



Swansea University  
Prifysgol Abertawe

# Testing the LLAMA aptitude tests

Vivienne Rogers



# Outline

- Background
- What are the LLAMA tests?
- Research Questions.
- Methodology (general data collection)
- Relevant background
- Results and Discussion
- Overall conclusions
- pilot Latin data?



# Aptitude tests

- MLAT: Modern Language Aptitude test
  - Carroll & Sapon (1959)
  - Four components:
    - the ability to learn words out of context,
    - grammatical sensitivity,
    - phonetic sensitivity
    - inductive learning ability
- PLAB: *Pimsleur Modern Language Aptitude Battery*
  - Pimsleur (1966)
  - vocabulary size in English is taken as a measure of overall verbal ability,
  - language analysis measures
  - sound discrimination measures auditory skills and sound-symbol association
  - a measure of general interest in languages (motivation)
- DLAB: *Defense Language Aptitude Battery*
  - Peterson & Al-Haiq (1976)

# Rationale



- LLAMA = free, loosely based on MLAT.
- Developed by Prof Paul Meara
- [www.lognostics.co.uk/tools/llama/index.htm](http://www.lognostics.co.uk/tools/llama/index.htm)
- Increasingly used in research projects.
  - Google Scholar lists nearly a thousand papers which have cited the LLAMA tests
- Has not been validated.



# What is LLAMA?

Not designed only for English L1.

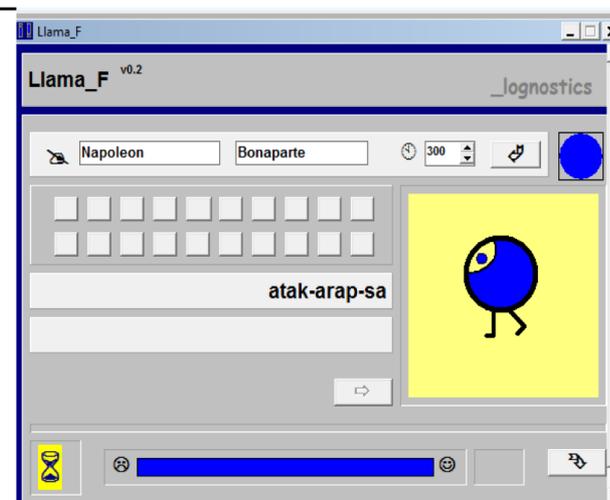
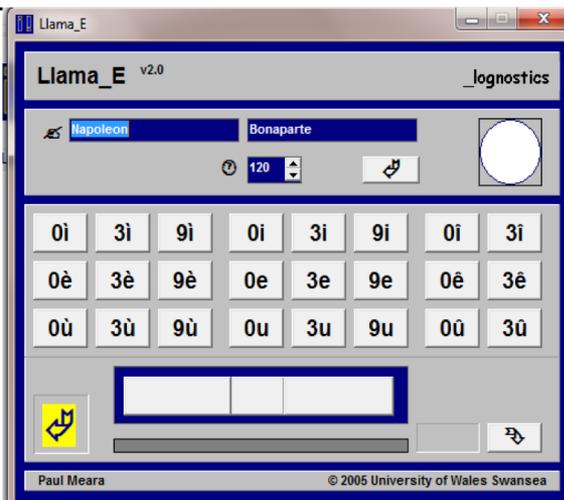
Four components:

- LLAMA B = vocabulary measure
  - MLAT paired associates task
- LLAMA D = sound recognition (implicit learning)
  - Not in MLAT, based on Service's work
- LLAMA E = sound-symbol correspondence
  - MLAT phonetic script subtest
- LLAMA F = grammatical inferencing
  - Explicit inductive learning ability



(not a LAMA)

# LLAMA Subtests



# Previous validation research

- Grañaena (2013):
- Internal consistency, Gender and Language neutrality
- n=187 aged 18-39
- Spanish, Chinese and English
- internal consistency but two forms of aptitude
  - LLAMA D measuring something different to the others

# 2013-14 students

Louise Fallon, Rosa Thomas & Emily Keey



Tom Goss & Rachel Aspinall



# Research Questions 2013-14

1. What is the role of gender?
2. Are the LLAMA tests language neutral?
3. What is the role of age?
4. What is the role of formal education qualifications?
5. Does playing logic puzzles affect LLAMA scores?
6. What difference would changing the test timings make to scores?



# Methodology



- 164 participants at standard length
- 65 participants at altered lengths
- Aged 10-75
- Limesurvey background questionnaire
- Data collected via individual and drop in sessions.

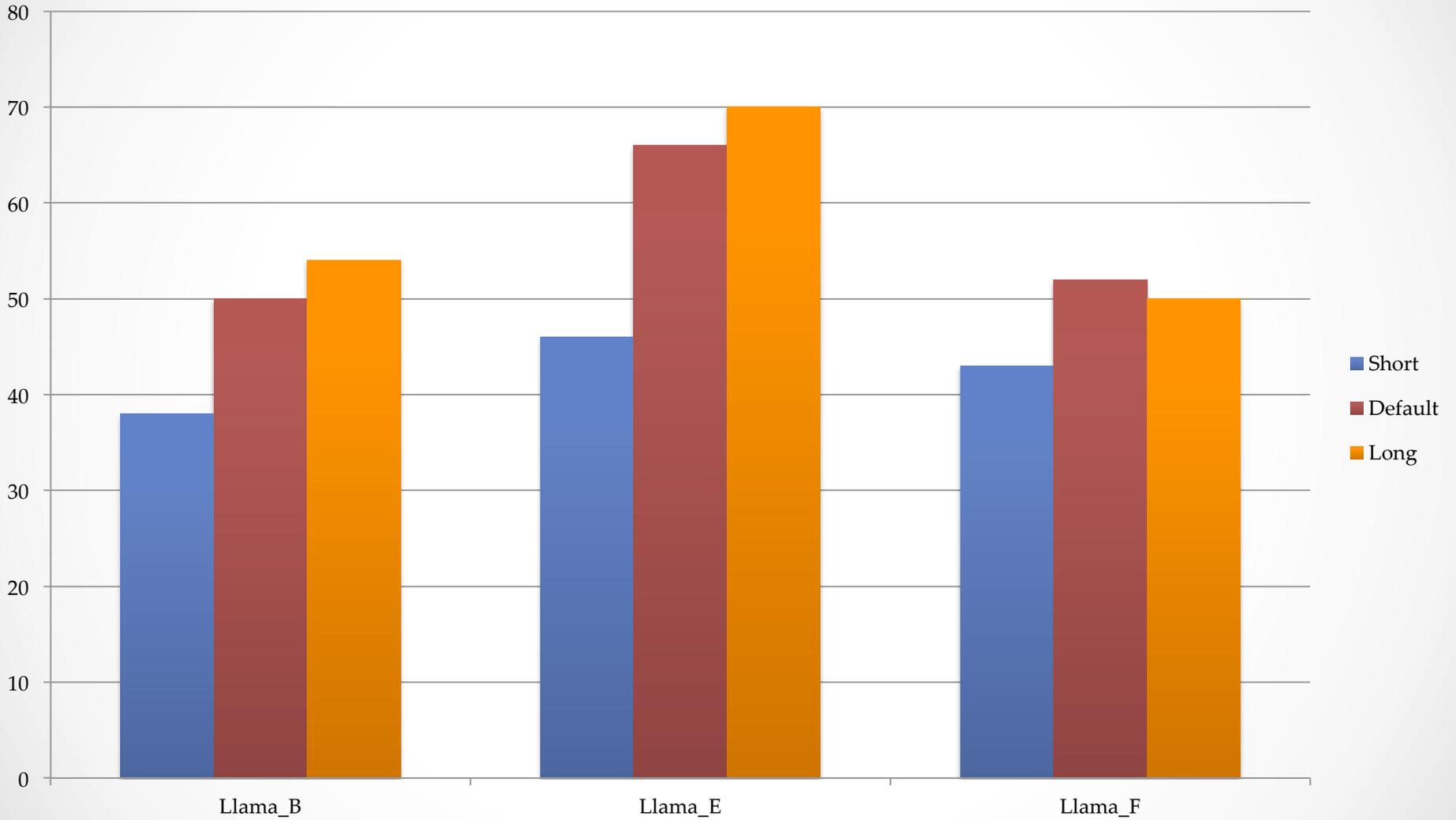


# RQ6: Timings

- Default timings:
  - LLAMA B, D & E = 2 mins
  - LLAMA F = 5 mins
- LLAMA D not included (recording)
- Shorter condition: minus 1 min
- Longer condition: plus 1 min
- Participants (n=98):
  - 32 short timing
  - 33 default timing
  - 33 longer timing



## Group Differences in Average Test Score



# Stats for Timings

- not normally distributed (non-parametric)
- Overall timing effects for:
  - LLAMA B (vocab)  $p=.011$
  - LLAMA E (sound symbol)  $p=.004$
- Within groups
  - Significant difference between default time and shorter time (LLAMA B & E)
  - Significant difference between shorter time and longer time (LLAMA E)
- No effect of timing on LLAMA F
  - even 4 mins may be too long
  - students seem to have finished early



# Co-variates?

- Participants matched gender, age, education and LS status
- Effect for L2 status on changed times with LLAMA B (vocab) & E (sound-symbol)
  - Monolingual scores more affected in B & E.
- Males more affected by changes than females for B & E.



# Overall 2013-4 results

- Results:
  - Comparable results to Grañena (2013)
    - Age  but Language neutrality  ? (LLAMA E)
  - Significant effect of formal education and playing logic puzzles on LLAMA E (sound-symbol)
  - Default timings for B & E appear optimal.
  - LLAMA F timing could be decreased.
- Limitations
  - Over-dominance of UG, monolingual participants.
  - Some of the groups were small, e.g. age effects, language neutrality.



# 2014-15 Students



Clare Curry



Tom Barnett-Legh



Emma Davie

# Research Questions 2014-15

Follow-up to previous study:

1. Are the LLAMA tests language neutral?
  - a. i.e. Does your L1 have an influence on your final scores?
2. What effect does L2/bilingual status have on LLAMA scores?
3. Does age affect aptitude as measured by LLAMA?



# Methodology

- Most of the data collected by final year BA students for their dissertations.
- Data also from international students on our pre-session course and by Khaled Alamri (PhD student).
- Data collected individually or in large computer sessions.
- Background questionnaire.
- Total number of participants = 240.

# RQ3: Previous research

- Several different views on age and aptitude:
- Abrahamsson & Hyltenstam (2008) argue that aptitude is only a relevant factor for learners over the age of 15.
  - Grañena and Long (2013a) show age-effects first influence L2 phonology, then lexis, collocation and morphosyntax.
- Muñóz (2014) investigated 48 bilingual Spanish-Catalan Primary school learners of English aged 10-11 and 11-12.
  - significant correlations with all components.
  - Thus, providing support for the notion of language aptitude in younger learners.

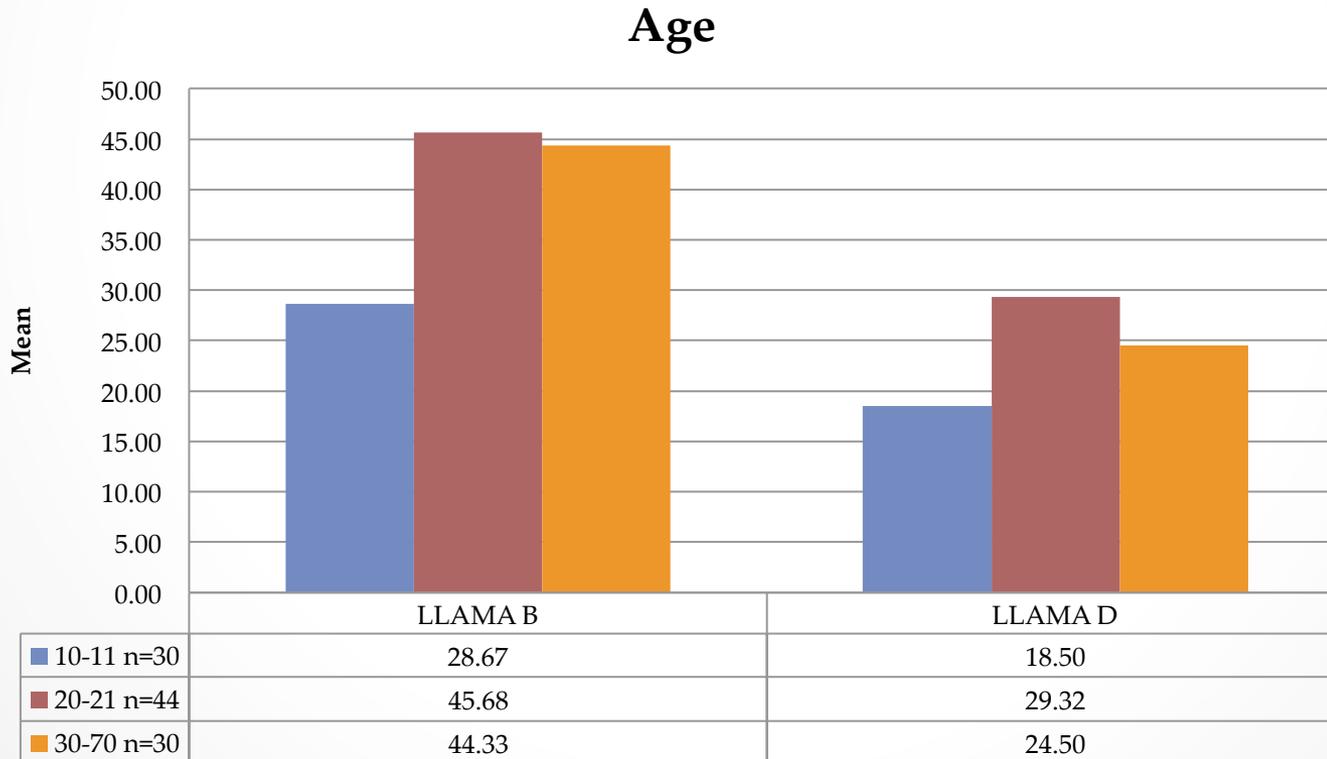
# RQ3: Background

Does age affect aptitude as measured by LLAMA?

- 2014 study on LLAMA B and LLAMA E found no significant differences but a different profile of results. This time looking at vocabulary and implicit learning (LLAMA D).
- LLAMA tests not originally designed for use with children (Meara p.c.)
- Separate MLAT for students aged 8-12
- Hypothesis 1: no difference on LLAMA B vocabulary scores (vocabulary learning is life-long).
- Hypothesis 2: younger participants will outperform older participants (implicit learning)

# RQ3: results

- Subset of 104 participants (matched, age and gender across age groups)



# RQ3: Results

- LLAMA B (vocabulary)
  - 10-11 year olds performed significantly worse than both older groups ( $p < .05$ )
  - No significant differences between 20-21s and 30-70s.
  - Hypothesis 1: Disconfirmed. Younger participants performed worse.
- LLAMA D (implicit)
  - 10-11 years olds performed significantly worse than 20-21s ( $p < .05$ ) but not than 30-70s.
  - No significant difference between older groups.
  - Hypothesis 2: disconfirmed. Younger group did not perform better than either of the two older groups.
- However, 10-11 year olds were able to do the tests. No conceptual or interface problems.
- But may need different norms?

# Bringing it all together

- Participants: n=404
  - (not including the different timings participants)
- Research questions:
  1. Are the LLAMA tests language neutral?
  2. What is the effect of monolingualism on LLAMA scores?
  3. How much of LLAMA test score variance do the individual factors measured account for?

# RQ1: Previous research

- Several studies suggest the degree of distance between an L1 and an L2 plays a fundamental role in word processing and retention in an L2
  - (Gholamain & Gera, 1999; Hamada & Koda, 2008; Green & Meara, 1987; Wong and Pyun , 2012)
- MLAT = designed for use with native English speakers.
  - used with a wide range of languages.
- If the language script of the L1 can influence the acquisition of the L2, then the question arises if the L1 script of the learner influences their aptitude scores.



# RQ1: Background

Does your L1 have an influence on your final scores?

- LLAMA B and LLAMA F have roman alphabet letters as part of the test.
- Chinese: morphosyllabic (Tolchinsky et al, 2011: 1598) or logographic (Baron, 2000: 2)
  - 您好
- Arabic: consonant alphabetic script (common ancestor with Roman scripts = North Semitic)
  - مرحبا



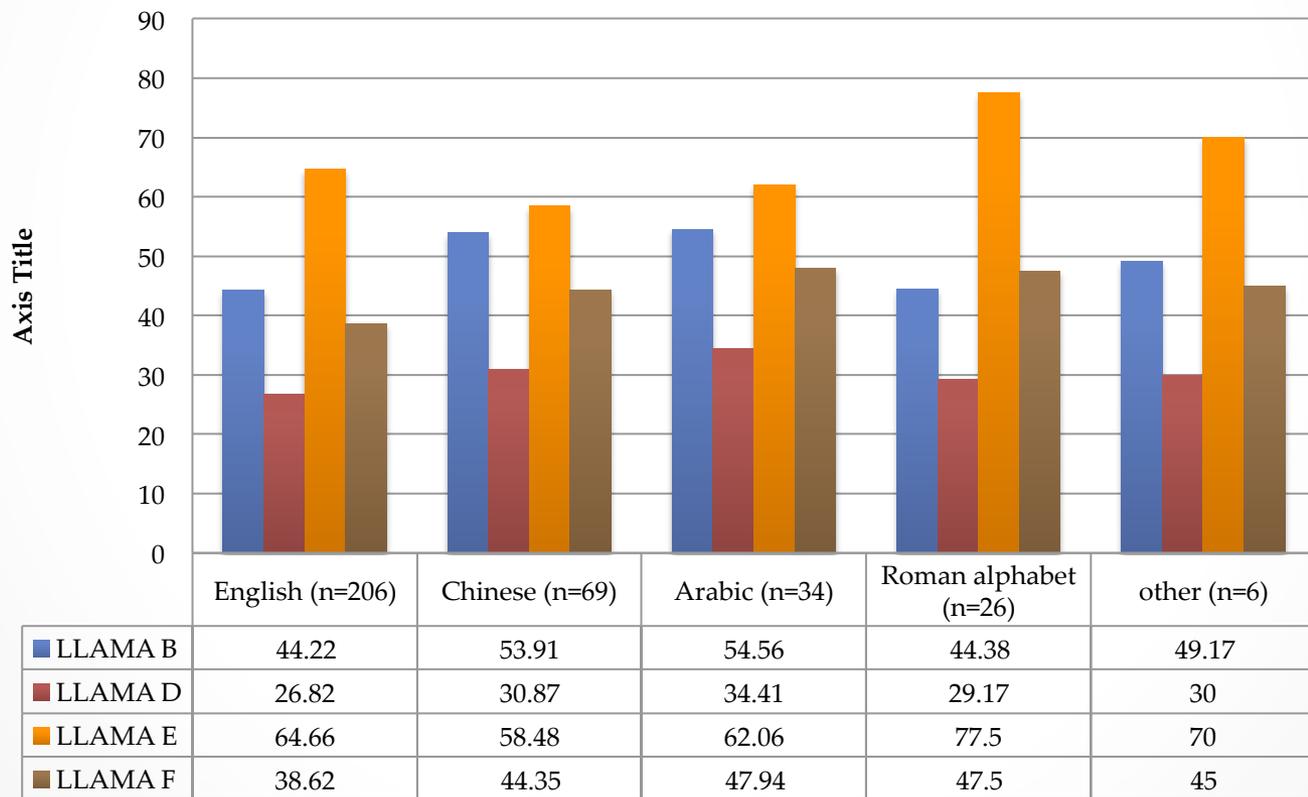
# RQ1: Hypotheses

1. English native speakers will outperform Chinese and Arabic native speakers on LLAMA B & F as the script will not require such a strong processing load for them.
  2. Arabic speakers will outperform Chinese speakers as it is an alphabetic script with a common ancestor to the Roman alphabet.
- Compare English (n=206), Chinese (n=69), Arabic (n=34), other Roman alphabet (n=24), other (n=6)
  - Total n=339



# RQ1: Language Neutrality

Chart Title



# RQ1: Language Neutrality

- No significant differences for any test.
- Possibly due to large differences in group size and large standard deviations.



# English speaker performance?

- English native speakers are outperformed in:
  - LLAMA B (vocabulary)
  - LLAMA D (incidental)
  - LLAMA F (grammatical inferencing)
- Is this because some of the English speakers were monolingual?

➔ RQ2



# RQ2: Previous research

- Training effect on aptitude
  - (Grigorenko et al, 2000; McLaughlin, 1990; Sternberg, 2002)
- Aptitude development significantly correlates to language experience
  - May change over time
  - (Eisenstein, 1980; Kormos, 2013; Sáfár & Kormos, 2008; Sawyer, 1992; Sparks, Ganschow, Fluharty & Little, 1995; Thompson, 2013).
- Multilinguals more able to adjust their L2 learning strategy to facilitate specific language components
  - but not more successful overall.
  - Nayak, Hansen, Krueger and McLaughlin (1990)



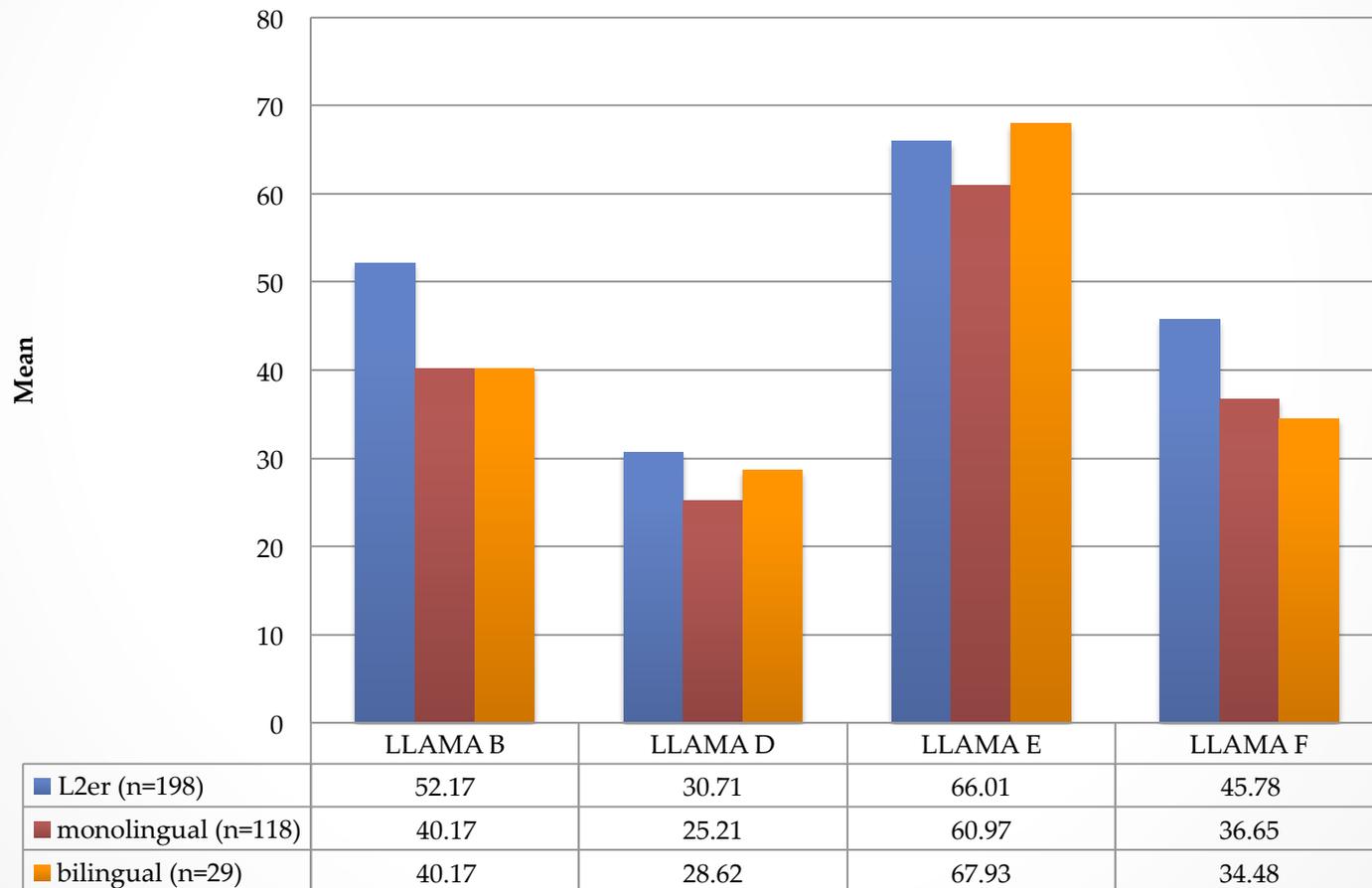
# RQ2: Background

What effect does L2/bilingual status have on LLAMA scores?

- Compare monolingual, L2ers and bilinguals
  - self identified as bilingual and began learning both languages before 5
- Hypothesis 1: L2 learners will outperform the other groups as they have developed conscious strategies
- Hypothesis 2: Bilinguals will outperform monolinguals as they are more aware of language



# RQ2: L2 status



# RQ2: L2 status

- LLAMA B (vocabulary)
  - L2ers significantly outperformed monolinguals and bilinguals ( $p < .05$ )
  - No difference between mono- and bilinguals.
- LLAMA D (implicit learning)
  - L2ers significantly outperform monolinguals ( $p < .05$ )
    - but not the bilinguals ( $p = .065$ )
- LLAMA E (sound-symbol)
  - No significant differences between any groups.
- LLAMA F (grammatical inferencing)
  - L2ers significantly outperformed the monolinguals ( $p < .05$ )
    - but not the bilinguals ( $p = .467$ ).
  - No difference between the mono- and bilinguals.
- Hypothesis 1 confirmed for LLAMA B and LLAMA F.
  - Not surprising as vocabulary and grammar learning form part of L2 curriculum.
- Hypothesis 2: not confirmed
  - Bilinguals outperformed monolinguals in all tests (except B) but not significant.



# RQ3: Background

How much of the variance in the scores do the individual differences identified account for?

Examined: Gender, L1 (language neutrality), L2 status, education level, logic puzzles, age

- Information collected through background questionnaire (both years).



# RQ3: LLAMA B

- Multiple regression,  $n=404$
- Overall factors:  $R^2 = 9.1\%$  of overall variance
  - Adjusted  $R^2 = 7.6\%$
- Individual independent variables:
  - Only L2 status reaches significance.
  - Beta value =  $-.250$ ,  $p < .05$
  - Contribution to overall variance =  $6.0\%$



# RQ3: LLAMA D

- Multiple regression,  $n=375$
- Overall factors:  $R^2 = 4.8\%$  of overall variance
  - Adjusted  $R^2 = 3.1\%$
- Individual independent variables:
  - L2 status and gender reach significance. Language neutrality ( $p=.055$ )
  - L2 status :
    - Beta value = .136,  $p = .012$
    - Contribution to overall variance = 1.8%
  - Gender
    - Beta value = .116,  $p = .030$
    - Contribution to overall variance = 1.3%



# RQ3: LLAMA E

- Multiple regression,  $n=370$
- Overall factors:  $R^2 = 3.4\%$  of overall variance
  - Adjusted  $R^2 = 1.8\%$
- Individual independent variables:
  - Playing logic games reaches significance.
    - Highest qualification:  $p=.056$
  - Logic games:
    - Beta value =  $.152$ ,  $p = .004$
    - Contribution to overall variance =  $2.3\%$



# RQ3: LLAMA F

- Multiple regression,  $n=346$
- Overall factors:  $R^2 = 6.6\%$  of overall variance
  - Adjusted  $R^2 = 4.9\%$
- Individual independent variables:
  - L2 status and language neutrality reach significance.
  - L2 status:
    - Beta value =  $-.165$ ,  $p = .002$
    - Contribution to overall variance =  $2.6\%$
  - Language neutrality
    - Beta value =  $.114$ ,  $p = .036$
    - Contribution to overall variance =  $1.3\%$



# RQ3: implications

- The factors examined so far do not account for much of the variance between scores either together or individually.
- Learning a L2 seems to be advantageous for the tests.
  - Need to be aware if using for projects.
- Need to consider IQ and WM.
  - Previous research (Wesche, 1981) has found overlap between MLAT and IQ.



# Next steps



1. LLAMA B is now online but others are in development.
  - a. LLAMA E is negatively skewed so presentation will be tweaked.
2. Examine WM and IQ scores.
  - o WM measure attempted in 2014-15 with 15 participants but incorrectly administered.
3. Pilot data collected to examine if LLAMA scores predict outcomes in intensive 2 week Latin class (6 participants).
  - o Includes motivation (LLOS) and anxiety (FLCAS) questionnaires
4. Extension to longer class (1 term/1 academic year, n=40+)





Thank you!

Any questions?



# 2013-14 results

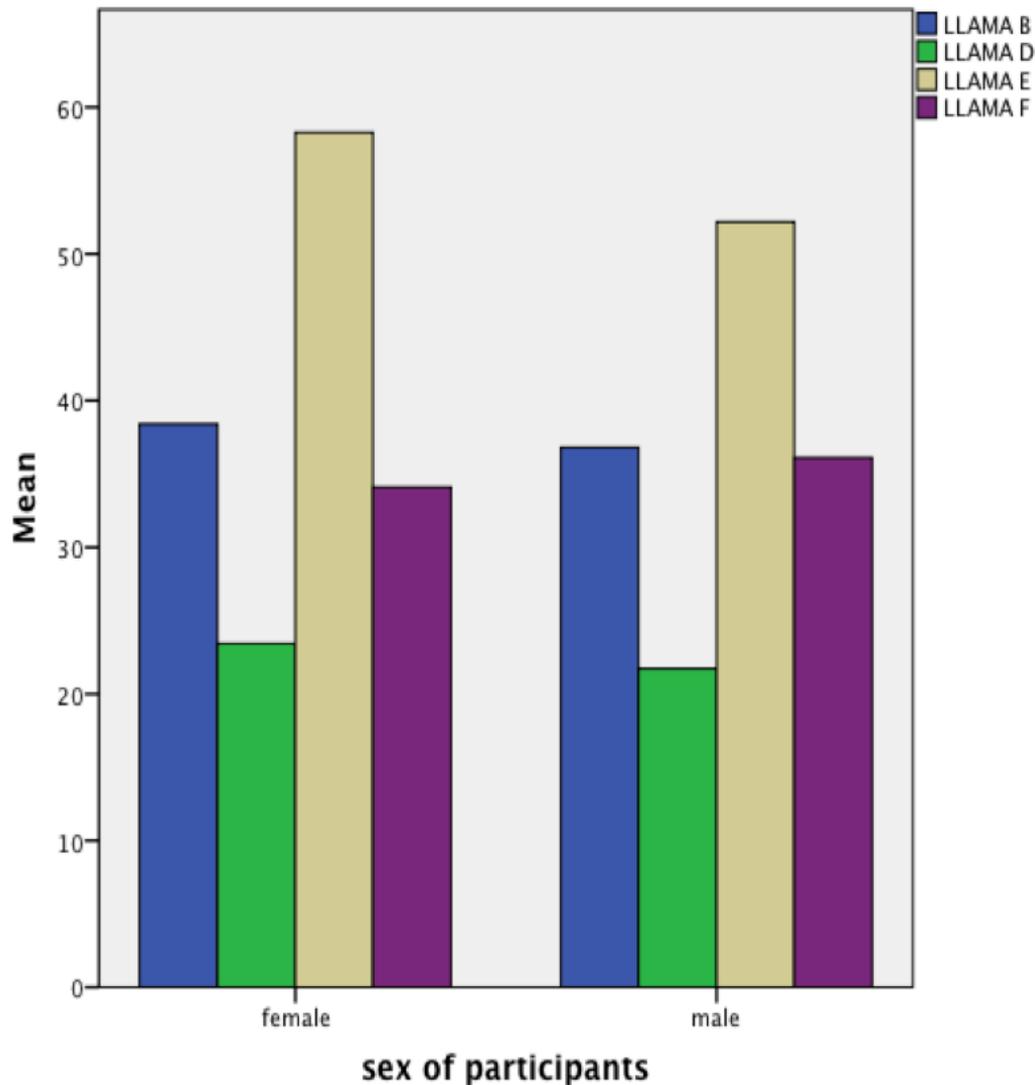
...

# Results: Gender

- Only students aged over 18 who had taken all the tests were included (n=135)

	LLAMA B	LLAMA D	LLAMA E	LLAMA F
Male (n=63)	36.80 (24.708)	21.71 (18.217)	52.18 (35.756)	36.11 (29.391)
Female (n=72)	38.40 (25.859)	23.44 (19.182)	58.27 (34.438)	34.10 (28.285)





## Gender differences chart

- No significant effect of gender on any component.
- LLAMA B (vocabulary)  $t(133)=.367$   $p=.729$
- LLAMA D (sound rec)  $t(133)=.536$   $p=.904$
- LLAMA E (sound/sym)  $t(133)=1.005$   $p=.488$
- LLAMA F (grammar)  $t(133)=-.404$   $p=.456$
- Same result at Grañena (2013)

# RQ2: Language neutrality

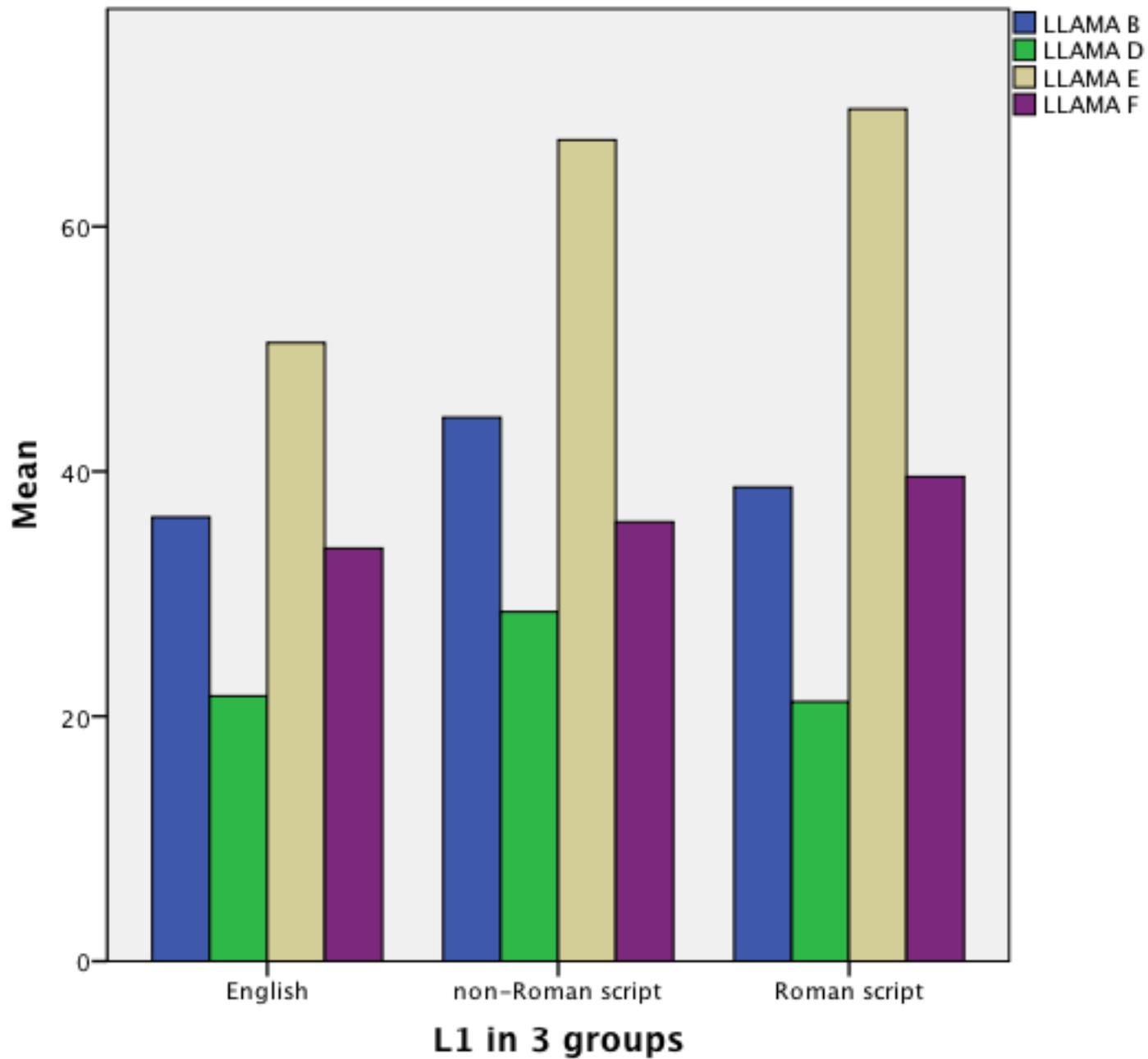
- 135 participants over 18, took all tests
- 18 different L1s
- Coded into 3 groups:
  - English (n=99)
  - Non Roman script (e.g. Chinese, Arabic) (n=17)
  - Roman script (e.g. French, German) (n=18)



# Language neutrality results

	LLAMA B	LLAMA D	LLAMA E	LLAMA F
English (n=99)	38.28 (25.013)	21.66 (18.133)	50.51 (35.623)	33.71 (28.962)
Non-Roman script (n=17)	44.71 (25.488)	28.53 (20.673)	67.06 (21.727)	35.88 (26.706)
Roman script (n=18)	38.70 (27.347)	21.19 (19.577)	69.61 (37.017)	39.58 (29.802)

- Significant effect of L1 type on LLAMA E (sound-symbol)  $F(2,131)=3.505$   $p=.033$
- Non-roman script significantly outperformed English group ( $p=.036$ )
- L1 interference?



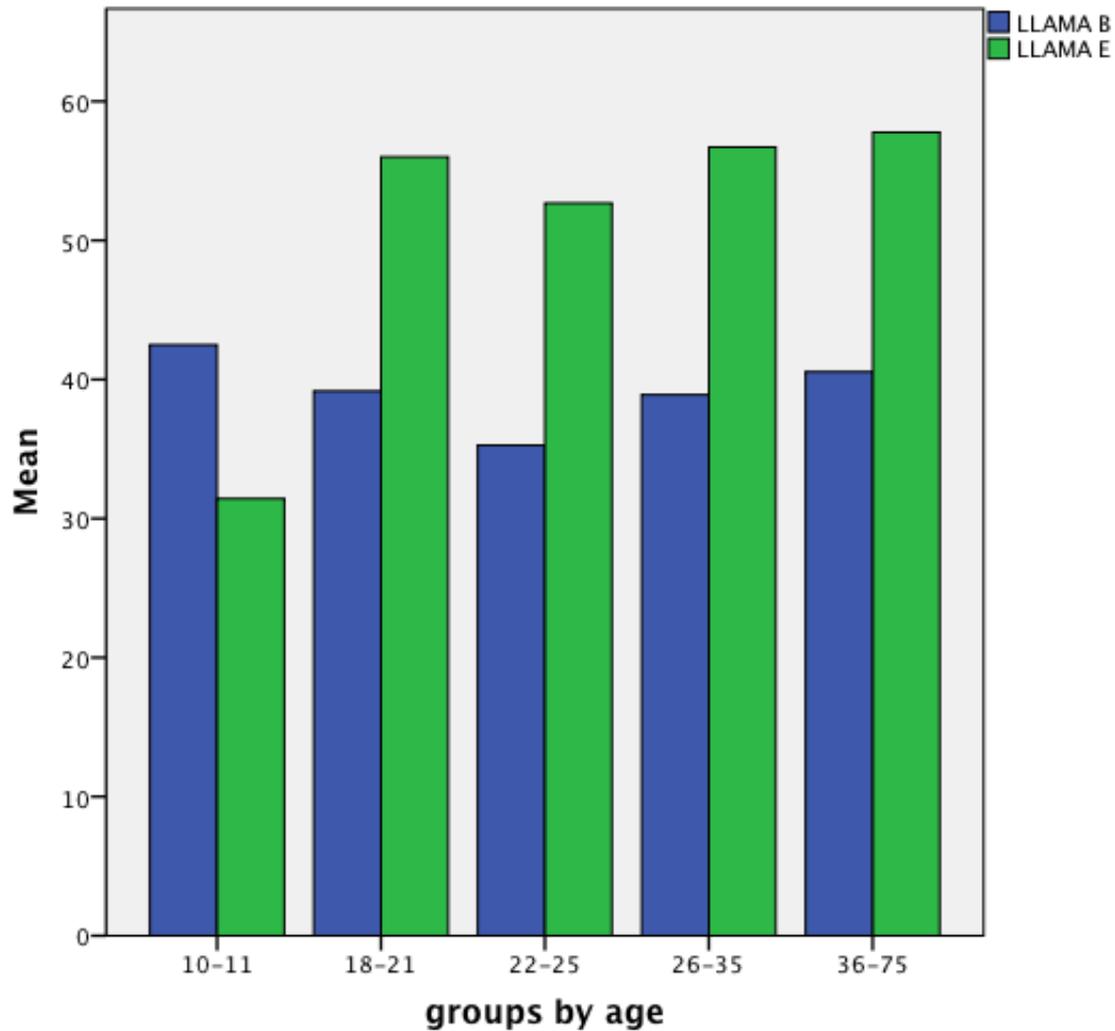
# RQ3: Age effects

- Total participants = 157
- LLAMA B (vocabulary) and E (sound/symbol) tested.
- NB: LLAMA not designed for use with children.

	10-11 (n=14)	18.21 (n=66)	22-25 (n=32)	26-35 (n=18)	36-75 (n=27)
LLAMA B	42.50 (17.623)	39.16 (26.685)	35.27 (28.158)	38.91 (23.769)	40.56 (30.551)
LLAMA E	31.43 (19.158)	56.01 (35.443)	52.67 (36.556)	56.72 (35.562)	57.78 (30.551)



# Results by age chart





# Stats for Age

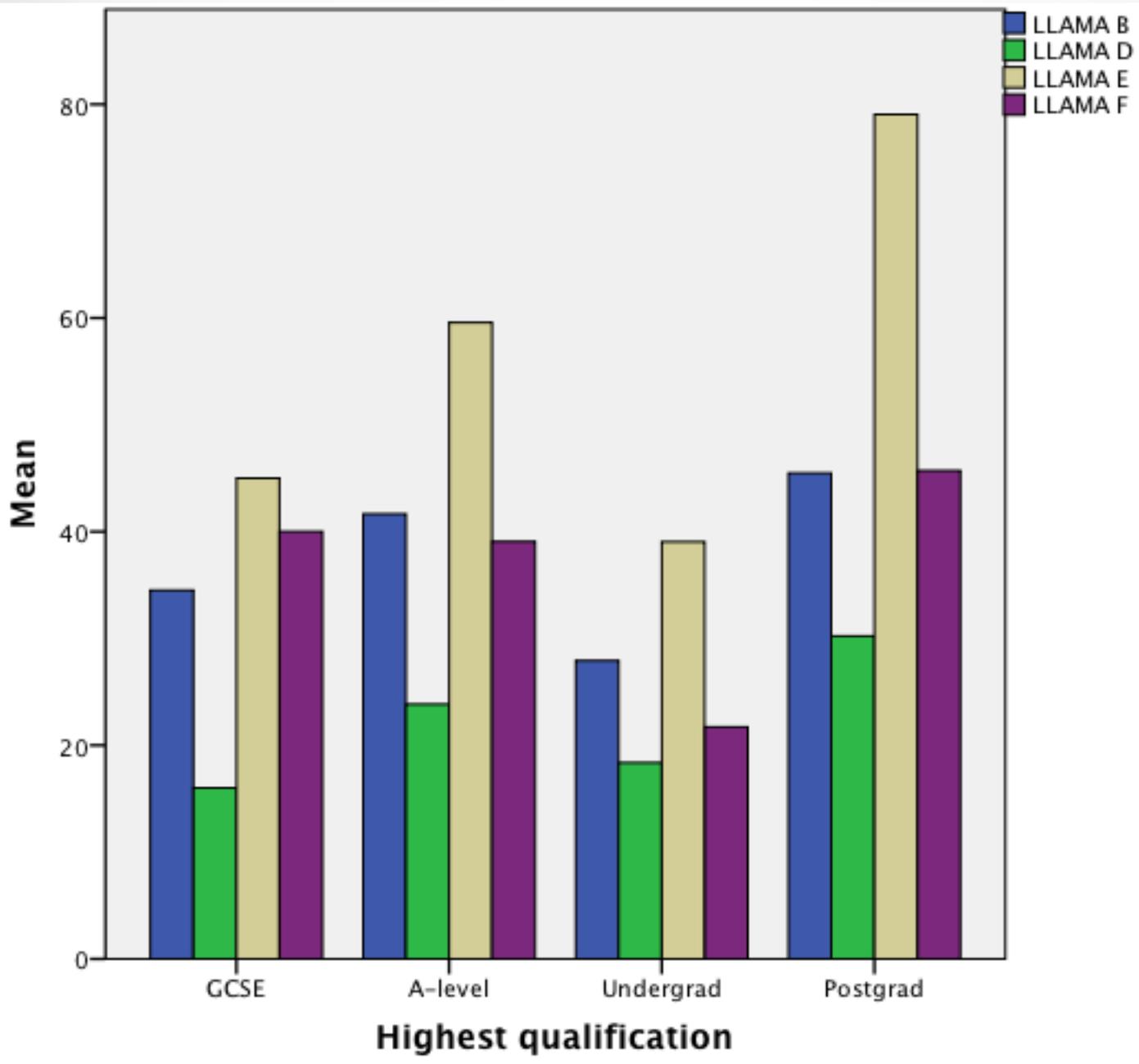
- No overall significant effects for age with vocabulary (B) or sound-symbol (E).
- However, younger groups profile differently.
- Post-hoc Games-Howell shows significant differences between 10-11s and 18-21s ( $p=.007$ ) and 36-75s ( $p=.014$ ).



# RQ4: Formal education

- 135 learners divided into 4 groups
- Asked for their highest qualification

	LLAMA B	LLAMA D	LLAMA E	LLAMA F
Aged 16 (n=10)	34.50 (18.174)	16.00 (13.499)	45.00 (38.658)	40.00 (18.856)
Aged 18 (n=64)	41.66 (25.120)	23.84 (18.345)	59.56 (31.237)	39.08 (29.934)
UG degree (n=40)	27.93 (25.819)	18.37 (19.682)	39.03 (38.220)	21.71 (25.892)
MA/ PhD (n=21)	45.48 (37.66)	30.24 (17.852)	79.05 (20.225)	45.71 (26.376)



# Formal qualifications results

- One way ANOVA
- Significant effect for highest formal qualification for:
  - Vocabulary (B)  $F(3,131)=3.413$   $p=.019$
  - Sound/symbol (E)  $F(3,131)=7.684$   $p=.000$
  - Grammar inferencing (F)  $F(3,131)=4.724$   $p=.004$
- Sound recognition (D) approached significance
  - $F(3,131)=2.439$   $p=.067$

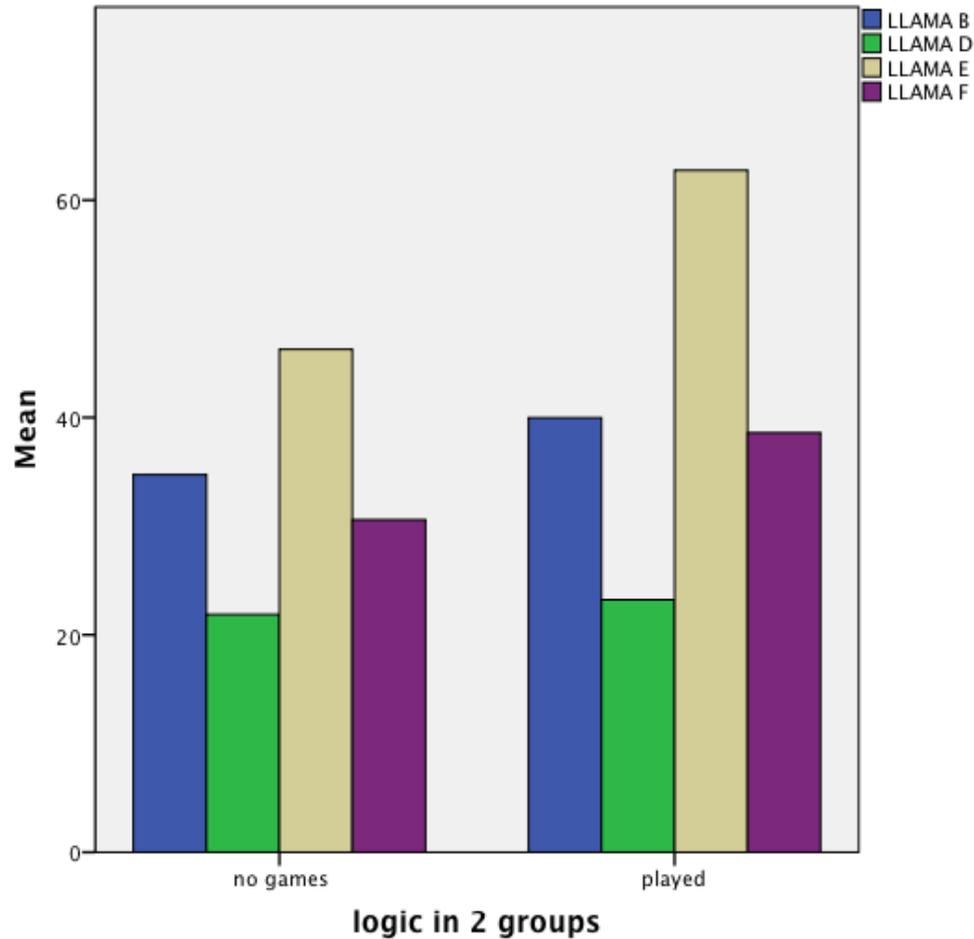


# RQ5: Logic games

- 135 participants over 18, all tests
- Divided into two groups

	LLAMA B	LLAMA D	LLAMA E	LLAMA F
No games (n=60)	35.75 (23.780)	21.88 (18.494)	46.28 (32.939)	30.58 (21.177)
Played (n=75)	39.98 (26.287)	23.24 (18.944)	62.75 (35.194)	38.60 (29.587)





- Significant effect for playing logic puzzles with LLAMA E (sound-symbol)  $t(133)=-2.781$   $p=.006$ .

# 2014-15 results

...

# RQ1: Language Neutrality

		LLAMA B	LLAMA D	LLAMA E	LLAMA F
English n=107	Mean	45.28	27.94	68.32	36.40
	s.d	(21.608)	(16.653)	(29.065)	(24.618)
Chinese n=56	Mean	55.89	31.16	56.34	46.96
	s.d	(27.288)	(24.458)	(28.034)	(25.984)
Arabic n=32	Mean	53.75	34.38	62.19	49.06
	s.d	(24.163)	(15.748)	(25.207)	(24.933)

# RQ2: L2 status

		LLAMA B	LLAMA D	LLAMA E	LLAMA F
L2er (n=142)	Mean	53.24	30.85	63.31	45.25
	s.d.	24.234	19.902	28.434	27.310
monolingual n=46	Mean	39.57	25.65	65.11	31.20
	s.d.	20.759	17.720	28.800	20.033
bilingual n=23	Mean	42.39	32.83	66.52	38.26
	s.d.	22.303	14.834	30.243	25.876

# RQ3: Age

		LLAMA B	LLAMA D
10-11 n=30	Mean	28.67	18.50
	s.d.	14.910	13.528
20-21 n=44	Mean	45.68	29.32
	s.d.	21.529	17.206
30-70 n=30	Mean	44.33	24.50
	s.d.	24.380	17.536