

# Multitudes v3 (2024/25) Prediction Models

Julian M. Siebert

2025-08-01

## Table of contents

<b>1</b>	<b>Preliminaries</b>	<b>2</b>
<b>2</b>	<b>ENGLISH MODELS</b>	<b>3</b>
2.1	Sample . . . . .	3
2.2	Descriptives . . . . .	5
2.2.1	Distributions of English Multitudes Tasks in Fall '23 . . . . .	5
2.2.2	Distributions of English Outcome Measures (Spring '24 and Spring '25)	7
2.3	Model Building & LOGO Cross-validated Evaluation . . . . .	8
2.4	Out-of-sample Evaluation of English Models Using AY 2024/25 Data . . . . .	8
<b>3</b>	<b>SPANISH MODELS</b>	<b>9</b>
3.1	Sample . . . . .	9
3.2	Descriptives . . . . .	10
3.2.1	Distributions of Spanish Multitudes Tasks . . . . .	10
3.2.2	Distributions of Spanish Outcome Measures . . . . .	11
3.3	Model Building & LOGO Cross-validated Evaluation . . . . .	13

# 1 Preliminaries

## Cautionary Note

This document is WORK IN PROGRESS and will be updated as new data (e.g., additional demographic information, score database updates/fixes) become available.

## To Do

- ~~recompute scores for universal screening tasks with updated 2025 item parameters once available~~
- ~~filter out implausible RAN score~~
- ~~add 23/24 data~~
- ~~align grades between years~~
- ~~align tasknames between years~~
- ~~re-run 23/24 models~~
- ~~run 24/25 models~~
- ~~run combined model~~
- ~~run two-year prediction models (K-G2)~~
- ~~recompute scores for remaining tasks with updated 2025 item parameters once available~~
- ~~isolate pooled models (pooled coefficients) for English model evaluations with 24/25 data~~
- as per decision in data meeting on 2025-07-25
  - ~~use combined (23/24 and 24/25) data for Spanish model~~
  - ~~check ELPD distribution for English samples to determine whether to use combined or only 23/24 data for English models~~

## 2 ENGLISH MODELS

- English models are built using data collected in the *2023/24* academic year only (sample described in left set of columns in Table 1);  $N = 1831$ , from 27 schools across 15 districts
- I am using re-calculated theta scores using the updated (MV3/2025) item parameters
- English models are evaluated using unseen data from *2024/25* academic year (sample described in right set of columns Table 1);  $N = 1317$ , from 26 schools across 12 districts

### 2.1 Sample

Table 1: Demographic Characteristics of the English Model Building and Evaluation Samples by Grade

Characteristic	Model Building - AY 23/24			Model Evaluation - AY 24/25		
	K N = 652	G1 N = 515	G2 N = 664	K N = 435	G1 N = 552	G2 N = 330
Gender						
Female	332 (51%)	240 (47%)	332 (50%)	180 (47%)	289 (55%)	144 (46%)
Male	320 (49%)	275 (53%)	330 (50%)	202 (53%)	233 (45%)	170 (54%)
Unknown	0	0	2	53	30	16
Race						
Asian	81 (13%)	65 (13%)	65 (9.9%)	24 (7.8%)	47 (9.6%)	33 (10%)
Black/African American	85 (14%)	53 (10%)	95 (14%)	41 (13%)	65 (13%)	40 (12%)
Not reported	137 (23%)	94 (18%)	134 (20%)	51 (17%)	100 (20%)	40 (12%)
Other	88 (14%)	71 (14%)	80 (12%)	34 (11%)	57 (12%)	34 (11%)
White	216 (36%)	229 (45%)	283 (43%)	159 (51%)	219 (45%)	175 (54%)
Unknown	45	3	7	126	64	8
Ethnicity						
Hispanic/Latin(o/a)	352 (55%)	255 (50%)	358 (54%)	295 (81%)	416 (77%)	228 (70%)
Not Hispanic/Latin(o/a)	289 (45%)	258 (50%)	301 (46%)	71 (19%)	124 (23%)	98 (30%)
Unknown	11	2	5	69	12	4
ELPD						
EL	148 (25%)	85 (17%)	117 (19%)	212 (59%)	238 (48%)	119 (38%)
EO	395 (66%)	379 (75%)	451 (72%)	123 (34%)	211 (43%)	155 (50%)
EP	37 (6.2%)	43 (8.5%)	56 (9.0%)	19 (5.2%)	37 (7.5%)	36 (12%)
TBD	20 (3.3%)	1 (0.2%)	1 (0.2%)	8 (2.2%)	10 (2.0%)	0 (0%)
Unknown	52	7	39	73	56	20
Ever IEP/504	43 (7.9%)	41 (9.1%)	56 (12%)	32 (9.2%)	33 (7.3%)	30 (10%)
Unknown	107	62	181	89	99	35
Home Language						
English	463 (72%)	391 (76%)	448 (72%)	156 (36%)	237 (44%)	170 (52%)
Other	63 (9.8%)	50 (9.8%)	62 (9.9%)	11 (2.6%)	16 (3.0%)	15 (4.6%)
Spanish	117 (18%)	71 (14%)	114 (18%)	261 (61%)	289 (53%)	140 (43%)
Unknown	9	3	40	7	10	5

## 2.2 Descriptives

### 2.2.1 Distributions of English Multitudes Tasks in Fall '23

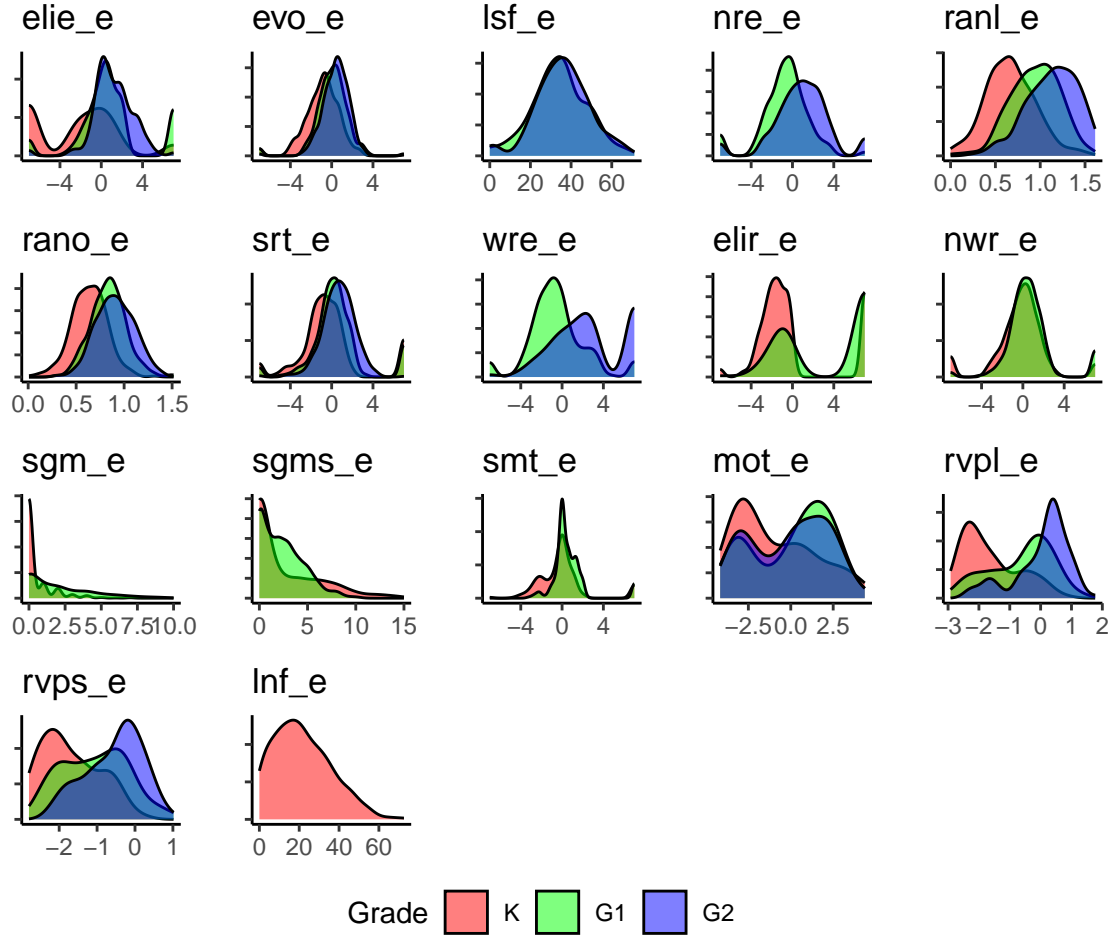


Figure 1: Distributions of English Multitudes Tasks by Grade (Fall '23)

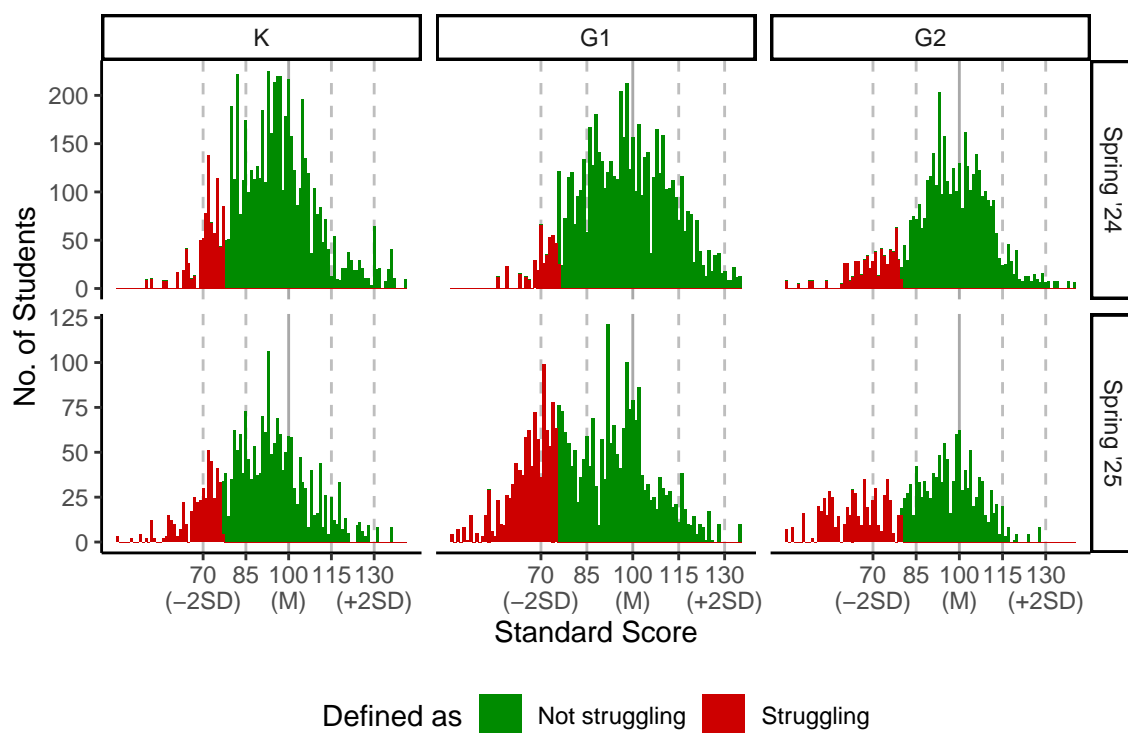


Figure 2: Distribution of Woodcock-Johnson Basic Reading Skills Cluster (for Kindergarten) Broad Reading Cluster (for Grades 1 and 2) by Grade and Language with Struggling Readers ( $\leq$  20th Percentile) Highlighted (Spring '24 and Spring '25).

### 2.2.2 Distributions of English Outcome Measures (Spring '24 and Spring '25)

Table 2: Comparison of Proportions of Struggling Readers ( $\leq$  20th Percentile) in the Study Sample Based on Overall Multitudes Sample and on Published Norms (Spring '24 and Spring '25).

Reference	Spring '24			Spring '25		
	K N = 652	G1 N = 515	G2 N = 664	K N = 435	G1 N = 552	G2 N = 330
Norm-based						
Not struggling	421 (65%)	372 (72%)	488 (73%)	228 (52%)	247 (45%)	192 (58%)
Struggling	231 (35%)	143 (28%)	176 (27%)	207 (48%)	305 (55%)	138 (42%)
Sample-based						
Not struggling	554 (85%)	472 (92%)	564 (85%)	332 (76%)	354 (64%)	229 (69%)
Struggling	98 (15%)	43 (8.3%)	100 (15%)	103 (24%)	198 (36%)	101 (31%)

## 2.3 Model Building & LOGO Cross-validated Evaluation

Table 3: Best (Balanced Accuracy) English-to-English Prediction Models Built Using Data from 2023/24 AY.

Grade	Predictors	All							EL							EO						
		n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.	n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.	n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.
K	LNF only	383	0.7866	0.7687	0.7451	0.7923	0.122	0.273	102	0.6887	0.7162	0.7619	0.6706	0.198	0.415	230	0.8103	0.7837	0.75	0.8173	0.103	0.241
K	LNF, RANO, ELIR	356	0.7088	0.733	0.7632	0.7029	0.098	0.343	89	0.6344	0.7231	0.8462	0.6	0.140	0.462	217	0.6941	0.7134	0.7368	0.69	0.087	0.347
K	LNF, RANO, ELIR, EVO	356	0.701	0.7522	0.8158	0.6886	0.098	0.361	89	0.6129	0.7428	0.9231	0.5625	0.140	0.505	217	0.7032	0.7184	0.7368	0.7	0.087	0.338
K	LNF, RANO, EVO	356	0.7526	0.7808	0.8158	0.7457	0.098	0.309	89	0.6452	0.7293	0.8462	0.6125	0.140	0.452	217	0.7808	0.7847	0.7895	0.78	0.087	0.269
G1	WRE only	487	0.7085	0.7456	0.7895	0.7018	0.077	0.336	75	0.6184	0.6615	0.7143	0.6087	0.092	0.421	364	0.7198	0.7591	0.8065	0.7117	0.085	0.332
G1	WRE, EVO	487	0.7247	0.7785	0.8421	0.7149	0.077	0.328	75	0.6053	0.7184	0.8571	0.5797	0.092	0.460	364	0.7418	0.7857	0.8387	0.7327	0.085	0.316
G1	WRE, RANO	282	0.7098	0.7747	0.8462	0.7033	0.046	0.322	40	0.625	NA	NA	0.625	0.000	0.375	214	0.7243	0.7813	0.8462	0.7164	0.061	0.318
G1	WRE, RANO, EVO	282	0.7203	0.8168	0.9231	0.7106	0.046	0.318	40	0.575	NA	NA	0.575	0.000	0.425	214	0.743	0.8272	0.9231	0.7313	0.061	0.308
G1	WRE, RANO, LSF	277	0.7829	0.8496	0.9231	0.7761	0.046	0.256	39	0.7436	NA	NA	0.7436	0.000	0.256	210	0.7952	0.8549	0.9231	0.7868	0.062	0.257
G1	WRE, RANO, LSF, EVO	277	0.7687	0.8055	0.8462	0.7649	0.046	0.263	39	0.641	NA	NA	0.641	0.000	0.359	210	0.7905	0.8165	0.8462	0.7868	0.062	0.252
G2	WRE only	617	0.8655	0.9076	0.9681	0.847	0.152	0.277	101	0.7723	0.8284	0.9667	0.6901	0.297	0.505	435	0.8736	0.9197	0.9831	0.8564	0.136	0.258
G2	WRE, RANL	309	0.8997	0.9417	1	0.8835	0.139	0.240	45	0.8444	0.9	1	0.8	0.222	0.378	241	0.9046	0.9452	1	0.8905	0.129	0.224
G2	WRE, RANL, EVO	309	0.8932	0.9282	0.9767	0.8797	0.139	0.240	45	0.8222	0.8857	1	0.7714	0.222	0.400	241	0.9046	0.9452	1	0.8905	0.129	0.224

Note. green:  $\geq .8$ ; black:  $\geq .7$ ; yellow:  $\geq .6$ ; red otherwise

8

## 2.4 Out-of-sample Evaluation of English Models Using AY 2024/25 Data

Table 4: Best (Balanced Accuracy) English-to-English Prediction Models Built Using Data from 2023/24 AY.

Grade	Predictors	All							EL							EO						
		n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.	n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.	n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.
K	LNF only	319	0.768	0.7151	0.6197	0.8105	0.223	0.285	123	0.7073	0.7228	0.7632	0.6824	0.309	0.455	123	0.8211	0.6643	0.4	0.9286	0.203	0.138
K	LNF, RANO, ELIR	219	0.6758	0.6949	0.725	0.6648	0.183	0.406	48	0.5833	0.6339	0.7273	0.5405	0.229	0.521	117	0.6923	0.6883	0.6818	0.6947	0.188	0.376
K	LNF, RANO, ELIR, EVO	199	0.7789	0.6636	0.4857	0.8415	0.176	0.216	31	0.7097	0.63	0.5	0.76	0.194	0.290	117	0.7863	0.6589	0.4545	0.8632	0.188	0.197
K	LNF, RANO, EVO	203	0.8177	0.689	0.4865	0.8916	0.182	0.177	35	0.7429	0.7014	0.625	0.7778	0.229	0.314	117	0.8205	0.6624	0.4091	0.9158	0.188	0.145
G1	WRE only	401	0.7622	0.8066	0.9557	0.6575	0.351	0.558	154	0.7711	0.7732	0.9512	0.5952	0.494	0.675	202	0.7304	0.804	0.9636	0.6443	0.270	0.520
G1	WRE, EVO	318	0.7705	0.812	0.9143	0.7097	0.298	0.476	88	0.7742	0.7927	0.925	0.6604	0.430	0.591	202	0.7647	0.8103	0.9091	0.7114	0.270	0.456
G1	WRE, RANO	345	0.7734	0.8152	0.9237	0.7068	0.307	0.487	112	0.7627	0.7777	0.9038	0.6515	0.441	0.593	196	0.7525	0.8074	0.9231	0.6918	0.263	0.470
G1	WRE, RANO, EVO	308	0.7742	0.814	0.9091	0.719	0.290	0.463	83	0.7727	0.7928	0.9189	0.6667	0.420	0.580	196	0.7677	0.8115	0.9038	0.7192	0.263	0.444
G1	WRE, RANO, LSF	333	0.8302	0.8221	0.8018	0.8423	0.299	0.350	104	0.7818	0.7798	0.766	0.7937	0.427	0.446	195	0.8477	0.8349	0.8077	0.8621	0.264	0.315
G1	WRE, RANO, LSF, EVO	304	0.8101	0.824	0.8571	0.7908	0.291	0.398	81	0.8023	0.8106	0.8611	0.76	0.419	0.500	195	0.8071	0.8196	0.8462	0.7931	0.264	0.376
G2	WRE only	257	0.8346	0.8674	0.9518	0.7831	0.305	0.441	78	0.8395	0.8774	1	0.7547	0.346	0.506	140	0.8214	0.8485	0.9388	0.7582	0.350	0.486
G2	WRE, RANL	231	0.8724	0.8854	0.9178	0.8529	0.300	0.379	58	0.8833	0.9	0.95	0.85	0.333	0.417	137	0.854	0.8686	0.9149	0.8222	0.343	0.431
G2	WRE, RANL, EVO	214	0.8756	0.8809	0.8939	0.8679	0.293	0.356	45	0.9348	0.933	0.9286	0.9375	0.304	0.326	137	0.8467	0.8579	0.8936	0.8222	0.343	0.423

Note. green:  $\geq .8$ ; black:  $\geq .7$ ; yellow:  $\geq .6$ ; red otherwise



### 3 SPANISH MODELS

- built using data collected in the *2023/24 and 2024/25* academic year
- using recalculated theta scores using the updated (MV3/2025) item parameters
- $N = 979$  unique students (with some students featuring in both AYs) from 23 schools across 9 districts

#### 3.1 Sample

Table 5: Demographic Characteristics of the Sample for the Spanish Prediction Model ( $N = 979$ ) by Grade

Characteristic	K N = 413	G1 N = 309	G2 N = 257
Gender			
Female	186 (49%)	153 (53%)	140 (55%)
Male	190 (51%)	137 (47%)	116 (45%)
Unknown	37	19	1
Race			
Asian	12 (3.7%)	7 (2.5%)	6 (2.4%)
Black/African American	2 (0.6%)	0 (0%)	1 (0.4%)
Not reported	191 (59%)	134 (49%)	119 (48%)
Other	22 (6.8%)	20 (7.3%)	7 (2.8%)
White	98 (30%)	114 (41%)	116 (47%)
Unknown	88	34	8
Ethnicity			
Hispanic/Latin(o/a)	379 (96%)	296 (96%)	242 (95%)
Not Hispanic/Latin(o/a)	17 (4.3%)	11 (3.6%)	14 (5.5%)
Unknown	17	2	1
ELPD			
EL	323 (82%)	237 (80%)	186 (74%)
EO	19 (4.8%)	13 (4.4%)	26 (10%)
EP	48 (12%)	46 (15%)	38 (15%)
TBD	5 (1.3%)	1 (0.3%)	1 (0.4%)
Unknown	18	12	6
Ever IEP/504	31 (8.8%)	22 (8.8%)	13 (6.0%)
Unknown	61	58	39
Home Language			
English	37 (9.1%)	18 (5.9%)	29 (12%)
Other	4 (1.0%)	1 (0.3%)	2 (0.8%)
Spanish	367 (90%)	288 (94%)	219 (88%)
Unknown	5	2	7

## 3.2 Descriptives

### 3.2.1 Distributions of Spanish Multitudes Tasks

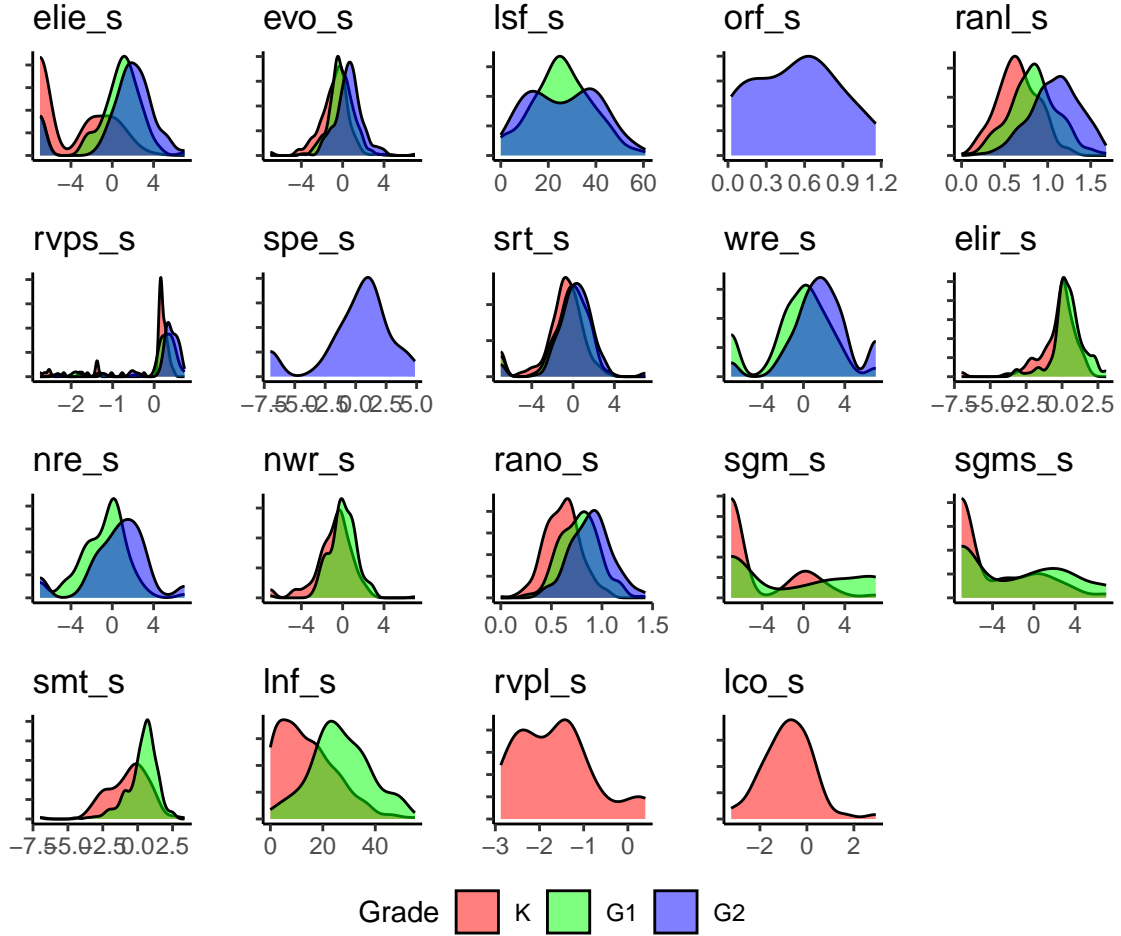


Figure 3: Distributions of Spanish Multitudes Tasks by Grade (Fall '23 & Fall '24 Combined)

### 3.2.2 Distributions of Spanish Outcome Measures

Table 6: Comparison of Proportions of Struggling Readers ( $\leq$  20th Percentile) in the Study Sample Based on Overall Multitudes Sample and on Norms (Spring 2024).

<b>Reference</b>	<b>K</b> N = 413	<b>G1</b> N = 309	<b>G2</b> N = 257
Norm-based			
Not struggling	193 (47%)	178 (58%)	138 (54%)
Struggling	220 (53%)	131 (42%)	119 (46%)
Sample-based			
Not struggling	331 (80%)	235 (76%)	194 (75%)
Struggling	82 (20%)	74 (24%)	63 (25%)

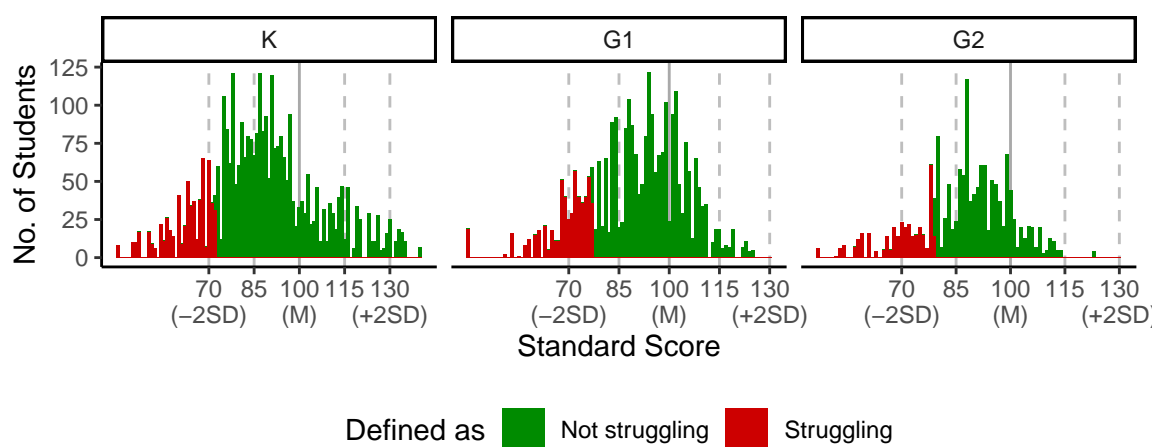


Figure 4: Distribution of Woodcock-Muñoz Basic Reading Skills Cluster (for Kindergarten) Broad Reading Cluster (for Grades 1 and 2) by Grade and Language with Struggling Readers ( $\leq$  20th Percentile) Highlighted (Spring '24 and Spring '25 Combined).

### 3.3 Model Building & LOGO Cross-validated Evaluation

Table 7: Best (Balanced Accuracy) Spanish-to-Spanish Prediction Models Built Using Data from 2023/24 AY.

Grade	Predictors	Method	All						
			n	Acc.	B.Acc.	Sens.	Spec.	Prev.	P.Prev.
K	LNF only	Youden	294	0.6327	0.6937	0.7826	0.6048	0.1565	0.4558
K	LNF, RANO, ELIR	Youden	224	0.5893	0.5938	0.6	0.5876	0.1339	0.4375
K	LNF, RANO, ELIR, EVO	Youden	224	0.6339	0.6619	0.7	0.6237	0.1339	0.4196
K	LNF, RANO, EVO	Youden	243	0.6091	0.652	0.7097	0.5943	0.1276	0.4444
G1	WRE only	Youden	216	0.7731	0.7994	0.8444	0.7544	0.2083	0.3704
G1	WRE, RANO, LSF	Youden	133	0.8571	0.8352	0.8	0.8704	0.188	0.2556
G1	WRE, RANO, LSF, EVO	Youden	132	0.8409	0.8252	0.8	0.8505	0.1894	0.2727
G2	WRE only	Youden	198	0.8535	0.8515	0.8478	0.8553	0.2323	0.3081
G2	WRE, RANL	Youden	150	0.82	0.8269	0.8387	0.8151	0.2067	0.32
G2	WRE, RANL, EVO	Youden	150	0.8533	0.836	0.8065	0.8655	0.2067	0.2733

Note. green:  $\geq .8$ ; black:  $\geq .7$ ; yellow:  $\geq .6$ ; red otherwise