Word-features influencing second-graders' word recognition in connected texts: Secondary analysis of oral reading fluency data using the Rasch & explanatory item-response models

BEAR seminar
April 29, 2025

Outline

1. Introduction
    a. Why examine word recognition in connected text by 2nd graders?
    b. Oral reading fluency assessment
    c. Word-feature predictors of interest:
        i. Letter ↔Sound consistency,
        ii. transparency rating,
        iii. decoding demand,
        iv. word position in text
2. Research Questions
3. Methods
4. Preliminary Findings
5. Next Steps

Introduction

Many factors affect word recognition

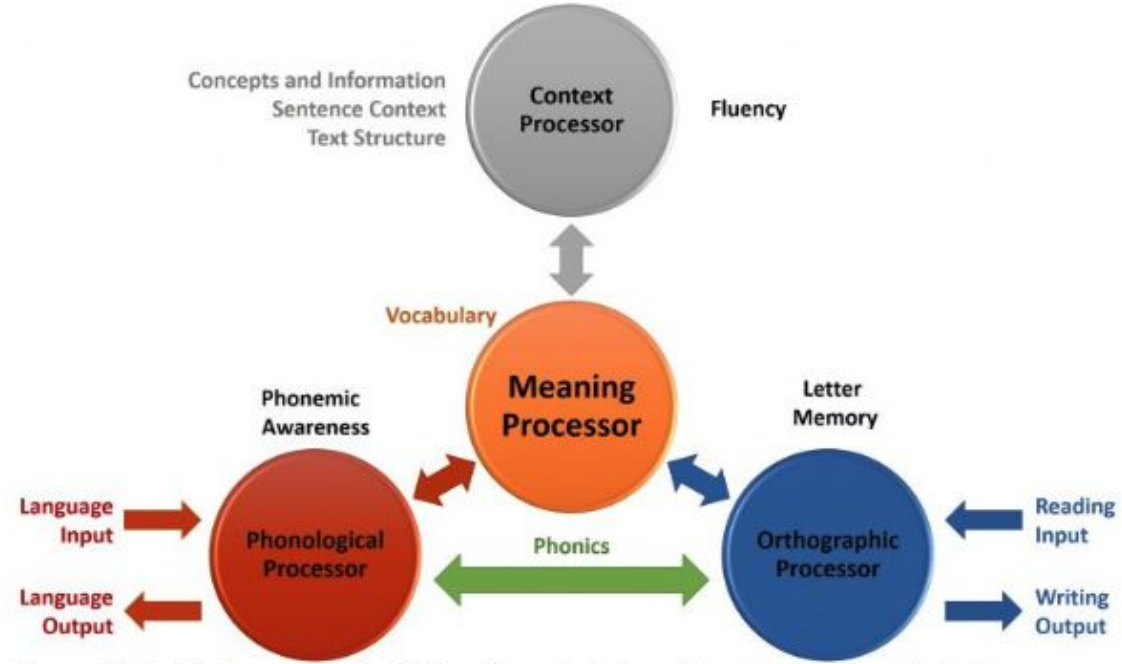
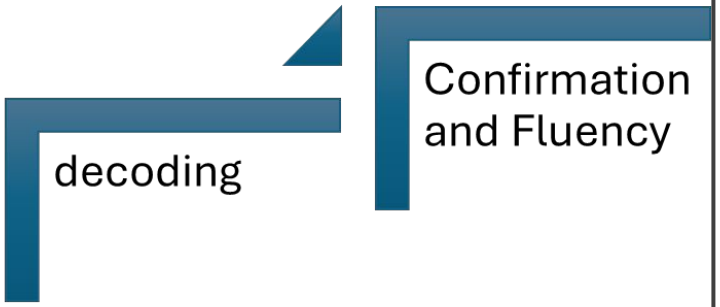


Figure 8.3 Seidenberg and McClelland's model of reading [Image description]

source: https://opentextbc.ca/psyclanguage/chapter/reading-models/#f8.4

Why second grade?



- **Theory**: Children should transition from laborious decoding to automatic word recognition during this period (Chall, 1983; Stahl et al., 2005).
- **Evidence**: Among students in the lowest 30% at end of Grade 1, the distinction between those who became proficient readers and those who continued to struggle was established by the end of second grade (Spira et al., 2005).

Oral Reading Fluency Assessments: Grade 2

Steven was good at ping pong. He even got good enough to beat his big brother and his uncle. His family was full of good ping pong players. Steven tried his best with each game, but he was always a good sport when he lost a game. Mostly he just had fun playing his best.

Typical ORF Information on Student Performance

Negligible Risk

Minimal Risk

Some Risk

At Risk

First grade			Second grade			Third grade		
B	M	E	B	M	E	B	M	E
Oral Reading Fluency (ORF) – Words Correct								
35+	57+	76+	85+	117+	128+	105+	141+	136+
34	56	75	84	116	127	104	140	135
10	**21**	**39**	**49**	**78**	**94**	**73**	**105**	**114**
9	20	38	48	77	93	72	104	113
5	10	26	29	59	77	55	85	96
4	9	25	28	58	76	54	84	95
0	0	0	0	0	0	0	0	0
Oral Reading Fluency (ORF) - Accuracy								
67+	**87+**	**91+**	**92+**	**96+**	**96+**	**96+**	**96+**	**96+**
66	86	90	91	95	95	95	95	95
41	54	85	84	91	91	91	91	91
40	53	84	83	90	90	90	90	90
0	0	0	0	0	0	0	0	0

Note: B = beginning of the year, M = middle, E = end

Oral Reading Fluency Assessments: Grade 2

Steven was good at ping pong. He even got good enough to beat his big brother and his uncle. His family was full of good ping pong players. Steven tried his best with each game, but he was always a good sport when he lost a game. Mostly he just had fun playing his best.

Vowel Digraphs:
Potential Features Influencing Recognition

Word	Ability Percentile 1st recognized	*U* function (frequency)	Morpho-logical Family Members	In oral language (Age of acquisition)	Concreteness	Likelihood to appear across texts (Dispersion)
each	25P	1231		4.9	2.03	.97
beat	35P	53	*beating, beats, beaten, beater: U* = 38	6.2	3.97	.86

Consistency of Orthographic patterns: Critical in a Quasi-Orthographic Language

Word	Forward (Spelling-Sound)		Backward (Sound-Spelling)	
	Onset	rime	Onset	Rime
each	.98	1.0	1.0	.63
beat	.99	.74	1.0	.50

→ cells indicate consistency measu
ranging from 0 (very inconsistent) t
(very consistent) (Chee et al., 2020

		Spelling-Sound	Sound-Spelling
	Rimes--Friends	Rimes: enemies	Rimes--enemies
each	beach, leach, peach, **reach, teach,** bleach, breach, preach		leech, beech, breech, screech, speech
beat	**eat,** feat, **heat,** meat, neat, peat, seat, bleat, cheat, cleat, pleat, treat, wheat	great	beet**, feet, meet,** fleet, greet, sheet, skeet, sleet, **street,** sweet**,** tweet

Boldfaced words are predicted to appear 100+ times in a million words of text—i.e., among the 1,000 most frequent words.

Illustrations of Vowel Patterns: High- and Low-Frequency Words

Vowel Pattern	High Frequency	Low Frequency
short	big, set	ping, shed
long	time, see	vine, weed
diphthong	out, saw	sprout, paws
r-controlled	hers, first	soared, chore
variant	some, done	cheese, dance

Cox, O. and Briggs, D., 2023. Development of a Reading Foundational Skills Learning Progression. CU-Boulder, CADRE.

Human transparency ratings (Edwards et al., 2024)

Spelling-to-pronunciation (forward) transparency ratings

20,000 most frequent English words rated on the difficulty of mapping decoded pronunciation to actual pronunciation

6 point scale, 1 = very easy to match, 6 = very difficult

2,623 undergraduates as raters

40 linking items w/expert rating

250 words randomly assembled as a set

Each word rated by ~30 raters; their average = transparency rating

Transparency ratings (conti.)

word	rating	sd
men	1	0
van	1	0
so	1.03	0.18
bill	1.03	0.18
tap	1.03	0.19

word	rating	sd
bourgeoisie	5.50	1.10
bureau	5.44	0.96
pneumatic	5.42	0.97
xerox	5.35	1.38
xenophobia	5.31	1.29

Decoding System Measure (DSyM, Saha et al., 2020)

1) Word Frequency

How common is the word in text?

Common words (e.g., "the") get lower scores.

Rare words (e.g., "quinoa") get higher scores.

2) Letter-Sound Discrepancy

Is there a 1:1 mapping of letters to sounds?

Words with more letters than sounds (e.g., "night") get higher scores.

Calculated as |letters - phonemes|.

Word Frequency + **Letter-Sound Discrepancy** + **Number of Blends** + **Vowel GPC Complexity**

3) Number of Blends

How many consonant clusters does the word contain?

More blends (e.g., "str" in "string") equals a higher score.

Blends are challenging for beginning readers.

4) Vowel Grapheme-Phoneme Complexity

How predictable are the vowel sound relationships?

Vowels have more complexity than consonants (e.g., 'a' has 8 different sounds).

Rare vowel mappings yield higher scores than common ones.

Higher values indicate greater decoding difficulty

Position in text

Indexes the position of each word within each passage, which helps to look at position effects independent of intrinsic word features

Ecologically valid because reading happens in text, not with isolated word lists and this variable may:

- Capture reading fatigue/stamina effects as readers progress through text
- Reflect contextual buildup and cognitive load
- Show differential processing (e.g., enhanced focus on sentence-initial words
    - Eye-tracking studies show position-based attentional differences (Ashby et al., 2006)

Our sample: 1 to 103 words (Mean = 30.31, SD = 19.79)

Item position effect in psychometric research

$$logit[Y_{pik} = 1] = \theta_p - [\beta_i + \gamma(k - 1)]$$

- θ_p ability of person *p*
- β_i difficulty of item *i* when placed at the first position (k = 1)
- γ linear position effect
  - $\gamma > 0$ → learning/practice effect
  - $\gamma < 0$ → fatigue effect

Individual differences in the position effect can be examined by using γ_p

→ random weight linear logistic model (Rijman & DeBoeck, 2003)

→ with normal distribution assumption for γ_p , correlation between θ_p and γ_p can be estimated.

negative correlation → higher the ability, lower the position effect

(e.g., Debeer & Jansen, 2013)

Research Questions

RQ1. What is the difficulty of the words in untimed ORF passages for second graders?

RQ2. What are the features that influence word difficulty? In particular, do the following features have an explanatory power above and beyond more traditional features (e.g., frequency, length, AoA, concreteness) known to influence word recognition?

- word's position in a passage
- decodability, spelling↔sound consistency, spelling-sound transparency
- vowel patterns in the first syllable

AoA = age of acquisition (index for familiarity)

Computerized Oral Reading Evaluation (CORE)

Study Overview

- Large-scale project (Nese, 2022)
- Students read passages without typical one minute constraint
- Utilized Automatic Speech Recognition (ASR; Bavieca) to collect detailed reading data

Passage Characteristics

- Original works of fiction
- Target length ±5 words (medium = 50, long = 85)
- Consistent structure with beginning, middle, and end
- Corpus of 150 passages (50 each for Grades 2-4; 20 long & 30 medium)

Administration Procedures

1. Teacher administered:
    - Quick overview for students
    - Ensured microphones were positioned properly
    - Verified mute buttons were off and volume was up
    - Instructed students not to touch microphone while reading
2. Student process:
    - Students accessed study website
    - Received audio and text instructions
    - Randomly assigned to multiple passages (read an average of 8.40 passages, range 1-12)
    - Self-paced progression through assigned passages

Analytic sample:

Original dataset included 706 2nd grade students in winter of 2017-18 & 2018-19:

- at 7 public schools in 4 districts in 2 pacific northwest states (OR & WA)
- over 85% of students read 10 passages each
- a total of 71 passages were administered with a linking design

To reduce fatigue effect, for each student, we took up to top 5 passages in terms of words read correctly per minute.

Excluded:

- very high-frequent & 2-letter words & words everyone read correctly (e.g., "the", "to", "and", "an", “they”),
- proper nouns, contractions, possessives, interjections, hyphenated (“two-legged”)
- items with less than 30 students attempted,
- items that were missing word-feature variables,
- underfitting items from initial Rasch calibrations,
- outlier-students in terms of number of items attempted

Analytic sample: **650** students, **1,267** items (or word-tokens), **606** unique words

Analytic sample: students

Students (n=650)

- 51% Female
- 65% White, 25% Hispanic, 10% Other
- 75% Free/reduced lunch
- 11% Students w/ disabilities

- # of passages read:

n_psg	n	percent
5	290	44.6%
4	271	41.7%
3	84	12.9%
2	5	0.8%

- # of items attempted:

  range: 3 - 170, mean: 100 sd: 27.4

- Median WCPM (among 2-5 psgs):

  range: 2 - 186, mean: 68 sd: 43.40

- Median accuracy (among 2-5 psgs):

  range: 15% - 100%, mean: 80.63% sd: 24.02%

Analytic sample: passages & items (word-tokens)

Passages (n=50)

- Original passage length
  62% medium (46-57 words)
  38% long (69-103 words)
- # of target “items” (word-tokens) per passage
  range: 17- 48 mean: 25 sd: 7.5

Items (n = 1,267) & unique words (n = 606)

n.occurance	n	percent
1	368	61%
2	103	17%
3	45	7%
4	31	5%
5	16	3%
6+	43	7%
	606	100%

Sample passage
words in black counted as “item”

Steven was good at ping pong. He even got good enough to beat his big brother and his uncle. His family was full of good ping pong players. Steven tried his best with each game, but he was always a good sport when he lost a game. Mostly he just had fun playing his best.

Word / item characteristics

	mean	sd	min	median	max
position	20.17	18.57	0	28	102
length	5.24	1.55	3	5	11
u-function	396	695	1	153	7600
dispersion	0.83	0.14	0.25	0.87	1.00
AoA	5	1.32	2.22	4.78	9.58
concreteness	3.43	1.08	1.12	3.48	5.00
decodability[1]	2.45	1.12	0.28	2.29	6.41
n.phonemes	4	1	2	4	9
n.senses	10	10	0	7	60
ff onset[2]	0.91	0.17	0.00	0.97	1.00
ff rime[2]	0.76	0.25	0.01	0.84	1.00
fb onset[2]	0.86	0.21	0.00	0.95	1.00
fb rime[2]	0.57	0.29	0.00	0.59	1.00
transparency[3]	1.91	0.49	1.07	1.81	4.22

[1] DsyM (Decoding System Measure, Saha et al., 2014);

[2] Consistency measures (Chee et al., 2020); ff = feed-forward; fb = feed-backward

[3] Spelling-to-pronunciation transparency rating (Edwards et al., 2024)

	N = 606	
1st syllable vowel pattern		
diphthong	59	10%
long	175	29%
r-controlled	84	14%
short	248	41%
variant	40	7%
Part of Speech		
Adjective	72	12%
Adverb	54	9%
Function	37	6%
Noun	232	38%
Verb	211	35%
N_syllables		
1	361	60%
2	212	35%
3+	33	5%

Correlation matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	difficulty*	1.00														
2	position	0.18	*1.00*													
3	length	0.21	*0.02*	1.00												
4	u-function	-0.26	*0.00*	-0.23	1.00											
5	dispersion	-0.20	*0.04*	*-0.08*	0.34	1.00										
6	AoA	0.23	*0.05*	0.25	-0.14	*0.05*	1.00									
7	concreteness	0.11	*-0.04*	*-0.06*	-0.29	-0.37	-0.25	1.00								
8	decodability[1]	0.17	*0.02*	0.84	-0.13	*0.00*	0.20	-0.09	1.00							
9	n.phonemes	0.20	*0.03*	0.86	-0.25	-0.09	0.28	*-0.06*	0.64	1.00						
10	n.senses	-0.09	*0.04*	-0.20	*0.03*	*0.24*	-0.15	*0.03*	-0.19	-0.20	1.00					
11	ff onset[2]	*-0.01*	*0.03*	-0.08	-0.17	-0.10	*-0.06*	0.10	*-0.08*	*-0.07*	0.16	1.00				
12	ff rime[2]	*-0.03*	*0.02*	-0.11	-0.19	*-0.07*	*-0.06*	0.14	-0.16	-0.14	0.19	0.09	1.00			
13	fb onset[2]	*-0.02*	*0.04*	-0.17	*0.03*	0.04	*-0.03*	*-0.01*	-0.14	-0.14	0.11	0.39	0.09	1.00		
14	fb rime[2]	*-0.07*	*-0.04*	*0.03*	-0.13	-0.09	*-0.02*	*0.07*	-0.10	0.18	0.08	*0.04*	0.31	*0.00*	1.00	
15	transparency[3]	0.20	*0.04*	0.36	*-0.01*	0.03	0.15	*-0.08*	0.42	0.10	-0.21	-0.26	-0.16	-0.18	-0.32	1.00

Note. * estimated from the Rasch model; Italisized = not significant at $p = .05$; [1, 2, 3] see the previous table

Modeling approach

Q1. Rasch model (with TAM package)

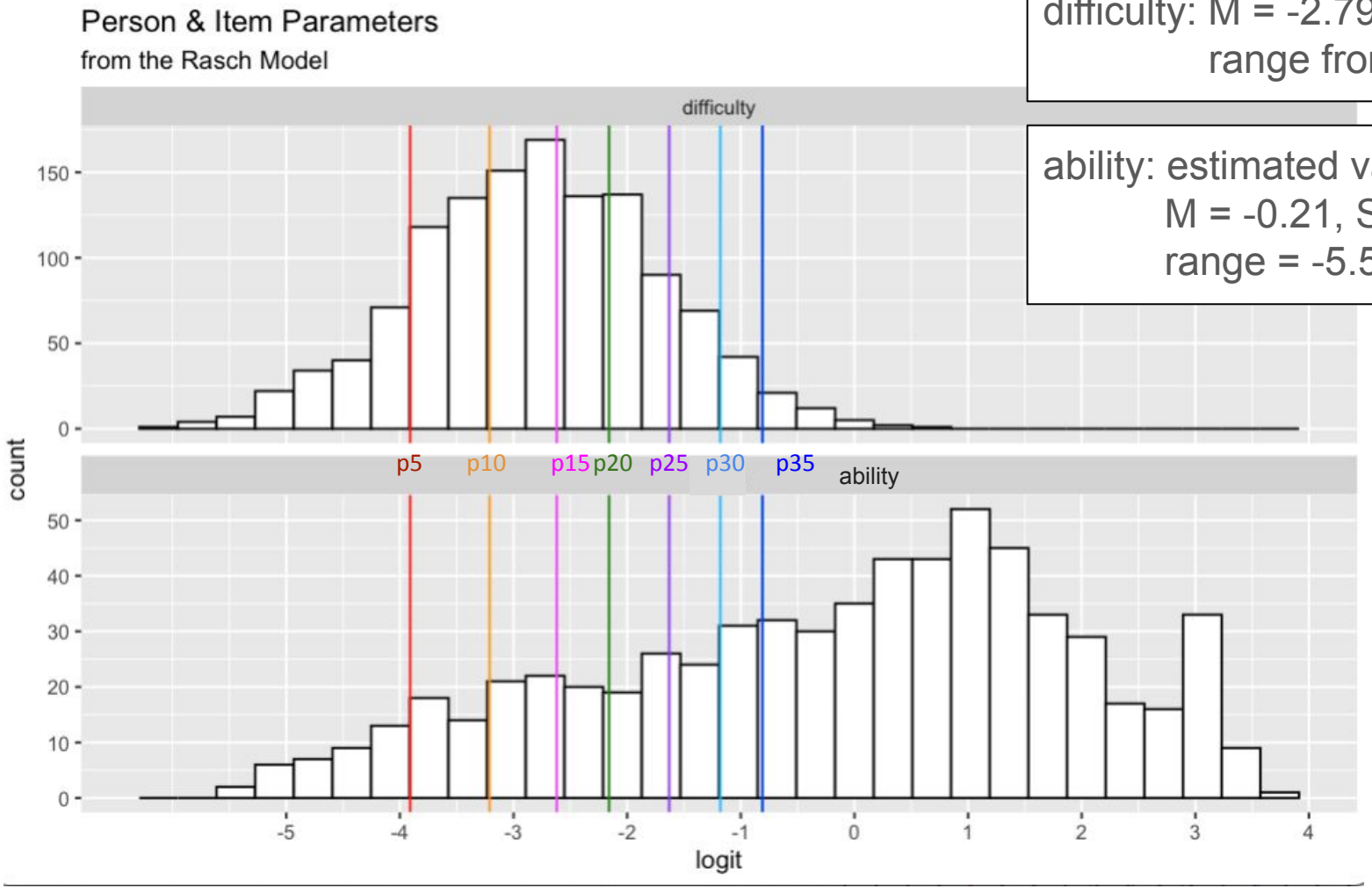
Q2. LLTMe (Janssen et al., 2004) plus random effect for passages (with lme4 package)

$$\text{logit}\,(x_{ri} = 1\,|\theta_r) = \theta_r - \delta_i$$

$$\delta_i = \sum_{k=1}^{K} \beta_i X_k + \gamma_{p(i)} + \varepsilon_i$$

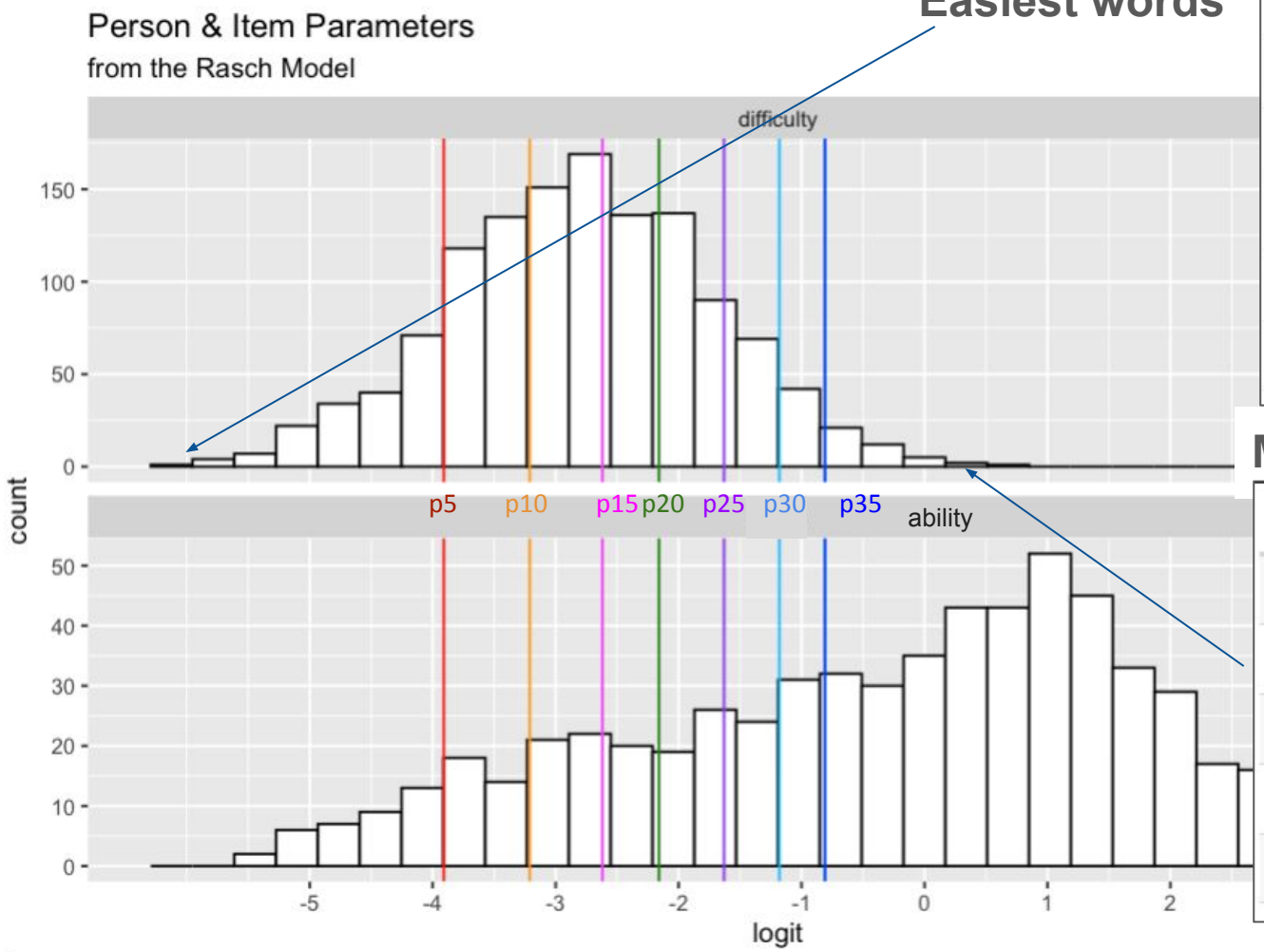
Indices:
r = respondent
i = item / word
k = item/ word feature
p = passage

Results from the Rasch Model



difficulty: M = -2.79, SD = 1.07
range from -6.29 to 0.68

ability: estimated variance = 5.1
M = -0.21, SD = 2.1
range = -5.53 to 3.58



Easiest words

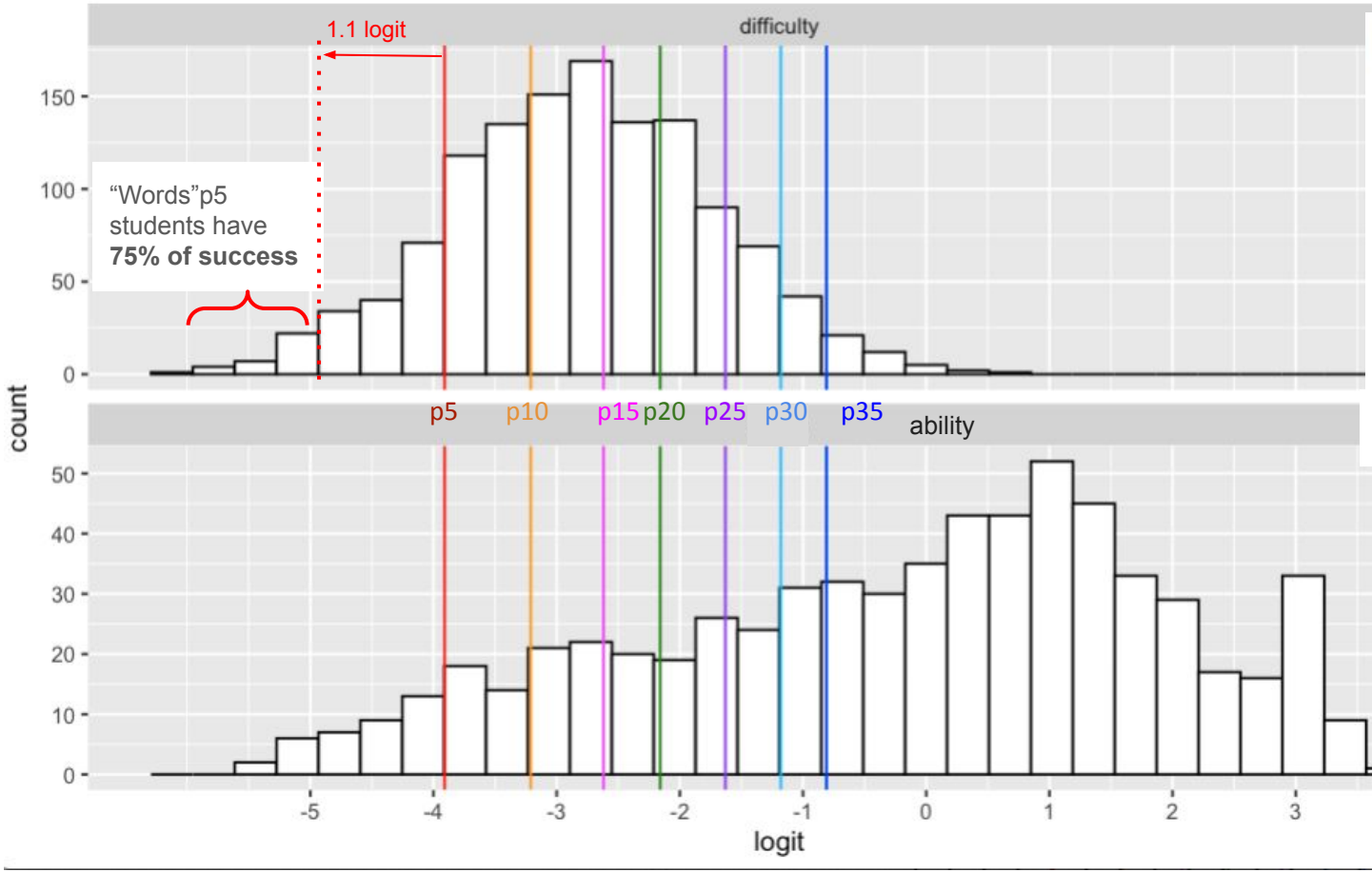
word	difficulty	position
good	-6.29	2
room	-5.93	14
good	-5.88	24
fish	-5.84	1
liked	-5.73	23

Most difficult words

word	difficulty	position
secrets	0.68	102
soon	0.28	48
eager	0.26	26
bean	0.15	30
compromise	0.09	39

Person & Item Parameters
from the Rasch Model

ability pctl	#.words	M.position	M.ufunc	M.disp	M.aoa	M.deco-dability	M.length	M.#. sense	M.ff_ rime	M.fb_ rime	M_trans parency
p05	19	19.58	1,164.05	0.89	4.01	1.70	4.05	11.47	0.67	0.63	1.84



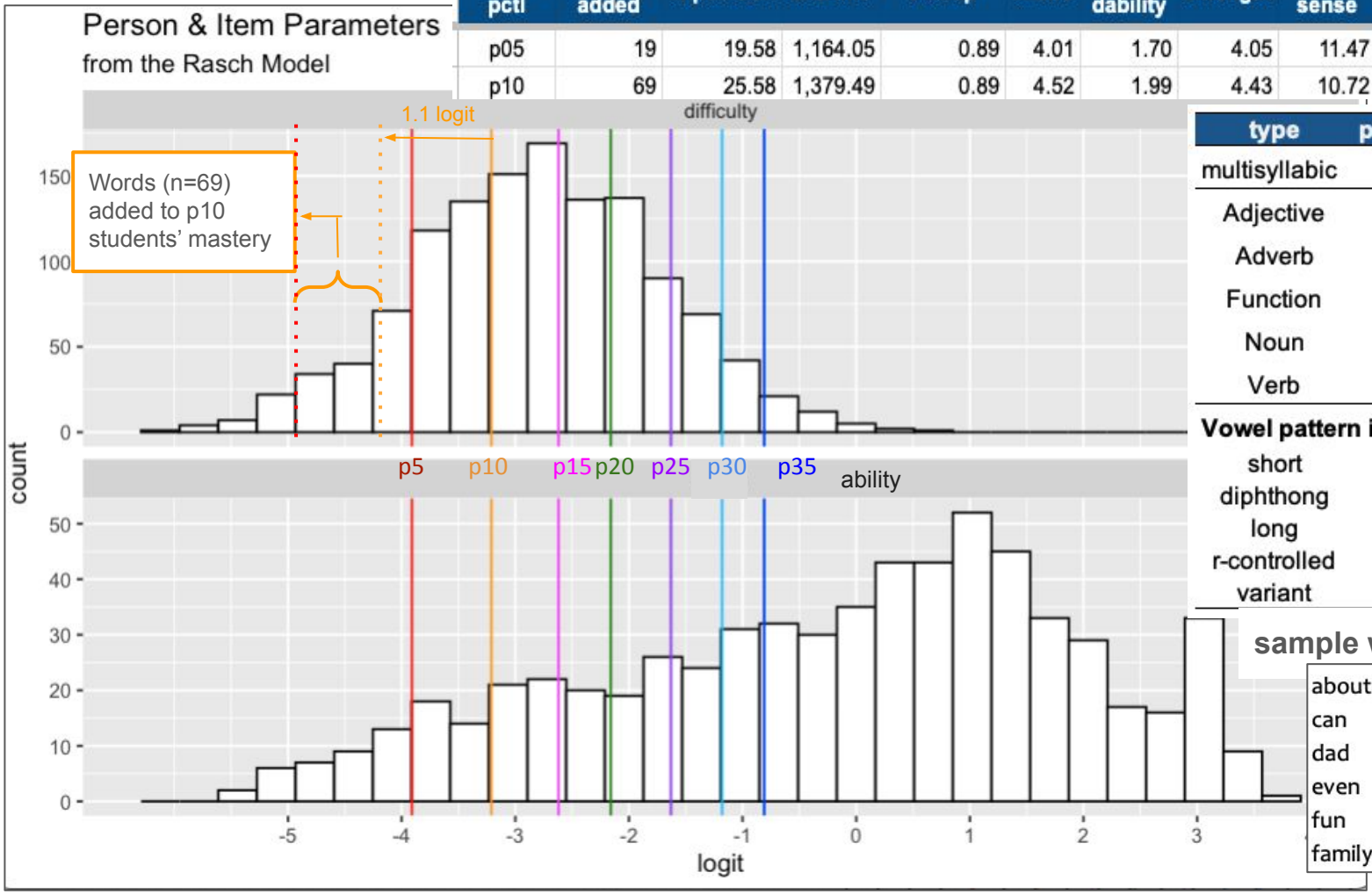
type	n	%
monosyllabic	16	84%
Adjective	4	21%
Adverb	5	26%
Noun	5	26%
Verb	5	26%
Vowel pattern in first syllable		
short	8	42%
diphthong	5	26%
long	4	21%
variant	2	11%

words (n=19)

end	how (2)	right
family	knew	room
fish	know	too
good (3)	liked	wanted
happy	mom	
have	not	

Note. "how" and "good" were double- and triple-counted as each occurrence in passage counted an different "item"

ability pctl	#.words added	M.position	M.ufunc	M.disp	M.aoa	M.deco-dability	M.length	M.#. sense	M.ff_ rime	M.fb_ rime	M_trans parency
p05	19	19.58	1,164.05	0.89	4.01	1.70	4.05	11.47	0.67	0.63	1.84
p10	69	25.58	1,379.49	0.89	4.52	1.99	4.43	10.72	0.72	0.61	1.83



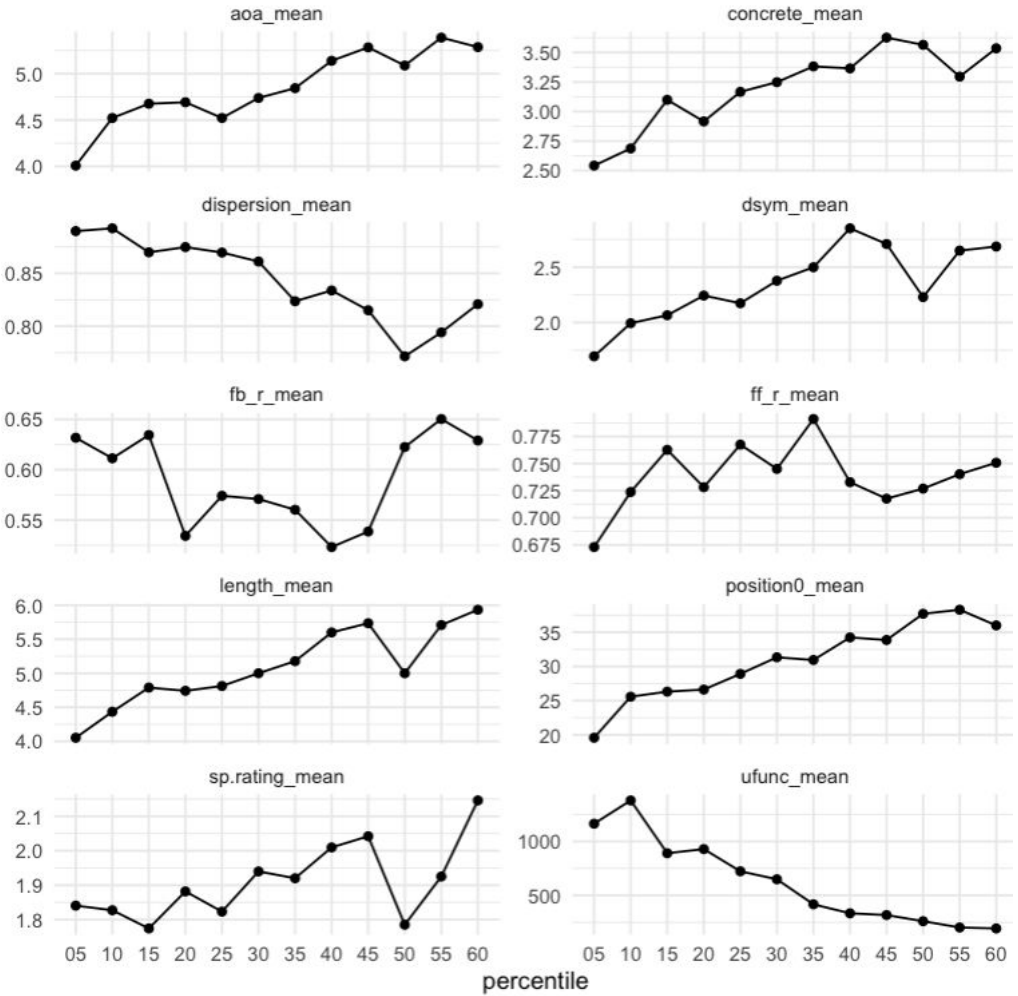
type	p5 (n=19)	p10 (n=69)
multisyllabic	16%	30%
Adjective	21%	13%
Adverb	26%	17%
Function	0%	16%
Noun	26%	20%
Verb	26%	33%
Vowel pattern in first syllable		
short	42%	51%
diphthong	26%	10%
long	21%	23%
r-controlled	0%	10%
variant	11%	6%

sample words p10

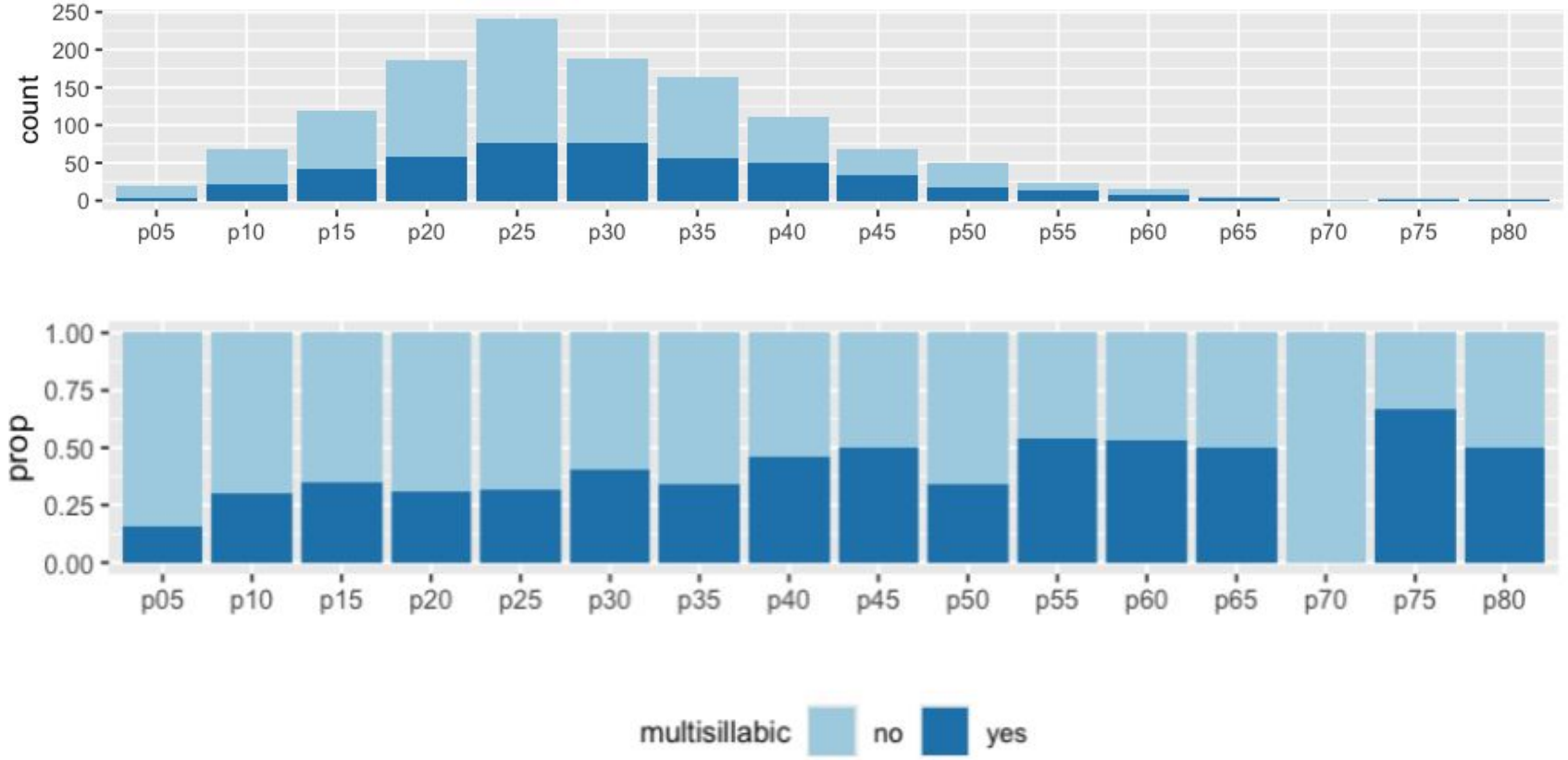
about	first	learned
can	good	new
dad	going	started
even	great	time
fun	into	too
family	just	would

Features of words added to mastery at different ability percentiles

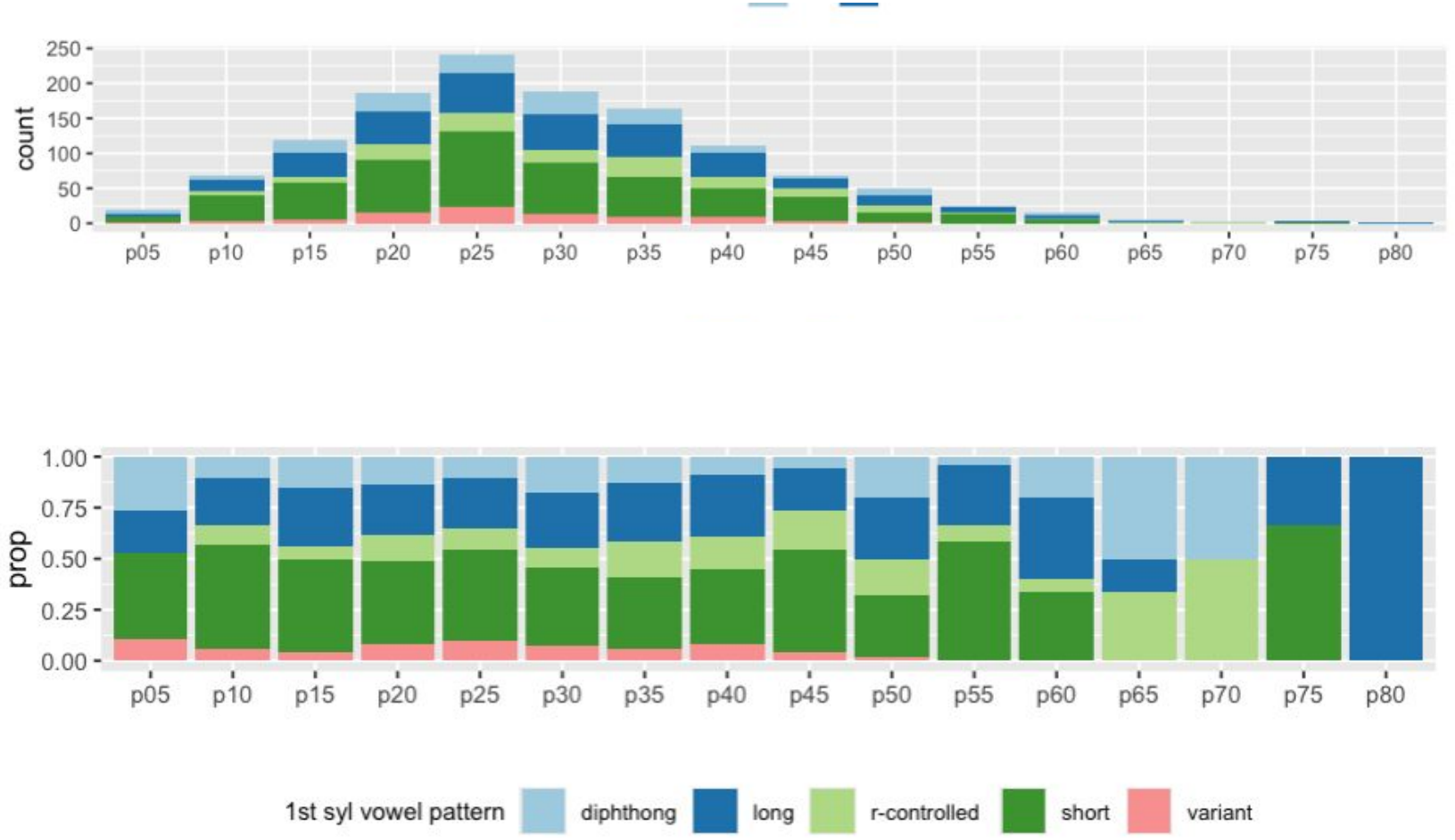
ability pctl	#.words added
p05	19
p10	69
p15	119
p20	186
p25	241
p30	189
p35	163
p40	110
p45	68
p50	50
p55	24
p60	15
p65	6
p70	2
p75	3
p80	2
TOTAL	1,266



Features of words added to mastery at different ability percentiles (conti.)



Features of words added to mastery at different ability percentiles (conti.)



Results from explanatory item response models

Identifying the base model

	Null		Rasch.rirp†		Null.tlt^		Rasch.rirp.tlt	
Fixed Effects								
(Intercept)	2.414	***	2.633	***	2.411	***	2.628	***
Random Effects								
Var(item)			0.700				0.666	
Var(student)	3.938		4.683		3.969		4.689	
Var(passage)					0.058		0.046	
BIC	38695.69		36855.60		38553.85		**36850.60**	
AIC	38677.558		36828.403		38526.656		**36814.34**	
log-Likelihood	-19336.779		-18411.202		-19260.328		**-18403.17**	

Note. † rirp = random item random person; ^ tlt = testlet

Model Fit

		# para	BIC	AIC	log-Likelihood	R^2	ΔR^2		LR-test
m0	Rasch.rr.tlt	3	36850.60	36814.34	-18403.17	0.0000			
m1	ctrl vars only	10	36636.42	36545.77	-18262.89	0.2943	0.294		
m2	m1 + position	11	36586	36486.29	-18232.14	0.3498	0.056	vs. m1	***
m3	m2 + consistency (ff/fb)	15	36618.1	36482.12	-18226.06	0.3604	0.011	vs. m2	*
m4	**m2 + transp.rating**	12	**36579.83**	36471.05	-18223.53	0.3664	0.017	vs. m2	***
m5	m2 + decoding	12	36588.26	36479.48	-18227.74	0.3574	0.008	vs. m2	**
m9	**m2 + consis + transp.r + decode**	17	36624.26	36470.16	**-18218.08**	**0.3739**	0.024	vs. m2	***
m9	**m2 + consis + transp.r + decode**	17	36624.26	36470.16	**-18218.08**	**0.3739**	0.008	vs. m4	NS^

Note. * $p<0.05$ ** $p<0.01$ *** $p<0.001$ ^ $p=0.053$

Parameter Estimates: models examining consistency / decoding measures

	M0 Rasch.rr.tlt	M1 ctrl vars only	M2 m1 +position	M3 m1+ consistency	M4 m1+ transp.rating	M5 m1+ decoding	M9 m1+consis+ trans.r+ decode
	Est.	*Est.*	*Est.*	*Est.*	*Est.*	*Est.*	*Est.*
Fixed Effects							
(Intercept)	2.63	2.557 ***	2.553 ***	2.557 ***	2.555 ***	2.559 ***	2.558 ***
positionZ			-0.197 ***	-0.194 ***	-0.198 ***	-0.198 ***	-0.196 ***
lengthZ		-0.141 ***	-0.146 ***	-0.147 ***	-0.103 ***	-0.03	-0.044
ufuncZ		0.213 ***	0.216 ***	0.239 ***	0.233 ***	0.226 ***	0.251 ***
dispZ		0.093 ***	0.096 ***	0.106 ***	0.100 ***	0.106 ***	0.110 ***
aoaZ		-0.221 ***	-0.213 ***	-0.204 ***	-0.207 ***	-0.211 ***	-0.203 ***
concZ		-0.207 ***	-0.206 ***	-0.202 ***	-0.205 ***	-0.203 ***	-0.205 ***
noun bi [yes]		0.253 ***	0.257 ***	0.246 ***	0.251 ***	0.239 ***	0.240 ***
consistency							
forward+ onset				0.058 *			0.039
forward+ rime				0.009			0.012
backward^ onset				-0.055 *			-0.068 *
backward^ rime				0.051			0.008
transparency rating					-0.111 ***		-0.097 **
decoding demand						-0.136 **	-0.082
Random Effects							
Var(item)	0.666	0.47	0.433	0.426	0.422	0.428	0.417
Var(student)	4.689	4.692	4.697	4.695	4.695	4.696	4.694
Var(passage)	0.046	0.047	0.044	0.043	0.043	0.045	0.043
R^2		29.43%	34.98%	36.04%	36.64%	35.74%	**37.39%**

Note. + forward = from spelling to sound; ^ backward = from sound to spelling; * $p<0.05$ ** $p<0.01$ *** $p<0.001$
Item easiness is modeled so negative coefficient means the predictor makes the word reading more difficult.

Model Fit

		# para	BIC	AIC	log-Likelihood	R^2	ΔR^2		LR-test
m0	Rasch.rr.tlt	4	36850.60	36814.34	-18403.17	0.0000			
m1	ctrl vars only	11	36636.42	36545.77	-18262.89	0.2943	0.294		
m2	m1 + position	12	36586.00	36486.29	-18232.14	0.3498	0.056	vs. m1	***
m3	m2 + consistency (ff/fb)	16	36618.10	36482.12	-18226.06	0.3604	0.011	vs. m2	*
m4	**m2 + transp.rating**	13	**36579.83**	36471.05	-18223.53	0.3664	0.017	vs. m2	***
m5	m2 + decoding	13	36588.26	36479.48	-18227.74	0.3574	0.008	vs. m2	**
m9	**m2 + consis + transp.r + decode**	18	36624.26	36470.16	-18218.08	0.3739	0.024	vs. m2	***
m9	**m2 + consis + transp.r + decode**	18	36624.26	36470.16	-18218.08	0.3739	0.018	vs. m4	NS^
m10	**m4 + 1st syllable vowel patterns**	16	36604.41	36459.36	-18213.68	**0.3844**	0.018	vs. m4	***
m11	**m4 + vow.pat x freq**	20	36647.72	36466.42	-18213.21	0.3844	0.018	vs. m4	**
m13	**m4 + vow.pat x position**	20	36640.56	**36459.25**	-18209.63	0.3919	0.026	vs. m4	***
m14	**m4+ vow.pat x freq x position**	29	36736.93	36474.04	**-18208.02**	**0.3930**	0.027	vs. m4	*
m14	**m4+ vow.pat x freq x position**	29	36736.93	36474.04	**-18208.02**	**0.3930**	0.036	vs. m13	NS

Note. * $p<0.05$ ** $p<0.01$ *** $p<0.001$ ^ $p=0.053$

Parameter Estimates:
models vowel patterns in 1st syllable & interactions w/ frequency & position

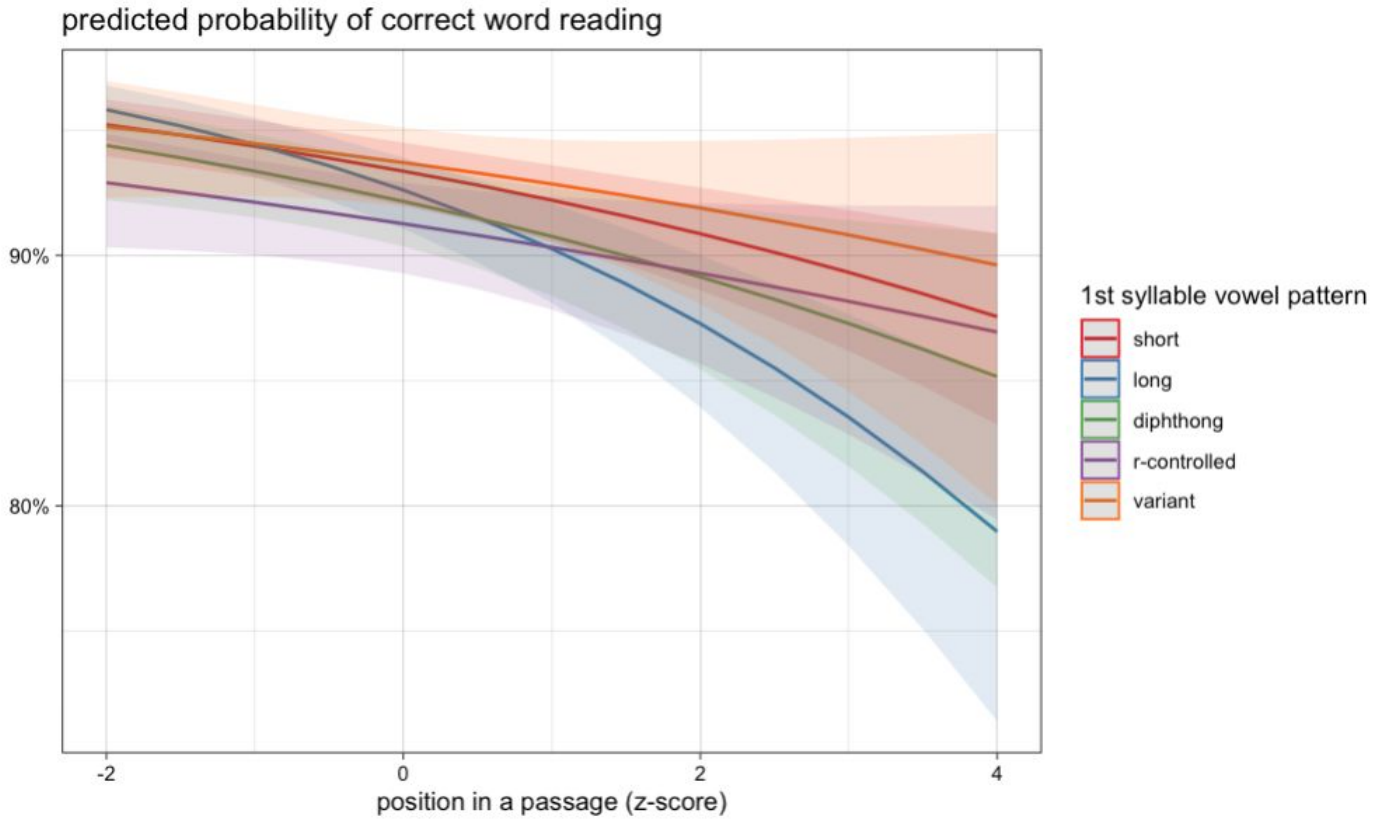
→ short-vowel served as a reference category

Vowel Pattern	High Frequency	Low Frequency
short	big, set	ping, shed
long	time, see	vine, weed
diphthong	out, saw	sprout, paws
r-controlled	hers, first	soared, chore
variant	some, done	cheese, dance

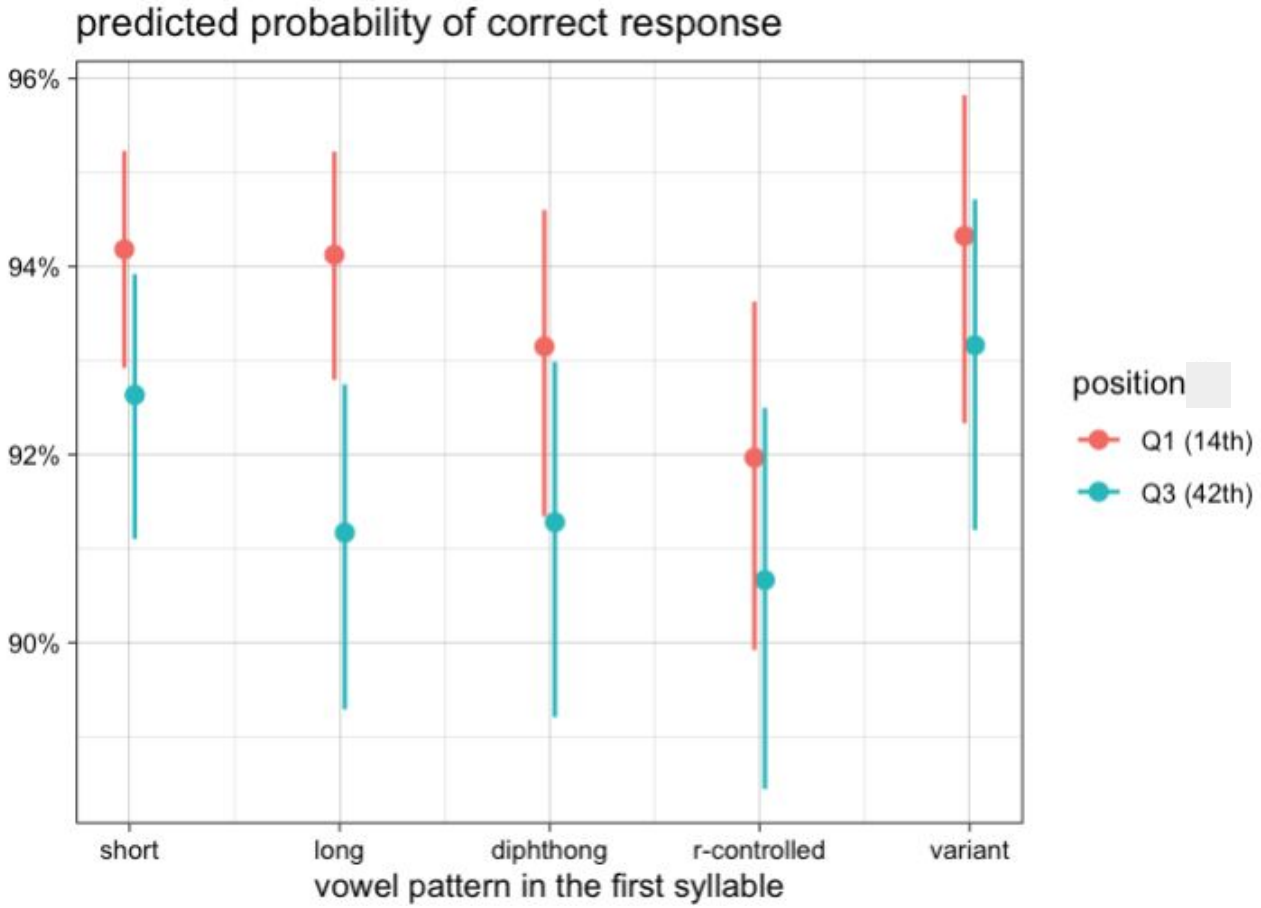
	M10 vwl.pat. main effect	M11 vwl.pat x freq	M12 vwl.pat x position	M13 vwl.pat x freq x position
	Est.	*Est.*	*Est.*	*Est.*
(Intercept)	2.648 ***	2.649 ***	2.642 ***	2.642 ***
positionZ	-0.198 ***	-0.198 ***	-0.173 ***	-0.174 ***
lengthZ	-0.111 ***	-0.113 ***	-0.109 ***	-0.109 ***
ufuncZ	0.232 ***	0.217 ***	0.231 ***	0.219 ***
dispZ	0.106 ***	0.106 ***	0.107 ***	0.106 ***
aoaZ	-0.205 ***	-0.204 ***	-0.206 ***	-0.206 ***
concZ	-0.193 ***	-0.191 ***	-0.196 ***	-0.196 ***
noun	0.234 ***	0.231 ***	0.246 ***	0.246 ***
spelling-sound consistency (human	-0.099 ***	-0.095 ***	-0.102 ***	-0.098 ***
[long]	-0.12	-0.119	-0.115	-0.114
[diphthong]	-0.181 *	-0.186 *	-0.180*	-0.188 *
[r-controlled]	-0.295 ***	0.297 ***	-0.297 ***	-0.300 ***
[variant]	0.066	0.065	0.057	0.06
[long] × ufunc		0.041		0.037
[diphthong] × ufunc		0.058		0.07
[r-controlled] × ufunc		-0.005		-0.002
[variant] × ufunc		-0.042		-0.035
[long] × position			-0.128 *	-0.129 *
[diphthong] × position			-0.006	-0.007
[r-controlled] × position			0.06	0.057
[variant] × position			0.037	0.05
position x ufuncZ				-0.01
(positionZ × ufuncZ) × [long]				-0.007
(positionZ × ufuncZ) × [diphthong]				-0.122
(positionZ × ufuncZ) × [r-controlled]				-0.013
(positionZ × ufuncZ) × [variant]				-0.08
Random Effects				
variance(items)	0.410	0.410	0.405	0.404
variance(students)	4.694	4.694	4.693	4.692
variance(passages)	0.042	0.041	0.042	0.041
R^2	0.384	0.384	0.392	**0.393**

*$p<0.05$ **$p<0.01$ ***$p<0.001$

Vowel pattern x position interaction



Vowel pattern x position interaction

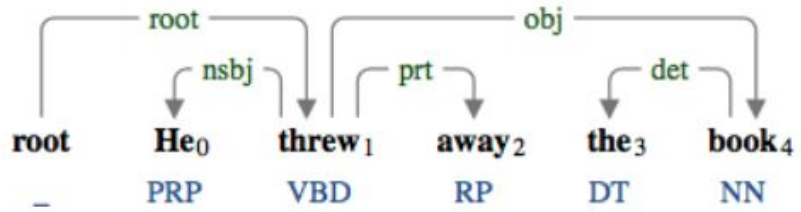


Summary

- Words in passages varied substantially in their difficulty (~6 logits!).
- The great majority of words were located bottom 35% of the ability distribution, which helps us understand which words these students can read and which they have difficulty with.
- Well-studied features known to affect word reading on list (e.g., frequency, AoA, dispersion) collectively accounted for ~30% of variance in difficulty of words in connected text.
- Word's position in passage had a large and significant effect; the variance explained increased ~5% points.
- Human judgements on spelling to sound transparency (Edwards et al., 2024) had a slightly larger and unique explanatory power than the measures based solely on word properties.
- Vowel patterns in the first syllable had small & unique effects on difficulty with r-controlled and diphthong being more difficult than short-vowel.
- Differentially larger impact of position was observed for words with long-vowel.

Next Steps

- The final model (vowel patterns x position) explained ~ 40% of variance in difficulty, which is typical of this type of study.
- Word-chunks, sentence- and text-level features should be brought in as additional predictors.
    - Kara et al (2023) used NLP to extract word-chunks in identifying meaningful pauses captured in CORE that predicted student's word read correctly per minute.



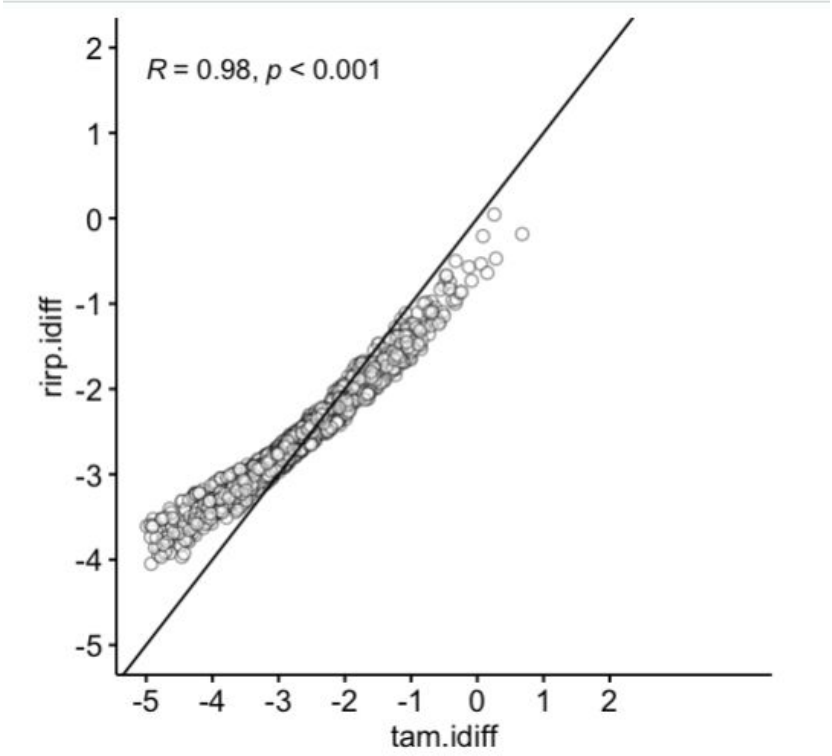
Next Steps (conti.)

- Differential impact of position among students and passages should be investigated.
  - With random person- or passage-weights on the position facet (Rijmen F, De Boeck P, 2002)
  - Different modeling options appear to be available for learning/practice & fatigue.
- Oral reading fluency data provide stimulating challenges for modeling:
  - Multiple occurrences of “items” (words) within/between passages.
  - letters > words > sentences > paragraphs > passages
  - CORE provides response time (RT), which captures pause between words.
- CORE data provides word reading data for grades 2-4 at 4 time points (students can be linked within the same grade).
  - Caveat: passage order within administration is not available.

Thank you!
ありがとう

Additional slides

Item difficulty estimates from TAM & lme4



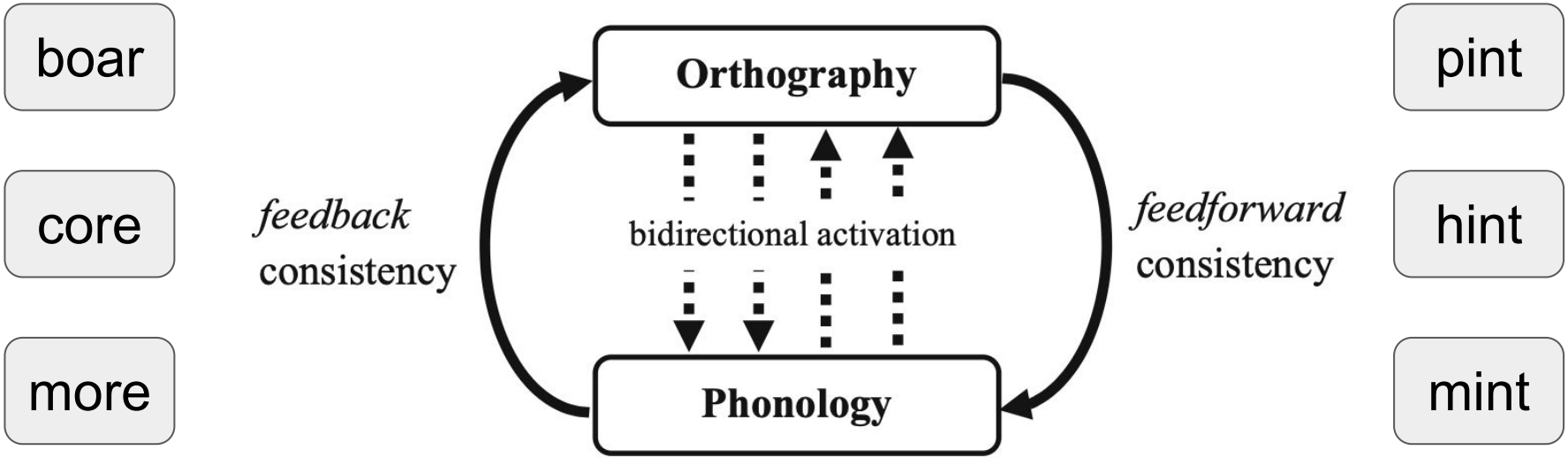
TAM: fixed effect for items
lme4: random effect for items

```
(1|id_item)
```

Highly correlated but why there is a curvature at the left bottom corner?

Consistency Measures

Spelling-sound (feedforward) & sound-spelling (feedbackward) consistency (Chee et al., 2020)



Consistency Measure

moire/once/one/ones

(feedforward onset)

drowsed/leagued/flyer

(feedforward rime)

bairn/dearth/does

(feedbackward onset)

leagued/drowsed/ewe/aye/coup/

(feedbackward rime)

drudgery/robbery/rubbery
(feedforward onset)

frankfurter/skateboarder/stepdaughter

(feedforward rime)

abstract/amide/anthrax/asphalt

(feedbackward onset)

addicts/adjunct/aspects/challenged/massive

(feedbackward rime)