Regression Analysis of Longitudinal Extensive Reading on Writing Fluency in Japanese High School

Victoria University of Wellington,
Keio Shiki Senior High School
Noriyuki MORIYAMA
moriyaman@shiki.keio.ac.jp

1

Opening Question

What kind of things do bodybuilders tend to do at home?

2

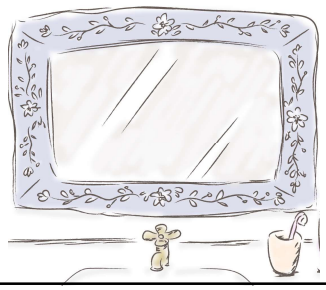
3

Visibility

Bodybuilders can see progress
Language learners usually can't

4

Need a mirror!



5

Why a "Mirror"?

Learners need a way to see progress
Recognize it, celebrate it

6

Not a Literal Mirror

"Not glass—" a tool that shows growth

7

Current Challenges 1

Reading helps
How much to read? → Unclear
Without outcomes → habits are hard

8

Current Challenges 2

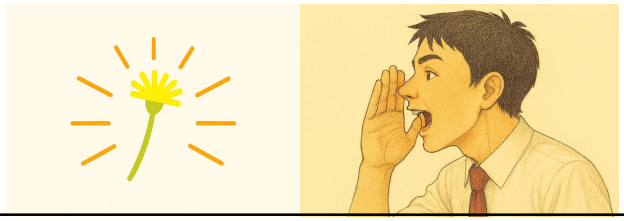
"[R]ule-breaking behaviors can be a real threat to the effectiveness of credit-based extensive reading" (Mikami and Shinozawa, 2023, p. 11).

9

What students need……

10

Small wins & encouragement
(from many angles)



11

Visible and measurable progress

12

"Reading and writing are no longer unrelated activities"
(Hirvela, 2016, pp.104-105).

ER would serve as a foundation for writing (Hedgcock & Ferris, 2018).

13

ER & Writing Connection

Key studies including writing fluency since 1990

- Mermelstein (2015)
- Lee & Hsu (2009)
- Lai (1993)
- Hafiz & Tudor (1990)

14

Research Question

To what extent do reading volume (total words read), reading breadth (books read), and essay-writing frequency (number of sessions) predict Japanese EFL students' essay length?

15

Methodology

16

Participants

N=170,　17-year-old boys
Private boys' high school
Neighboring prefecture of Tokyo
CEFR B1 level

17

Learning Context

EFL setting
Gradual shift:
General English → EAP
1/5 classes: writing
step-by-step five-paragraph essay

18

Materials (books)

Digital (Xreading) and physical books

Simplified (graded readers) and authentic English children's books

19

Materials (books)

YL (Yomiyasusa Level)
book rating system for ER

"[T]he most useful" guide for Japanese students to choose books
(Holster et al., 2017, p.238)

20

Materials (writing)

Criterion:
Online writing evaluation system by ETS

Overall score, total words, grammar & errors, etc.

*Discontinued in August

21

Procedure

ER: Two years

Writing: One year
Five-paragraph essay writing (step by step)

22

Procedure

The first year

ER
in class: 10-15 mins once a week
outside the class: voluntarily reading

23

Procedure

The second year

ER
in class: 10-15 mins once a week
outside the class: voluntarily reading

Writing class
Step-by-step five-paragraph essay writing
+ 30-minute essay writing 10 times a year

24

Data Analysis

ER: Xreading LMS

(both digital and physical books)

Writing: Criterion (ETS); online app

Regression analysis: Jamovi;
open-source statistical software

25

Writing Session vs ER

Sample / Data:
N_students = 170; 1,686 essays used
Missing < 3%, listwise deletion

Index:
Books & words were strongly correlated → combined into an Extensive Reading Index (a single composite).

26

Writing Session vs ER

Model:
Mixed-effects model allowing person-specific baseline and growth.
words ~ session + ER_index + (1 + session | ID)

Estimation:
z-scale; ML for comparison (AIC/BIC, R^2m) → report REML coefficients.

27

Words Read model

Outcome: Words written
Predictor: Words read
Cohort: 170 students;
Session-10 observed: 158 (93%).
Analysis sample: N=158, Session-10 only, listwise complete on Words Read/Written (missing within S10 = 0).

28

Books Read model

Outcome: Words written
Predictor: Books read
Cohort: 170 students;
Session-10 observed: 158 (93%)
Analysis sample: N=158, Session-10 only, listwise complete on Books Read/Written (missing within S10 = 0).

29

Results

30

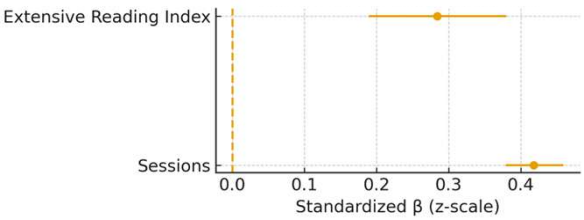
Sessions lead, ER lifts

session β = 0.42, ER index β = 0.28
(both $p < .001$)

Model fit:
adding the ER index raises R^2m 0.18 → 0.27
(Δ≈ +0.10) and lowers AIC/BIC (LRT $p < .001$).

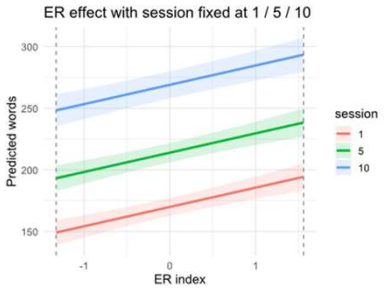
31

Effect sizes with 95% CIs



Note. Standardized (z) coefficients from LMM: words ~ session + ER_index + (1 + session | ID); ΔR^2m = +.096; both predictors $p < .001$.

32



Note. Predictions are conditional means: E(words | session, ER_index). ER_index is a person-level covariate measured at term end (retrospective grouping). Between-person association; not within-session causality. Shaded = 95% CI/PI; dashed lines = observed 10–90% ER range.

33

Words Read Model

	R^2	Adj. R^2	F (df1, df2)	p
Words read	0.21	0.21	41.5 (1,156)	<.001

34

Words Read Model

	N	*M*	*SD*	Missing
Words Written	158	275.8	77.6	0
Words Read	158	196216	188439	0

Note: The drop from 170 to 158 is due to students without a Session-10 record (attrition), not item-level missing.

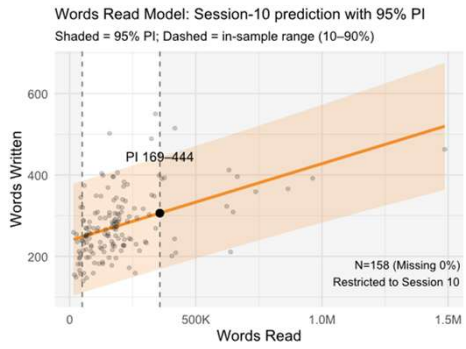
35

Words Read Model

Predictor	Estimate	*p*	Std. Estimate(β)	95% CI
Words Read	0.00019	<.001	0.46	[0.00013, 0.00025]

Note. **+100,000 words → ≈ +19 words (95%CI:13-25)**
(Intercept) Estimate = 238.8 (95%CI: 223.1–254.5, $p < .001$)

36



Orange line: Predicted mean

Scatter: Observed data points (plotted lightly).

Simple linear prediction (in-sample): association, not causation.

90th pct = 357,540 words (upper safe-zone limit)

37

Words Read Model

Prediction formula (mean response)

$\hat{Y} \approx 238.78 + (0.000189 \times \text{Words Read})$

Ex. Words Read 357,540 → ≈ 306 words

$238.78 + (0.000189 \times 357{,}540) = 306.3$

Note. Session-10 average association; valid within the observed 10–90% of Words Read. (outside = extrapolation)
Slope = +0.00019 (~+19/100k); 95% PI ≈ $\hat{y}$ ±136. Non-causal; CI (mean) ≠ PI (individual).

38

Books Read Model

	R^2	Adj. R^2	F (df1, df2)	p
Books read	0.22	0.22	44.4 (1,156)	<.001

39

Books Read Model

	N	*M*	*SD*	Missing
Words Written	158	275	77.6	0
Books Read	158	242	132	0

Note: The drop from 170 to 158 is due to students without a Session-10 record (attrition), not item-level missing.

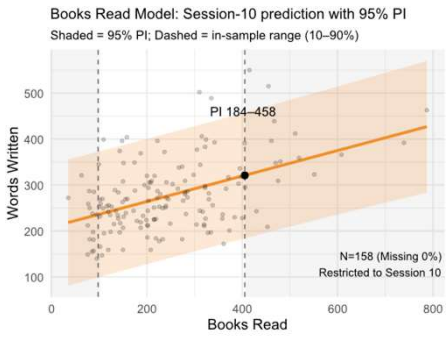
40

Books Read Model

Predictor	Estimate	*p*	Std. Estimate(β)	95% CI
Books Read	0.28	<.001	0.47	[0.20, 0.36]

Note. **100 books ≈ +28 words (95%CI: 20-36)**
(Intercept) Estimate = 275.8 (95%CI: 265 – 286.6, $p < .001$)

41



Orange line: Predicted mean

Scatter: Observed data points (plotted lightly).

Simple linear prediction (in-sample): association, not causation.

90th pct = 405 books (upper safe-zone limit)

42

Books Read Model

Prediction formula (mean response)
$\hat{Y} \approx 208.72 + (0.2773 \times \text{the number of books read})$

Ex. 405 books → ≈ 321 words
208.72 + (0.2773 x 405) = 321

Note. Session-10 average association; valid within the observed 10–90% of Books Read (outside = extrapolation). Slope ≈ 0.28 words/book (95% CI 0.20–0.36). 95% PI = $\hat{y}$ ±135. Non-causal; CI (mean) ≠ PI (individual).

43

Conclusion

44

Regression gives predictions & prediction formula

⇩

Small concrete & measurable goals for students

⇩

Students can celebrate these small successes.

45

Writing sessions > ER (Std. β)
However, both contribute
Collect & calibrate your own data
Mirror progress back to students!

46

Limitations & Future Work

47

- Bookshelf heavily relies on online ER system
- Limited to writing fluency alone
- Need to collect more data of avid readers

48

Acknowledgements

I am deeply grateful to Dr. Skalicky Stephen (Victoria University of Wellington) for his invaluable guidance throughout this study.

49

Reference

50

Fox, J., & Weisberg, S. (2023). *car: Companion to Applied Regression*. [R package]. Retrieved from https://cran.r-project.org/package=car.

Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Retrieved from https://gamlj.github.io/.

Gallucci, M. (2020). *Model goodness of fit in GAMLj*.

Hafiz, F. M., & Tudor, I. (1990). Graded readers as an input medium in L2 learning. *System*, *18*(1), 31–42. https://doi.org/10.1016/0346-251X(90)90026-2

Hedgcock, J. S., & Ferris, D. R. (2018). *Teaching readers of English: Students, texts, and contexts* (2nd ed., Vol. 1). Routledge. https://doi.org/10.4324/9781315465579

Hirvela, A. (2016). *Connecting reading & writing in second language writing instruction* (2nd ed.). University of Michigan Press.

Holster, T. A., Lake, J. W., & Pellowe, W. R. (2017). Measuring and predicting graded reader difficulty. *Reading in a Foreign Language*, *29*(2), 218–244.

Lai, F.-K. (1993). The effect of a summer reading course on reading and writing skills. *System*, *21*(1), 87–100. https://doi.org/10.1016/0346-251X(93)90009-6

51

Lee, S.-Y., & Hsu, Y. (2009). *Determining the crucial characteristics of extensive reading programs: The impact of extensive reading on EFL writing*. https://ijflt.com/wp-content/uploads/2021/03/IJFLTSummer-09.pdf

Lüdecke, Ben-Shachar, Patil & Makowski (2020). *Extracting, Computing and Exploring the Parameters of Statistical Models using R*. CRAN.

Mermelstein, A. D. (2015). Improving EFL learners' writing through enhanced extensive reading. *Reading in a Foreign Language*, *27*(2), 182–198.

Mikami, H., & Shiozawa, T. (2023). Cheating in Extensive Reading: Myth or Reality? SAGE Open, 13(2). https://doi.org/10.1177/21582440231168809

The jamovi project (2025). *jamovi*. (Version 2.7) [Computer Software]. Retrieved from https://www.jamovi.org.

R Core Team (2025). *R: A Language and environment for statistical computing*. (Version 4.5) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from CRAN snapshot 2025-05-25).

Revelle, W. (2025). *psych: Procedures for Psychological, Psychometric, and Personality Research*. [R package]. Retrieved from https://cran.r-project.org/package=psych.

52

Appendices

53

Writing session & ER

	session	words	ER_index	books_final	read_final
N	1686	1686	1700	1700	1700
Missing	14	14	0	0	0
Mean	5.4638	219.08	-8.0665e−18	236.30	190201
Median	5.0000	211.00	-0.17392	215.00	158101
Standard deviation	2.8572	75.215	1.3654	132.01	184376
Minimum	1	40	-1.6079	35	14931
Maximum	10	550	7.9226	787	1486825

54

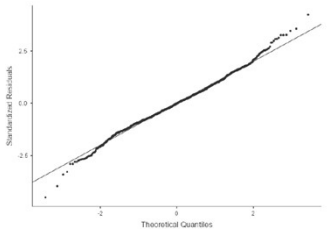
Writing session & ER

Names	Estimate	SE	95% Confidence Intervals Lower	Upper	df	t	p
(Intercept)	-8.5382e−4	0.054122	-0.10701	0.10530	163.84	-0.015776	0.987
session	0.41820	0.020326	0.37833	0.45807	169.31	20.574457	<.001
ER_index	0.28428	0.048768	0.18863	0.37993	167.94	5.829221	<.001

Note. Predictors (session, Extensive Reading Index) and the outcome (words) were z-scored (columnwise) prior to fitting in GAMLj. Therefore, the fixed-effect estimates are standardized coefficients (β): the change in the dependent variable (in SD units) for a 1-SD increase in each predictor. Random effects: (1 + session | ID). Final coefficients reported with REML; ML was used for model comparison (AIC/BIC, marginal R^2, LRT). Satterthwaite *df*.

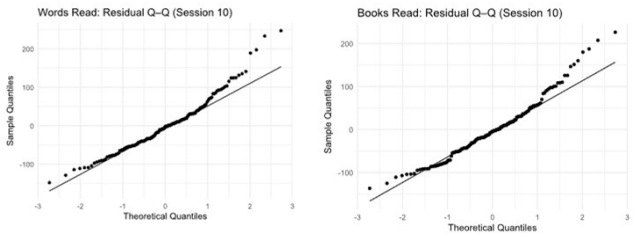
55

Q-Q plots of writing session & ER_index



56

Q-Q plots (Words Read & Books Read model)



Q–Q shows right-tail heaviness; 95% PI may under-cover at high outcomes. Interpret high-end predictions with caution.

57

Words Read Model

	Words Written	Words Read
N	158	158
Missing	0	0
M	275.8	196216
SD	77.6	188439
Minimum	140	14931
Maximum	550	1486825

58

Words Read Model

Names	Estimate	SE	95% Confidence Intervals Lower	Upper	β	df	t	p
(Intercept)	238.77772	7.23231	223.05037	254.50507	0.00000	156	33.01542	<.001
Words Read	0.00019	0.00003	0.00013	0.00025	0.45852	156	6.09937	<.001

Note. Inferential tests and p-values of the effects are adjusted for heteroschedasticity.

59

Books Read Model

	Words Written	Books Read
N	158	158
Missing	0	0
M	275.8	242.1
SD	77.6	131.8
Minimum	140	35
Maximum	550	787

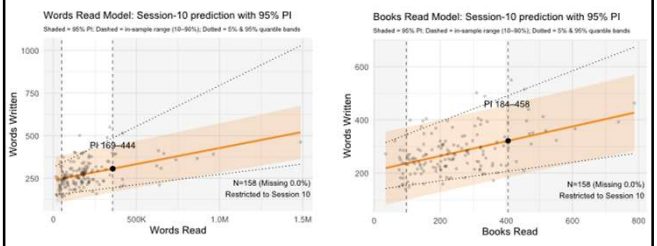
60

Books Read Model

Names	Estimate	SE	95% Confidence Intervals Lower	Upper	β	df	t	p
(Intercept)	275.84177	5.49805	265.04516	286.63838	0.00000	156	50.17083	<.001
Books Read	0.27730	0.04030	0.19514	0.35947	0.47086	156	6.88106	<.001

Note. Inferential tests and p-values of the effects are adjusted for heteroschedasticity.

61



Shaded = 95% PI; Dashed = in-sample range (10–90%); Dotted = 5–95% conditional quantile bands. Residual Q–Q indicates mild right-tail heaviness; coefficient inference uses HC3 robust SE; interpret high-end predictions with caution.

62