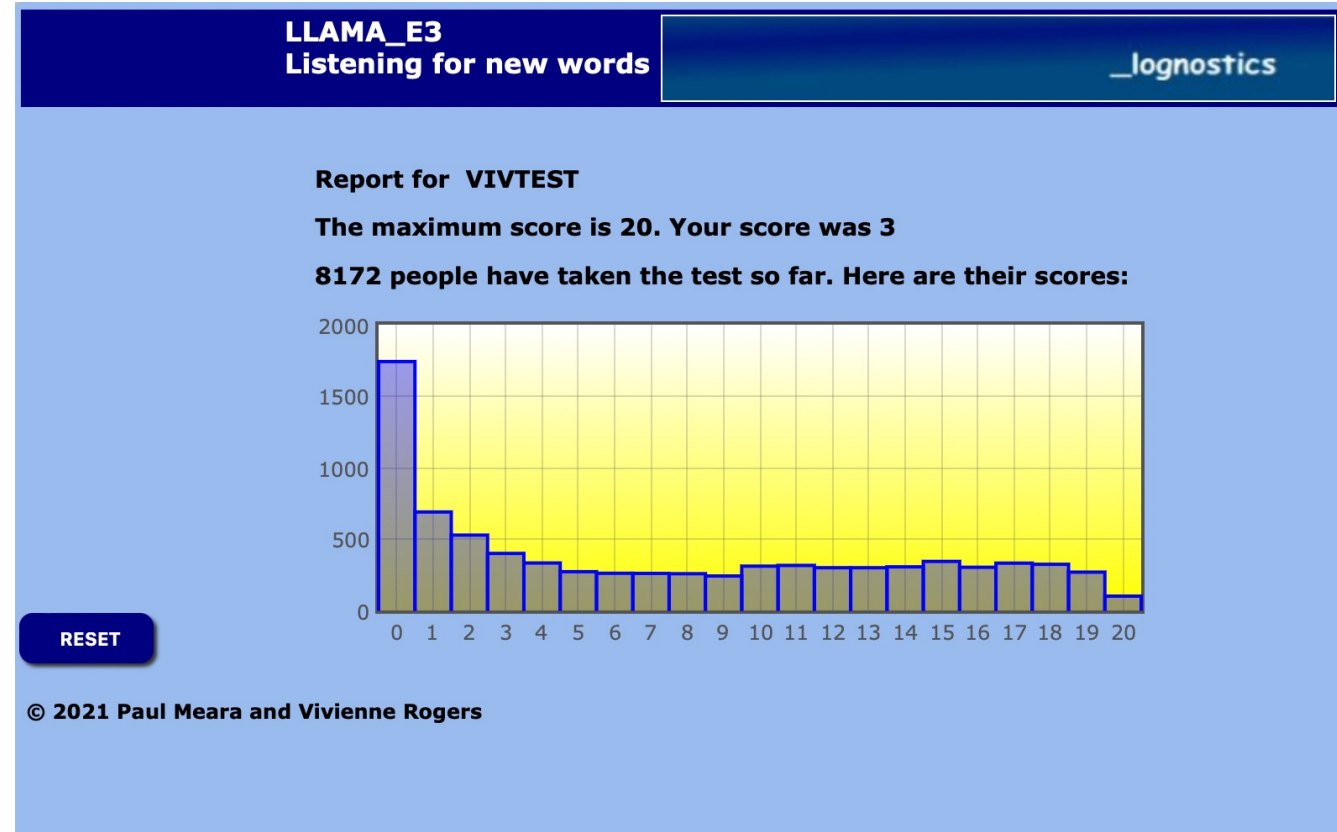


LLAMA E (sounds and symbols) changes



Swansea University
Prifysgol Abertawe

- Learning screen is unchanged
- Test screen changed
 - Used to be a binary choice.
 - Now choice out of 20.
- 20 test items, each scores 1.
- No penalty for guessing

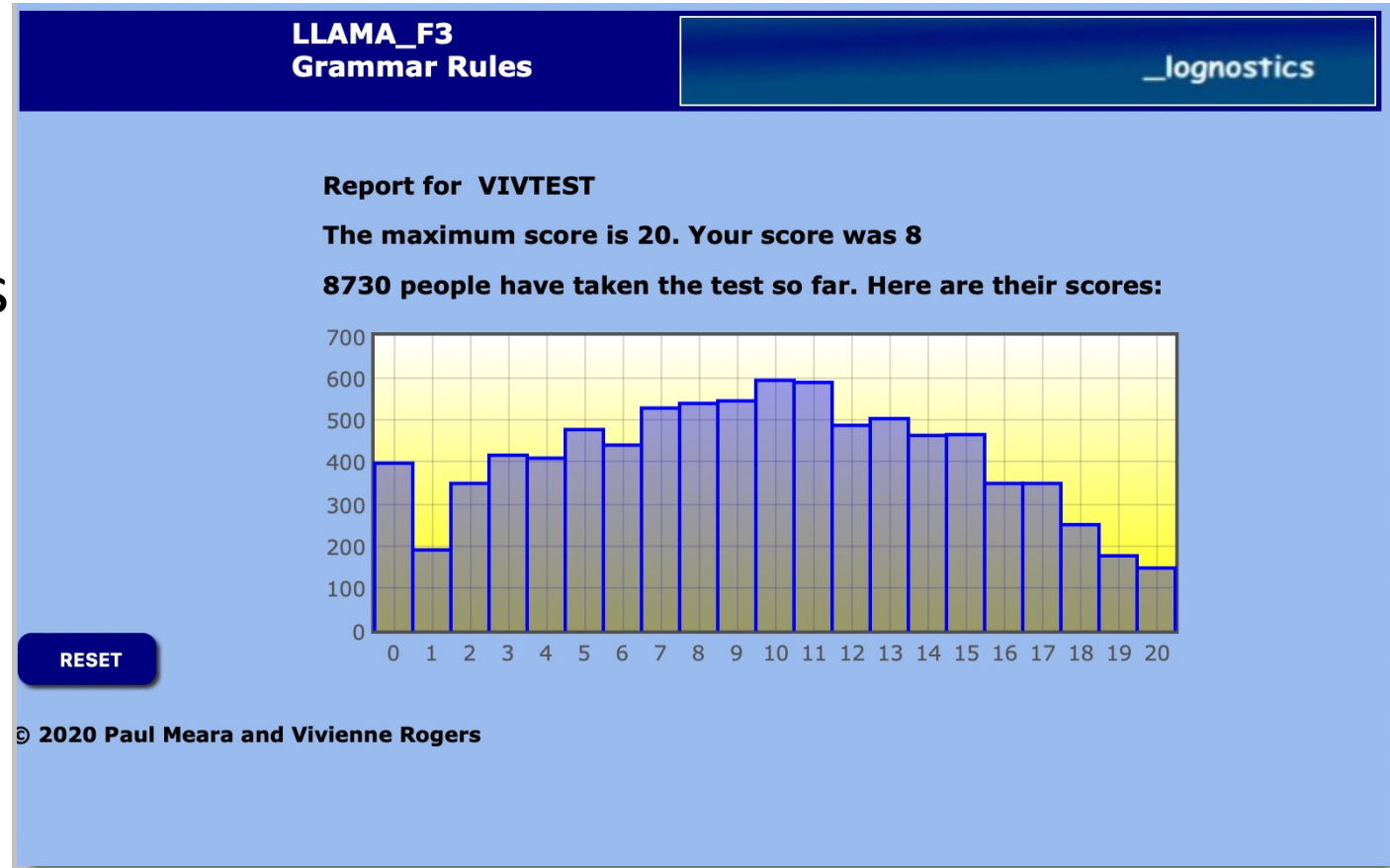


LLAMA F (grammatical inferencing) changes



Swansea University
Prifysgol Abertawe

- Learning screen same
 - 20 images
- 4 mins (1 min shorter)
- Test phase significant changes
- 10 items
- Click on words to make sentences
- Each sentence scored twice.
- Each rule tested 4 times.
- Displayed as a graph



**For more information, please
see:**

Rogers, Vivienne, Meara, Paul & Rogers, Brian (2023) “Testing Language Aptitude: LLAMA evolution and refinement” In Wen, Z.E., Skehan, P. & R. Sparks (eds) Language Aptitude: Theory and Practice. CUP.



Swansea University
Prifysgol Abertawe



Swansea University
Prifysgol Abertawe

Has it made any difference?



Internal reliability across versions

	Version 1 * (n= 350)	Version 2 (n= 123)	Version 3 ** (n=640)
LLAMA B	.81	.850	.897
LLAMA D (scoring 1)	.54	.385 (all items)	.702
LLAMA D (scoring 2)	n/a	.544 (yes only)	.875
LLAMA E	.74	.833	.903
LLAMA F	.60	.617	.864

* Data for version 1 from Bokander & Bylund (2020, table 4)

** Data for version 3 from Rogers, Bokander, Meara & Rogers (in prep)

Revisions in yellow

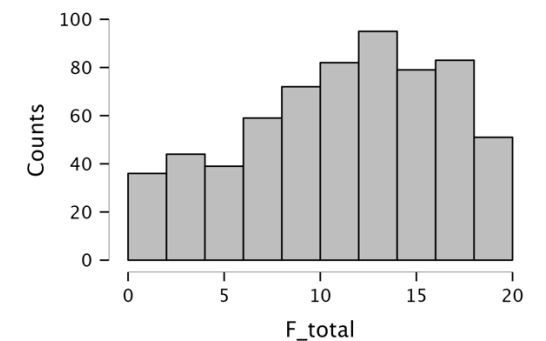
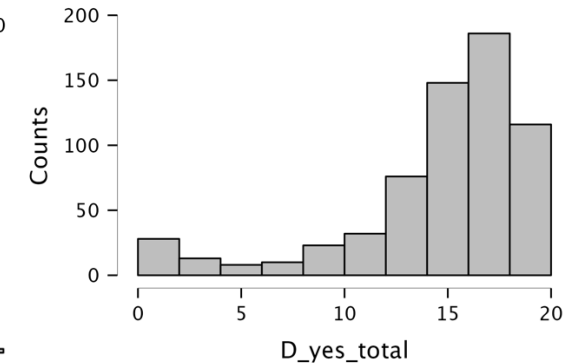
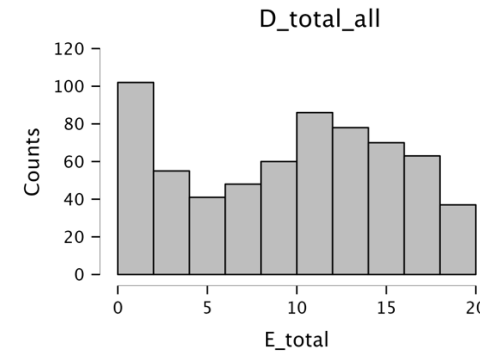
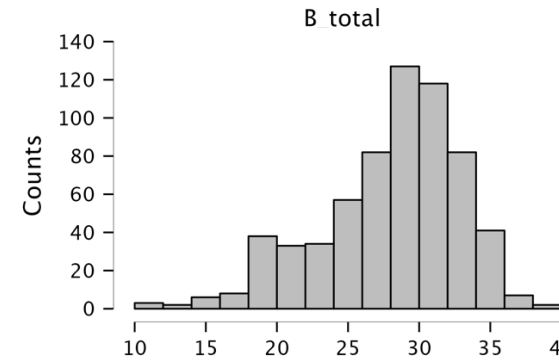
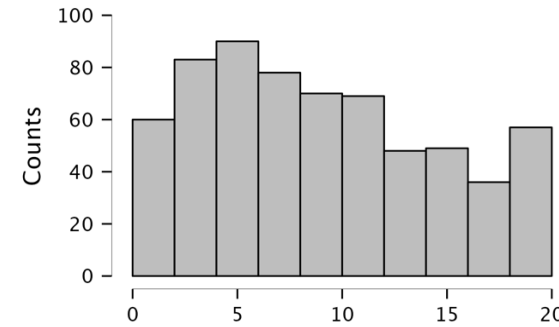
Rogers, Bokander, Meara & Rogers (in prep)



Swansea University
Prifysgol Abertawe

- Files matched across conditions (11th May 2022)
 - All incomplete tests removed
 - Duplicate IDs – first complete test used
 - Blank removed
 - IDs matched
- N= 640

Descriptive Statistics					
	B	D_all	D_yes	E	F
Mean	9.455	28.566	15.048	9.989	11.653
Std. Deviation	5.621	4.896	4.581	5.821	5.167





Conclusion

- LLAMA tests are a work in progress.
 - not extensively validated – caveats still apply.
- However, new versions are an improvement in reliability.
- Work online across platforms.
- Very popular (figures from 29th May 2023):
 - LLAMA B = 15483
 - LLAMA D = 10487
 - LLAMA E = 8172
 - LLAMA F = 8730
- Please use the new tests not the old ones.



New website coming
soon!

www.llamatests.org

Thank you!

Diolch yn fawr!

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

website email: info@llamatests.org

twitter: @RogersVivienne

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

Brian Rogers: b.c.rogers@swansea.ac.uk

Twitter: @dh_bri

New website coming
soon!

www.llamatests.org



Advance notice:
ISBPAC2024 – May 8th-10th 2024 in SWANSEA
CfP coming soon!

