

Comparison of Literacy Screener Risk Selection between English Proficient Students and
English Learners

Supplemental File

Tables A1-A4 list the optimal decision thresholds for the D6 measures, LNF, PSF, NWF, and ORF, as well as the accompanying statistics from the ROC curve, values for A , decision thresholds, classification statistics, base rates, ρ , and proportions screened positive (τ). These statistics were defined for students with criterion values at the 20th normative percentile (at risk), 40th percentile (benchmark or some risk), and 60th percentile (target) on comprehensive tests collected at the end of the same year each screener was administered. For estimates of A , sensitivity, and specificity, we also describe the level of precision using a 95% confidence bound.

Results Illustration: LNF

In Table A1, we report the results for LNF; the remaining tables would be interpreted similarly. For the at-risk level, the accuracy for D6 LNF in the fall of kindergarten was low, $A = .68$ with 95% confidence interval of $[.65, .71]$, well below our a priori $A \geq .75$ value chosen as minimally acceptable. Students who were truly at risk of reading failure on the SAT10 had an 84% chance (sensitivity, $[.81, .86]$) of being identified as at-risk on the LNF screener if they scored below 3 (threshold), the lowest score for which specificity exceeded $.80$. But as expected given the value for A , corresponding specificity was poor; of students at or above the 20th percentile on the SAT10, just 46% $[.41, .51]$ were identified as true negatives. This implies that 54% of students whose scores were not below the 20th percentile on the SAT10 were falsely identified as positive. A different threshold for LNF could improve specificity but only at the expense of reduced sensitivity. The winter administration of LNF in kindergarten had a higher level of accuracy for the at-risk decision threshold, $A = .83$ $[.80, .85]$, and consequently a more

acceptable specificity value, .65 [.61, .70], for our chosen level of sensitivity. Nonetheless, the overall accuracy of LNF failed to exceed the moderate range, A from .75 to .85, for the at-risk threshold. Accuracy was better for the some-risk and target thresholds.

Predictive values suggest the practical significance of the screener (Pepe, 2003). Among the 74% of students (τ) who scored below 3 on the fall assessment of LNF in kindergarten and thus screened positive, the positive predictive value (PPV) shows that 75% will likely fall below the 20th percentile on the SAT10 in the spring. The negative predictive value (NPV) indicates that 59% of students who screened negative will likely score at or above the 20th percentile on the SAT10 in the spring. Given the poor overall accuracy in the fall, however, these values are not necessarily meaningful, and they improve for other levels of risk and assessment times. Note also that unlike A , sensitivity, and specificity, predictive values depend on the base rate (ρ); PPV ranges between ρ and 1; NPV ranges between $\rho - 1$ and 1. This limits the utility of predictive values when establishing cut scores in samples that are otherwise comparable but with different base rates. Authors (2015, 2016a), however, show how to estimate predictive values for different base rates.

Table A1

Optimal Decision Thresholds for Letter Naming Fluency

Statistic		Kindergarten			1st	
		F	W	S	F	
At Risk	A	.68	.83	.81	.81	
	95% CI	[.65, .71]	[.80, .85]	[.78, .83]	[.79, .83]	
	Threshold	3	21	38	29	
	Sensitivity	.84	.81	.81	.82	
	95% CI	[.81, .86]	[.78, .83]	[.79, .84]	[.79, .84]	
	Specificity	.46	.65	.62	.62	
	95% CI	[.41, .51]	[.61, .70]	[.57, .66]	[.58, .65]	
	NPV	.59	.62	.60	.69	
	PPV	.75	.83	.82	.76	
	ρ	.66	.68	.69	.60	
	τ	.74	.66	.68	.64	
	Some Risk	A	.74	.86	.85	.83
		95% CI	[.70, .79]	[.83, .89]	[.82, .88]	[.80, .85]
Threshold		4	27	42	34	
Sensitivity		.82	.81	.81	.80	
95% CI		[.80, .84]	[.79, .84]	[.78, .83]	[.78, .82]	
Specificity		.58	.76	.71	.65	
95% CI		[.50, .67]	[.69, .83]	[.64, .79]	[.60, .70]	
NPV		.29	.33	.30	.40	
PPV		.94	.97	.96	.92	
ρ		.89	.89	.90	.83	
τ		.78	.75	.75	.73	

	Statistic	Kindergarten			1st
		F	W	S	F
Target	A	.80	.89	.88	.87
	95% CI	[.74, .86]	[.85, .93]	[.84, .91]	[.84, .90]
	Threshold	4	28	43	36
	Sensitivity	.80	.80	.80	.80
	95% CI	[.78, .83]	[.78, .83]	[.78, .82]	[.78, .82]
	Specificity	.69	.80	.81	.75
	95% CI	[.57, .80]	[.71, .90]	[.71, .90]	[.67, .83]
	NPV	.17	.18	.17	.22
	PPV	.98	.99	.99	.98
	ρ	.95	.95	.95	.93
	τ	.78	.77	.77	.77

Note: Thresholds based on SAT10 criterion values: 20th, 40th, and 60th percentile for at risk, some risk, and target. *A* represents the area under the ROC curve; NPV = negative predictive value; PPV = positive predictive value; ρ = base rate; and τ = proportion screened positive. Thresholds bolded if $A \geq .75$.

Table A2

Optimal Decision Thresholds for Phoneme Segmentation Fluency

Statistic		Kindergarten		F	1st	
		W	S		W	S
At Risk	A	.75	.73	.71	.67	.60
	95% CI	[.72, .78]	[.71, .76]	[.69, .74]	[.64, .69]	[.58, .63]
	Threshold	21	48	37	53	59
	Sensitivity	.80	.80	.80	.80	.83
	95% CI	[.78, .83]	[.78, .83]	[.78, .83]	[.78, .82]	[.81, .85]
	Specificity	.52	.48	.44	.39	.24
	95% CI	[.47, .57]	[.44, .53]	[.41, .48]	[.35, .42]	[.21, .27]
	NPV	.56	.53	.60	.56	.49
	PPV	.78	.77	.68	.66	.63
	ρ	.68	.69	.60	.60	.60
	τ	.70	.71	.70	.73	.80
	Some Risk	A	.76	.75	.71	.65
95% CI		[.72, .80]	[.72, .79]	[.69, .74]	[.62, .69]	[.55, .62]
Threshold		26	50	40	55	59
Sensitivity		.81	.81	.80	.81	.81
95% CI		[.79, .83]	[.78, .83]	[.78, .82]	[.79, .83]	[.79, .83]
Specificity		.54	.54	.44	.35	.26
95% CI		[.45, .62]	[.46, .62]	[.38, .49]	[.30, .40]	[.22, .31]
NPV		.26	.25	.32	.27	.22
PPV		.94	.94	.87	.86	.85
ρ		.89	.90	.83	.83	.83
τ		.77	.77	.76	.78	.80

Statistic	Kindergarten		F	1st	
	W	S		W	S
Target					
A	.79	.75	.74	.65	.59
95% CI	[.74, .84]	[.69, .80]	[.71, .78]	[.61, .70]	[.55, .64]
Threshold	27	51	41	56	59
Sensitivity	.80	.80	.80	.81	.81
95% CI	[.78, .82]	[.78, .83]	[.78, .82]	[.80, .83]	[.79, .82]
Specificity	.58	.48	.44	.30	.27
95% CI	[.46, .70]	[.36, .60]	[.35, .53]	[.22, .39]	[.19, .34]
NPV	.14	.11	.14	.10	.09
PPV	.97	.97	.95	.94	.94
ρ	.95	.95	.93	.93	.94
τ	.78	.79	.79	.81	.80

Note: Thresholds based on SAT10 criterion values: 20th, 40th, and 60th percentile for at risk, some risk, and target. A represents the area under the ROC curve; NPV = negative predictive value; PPV = positive predictive value; ρ = base rate; and τ = proportion screened positive. Thresholds bolded if $A \geq .75$.

Table A3

Optimal Decision Thresholds for Nonsense Word Fluency

		Kindergarten			1st		2nd
Statistic		W	S	F	W	S	F
At Risk	A	.82	.82	.82	.82	.80	.78
	95% CI	[.79, .84]	[.80, .85]	[.81, .84]	[.80, .84]	[.78, .82]	[.76, .81]
	Threshold	9	30	18	49	67	55
	Sensitivity	.81	.82	.81	.81	.80	.81
	95% CI	[.79, .84]	[.79, .84]	[.78, .83]	[.79, .83]	[.78, .83]	[.78, .83]
	Specificity	.69	.64	.67	.65	.62	.60
	95% CI	[.64, .73]	[.60, .69]	[.63, .70]	[.61, .68]	[.59, .65]	[.56, .63]
	NPV	.63	.61	.70	.69	.67	.70
	PPV	.84	.83	.78	.78	.76	.72
	ρ	.68	.69	.60	.60	.60	.57
	τ	.65	.67	.62	.63	.64	.63
	Some Risk	A	.85	.88	.84	.83	.81
95% CI		[.82, .89]	[.85, .91]	[.82, .86]	[.80, .85]	[.78, .83]	[.77, .82]
Threshold		13	34	24	55	75	64
Sensitivity		.82	.80	.82	.80	.80	.80
95% CI		[.80, .84]	[.78, .83]	[.80, .84]	[.78, .82]	[.78, .82]	[.78, .82]
Specificity		.76	.78	.69	.70	.63	.59
95% CI		[.68, .83]	[.71, .85]	[.64, .74]	[.65, .75]	[.57, .68]	[.54, .65]
NPV		.34	.32	.43	.41	.38	.39
PPV		.97	.97	.93	.93	.92	.90
ρ		.89	.90	.83	.83	.83	.82
τ		.76	.74	.73	.72	.73	.73

	Statistic	Kindergarten		1st		2nd	
		W	S	F	W	S	F
Target	A	.88	.90	.87	.85	.84	.79
	95% CI	[.83, .93]	[.87, .93]	[.84, .90]	[.82, .89]	[.81, .88]	[.75, .83]
	Threshold	14	36	27	58	79	70
	Sensitivity	.81	.80	.81	.80	.80	.81
	95% CI	[.79, .84]	[.78, .83]	[.80, .83]	[.78, .82]	[.78, .82]	[.79, .83]
	Specificity	.83	.82	.76	.74	.73	.63
	95% CI	[.74, .92]	[.73, .91]	[.69, .84]	[.66, .81]	[.66, .81]	[.54, .73]
	NPV	.19	.18	.23	.21	.21	.17
	PPV	.99	.99	.98	.98	.98	.97
	ρ	.95	.95	.93	.93	.94	.94
	τ	.78	.77	.78	.77	.77	.78

Note: Thresholds based on SAT10 criterion values: 20th, 40th, and 60th percentile for at risk, some risk, and target. *A* represents the area under the ROC curve; NPV = negative predictive value; PPV = positive predictive value; ρ = base rate; and τ = proportion screened positive. Thresholds bolded if $A \geq .75$.

Table A4

Optimal Decision Thresholds for Oral Reading Fluency

Statistic	1st		2nd			3rd		S
	W	S	F	W	S	F	W	
At Risk A	.91	.93	.86	.89	.88	.78	.78	.77
95% CI	[.89, .92]	[.92, .94]	[.84, .88]	[.87, .90]	[.87, .90]	[.76, .81]	[.75, .81]	[.75, .80]
Threshold	13	32	30	58	82	59	77	102
Sensitivity	.82	.81	.80	.80	.80	.80	.81	.80
95% CI	[.80, .84]	[.79, .83]	[.78, .83]	[.78, .83]	[.78, .83]	[.77, .84]	[.77, .84]	[.77, .84]
Specificity	.83	.90	.75	.82	.77	.58	.60	.57
95% CI	[.80, .85]	[.88, .92]	[.72, .78]	[.79, .84]	[.74, .80]	[.55, .62]	[.56, .63]	[.53, .61]
NPV	.76	.76	.74	.75	.74	.82	.82	.81
PPV	.88	.93	.81	.86	.83	.55	.57	.55
ρ	.60	.60	.57	.58	.58	.39	.40	.40
τ	.56	.53	.57	.54	.56	.57	.56	.58
Some A	.92	.93	.87	.89	.87	.76	.76	.76
Risk 95% CI	[.91, .94]	[.92, .94]	[.85, .89]	[.87, .91]	[.85, .89]	[.73, .79]	[.73, .79]	[.73, .79]
Threshold	19	48	40	75	97	71	91	111
Sensitivity	.82	.81	.80	.80	.81	.80	.81	.81
95% CI	[.80, .84]	[.79, .83]	[.78, .83]	[.78, .82]	[.79, .83]	[.78, .83]	[.78, .83]	[.78, .84]
Specificity	.86	.90	.74	.81	.76	.56	.55	.57
95% CI	[.83, .90]	[.86, .93]	[.69, .79]	[.77, .85]	[.71, .81]	[.51, .61]	[.50, .60]	[.52, .62]
NPV	.49	.49	.45	.46	.45	.58	.57	.59
PPV	.97	.98	.94	.95	.94	.79	.79	.80
ρ	.83	.83	.82	.83	.83	.67	.68	.68
τ	.70	.69	.71	.70	.71	.69	.69	.69

Statistic	1st		2nd			3rd		
	W	S	F	W	S	F	W	S
Target A	.94	.94	.87	.88	.87	.80	.80	.79
95% CI	[.92, .95]	[.93, .96]	[.84, .90]	[.85, .91]	[.84, .90]	[.75, .84]	[.76, .84]	[.74, .83]
Threshold	22	54	47	81	104	76	97	117
Sensitivity	.81	.81	.80	.80	.80	.81	.80	.80
95% CI	[.79, .82]	[.79, .83]	[.78, .82]	[.78, .82]	[.79, .82]	[.78, .83]	[.78, .83]	[.78, .82]
Specificity	.91	.96	.77	.79	.77	.63	.66	.60
95% CI	[.86, .96]	-	[.69, .85]	[.71, .87]	[.68, .85]	[.55, .71]	[.58, .74]	[.52, .69]
NPV	.25	.26	.20	.20	.19	.28	.28	.27
PPV	.99	1.00	.98	.98	.98	.95	.95	.94
ρ	.93	.94	.94	.94	.94	.89	.89	.89
τ	.76	.76	.77	.77	.77	.76	.75	.76

Note: Thresholds based on SAT10 criterion values: 20th, 40th, and 60th percentile for at risk, some risk, and target. A represents the area under the ROC curve; NPV = negative predictive value; PPV = positive predictive value; ρ = base rate; and τ = proportion screened positive. Thresholds bolded if $A \geq .75$.