



# Internal validity of the new LLAMA (v.3) aptitude tests

Lars Bokander, Vivienne Rogers, Paul Meara & Brian Rogers

JÖNKÖPING UNIVERSITY

[lars.bokander@ju.se](mailto:lars.bokander@ju.se) / [v.e.rogers@swansea.ac.uk](mailto:v.e.rogers@swansea.ac.uk)



Prifysgol Abertawe  
Swansea University

## Background

- Language learning aptitude has featured intermittently in the spotlight since Henmon's work in the 1930s.
- Meara (2005) developed the original LLAMA tests.
- Subject to a number of criticisms (e.g. Bokander & Bylund, 2020).
- Version 3 released in 2019, addresses some of these validity concerns (Rogers, Meara & Rogers, 2023).

## Research Question

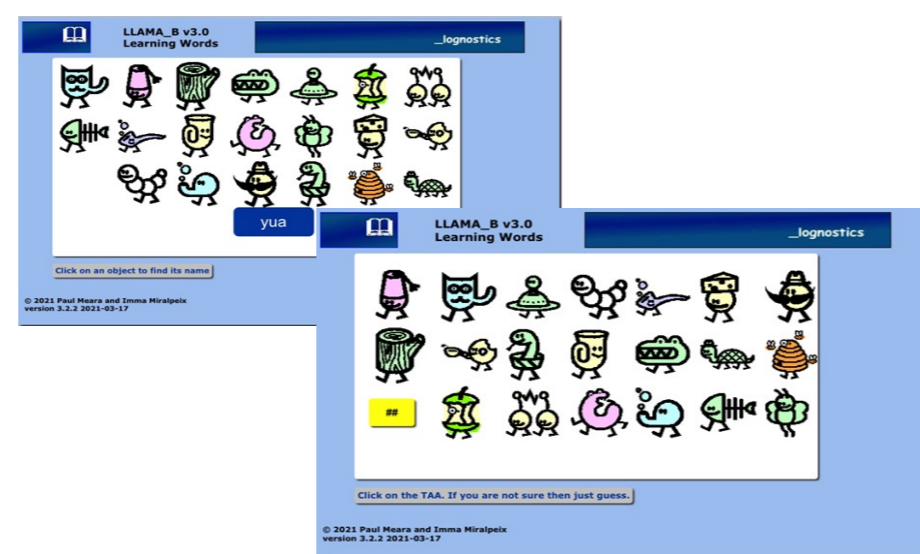
Are the LLAMA v.3 tests (more) reliable than v.1?

## Changes to LLAMA tests (v.1 – v.3)

- Tests are web based rather than Windows download.
- Common ID screen.
- Records individual answers.

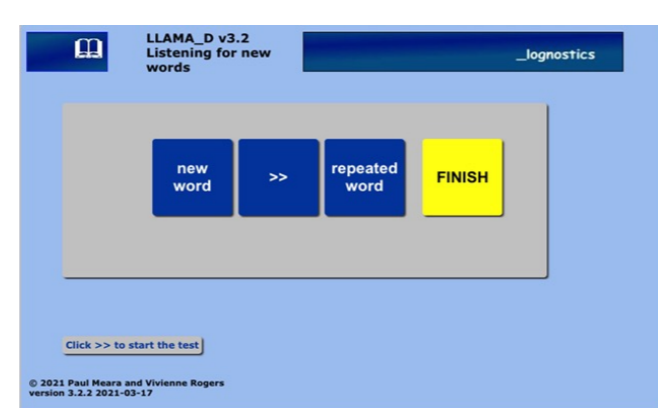
## LLAMA B (vocabulary test)

- Some new pictures and item names.
- No changes to learning or test phase.



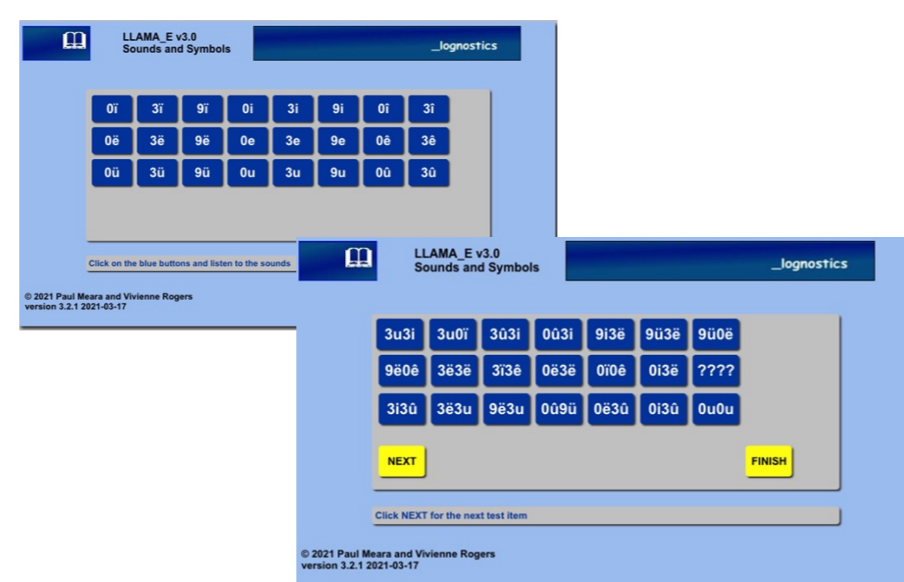
## LLAMA D (sound recognition)

- Learning and test phases combined.
- Initial 10 items = learning.
- 40 test items: 10 learning items repeated twice.
- Change to one difficult sound.



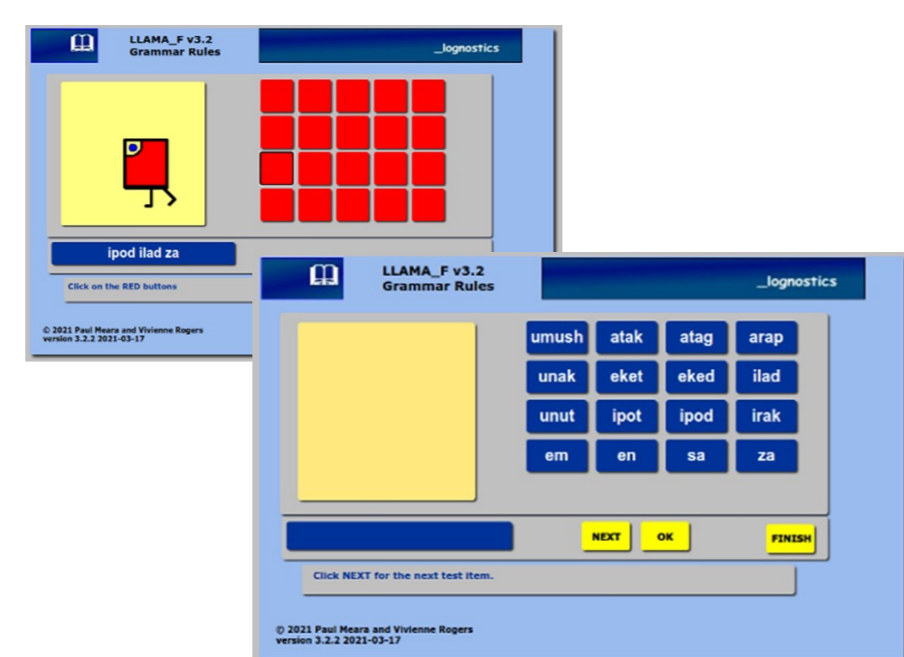
## LLAMA E (sound symbol correspondence)

- No change to learning phase.
  - 20 items with sounds
- Test phase: changed from binary choice to 20 combined sounds.



## LLAMA F (grammatical inferencing)

- No change to learning phase.
- Test phase = not binary choice.
- construct sentence from words.
- 10 items
- Scored 4 times against 5 rules.



## Methodology: Participants

- n= 640
- Data taken from participants matched across all four LLAMA tests on lognostics website from 2021 to 2022.
- [https://www.lognostics.co.uk/tools/LLAMA\\_3/index.htm](https://www.lognostics.co.uk/tools/LLAMA_3/index.htm)
- Inclusion/ Exclusion criteria
  - All incomplete tests removed
  - Duplicate IDs – first complete test used
  - Blank/ 'anon' removed
  - IDs matched

LLAMA test	Downloaded total	Invalid test answers removed	Duplicate IDs removed	Manual check (blank/anon)	IDs matched
B	5813	211	3069	2	640
D	4655	108	2586	2	640
E	2347	418	402	2	640
F	2118	64	707	2	640

## Results:

Table 1. Comparison of internal consistency (Cronbach's  $\alpha$ ) in LLAMA v.1 and v.3.

LLAMA Subtest	Original tests Bokander & Bylund (2020, table 4)	New LLAMA tests v.3
LLAMA B	.81	.90
LLAMA D	.54	.70
LLAMA E	.74	.90
LLAMA F	.60	.88

Table 2. Computed coefficient  $\omega$ , an internal consistency coefficient based on the factor structure of the items in each subtest.

LLAMA v.3 Subtest	$\omega_{total}$	$\omega_{hierarchical}$
LLAMA B	.91	.77
LLAMA D	.88	.69
LLAMA E	.91	.64
LLAMA F	.90	.66

Note:  $\omega_{total}$  is based on loadings on all factors and estimates overall reliability;  $\omega_{hierarchical}$  is based on item loadings on a common factor.

## Item analysis - comparison of LLAMA v.1 and v.3

- The number of items with poor discrimination ( $DI < .10$ ) and large Rasch item misfit ( $infit-t > 2$ ) was lower in the new LLAMA v.3, compared to the original v. 1 (reported in Bokander & Bylund, 2020).
- In LLAMA D, such items decreased from seven in v.1 to five in v.3 (even though the total number of items has increased from 30 to 40 in the new version).
- In LLAMA F, malfunctioning items dropped from three to zero.
- No changes to LLAMA B and E; all items performed well.

## Discussion

- Both the  $\alpha$  and  $\omega$  coefficients of internal consistency reliability suggest improvements in LLAMA v.3 over the original LLAMA v.1.
- The overall reliability as estimated with  $\omega_{total}$  is good in all subtests. The lower  $\alpha$  in LLAMA D (Table 1) might underestimate reliability, possibly due to dimensionality issues (group factors related to "old" and "new" sound stimuli).
  - LLAMA D  $\alpha$  improves to .88 for "old" only items.
- The lower  $\omega_{hierarchical}$  in all tests suggests systematic variance that is not due to a common factor. We are unsure as to what might explain such variance.
- The item analysis shows that there is still room for improvement of LLAMA items, which may in turn strengthen reliability further.

## Conclusion

- New LLAMA tests are more internally consistent.
- New LLAMA website coming soon: [www.llamatests.org](http://www.llamatests.org)

## Limitations:

- No background information on participants.

## References

- Bokander, L., & Bylund, E. (2020). Probing the internal validity of the LLAMA language aptitude tests. *Language learning*, 70(1), 11-47.
- Meara, P. (2005). LLAMA language aptitude tests: The manual. *Swansea: Lognostics*.
- Meara, P. and Rogers, V.E. *The LLAMA Tests v3*. Cardiff: Lognostics. 2019.
- Rogers, V.E., Meara, P., & Rogers, B.C. (2023) "Testing Language Aptitude: LLAMA evolution and refinement" In Wen, Z.E., Skehan, P. & R. Sparks (eds) *Language Aptitude: Theory and Practice*. CUP.