



Swansea University
Prifysgol Abertawe

Validating the LLAMA aptitude tests

Vivienne Rogers, Tom Barnett-Legh, Clare Curry
& Emma Davie with Paul Meara

Outline

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



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
- Increasingly used in research projects.
- Has not been validated.
- Grañena (2013): internal consistency but two forms of aptitude
 - Gender and Language neutrality



What is LLAMA?

Not designed only for English L1.

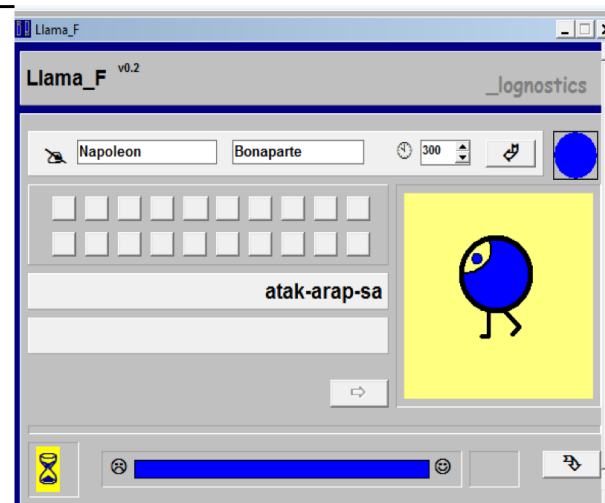
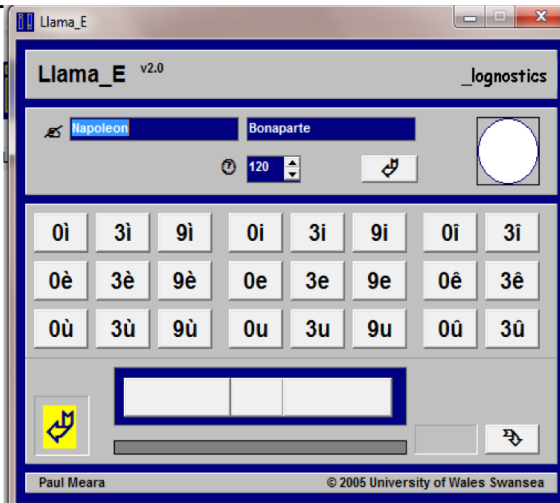
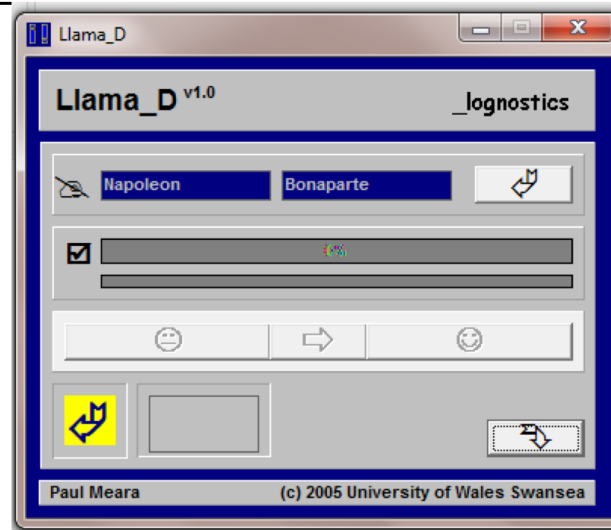
Four components:

- 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



Our 2013-4 Study

- Examined:
 - gender, age, formal education, playing logic puzzles, language neutrality and differences in test timings.
- Methodology:
 - 164 participants at standard length
 - 65 participants at altered lengths
 - Aged 10-75
- 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.



Research Questions

Follow-up to Rogers et al (2014) 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?
4. How much of the variance in the scores do the individual differences identified account for?
 - a. Gender, L1, L2 status, education level, logic puzzles, age



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.



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.
- Compare English (n=102), Arabic (n=32) and Chinese (n=57) speakers.
- Chinese: morphosyllabic (Tolchinsky et al, 2011: 1598) or logographic (Baron, 2000: 2)
 - 您好
- Arabic: consonant alphabetic script (common ancestor with Roman scripts = North Semintic)
 - مرحبا



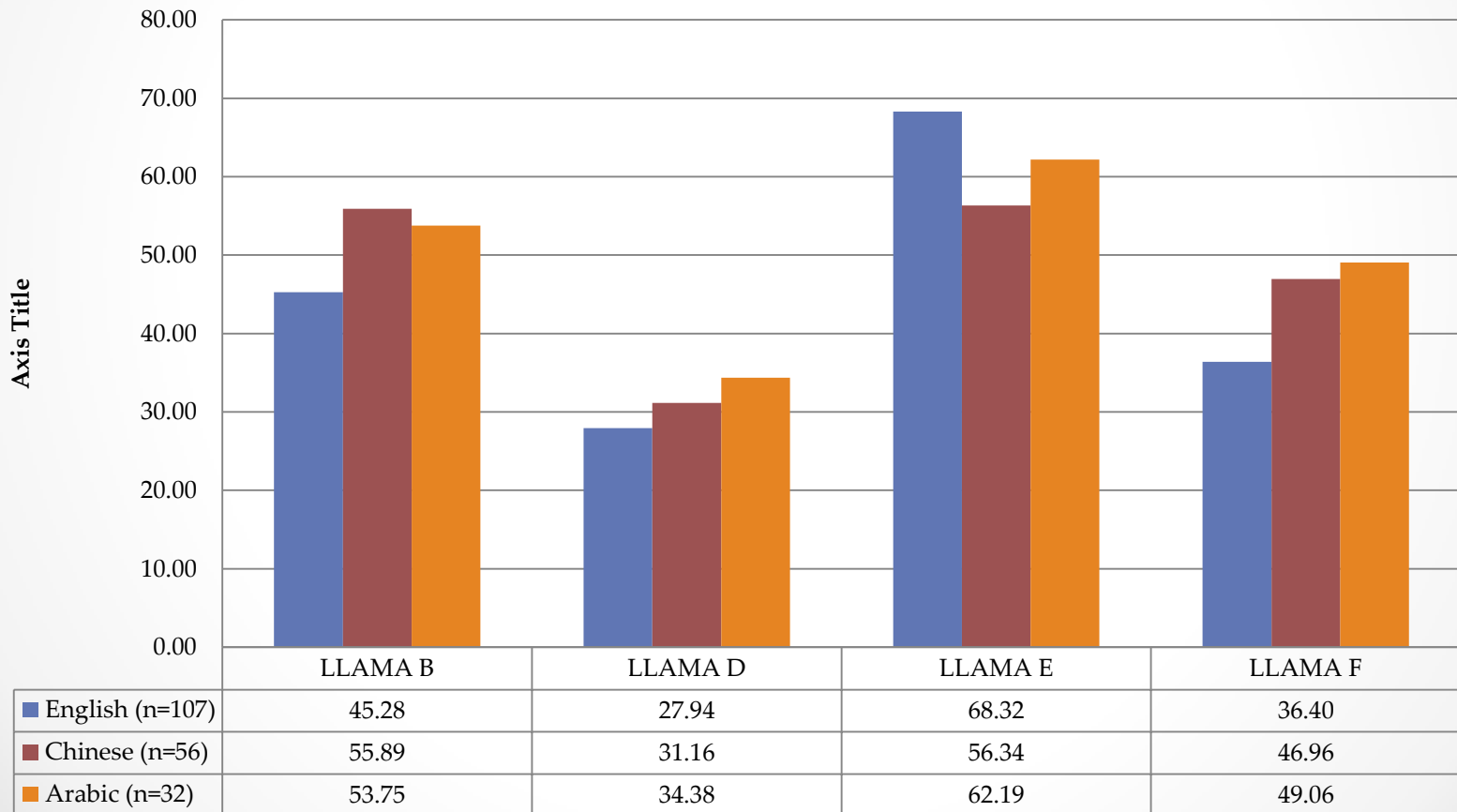
RQ1: Hypotheses

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.



RQ1: Language Neutrality

Language Neutrality



RQ1: Language Neutrality

- LLAMA B (vocabulary):
 - Chinese and Arabic participants outperformed the English participants ($p < .05$).
 - Hypothesis 1 disconfirmed: non-Roman alphabet participants are not negatively affected.
 - No difference between Chinese and Arabic participants
 - Hypothesis 2 disconfirmed as Arabic participants were not advantaged over Chinese participants.
- LLAMA F (grammatical inferencing)
 - Same results as for LLAMA B.
 - Hypotheses disconfirmed.
- LLAMA D (implicit learning)
 - Arabic participants significantly outperformed English participants.
- LLAMA E (sound-symbol correspondence)
 - Chinese participants performed significantly worse than English.
 - Also lower than Arabic but not significant.



LLAMA E question?

- Lower scores by Chinese participants in LLAMA E.
- Is there a problem with the test?
- Are there contrasts that are allophonic in Chinese?
- Consonants in LLAMA E:
 - [p],[b],[t],[d],[k],[g],[m],[n]
- Vowels in LLAMA E
 - a:, i:, u:
 - 0ì
 - 0è
 - 0ù
- None of these are allophonic in Chinese (Swan & Smith (2001)).



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)



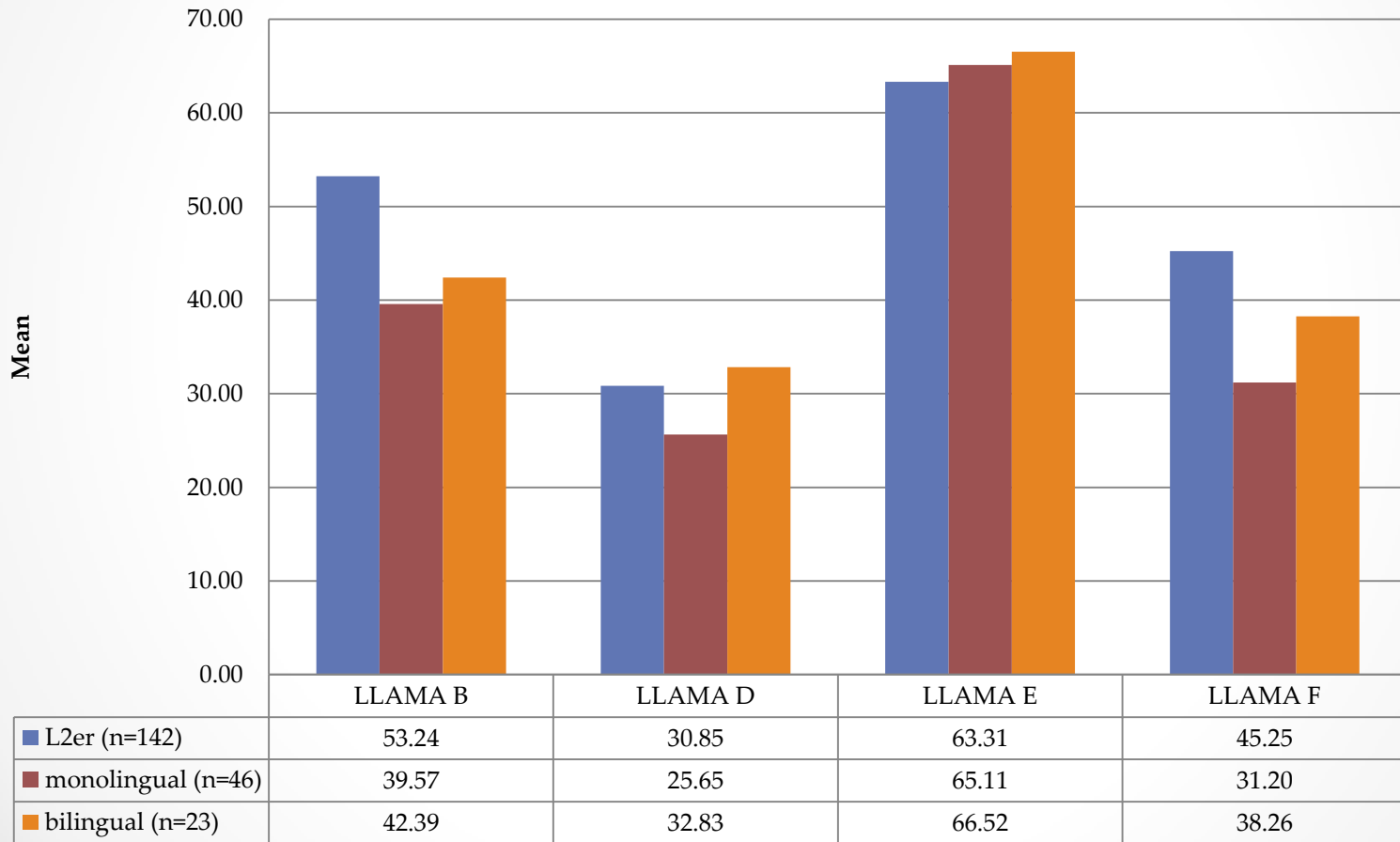
RQ2: Background

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)
 - No significant differences between any groups.
- 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 but not significant.



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.

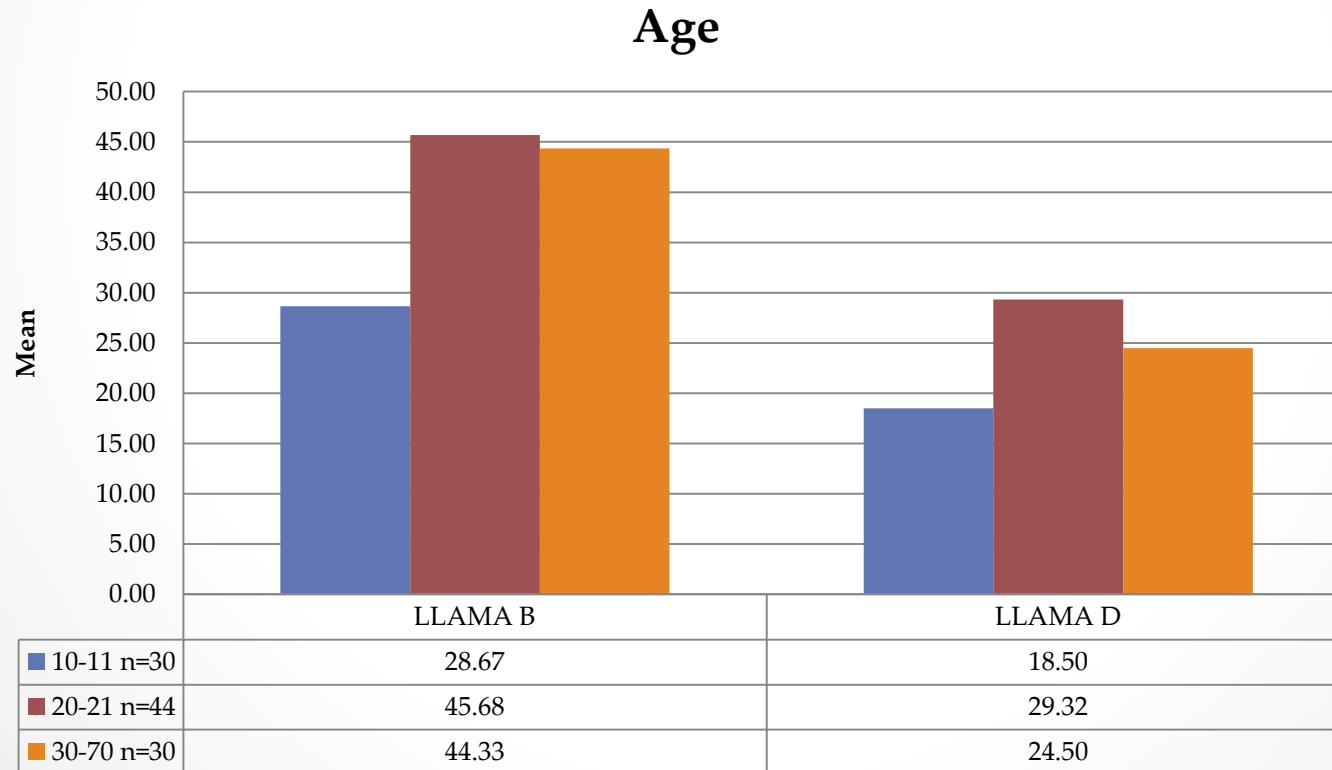
RQ3: Background

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?

RQ4: Background

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

- a. Gender, L1, L2 status, education level, logic puzzles, age
- These additional factors were examined in the 2014 study.
- Information collected through background questionnaire.



RQ4: LLAMA B

- Multiple regression, $n=240$
- Factors: L1, age, L2 status, educational level, gender, logic games
- Overall factors: $R^2 = 12.6\%$ of overall variance
 - Adjusted $R^2 = 9.9\%$
- Individual independent variables:
 - Only L2 status reaches significance.
 - Beta value = $-.240$, $p = .001$
 - Contribution to overall variance = 4.8%



RQ4: LLAMA D

- Multiple regression, $n=240$
- Factors: L1, age, L2 status, educational level, gender, logic games
- Overall factors: $R^2 = 8\%$ of overall variance
 - Adjusted $R^2 = 5.2\%$
- Individual independent variables:
 - Language neutrality and gender reach significance.
 - Language neutrality:
 - Beta value = .144, $p = .046$
 - Contribution to overall variance = 1.9%
 - Gender
 - Beta value = .178, $p = .010$
 - Contribution to overall variance = 3.2%



RQ4: LLAMA E

- Multiple regression, $n=211$
- Factors: L1, age, L2 status, educational level, gender, logic games
- Overall factors: $R^2 = 4.5\%$ of overall variance
 - Adjusted $R^2 = 1.6\%$
- Individual independent variables:
 - No variable reaches significance.
 - Highest beta value = education level
 - 1.6% of variance



RQ4: LLAMA F

- Multiple regression, $n=211$
- Factors: L1, age, L2 status, educational level, gender, logic games
- Overall factors: $R^2 = 8.1\%$ of overall variance
 - Adjusted $R^2 = 5.3\%$
- Individual independent variables:
 - L2 status and education level reach significance.
 - L2 status:
 - Beta value = $-.201$, $p = .008$
 - Contribution to overall variance = 3.4%
 - Education level
 - Beta value = $.186$, $p = .016$
 - Contribution to overall variance = 2.8%



RQ4: 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 eschewed so presentation will be tweaked.
2. Examine WM and IQ scores.
 - o WM measure attempted 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?



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