

# Multitudes v3 (2024/25) Prediction Models

Julian M. Siebert

2025-07-29

## Table of contents

<b>1 Preliminaries</b>	<b>2</b>
<b>2 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 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

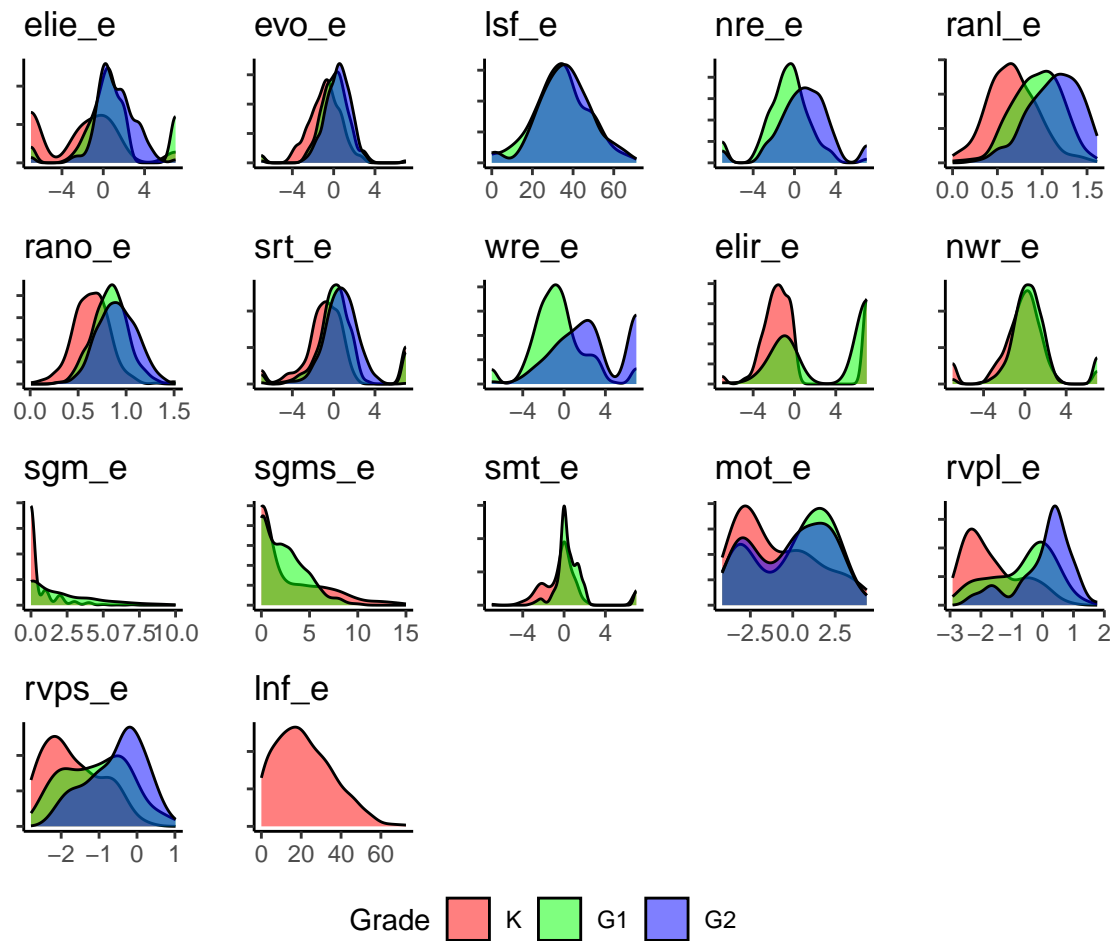


Figure 1: Distributions of English Multitudes Tasks by Grade (Fall '23)

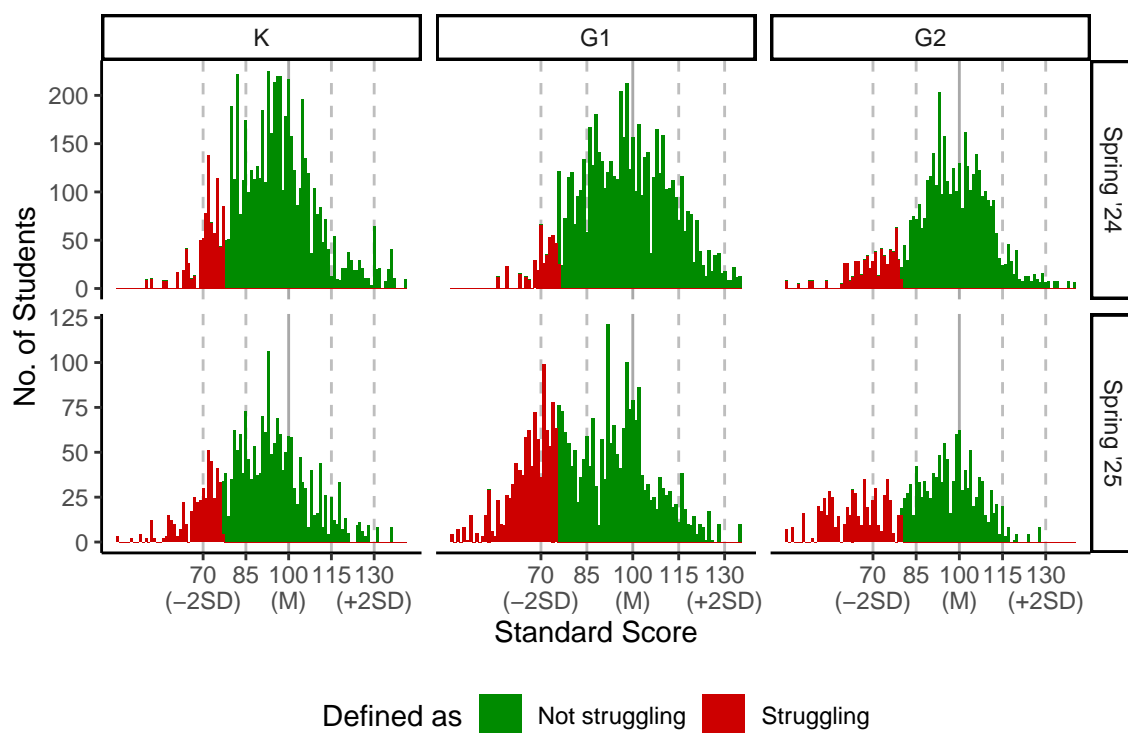


Figure 2: Distribution of Woodcock-Johnson Basic Reading Skills Cluster (for Kindergarten) Broad Reading Cluster (for Grades 1 and 2) by Grade and Language with Struggling Readers ( $\leq$  20th Percentile) Highlighted (Spring '24 and Spring '25).

### 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	Method	All					EL					EO				
			n	Acc.	B.Acc.	Sens.	Spec.	n	Acc.	B.Acc.	Sens.	Spec.	n	Acc.	B.Acc.	Sens.	Spec.
K	LNF only	Youden	652	0.8165	0.7822	0.7333	0.8311	166	0.7571	0.7975	0.8684	0.7266	410	0.8402	0.7693	0.6727	0.8659
K	LNF, RANO, ELIR	Youden	652	0.7681	0.7773	0.7905	0.7642	166	0.6893	0.7448	0.8421	0.6475	410	0.7918	0.8029	0.8182	0.7877
K	LNF, RANO, ELIR, EVO	Youden	652	0.7112	0.7675	0.8476	0.6873	166	0.6102	0.7136	0.8947	0.5324	410	0.7385	0.7953	0.8727	0.7179
K	LNF, RANO, EVO	Youden	652	0.7738	0.7846	0.8	0.7692	166	0.6893	0.7639	0.8947	0.6331	410	0.8015	0.7778	0.7455	0.8101
G1	WRE only	Youden	515	0.7247	0.7671	0.8182	0.7161	85	0.6512	0.7224	0.8182	0.6267	379	0.7335	0.7718	0.8182	0.7254
G1	WRE, RANO, LSF	MaxSpSe	515	0.8432	0.8422	0.8409	0.8434	85	0.7674	0.7891	0.8182	0.76	379	0.8549	0.852	0.8485	0.8555
G1	WRE, RANO, LSF, EVO	Youden	515	0.8279	0.8441	0.8636	0.8246	85	0.7558	0.86	1	0.72	379	0.8364	0.8282	0.8182	0.8382
G2	WRE only	Youden	664	0.878	0.9117	0.96	0.8635	117	0.8205	0.8658	0.9697	0.7619	451	0.878	0.9157	0.9677	0.8638
G2	WRE, RANL	Youden	664	0.8946	0.9256	0.97	0.8812	117	0.8462	0.8837	0.9697	0.7976	451	0.8958	0.9328	0.9839	0.8817
G2	WRE, RANL, EVO	Youden	664	0.9142	0.933	0.96	0.906	117	0.8547	0.8896	0.9697	0.8095	451	0.9224	0.955	1	0.91

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	Method	All					EL					EO				
			n	Acc.	B.Acc.	Sens.	Spec.	n	Acc.	B.Acc.	Sens.	Spec.	n	Acc.	B.Acc.	Sens.	Spec.
K	LNF only	Youden	435	0.768	0.7252	0.6479	0.8024	362	0.7689	0.7348	0.6667	0.803	362	0.7689	0.7348	0.6667	0.803
K	LNF, RANO, ELIR	Youden	435	0.7534	0.655	0.5	0.8101	362	0.7371	0.6596	0.5278	0.7914	362	0.7371	0.6596	0.5278	0.7914
K	LNF, RANO, ELIR, EVO	Youden	435	0.7638	0.6544	0.4857	0.8232	362	0.7613	0.6694	0.5161	0.8226	362	0.7613	0.6694	0.5161	0.8226
K	LNF, RANO, EVO	Youden	435	0.8177	0.71	0.5405	0.8795	362	0.8176	0.7284	0.5758	0.881	362	0.8176	0.7284	0.5758	0.881
G1	WRE only	Youden	552	0.8044	0.8057	0.8101	0.8014	496	0.8069	0.8113	0.8264	0.7962	496	0.8069	0.8113	0.8264	0.7962
G1	WRE, RANO, LSF, EVO	Youden	552	0.8516	0.8382	0.8061	0.8703	496	0.8519	0.8396	0.809	0.8702	496	0.8519	0.8396	0.809	0.8702
G2	WRE only	Youden	330	0.8162	0.8542	0.9518	0.7566	310	0.8056	0.8446	0.9494	0.7399	310	0.8056	0.8446	0.9494	0.7399
G2	WRE, RANL	Youden	330	0.8683	0.8746	0.8904	0.8588	310	0.8565	0.8641	0.8841	0.8442	310	0.8565	0.8641	0.8841	0.8442
G2	WRE, RANL, EVO	Youden	330	0.8844	0.8828	0.8788	0.8868	310	0.8786	0.8765	0.871	0.8819	310	0.8786	0.8765	0.871	0.8819

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

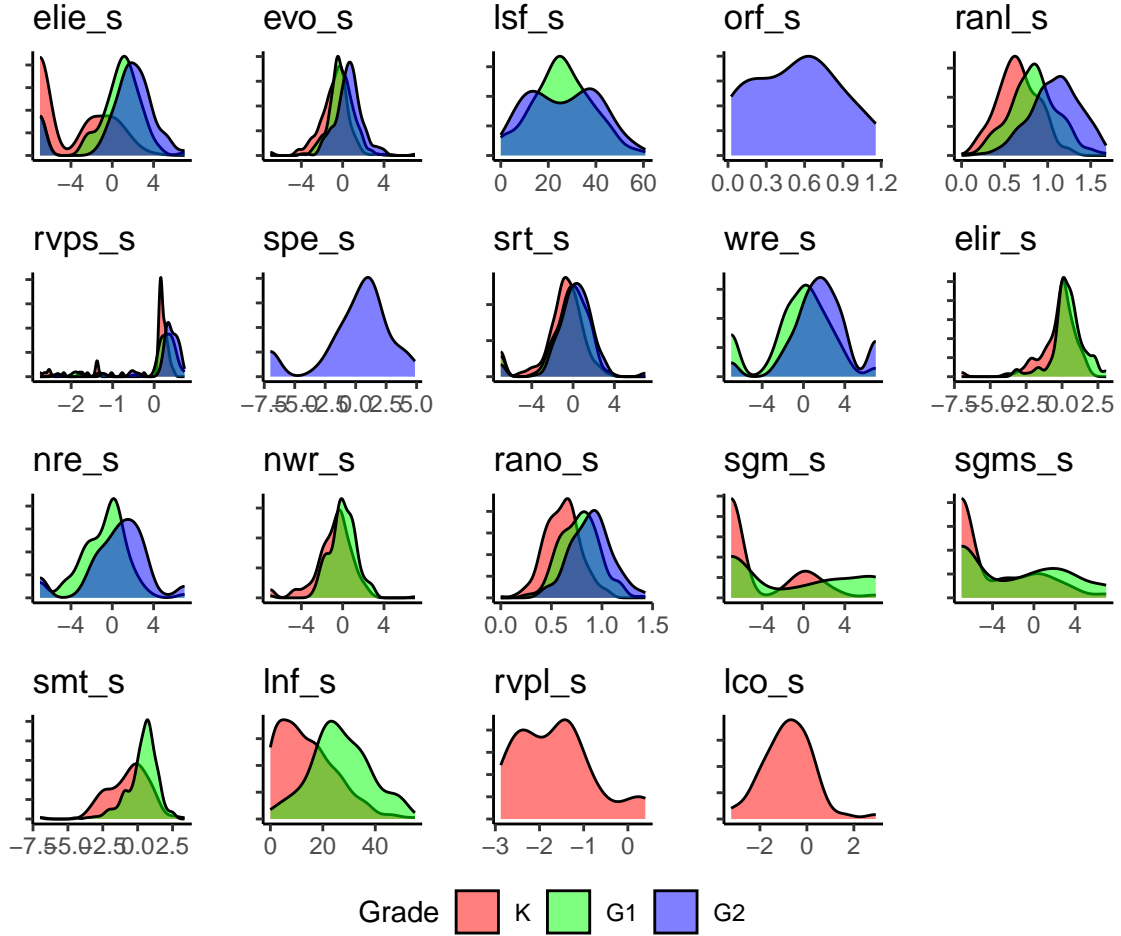


Figure 3: Distributions of Spanish Multitudes Tasks by Grade (Fall '23 & Fall '24 Combined)

### 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%)

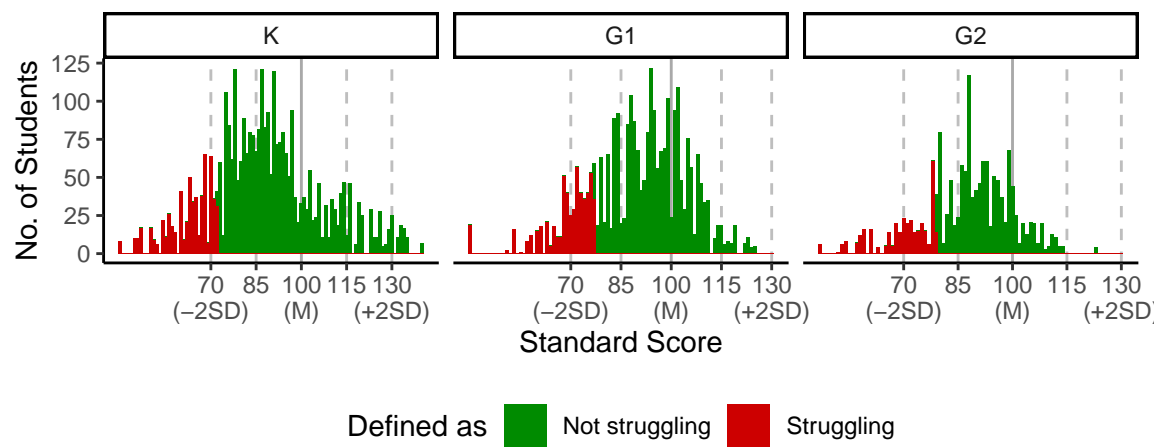


Figure 4: Distribution of Woodcock-Muñoz Basic Reading Skills Cluster (for Kindergarten) Broad Reading Cluster (for Grades 1 and 2) by Grade and Language with Struggling Readers ( $\leq$  20th Percentile) Highlighted (Spring '24 and Spring '25 Combined).

### 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.
K	LNF only	Youden	413	0.707	0.7255	0.7561	0.6949
K	LNF, RANO, ELIR	Youden	413	0.678	0.7211	0.7927	0.6495
K	LNF, RANO, ELIR, EVO	Youden	413	0.661	0.7014	0.7683	0.6344
K	LNF, RANO, EVO	Youden	413	0.6828	0.7012	0.7317	0.6707
G1	WRE only	PrevalenceMatching	309	0.8867	0.8422	0.7568	0.9277
G1	WRE, RANO, LSF	Youden	309	0.8738	0.8754	0.8784	0.8723
G1	WRE, RANO, LSF, EVO	Youden	309	0.877	0.8775	0.8784	0.8766
G2	WRE only	Youden	257	0.8872	0.8877	0.8889	0.8866
G2	WRE, RANL	SpEqualSe	257	0.8911	0.8903	0.8889	0.8918
G2	WRE, RANL, EVO	SpEqualSe	257	0.8911	0.8903	0.8889	0.8918

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