

Eyeing acceptability judgment task

Ryo Maie
Second Language Studies
Michigan State University, MI

2020 Annual Conference for AAAL
Denver, Colorado

SLRF2019 → AAAL2020



L1 speakers: $n = 27$

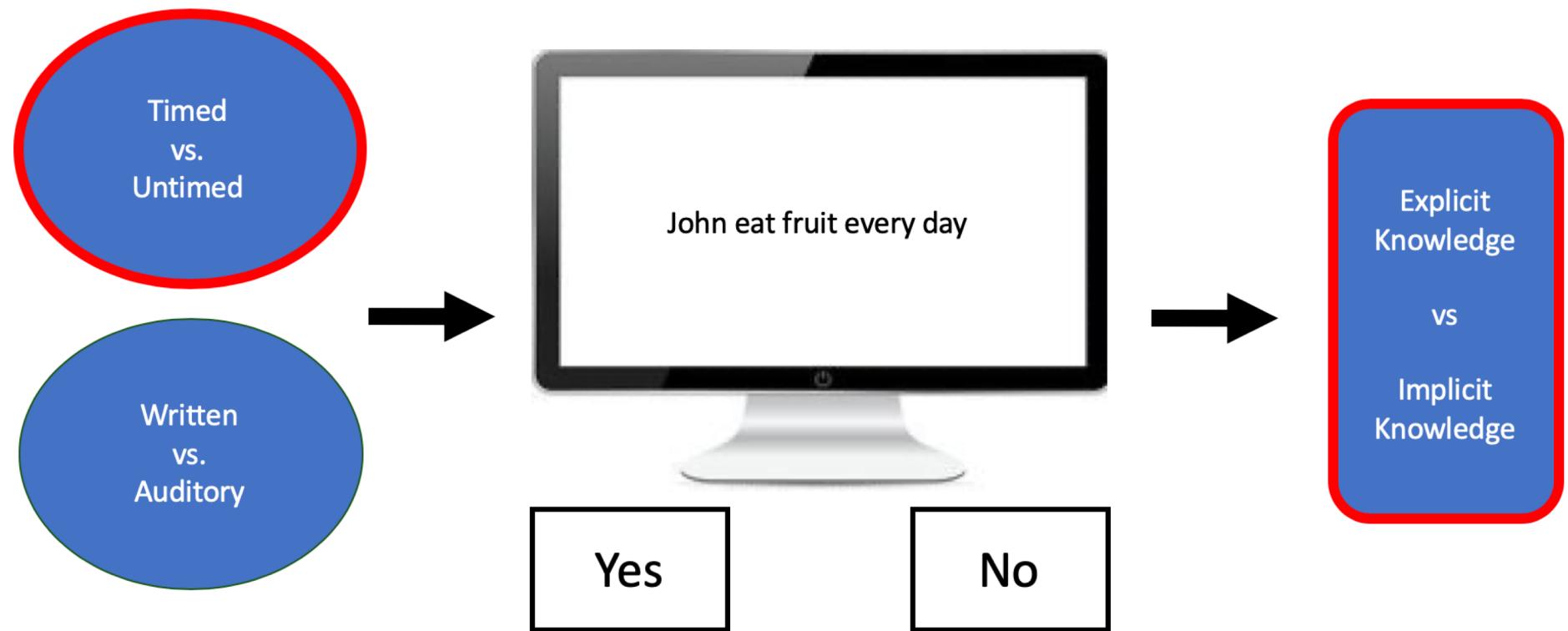
L2 speakers: $n = 27$



L1 speakers: $n = 31$

L2 speakers: $n = 40$

Acceptability judgment task



AJT and explicit and implicit knowledge – Ellis and others

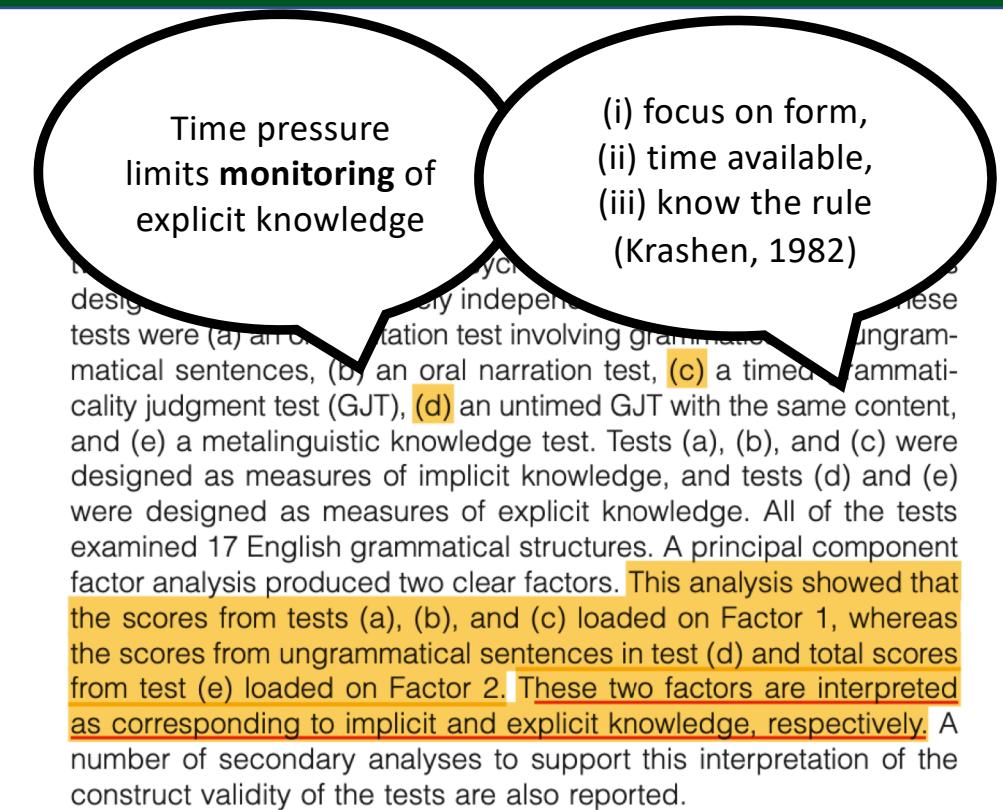
SSLA, 27, 141–172. Printed in the United States of America.
DOI: 10.1017/S0272263105050096

MEASURING IMPLICIT AND EXPLICIT KNOWLEDGE OF A SECOND LANGUAGE

A Psychometric Study

Rod Ellis

University of Auckland



Untimed = **Explicit**, Timed = **Implicit**

AJT and explicit and implicit knowledge – Vafaee et al. and others

Studies in Second Language Acquisition, 2016, page 1 of 37.
doi:10.1017/S0272263115000455

VALIDATING GRAMMATICALITY JUDGMENT TESTS

Evidence from Two New Psycholinguistic Measures

Payman Vafaee
University of Maryland

Yuichi Suzuki*
Kanagawa University

Ilina Kachisnke*
University of Maryland

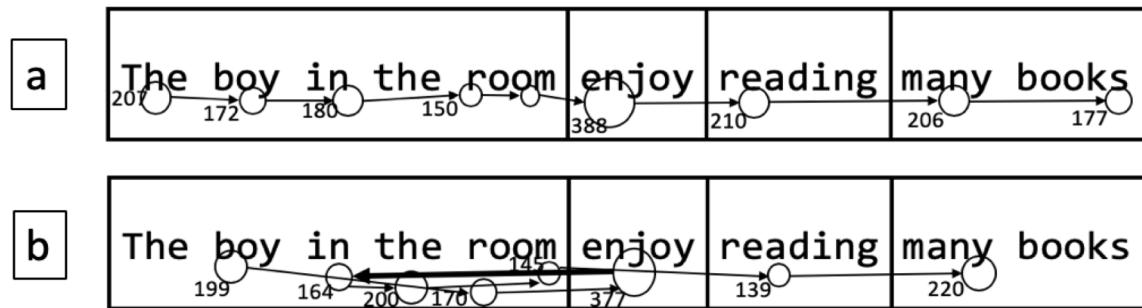
Time pressure merely makes it hard, not impossible (DeKeyser, 2003)
If automatized, time pressure cannot exclude EK

Self-paced reading tasks measure explicit knowledge (Ellis, 2007).
Word-monitoring tasks measure implicit knowledge (Gutiérrez, 2013).
It has also been shown that, irrespective of the time condition chosen, GJTs' grammatical sentences tap into IK, whereas their ungrammatical ones invoke EK (Gutiérrez, 2013). The current study examined these conclusions by employing two more fine-grained measures of IK: that is, a self-paced reading task and a word-monitoring task. The results of a confirmatory factor analysis revealed that manipulating GJTs' time conditions and/or the grammaticality of the sentences does not render them distinct measures of EK and IK. The current work shows that GJTs are too coarse to be measures of IK, and that the different types of GJTs measure different levels of EK.

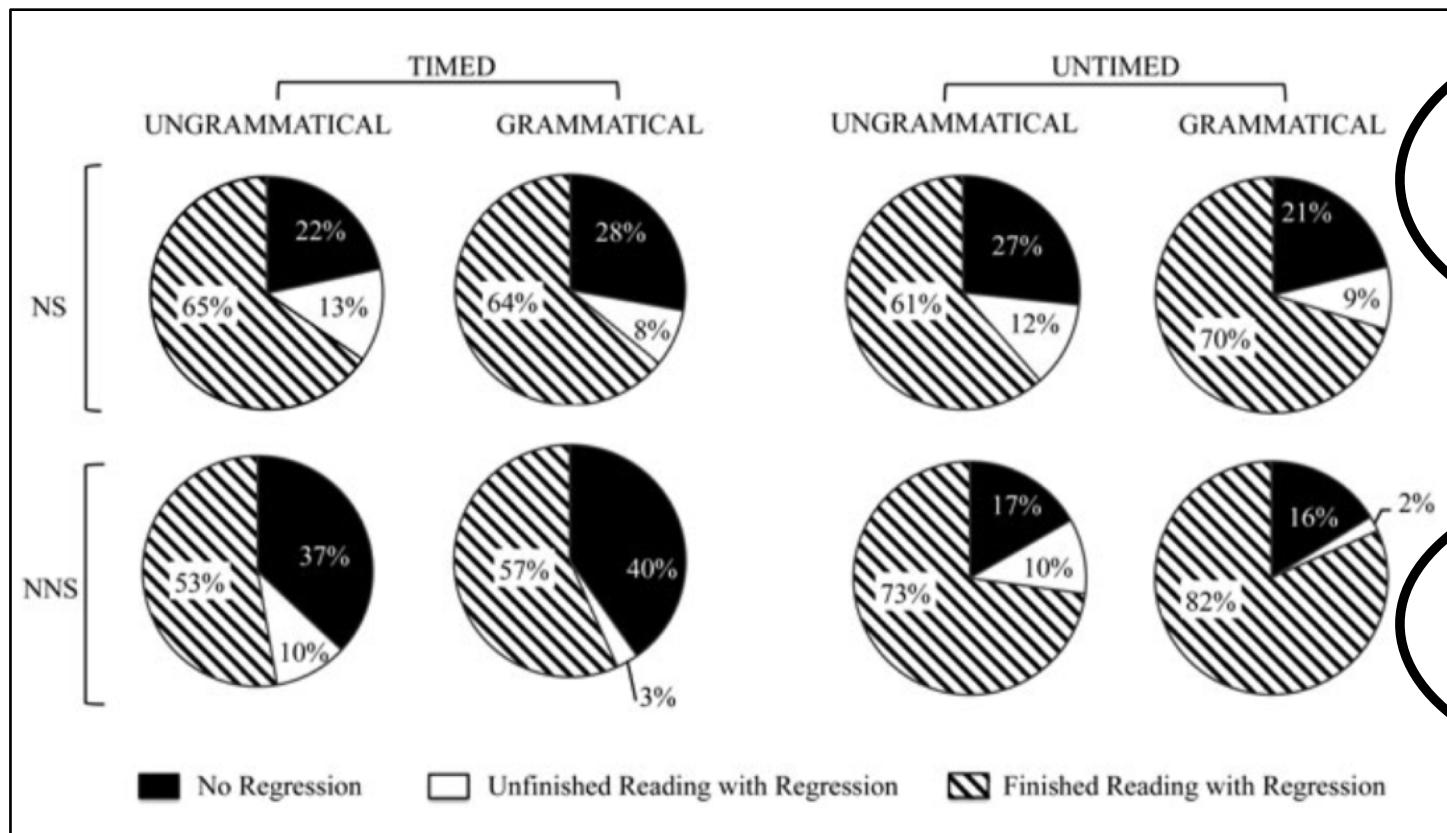
Untimed = **Explicit**, Timed = **Explicit**

Godfroid, Loewen, Jung, Park, Gass, & Ellis (2015)

- An eye-tracking study of Timed and Untimed AJT
 - My study is a conceptual replication and extension
 - Only one study of AJTs from a processing perspective
 - Virtually **nobody** looks at processing (as opposed to product) in AJTs
 - **Nobody** tests whether monitoring is restricted
 - Specifically looked at **regression** during reading
 - Regression while reading indicates **structural reanalysis** and **controlled processing** (Von der Malsburg & Vasishth, 2011)



Godfroid et al. (2015)



"[O]ur data suggested that timed and untimed GJTs ... measure different things." (p. 291)

Whether T-GJT measures implicit or automatized explicit knowledge depends on one's theoretical orientation

Why replication and extension?

- Conceptual and methodological verification needed
 - Regression conflates:
 - a. Controlled processing due to *processing difficulties* (Von der Malsburg & Vasishth, 2011)
 - b. Controlled processing due to *general conscious engagement in the task*
 - Godfroid et al. refers to (a), but R. Ellis (2005) and others refer (b) as **monitoring**.
 - **Monitoring by feel** = checking of one's production or input he/she perceives
 - **Monitoring by rule** = the process of drawing on explicit knowledge
 - Necessity to additional eye data that can complement regressions
- The role of (lexical) processing automaticity
 - If slow, L2 speakers can break down under time pressure (Hopp, 2010)
 - Time pressure effect is across the board (not just access to explicit knowledge)
 - To monitor by rule, readers must be able to read the sentence fast

Research questions

1. Analyzed both temporal and spatial measures of eye movements
 - Does time pressure consistently suppress fixations/regressions across the eye-movement measures, and if so, how does it differ between L1 and L2 speakers?
 - If L2 speakers not only regress less, but also fixate shorter and less often, there seems to be a good (or better) evidence that time pressure limits monitoring by rule
2. Examined the moderating role of (lexical) processing automaticity
 - Does processing automaticity moderate the degree to which L1 and L2 speakers are affected by time pressure, and if so, how does it differ between L1 and L2 speakers?
 - If those fast readers can perform similarly in timed GJT as in untimed GJT, they can outstrip the effect of time pressure, therefore can use explicit knowledge

Participants

	L1 speakers (<i>n</i> = 31)				L2 speakers (<i>n</i> = 40)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Age (years)	23.56	6.62	19	48	29.79	5.95	21	49
Age of arrival (years)	–	–	–	–	26.71	11.96	12	34
Length of residence (months)	–	–	–	–	48.64	39.64	12	216
TOEFL iBT	–	–	–	–	103.76	8.13	91	116

- All advanced speakers of English with
 - age of arrival at or after the age of 12
 - length of residence of at least a year (12 months)
 - TOEFL score higher than 90

Acceptability judgment task

- 96 sentences
 - (a) third-person singular -s, (b) mass vs. count, (c) passives, (d) verb complements, (e) embedded questions, and (f) comparatives
 - A half grammatical and the other half ungrammatical
 - Sentence length: *Mean* = 10.68 words, *Min* = 8, and *Max* = 13
- Time pressure
 - Pilot with 10 L1 English speakers
 - Median RT x 1.2 as time limit for L2 speakers (R. Ellis, 2005; Godfroid et al., 2015; Loewen, 2009)
 - Time limits ranged from 2172ms to 7850ms
- Internal consistency (KR-20)
 - L1 Timed = .79, L1 Untimed = .79, L2 Timed = .87, L2 Untimed = .88

Semantic classification task

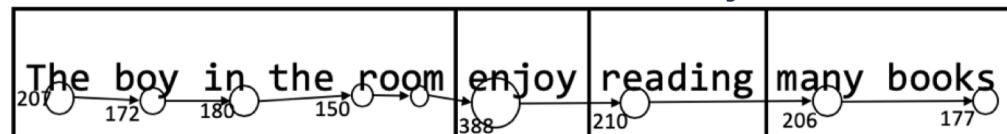
- A measure of lexical processing automaticity (speed)
- Judging a series of English words as animate or inanimate
 - 50 nouns (25 animate and 25 inanimate)
 - Mean frequency (BNC) = 181.50 per million ($SD = 208.99$)
 - Mean length = 5.45 letters ($SD = 3.07$)
- All are taken from sentence stimuli of AJTs (Lim & Godfroid, 2015)
- **Coefficient of variation** ($SD/Mean$) of RT was calculated
 - Shows the processing stability (one aspect of automaticity)
 - Hulstijn et al.'s (2009) five-step data cleaning procedure was applied
- Internal consistency (alpha)
 - .94 for L1 speakers and .92 for L2 speakers

Eye-tracking measures

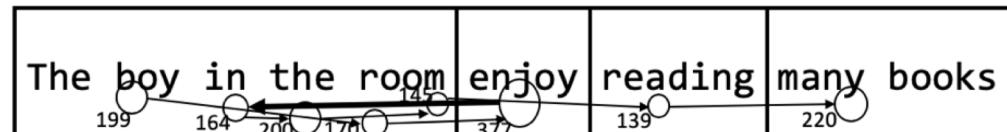
Eye-link 1000 plus



a



b



Mean Fixation Duration	Mean Fixation Count	Regression	Total Reading Time
<ul style="list-style-type: none">TemporalMean of all fixations across the sentenceLexical levelProcessing effort or depthLog10-transformed	<ul style="list-style-type: none">SpatialMean of the number of fixations across the sentenceSentence levelHow detailed readers process and intensity of processing	<ul style="list-style-type: none">SpatialSentence levelStructural analysis due to processing problemsBinary	<ul style="list-style-type: none">TemporalSum of all fixations across the sentenceSentence levelThe time readers read a sentenceLog10-transformed

Statistical analysis

- **Correlation analysis (skipped)**

- Whether and how accuracy of judgments relate to eye movements
- Whether and how eye-movement measures relate to each other

- **Statistical modeling**

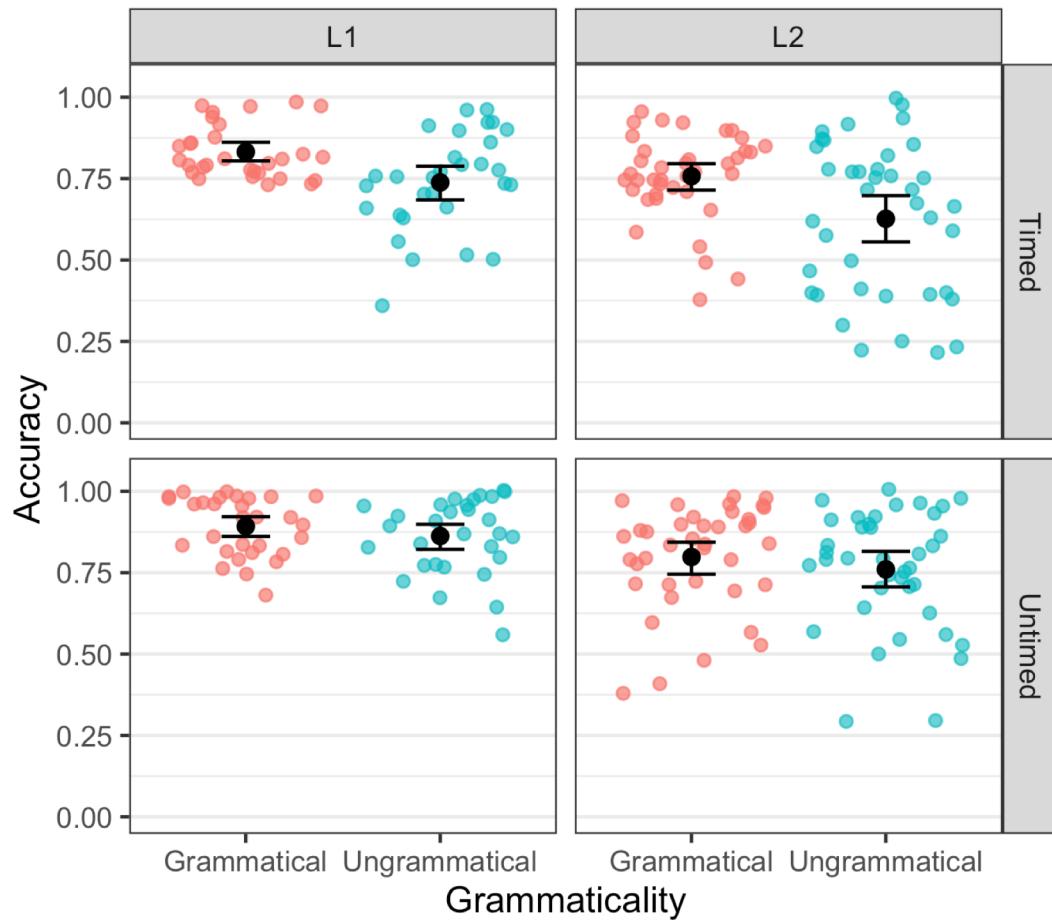
- Generalized linear mixed models using Bayesian inference
- Posterior \propto Prior \times Likelihood
 - Updating prior belief (Godfroid et al.) with likelihood (my data)
- Fixed effects:
 - Those that were significant in Godfroid et al. (2015) + CV (its interactions)
 - $Y \sim \text{Group} + \text{Condition} + \text{Grammaticality} + \text{CV} + \text{Group:Cond} + \text{Cond:Gram} + \text{Cond:CV} + \text{Group:Cond:CV}$
- Random effects:
 - $(1 + \text{Cond*Gram} | \text{Subject}) + (1 | \text{Item})$

Statistical models for eye-movement measures

Measures	Model	Link function	Note
Mean fixation duration	Gaussian	Identity	Log10-transformed
Mean fixation count	Poisson	Log_e	
Regression	Binomial (Bernoulli)	Logit	$n = 96$
Total reading time	Gaussian	Identity	Log10-transformed

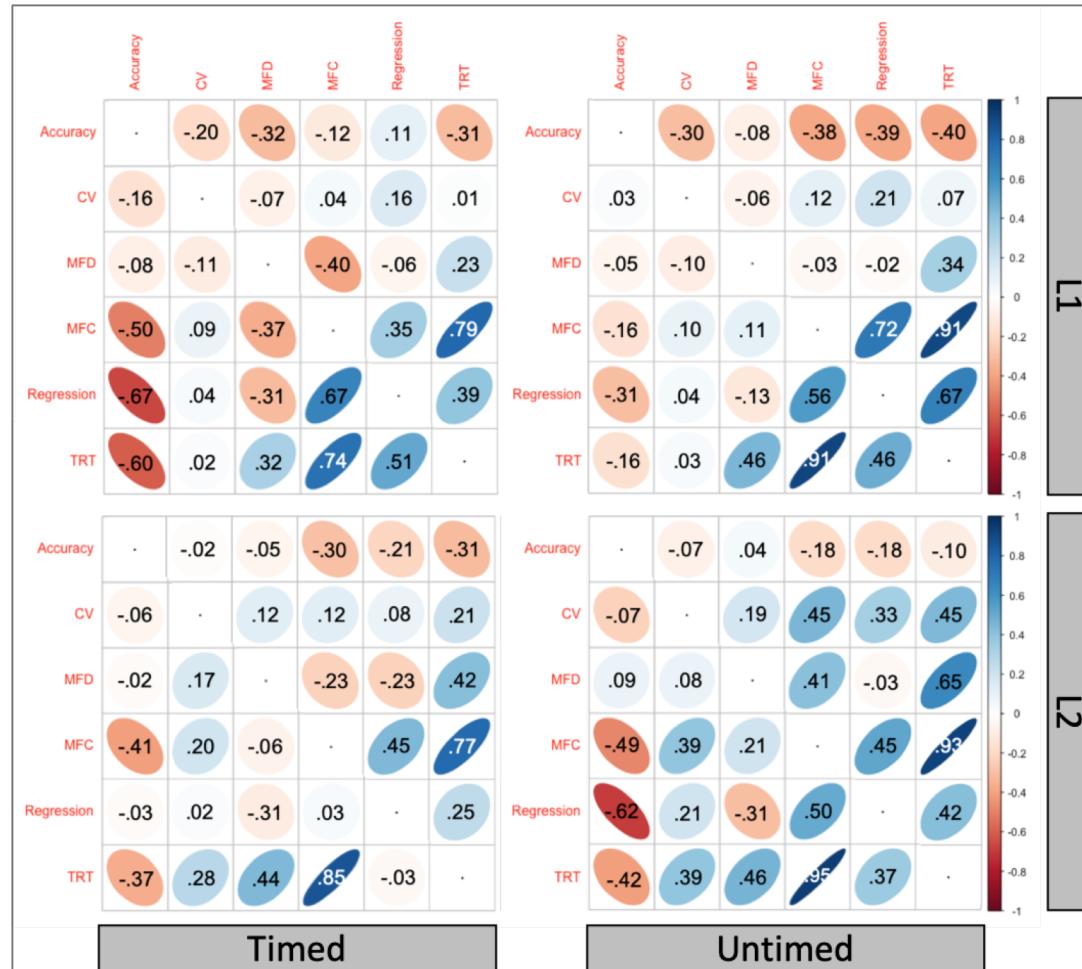
- Prior
 - For binomial model, informative prior based on Godfroid et al., where relevant
 - For others, weakly informative prior based on real-world knowledge (\neq non-informative)
- Estimation (point and interval)
 - Markov Chain Monte Carlo simulation (20000 iterations, 5000 burn-in, 3 thinning)
 - No-U-Turn Sampler, an extension of Hamiltonian Monte Carlo algorithm
 - Expected a posteriori = point estimates
 - 95% highest posterior density intervals = interval estimates (95 % credible)

Accuracy

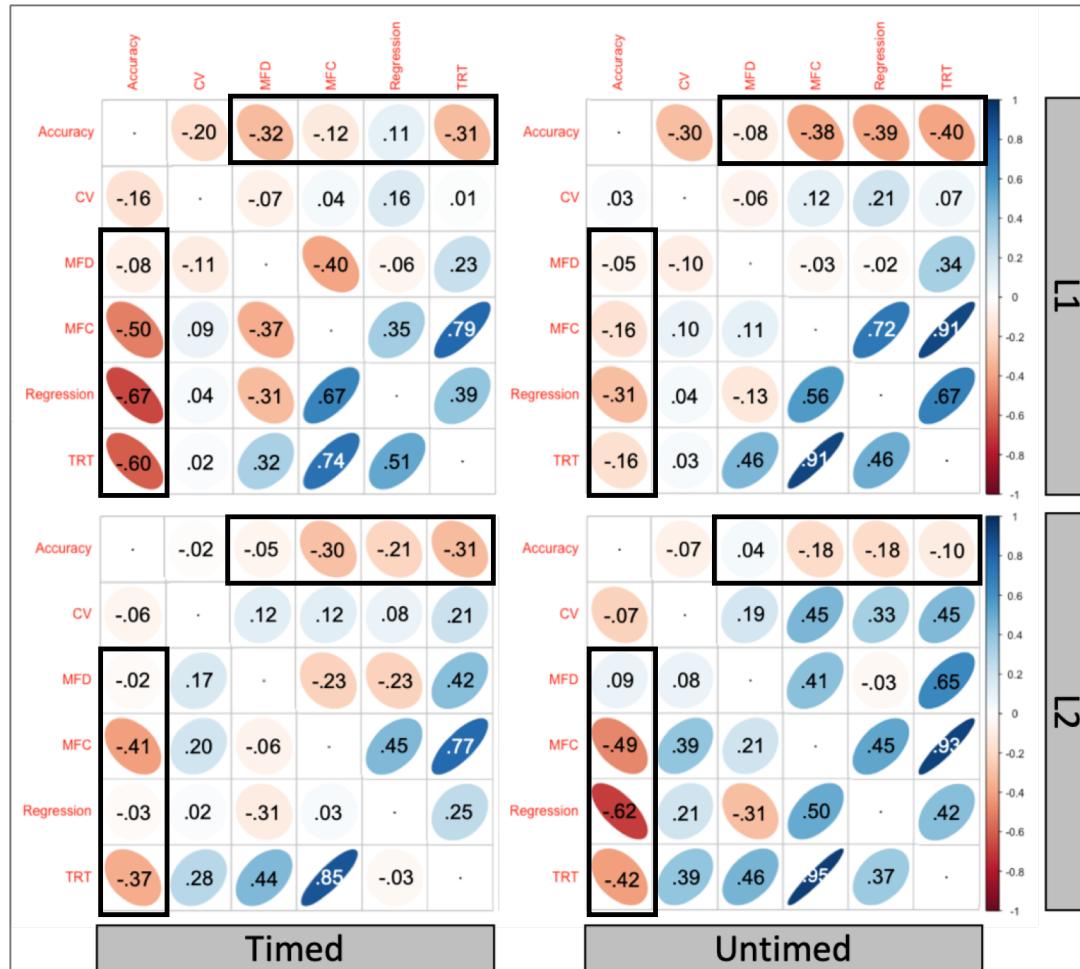


- Based on Bayesian ANOVA
 - Group ($L1 > L2$)
 - Condition (Untimed > Timed)
 - Grammaticality (Gram > Ungram)
 - Cond x Gram interaction

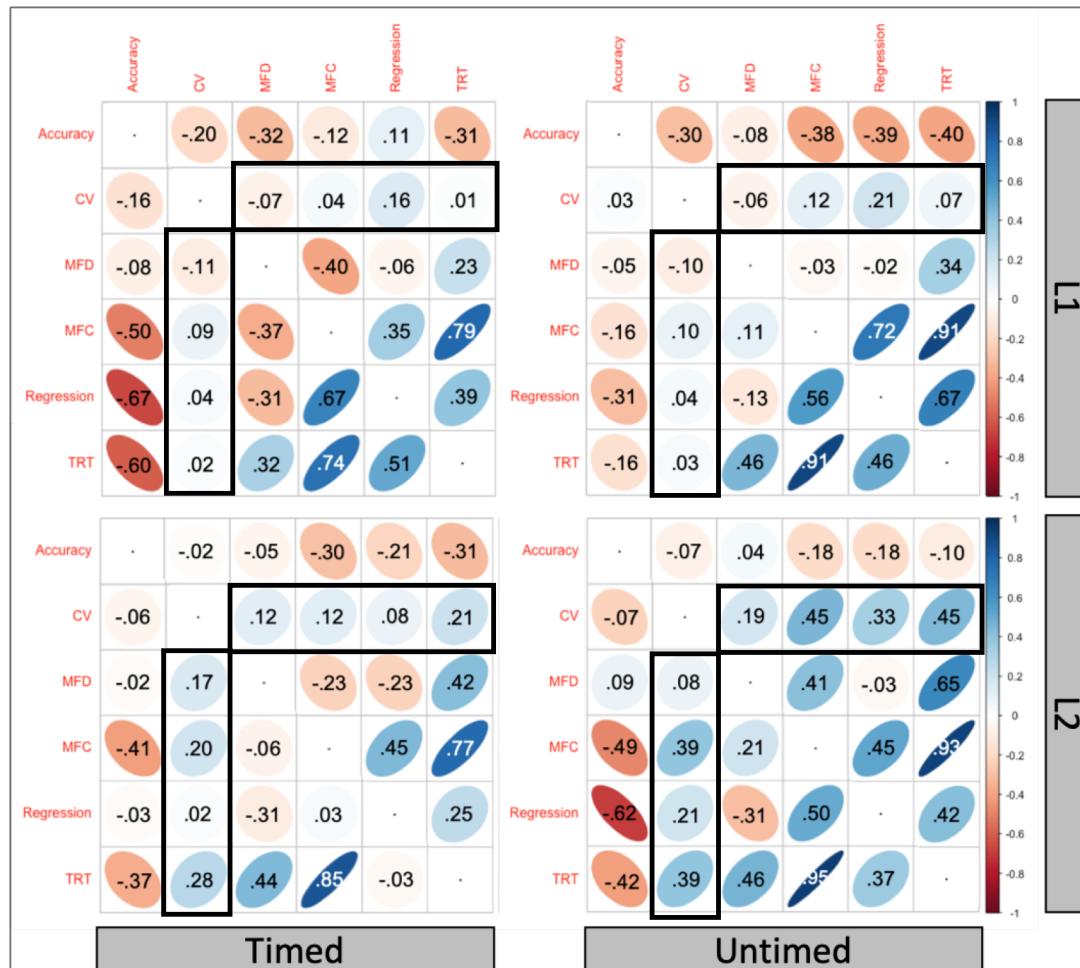
Correlational analysis



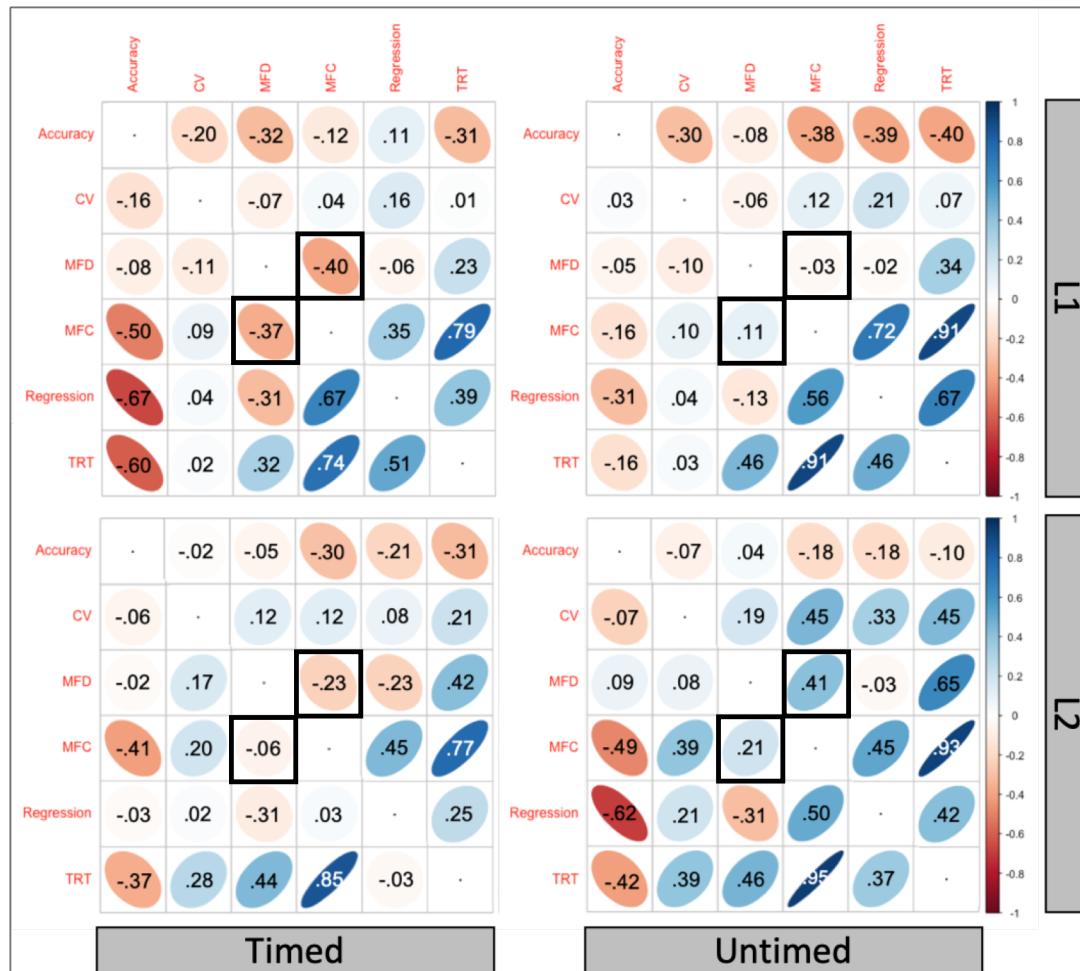
Correlational analysis



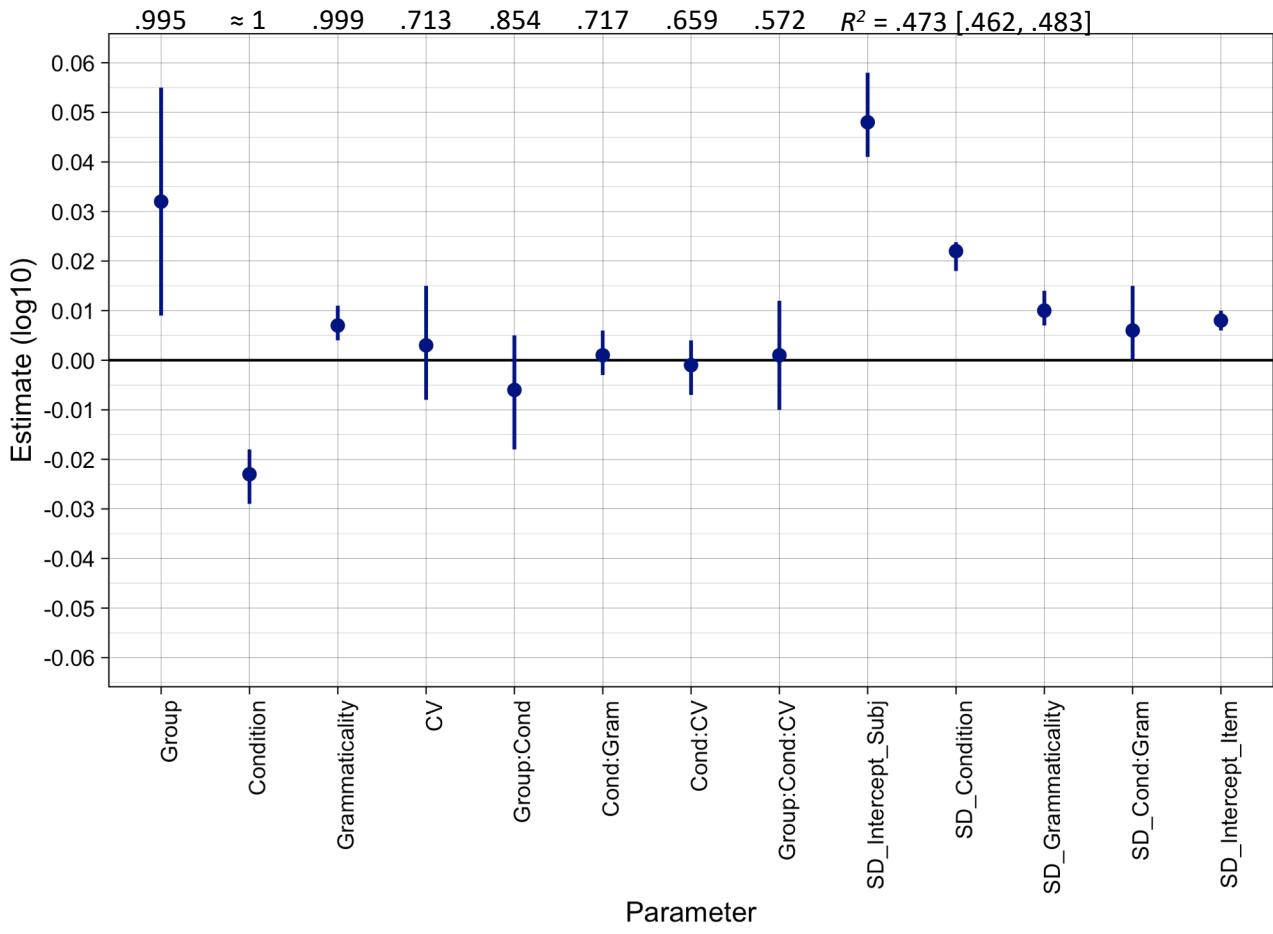
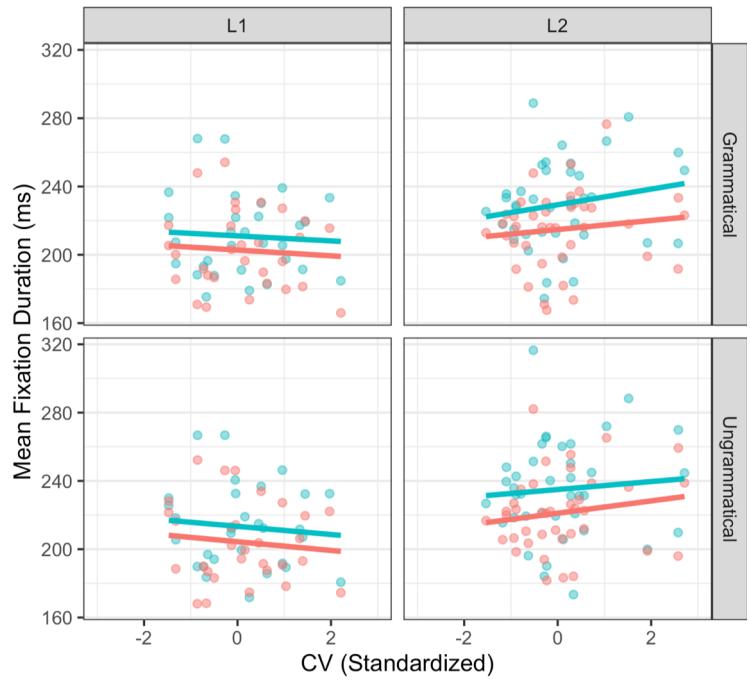
Correlational analysis



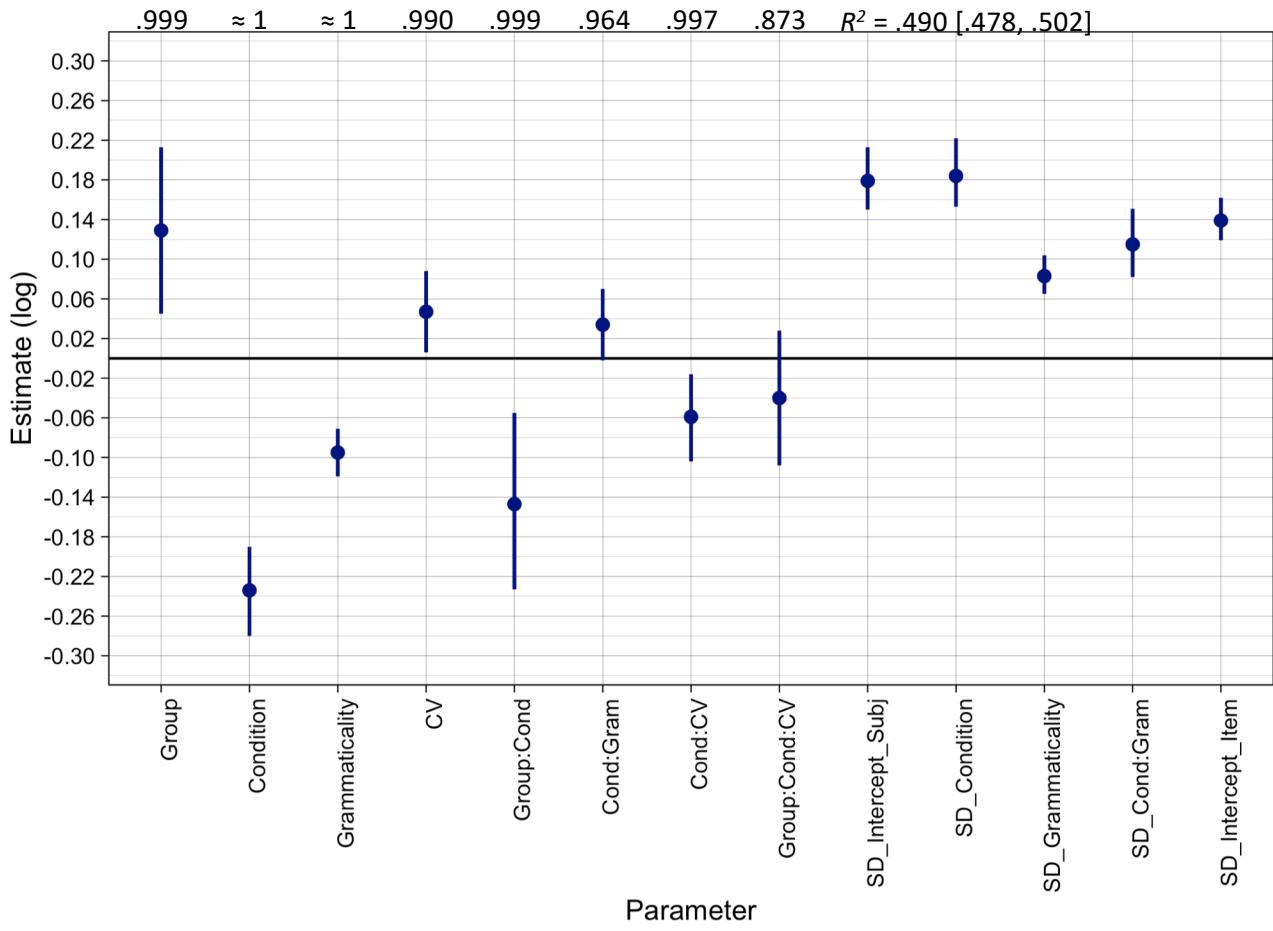
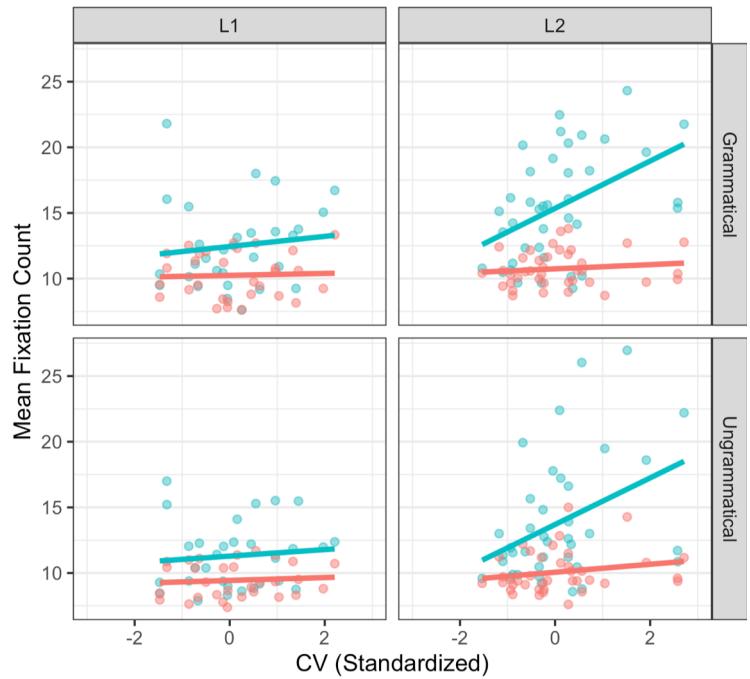
Correlational analysis



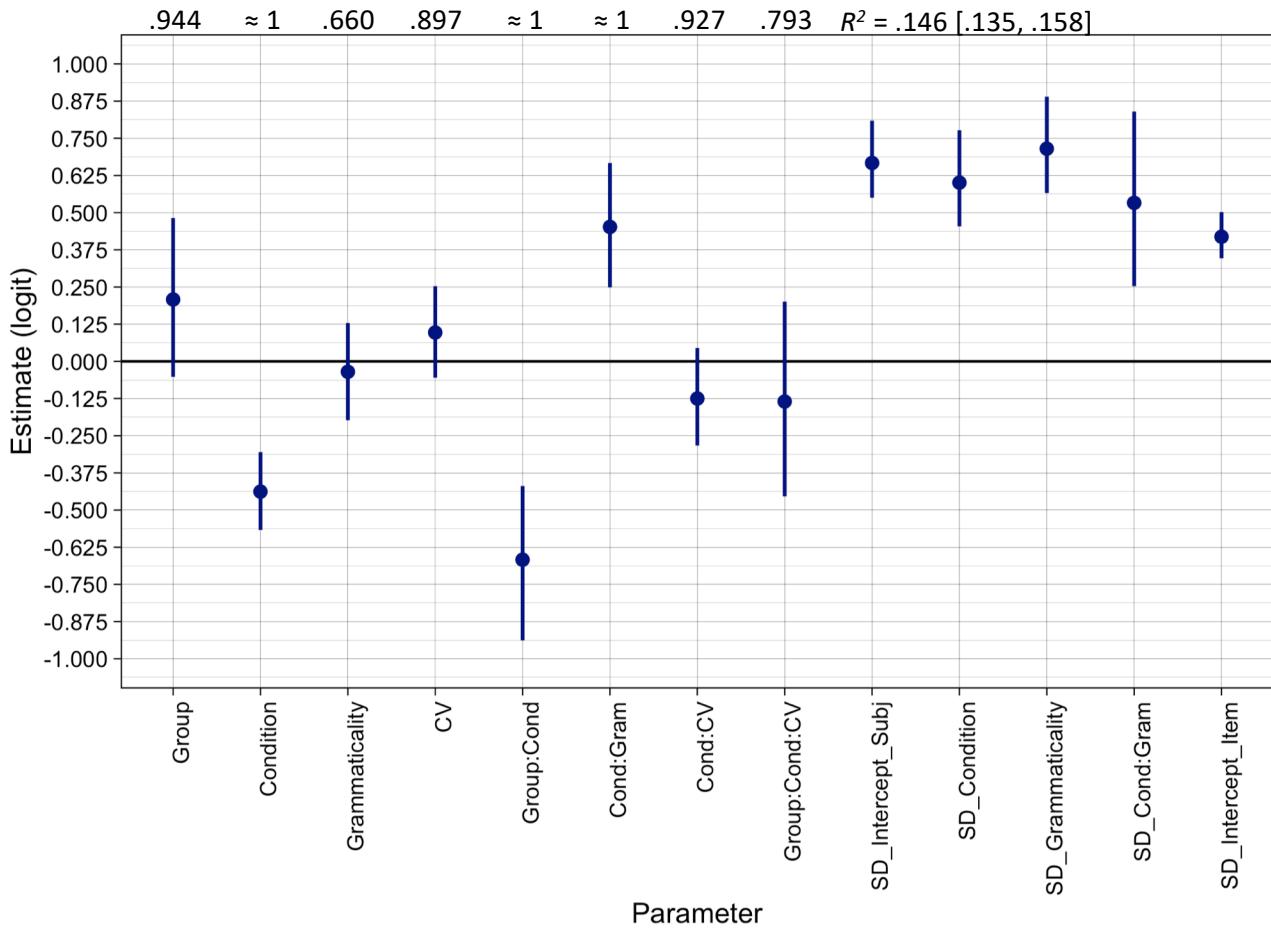
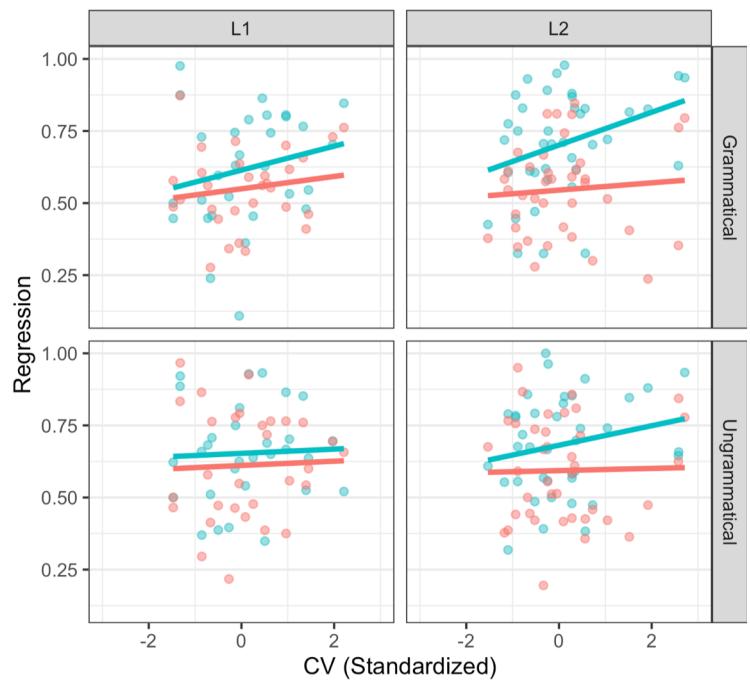
Mean fixation duration



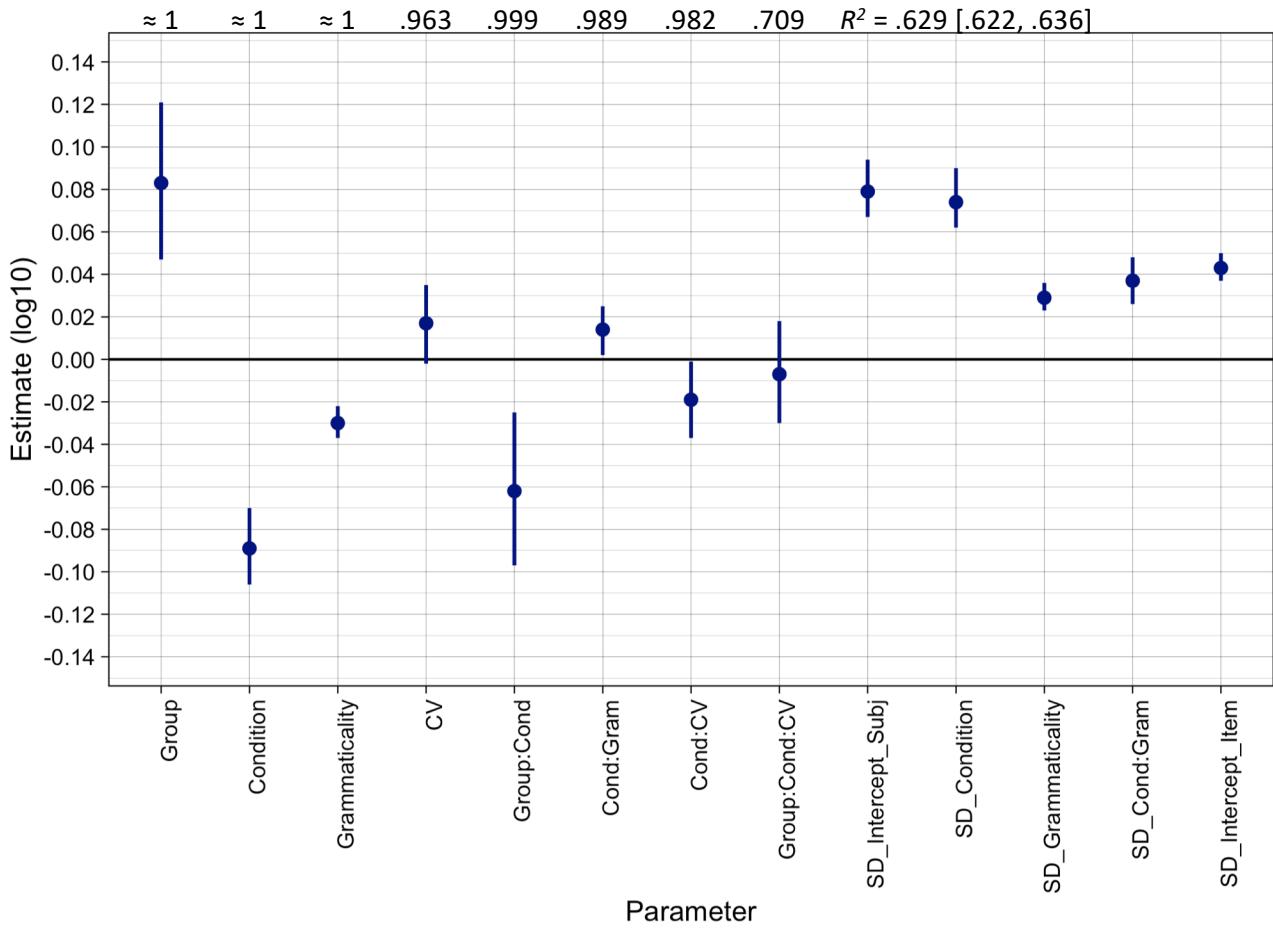
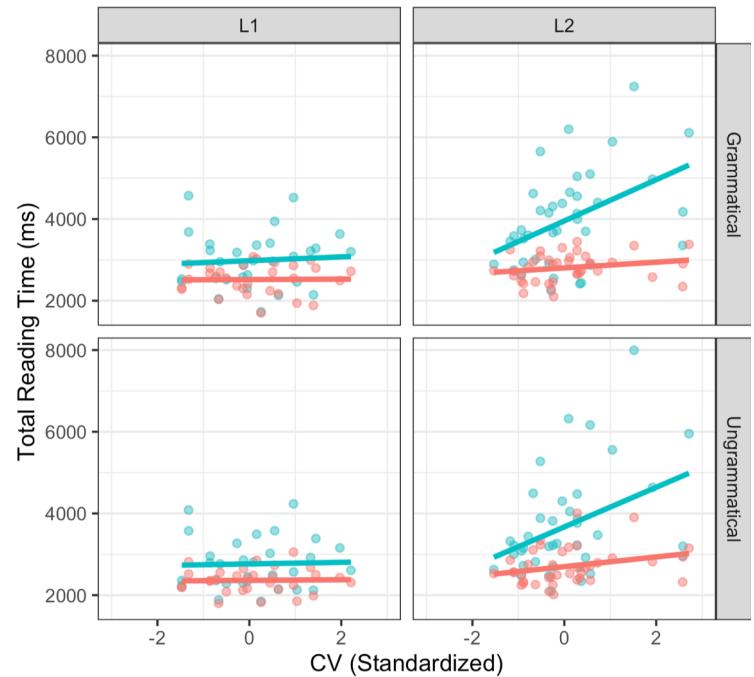
Mean fixation count



Regression



Total reading time

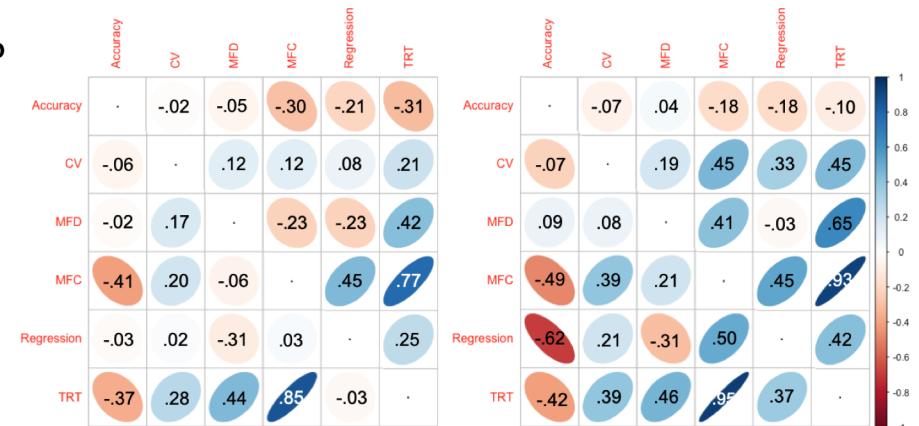
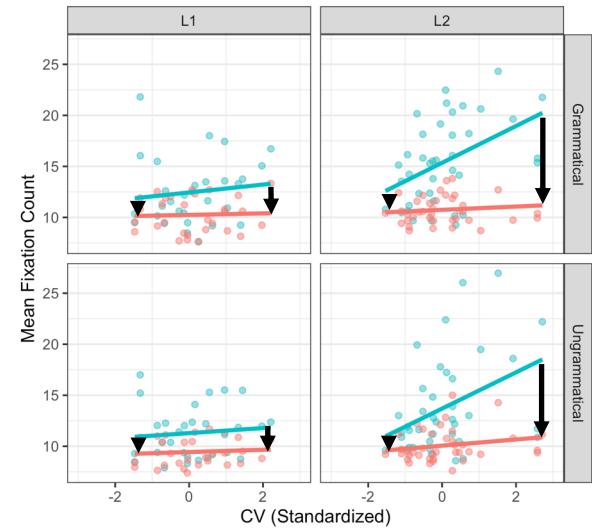


Discussion

- Time pressure
 - Both L1 and L2 speakers are affected on all four measures
 - On the sentence-level measures, a larger effect on L2 speakers
 - They are fixating and regressing not only because of processing difficulties but also of their general engagement in the task with conscious strategies (explicit).
- L1 vs L2 speakers
 - Correlational patterns (though not discussed in detail) suggests
 - Until hitting an ungrammatical element, L1 speakers monitor by feel (implicit knowledge), whereas L2 speakers monitor by rule (explicit knowledge)
 - After hitting an ungrammatical element, L1 speakers monitor by rule (only to fail), and L2 speakers also monitor by rule (explicit knowledge)

Discussion

- Moderation of time pressure
 - For L2 speakers
 - If with higher processing automaticity, T-AJT \approx U-AJT
 - Assuming U-AJT measures explicit knowledge, T-AJT also measures **explicit knowledge?**
 - If with slower processing automaticity, T-AJT < U-AJT
 - No room for monitoring \rightarrow **implicit knowledge?**
 - Both Ellis and Vafaee et al. are correct?
 - Can't cope with time pressure \rightarrow breakdown?
 - More likely
 - Any implication for non-advanced learners?
 - Instructed SLA
 - Explicit and implicit learning studies
 - Others?



Conclusion and Limitations

- Conclusion
 - Time pressure and processing automaticity moderate type of knowledge
 - Task features (e.g., time pressure) are instrumental but factors internal to learners (e.g., processing automaticity) must also be taken into account
- Limitations
 - No triangulation with awareness
 - Addition of subjective measures of awareness may help
 - Other individual difference factors unknown
 - Declarative vs. Procedural memory
 - Explicit vs. Implicit cognitive aptitudes
 - Different types of L1 scripts
 - Different L2 learning history

Acknowledgement

- Aline Godfroid
 - Michigan State University
- Shawn Loewen
 - Michigan State University
- Bronson Hui
 - Michigan State University
- Masaki Eguchi
 - University of Oregon



Thank you!

- For general inquiries, please email: maieryo@msu.edu
- Please visit my **ResearchGate** ([here](#)) for the slides
- Please visit **Open Science Framework** page ([here](#)) for this study