

SUPPLEMENTARY MATERIAL: NORMATIVE DATA (LONG-FORM ASSESSMENTS)

This document provides supplementary material for “Validation of two measures for assessing English vocabulary knowledge on web-based testing platforms: Long-form assessments,” including detailed demographic reporting, individual subject performance, and normative values for the web-based versions of the Vocabulary Size Test (VST) and Word Familiarity Test (WordFAM). The list of contents is shown below.

Tables

Table S1. Count of self-reported gender and age characteristics of the participant sample.

Table S2. Count of self-reported dialect characteristics of the participant sample.

Table S3. Count of self-reported race and ethnicity characteristics of the participant sample.

Table S4. Mean proportion correct (and *SD*) and point-biserial correlation coefficient (*r*) across the 100 participants for each item on the VST.

Table S5. Mean rating (and *SD*) and correlation coefficient (*r*) across the 100 participants for each item on the WordFAM.

Figures

Figure S1. For each subject, mean accuracy (proportion correct) by frequency bin on the long-form VST.

Figure S2. For each subject, mean familiarity rating by frequency bin on the long-form WordFAM.

Table S1. Count of self-reported gender and age characteristics of the participant sample.

Assessment	<i>n</i>	Gender			Age (years)		
		Men	Women	Unreported	<i>Mean</i>	<i>SD</i>	<i>Range</i>
VST	100	47	53	0	25	6	18 – 35
WordFAM	100	48	51	1	27	5	18 – 35

[Continued on next page.]

Table S2. Count of self-reported dialect characteristics of the participant sample.

Dialect	VST	WordFAM
American	—	1
Decline to state	2	1
Do not know	11	17
Do not know, Mid Atlantic	1	—
Do not know, Mid Atlantic (Pennsylvania)	1	—
Do not know, New England	1	—
Do not know, Southern, Southwestern	—	1
Eastern/Pittsburgh	1	—
Inland Northern	1	—
Mid Atlantic	3	2
Midwestern	27	22
Midwestern, Pacific Southwest	1	—
Midwestern, Southwestern	1	—
New England	10	13
New England, Midwestern	1	—
New England, New York	1	—
New York	1	1
Northeastern	1	2
Northeastern (Philadelphia)	1	—
Pacific Northwest	5	9
Pacific Southwest	7	10
Southern	12	16
Southern Californian	1	—
Southern, Midwestern	2	—
Southern, Southwestern	—	—
Southern, Texan	1	—
Southwestern	5	4
Southwestern, Texan, Non-regional Diction	1	—
Spanglish	1	—
Tristate	—	1

Table S3. Count of self-reported race and ethnicity characteristics of the participant sample.

Assessment	Race	Ethnicity		
		Hispanic or Latino	Not Hispanic or Latino	Unreported
VST	American Indian/Alaska Native	–	1	–
	American Indian/Alaska Native, White, More than One Race	–	1	–
	Asian	–	6	–
	Asian, Native Hawaiian or Other Pacific Islander, White	–	1	–
	Asian, White	–	2	–
	Black or African American	–	7	–
	Black or African American, White	1	–	–
	Black or African American, White, More than One Race	–	1	–
	Decline to state	–	–	2
	More than One Race	1	1	–
	White	3	73	–
WordFAM	American Indian/Alaska Native, White	–	1	–
	Asian	–	11	–
	Asian, White, More than One Race	–	3	–
	Black or African American	–	8	–
	Black or African American, White	–	2	–
	More than One Race	1	–	–
	White	1	72	1

[Continued on next page.]

Table S4. Mean proportion correct (and *SD*) and point-biserial correlation coefficient (*r*) across the 100 participants for each item on the VST. NA values for *r* indicate items for which accuracy was at ceiling for all participants. Number and group correspond to those used in the original VST Form A assessment, available at: <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests/the-vocabulary-size-test/VST-version-A.pdf>. Bin and version correspond to the divisions made in the current work as described in the main text.

Number	Group	Bin	Version	Item	Mean	SD	<i>r</i>
1	G.01000	High	A	see	1.00	0.00	NA
2	G.01000	High	B	time	1.00	0.00	NA
3	G.01000	High	A	period	0.99	0.10	0.02
4	G.01000	High	B	figure	0.85	0.36	0.18
5	G.01000	High	A	poor	1.00	0.00	NA
6	G.02000	High	B	microphone	0.92	0.27	0.22
7	G.02000	High	A	nil	0.94	0.24	0.19
8	G.02000	High	B	pub	1.00	0.00	NA
9	G.02000	High	A	circle	0.95	0.22	0.11
10	G.02000	High	B	dig	1.00	0.00	NA
11	G.03000	High	A	soldier	1.00	0.00	NA
12	G.03000	High	B	restore	0.99	0.10	0.12
13	G.03000	High	A	pro	0.97	0.17	0.10
14	G.03000	High	B	compound	0.94	0.24	0.06
15	G.03000	High	A	deficit	0.89	0.31	0.29
16	G.04000	High	B	strap	0.97	0.17	0.15
17	G.04000	High	A	weep	0.99	0.10	0.06
18	G.04000	High	B	haunt	0.99	0.10	0.11
19	G.04000	High	A	cube	1.00	0.00	NA
20	G.04000	High	B	butler	0.99	0.10	0.09
21	G.05000	High	A	nun	1.00	0.00	NA
22	G.05000	High	B	olive	1.00	0.00	NA
23	G.05000	High	A	shudder	0.98	0.14	0.25
24	G.05000	High	B	threshold	0.97	0.17	-0.02
25	G.05000	High	A	demography	0.86	0.35	0.33
26	G.06000	Mid-High	B	malign	0.85	0.36	0.21
27	G.06000	Mid-High	A	strangle	0.99	0.10	0.03
28	G.06000	Mid-High	B	dinosaur	0.98	0.14	0.05
29	G.06000	Mid-High	A	jug	1.00	0.00	NA
30	G.06000	Mid-High	B	crab	0.98	0.14	0.09
31	G.07000	Mid-High	A	quilt	0.99	0.10	0.07
32	G.07000	Mid-High	B	tummy	1.00	0.00	NA
33	G.07000	Mid-High	A	eclipse	1.00	0.00	NA
34	G.07000	Mid-High	B	excrete	0.96	0.20	0.05

35	G.07000	Mid-High	A	ubiquitous	0.61	0.49	0.41
36	G.08000	Mid-High	B	marrow	1.00	0.00	NA
37	G.08000	Mid-High	A	cabaret	0.92	0.27	0.26
38	G.08000	Mid-High	B	cavalier	0.49	0.50	0.35
39	G.08000	Mid-High	A	veer	0.74	0.44	0.30
40	G.08000	Mid-High	B	yogurt	1.00	0.00	NA
41	G.09000	Mid-High	A	octopus	1.00	0.00	NA
42	G.09000	Mid-High	B	monologue	0.96	0.20	0.32
43	G.09000	Mid-High	A	candid	0.83	0.38	0.32
44	G.09000	Mid-High	B	nozzle	0.98	0.14	0.12
45	G.09000	Mid-High	A	psychosis	0.97	0.17	0.09
46	G.10000	Mid-High	B	ruck	0.21	0.41	0.13
47	G.10000	Mid-High	A	rouble	0.63	0.49	0.19
48	G.10000	Mid-High	B	canonical	0.65	0.48	0.21
49	G.10000	Mid-High	A	puree	0.97	0.17	0.24
50	G.10000	Mid-High	B	vial	1.00	0.00	NA
51	G.11000	Mid-Low	A	counterclaim	0.95	0.22	0.30
52	G.11000	Mid-Low	B	refectory	0.31	0.46	0.12
53	G.11000	Mid-Low	A	trill	0.70	0.46	0.40
54	G.11000	Mid-Low	B	talon	0.99	0.10	0.22
55	G.11000	Mid-Low	A	plankton	0.98	0.14	0.17
56	G.12000	Mid-Low	B	soliloquy	0.75	0.44	0.26
57	G.12000	Mid-Low	A	puma	0.98	0.14	0.10
58	G.12000	Mid-Low	B	augur	0.32	0.47	0.35
59	G.12000	Mid-Low	A	emir	0.34	0.48	0.40
60	G.12000	Mid-Low	B	didactic	0.43	0.50	0.45
61	G.13000	Mid-Low	A	cranny	0.75	0.44	0.18
62	G.13000	Mid-Low	B	lectern	0.49	0.50	0.13
63	G.13000	Mid-Low	A	azalea	0.91	0.29	0.32
64	G.13000	Mid-Low	B	marsupial	0.87	0.34	0.36
65	G.13000	Mid-Low	A	bawdy	0.71	0.46	0.43
66	G.14000	Mid-Low	B	crowbar	0.97	0.17	0.14
67	G.14000	Mid-Low	A	spangled	0.76	0.43	0.34
68	G.14000	Mid-Low	B	aver	0.47	0.50	0.17
69	G.14000	Mid-Low	A	retro	0.88	0.33	0.30
70	G.14000	Mid-Low	B	rascal	0.99	0.10	0.22
71	G.15000	Mid-Low	A	tweezers	0.93	0.26	0.16
72	G.15000	Mid-Low	B	bidet	0.97	0.17	0.21
73	G.15000	Mid-Low	A	sloop	0.56	0.50	0.42

74	G.15000	Mid-Low	B	swingeing	0.36	0.48	0.04
75	G.15000	Mid-Low	A	cenotaph	0.39	0.49	0.38
76	G.16000	Low	B	denouement	0.50	0.50	0.15
77	G.16000	Low	A	bittern	0.48	0.50	0.19
78	G.16000	Low	B	reconnoitre	0.37	0.49	0.28
79	G.16000	Low	A	magnanimity	0.54	0.50	0.30
80	G.16000	Low	B	effete	0.75	0.44	0.32
81	G.17000	Low	A	rollick	0.87	0.34	0.33
82	G.17000	Low	B	gobbet	0.59	0.49	0.29
83	G.17000	Low	A	rigmarole	0.76	0.43	0.37
84	G.17000	Low	B	alimony	0.88	0.33	0.34
85	G.17000	Low	A	roughshod	0.14	0.35	0.27
86	G.18000	Low	B	copra	0.41	0.49	0.26
87	G.18000	Low	A	bier	0.36	0.48	0.27
88	G.18000	Low	B	torpid	0.41	0.49	0.40
89	G.18000	Low	A	dachshund	0.96	0.20	0.21
90	G.18000	Low	B	cadenza	0.51	0.50	0.32
91	G.19000	Low	A	obtrude	0.59	0.49	0.27
92	G.19000	Low	B	panzer	0.47	0.50	0.18
93	G.19000	Low	A	cyborg	1.00	0.00	NA
94	G.19000	Low	B	zygote	0.91	0.29	0.25
95	G.19000	Low	A	sylvan	0.50	0.50	0.37
96	G.20000	Low	B	sagacious	0.32	0.47	0.53
97	G.20000	Low	A	spatiotemporal	0.86	0.35	0.26
98	G.20000	Low	B	casuist	0.33	0.47	0.29
99	G.20000	Low	A	cyberpunk	0.88	0.33	0.25
100	G.20000	Low	B	pussyfoot	0.75	0.44	0.19

[Continued on next page.]

Table S5. Mean rating (and *SD*) and correlation coefficient (*r*) across the 100 participants for each item on the WordFAM. Number refers to the order in which items appear on the paper-and-pencil version of the WordFAM. The Hoosier vector shows the mean rating for each item in the original Hoosier sample. Bin and version correspond to the divisions made in the current work as described in the main text.

Number	Hoosier	Bin	Version	Word	Mean	SD	<i>r</i>
25	1.00	Low	A	inrush	2.65	1.87	0.43
46	1.00	Low	B	systolic	3.71	2.20	0.58
74	1.00	Low	A	oppress	6.82	0.70	0.24
75	1.00	Low	B	braggadocio	2.74	2.33	0.52
132	1.00	Low	A	seasonable	6.31	1.32	0.20
53	1.33	Low	B	jalousie	1.54	1.08	0.30
78	1.33	Low	A	dysprosium	1.89	1.43	0.48
105	1.33	Low	B	shibboleth	1.90	1.67	0.54
81	1.50	Low	A	chemurgic	1.53	0.96	0.49
62	1.67	Low	B	lek	1.45	1.14	0.33
24	1.75	Low	A	cenobitic	1.61	1.14	0.48
141	1.75	Low	B	jardiniere	1.93	1.47	0.42
131	1.83	Low	A	obi	2.59	2.19	0.50
26	1.91	Low	B	campanile	1.84	1.45	0.42
41	1.92	Low	A	malfeasance	3.97	2.26	0.48
35	2.00	Low	B	molybdenum	2.29	2.10	0.49
52	2.00	Low	A	bosh	2.86	1.81	0.34
56	2.00	Low	B	aileron	2.06	1.82	0.40
32	2.08	Low	A	batrachian	1.45	0.98	0.47
67	2.17	Low	B	encomium	2.11	1.87	0.49
87	2.17	Low	A	bemire	2.05	1.50	0.55
99	2.17	Low	B	tinct	2.35	1.80	0.53
9	2.18	Low	A	appanage	1.97	1.55	0.31
6	2.25	Low	B	palliation	2.58	2.08	0.55
80	2.25	Low	A	citify	2.52	1.74	0.53
150	2.25	Low	B	inchoate	1.93	1.60	0.55
51	2.33	Low	A	puttee	2.01	1.59	0.22
94	2.33	Low	B	parquetry	1.79	1.35	0.57
110	2.33	Low	A	arable	3.14	2.28	0.50
116	2.33	Low	B	duenna	1.63	1.26	0.44
18	2.42	Low	A	ferrule	2.01	1.61	0.42
21	2.42	Low	B	egregious	5.27	2.21	0.49
44	2.42	Low	A	meliorate	2.53	1.84	0.60
117	2.42	Low	B	equine	4.59	2.53	0.51
76	2.50	Low	A	capstan	1.79	1.39	0.53

113	2.50	Low	B	viceregal	1.93	1.42	0.53
137	2.50	Low	A	hidalgo	2.69	1.74	0.44
139	2.50	Low	B	alembic	1.85	1.55	0.44
45	2.67	Low	A	crozier	1.97	1.49	0.46
50	2.67	Low	B	aniline	1.91	1.58	0.50
65	2.67	Low	A	ennui	2.84	2.40	0.48
68	2.67	Low	B	gustatory	2.73	2.14	0.50
86	2.75	Low	A	exegesis	2.03	1.42	0.48
143	2.75	Low	B	sessile	2.21	1.72	0.40
8	2.80	Low	A	mullion	2.01	1.57	0.43
28	2.83	Low	B	flivver	1.62	1.44	0.40
134	2.83	Low	A	fief	3.00	2.26	0.58
27	2.91	Low	B	imprimatur	1.81	1.37	0.41
40	2.92	Low	A	triumvir	1.83	1.56	0.48
114	2.92	Low	B	torsion	4.14	2.22	0.52
123	3.00	Medium	A	mastoid	3.12	2.09	0.62
2	3.08	Medium	B	cacophony	4.32	2.51	0.54
120	3.08	Medium	A	hemolytic	2.79	2.12	0.64
135	3.17	Medium	B	fusillade	2.41	1.90	0.60
5	3.25	Medium	A	scintillate	3.36	2.36	0.54
57	3.25	Medium	B	undulant	3.28	2.00	0.63
63	3.25	Medium	A	czarina	2.37	2.12	0.51
133	3.25	Medium	B	transept	2.25	1.69	0.49
47	3.33	Medium	A	warder	3.55	2.16	0.44
13	3.42	Medium	B	overawe	2.91	1.97	0.56
16	3.42	Medium	A	darnel	2.00	1.39	0.56
70	3.42	Medium	B	diathermy	1.78	1.24	0.33
121	3.50	Medium	A	grandiloquence	2.34	1.84	0.60
72	3.55	Medium	B	titivate	1.97	1.59	0.43
85	3.67	Medium	A	coitus	4.88	2.43	0.41
88	3.67	Medium	B	expatriate	4.41	2.47	0.65
146	3.67	Medium	A	triennial	3.08	2.25	0.53
98	3.73	Medium	B	expletive	6.05	1.84	0.46
34	3.75	Medium	A	smirch	3.22	2.20	0.65
36	3.75	Medium	B	dactyl	3.28	2.02	0.54
55	3.75	Medium	A	concertina	2.67	2.00	0.59
66	3.75	Medium	B	mastodon	4.62	2.29	0.50
77	3.75	Medium	A	rasher	3.54	2.32	0.62
90	3.75	Medium	B	parboil	3.23	2.46	0.56

107	3.75	Medium	A	ruse	6.05	1.66	0.44
111	3.75	Medium	B	hullabaloo	5.05	2.29	0.41
122	3.75	Medium	A	audiophile	5.76	1.94	0.55
96	3.83	Medium	B	exonerate	6.13	1.80	0.57
11	3.92	Medium	A	perspicuous	3.57	2.08	0.47
12	3.92	Medium	B	philodendron	2.68	2.08	0.55
15	3.92	Medium	A	stolidly	2.51	1.91	0.61
92	4.00	Medium	B	vestry	2.57	1.81	0.55
43	4.17	Medium	A	Pullman	3.31	2.15	0.50
119	4.17	Medium	B	conjugal	5.42	1.89	0.61
125	4.17	Medium	A	vesture	3.68	1.97	0.53
95	4.25	Medium	B	crocus	2.84	2.17	0.43
30	4.33	Medium	A	underslung	3.03	2.08	0.57
129	4.33	Medium	B	mallow	4.01	2.14	0.56
147	4.42	Medium	A	refutation	4.38	2.30	0.52
14	4.50	Medium	B	histamine	5.22	1.94	0.48
3	4.58	Medium	A	briquette	3.34	2.41	0.45
84	4.58	Medium	B	gentry	4.43	2.20	0.65
48	4.67	Medium	A	pommel	4.53	2.34	0.62
101	4.75	Medium	B	assiduous	3.19	2.12	0.58
73	4.83	Medium	A	knobbed	4.84	1.92	0.48
130	4.83	Medium	B	tendrill	4.97	2.24	0.42
10	4.92	Medium	A	affray	3.26	1.99	0.55
54	4.92	Medium	B	Pakistani	6.66	1.12	0.40
82	4.92	Medium	A	genic	2.89	1.91	0.40
142	4.92	Medium	B	radioisotope	4.43	2.12	0.57
102	5.00	High	A	freedman	4.52	2.26	0.60
115	5.00	High	B	ply	5.80	1.76	0.54
140	5.17	High	A	autumnal	4.71	2.39	0.59
97	5.33	High	B	denature	4.28	2.32	0.56
59	5.50	High	A	goodwife	4.84	1.99	0.43
103	5.50	High	B	cessation	4.62	2.29	0.52
4	5.83	High	A	deluge	4.70	2.19	0.64
17	5.92	High	B	authenticate	6.94	0.28	0.12
38	5.92	High	A	defy	6.68	1.09	0.20
128	5.92	High	B	morbidity	6.01	1.56	0.54
106	6.00	High	A	invigorate	6.12	1.83	0.47
20	6.08	High	B	mutt	6.55	1.13	0.24
29	6.25	High	A	euphemism	6.16	1.67	0.48

1	6.33	High	B	pox	5.35	1.88	0.61
39	6.33	High	A	dike	5.27	1.99	0.38
118	6.33	High	B	accusal	5.69	1.96	0.41
89	6.42	High	A	forked	6.09	1.40	0.55
61	6.45	High	B	immobility	6.88	0.41	0.25
124	6.50	High	A	drag	6.96	0.24	0.24
49	6.55	High	B	drab	6.26	1.47	0.39
104	6.58	High	A	handrail	6.11	1.84	0.28
109	6.58	High	B	municipal	5.99	1.55	0.53
69	6.67	High	A	gab	5.17	2.22	0.53
91	6.67	High	B	hedge	6.70	0.98	0.22
22	6.75	High	A	destitute	5.60	2.06	0.58
33	6.75	High	B	fabrication	6.85	0.44	0.23
148	6.75	High	A	antler	6.44	1.60	0.27
7	6.83	High	B	misapplication	6.02	1.63	0.38
58	6.83	High	A	cannibal	6.97	0.17	0.18
60	6.83	High	B	index	6.80	0.60	0.19
71	6.83	High	A	quit	6.97	0.30	-0.07
42	6.92	High	B	inflexible	6.65	1.20	0.23
79	6.92	High	A	objectivity	6.61	0.94	0.36
127	6.92	High	B	impair	6.58	1.25	0.23
19	7.00	High	A	comforter	6.82	0.73	0.24
23	7.00	High	B	outcast	6.90	0.63	0.06
31	7.00	High	A	greed	6.90	0.63	-0.01
37	7.00	High	B	ash	6.88	0.46	0.14
64	7.00	High	A	grab	6.95	0.33	-0.07
83	7.00	High	B	educate	6.99	0.10	0.12
93	7.00	High	A	mother	7.00	0.00	NA
100	7.00	High	B	cop	6.93	0.33	0.01
108	7.00	High	A	central	6.95	0.22	0.32
112	7.00	High	B	monologue	6.80	0.71	0.04
126	7.00	High	A	glory	6.87	0.66	0.33
136	7.00	High	B	classmate	6.93	0.61	0.11
138	7.00	High	A	battery	6.97	0.17	0.06
144	7.00	High	B	affect	6.88	0.41	0.16
145	7.00	High	A	leaves	6.94	0.28	0.08
149	7.00	High	B	cancel	7.00	0.00	NA

Figure S1. For each subject, mean accuracy (proportion correct) by frequency bin on the long-form VST.

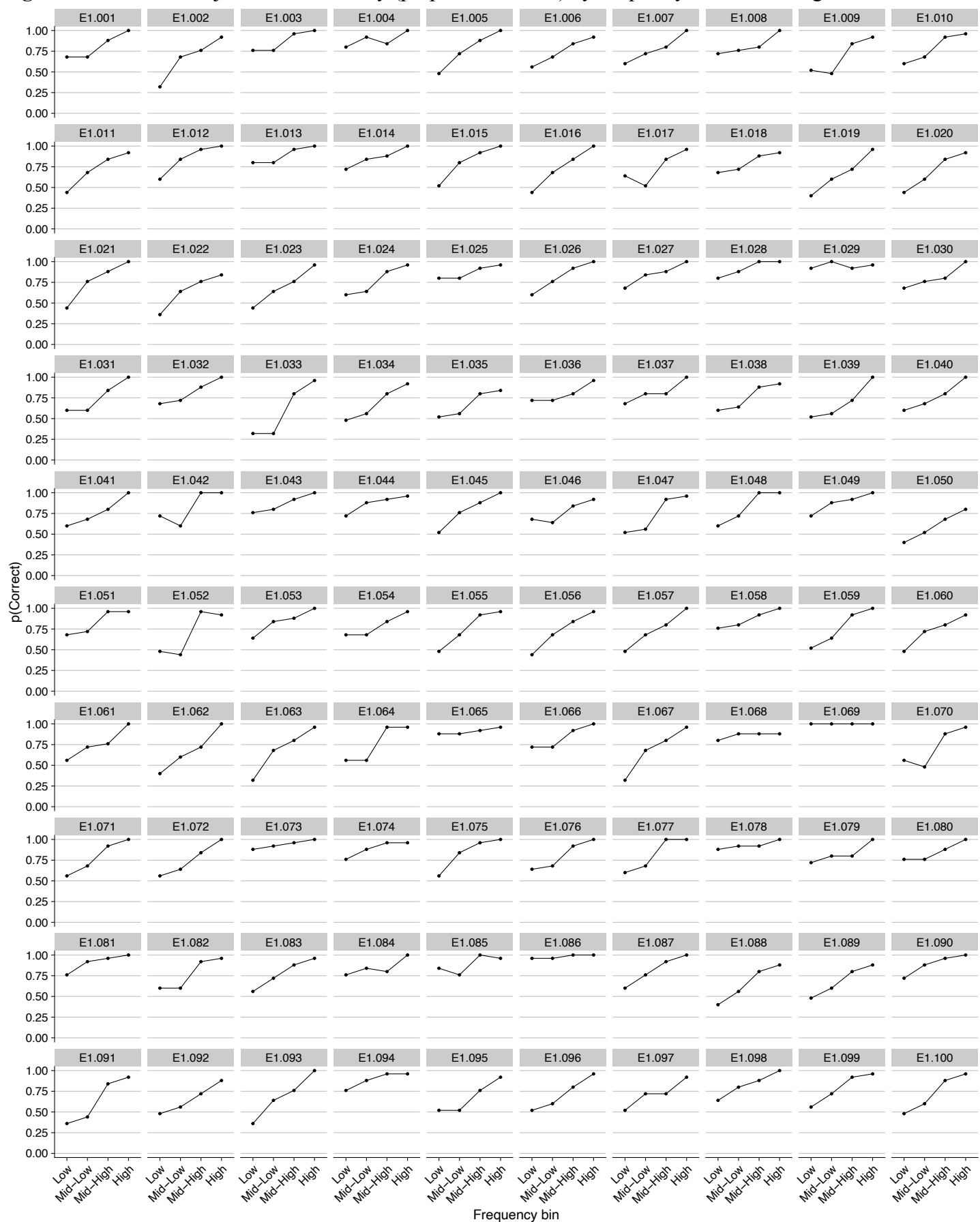


Figure S2. For each subject, mean familiarity rating by frequency bin on the long-form WordFAM.

