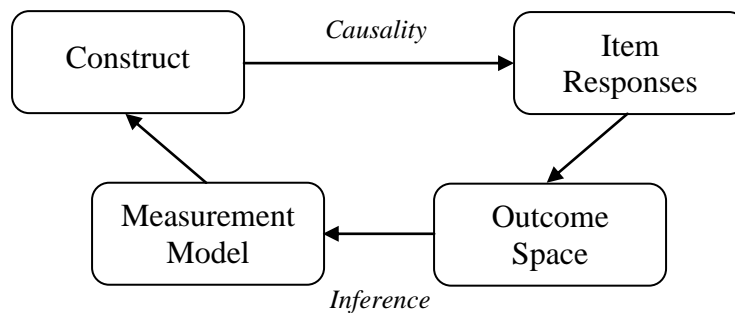


**Using Contrasting Groups Design to Build IGDI Identification Bundles**

An important functional step in the RTI framework is the identification of children who are likely to benefit from Tier 2 or 3 interventions. This identification process is improved through the collection and use of relevant information. That is, information relevant to the skills seen as important to successful literacy and language development through early childhood and early elementary school is important to assess and monitor to ensure successful progression. IGDI 2.0 were developed with the intention of contribution to this information gathering process. The IGDI 2.0s are being evaluated as identification and progress monitoring tools. The first component, described here, includes the identification process.



Wilson's (2005) item response model approach to measurement construction provides a framework through which the development of the IGDI 2.0 is supported. In this framework, the recognition of different purposes for the IGDI 2.0s (identification or progress monitoring) is embodied in the Outcome Space and related inferences through a Measurement Model to the Construct. The Outcome Space allows us to define the features of the response to assessment that we value – for the purpose of identification, it is range of performance that indicates likelihood of benefit from Tier 2 or 3 placement, given relevant location on the measurement or construct map. Through the Measurement Model (Rasch) we rely on the construct map (through the item map) to locate the tasks that most efficiently identify children at the point of concern.

To facilitate identification, cut scores or cut score ranges (as described below), need to be selected on each of the high performing IGDI 2.0s. To support this process, the Validity framework described a standard setting process known as Contrasting Groups Design. Through this process, we are able to identify a point on the IGDI score scale that optimally distinguishes or contrasts known groups – groups of children likely to benefit from Tier 2 or 3 placement, based on teacher judgment. This

process used in the CRtIEC initial standard setting study, based on end-of-year performance of 4 or 5 year old children prior to entering kindergarten, included the following steps.

1. Teachers of children at the end of the year prior to entering kindergarten were invited to complete a child-performance survey, without information on performance from the assessments.
2. Teachers were asked to place children into a Tier level (1, 2, or 3), based on their understanding of the performance level from the tier level descriptors (TLDs). These assignments were made for each of the domains independently, including (a) oral language, (b) phonological awareness, and (c) alphabet knowledge.
3. Children were assessed on the IGDIs and the distributions on the actual measures for each performance level were compared.
4. The points (cut-scores) that discriminate among children between tier levels were estimated using multiple methods to assess agreement and sensitivity to method, including
  - a. ROC analysis to achieve a balance between Sensitivity and Specificity, with a minimum Sensitivity of .70;
  - b. A check on ROC analysis through logistic regression and classification accuracy;
  - c. Analysis of the effect of the selected cut score on the distributions of scores.
5. This cut score estimation process was conducted on Wave 1 data directly based on the logit metric (measure from Winsteps) and the True Score (estimated from a test of 20 cards), based on Wave 4 Teacher Tier Placements of children. The process was replicated on Wave 4 data (setting the cut score at Wave 4), and then using the Wave 4 cut score to predict the Wave 1 cut score.

The following process provides the technical background and statistical results in the series of analyses employed to estimate the appropriate cut scores for the Identification Bundles to be used to inform Tier placement decisions.

Two major steps to complete this process include:

1. Estimate the length of Identification Bundles and cut score ranges to secure number correct scores that will inform decisions such that:
  - a. Below some number, we highly recommend Tier 2 placement;
  - b. Within some range of scores we recommend additional information to make the placement decision; and
  - c. Above some number, we recommend Tier 1 placement (business as usual).

This could be achieved by constructing a test with the number of items that yields a confidence interval around the cut scores that is as wide as the range of scores below that confidence interval.

2. Recommend Identification Bundle length and appropriate cards to include.

Through the use of IGDI 2.0 Identification bundles, information is obtained that is construct and instructionally relevant given the RTI framework developed for early childhood education. IGDI results should be used within a clearly defined decision-making process as one source of information in a multiple-measures decision-making framework. No important educational decisions should be made with a single piece of information and certainly not with a single test score.

*Note:* An important limitation of these analyses should be taken into consideration. The teacher placement decisions were completed at the end of the year based on end-of-year skill-level expectations. The cut scores are being set for fall decision making. To reduce potential bias, IGDI scores at both waves 1 and 4 were scaled on the Wave 1 scale through Rasch equating.

**Oral Language**  
Cut Score Set with Picture Naming

*Selection of Cut Score with Picture Naming Wave 1 Logit Scores*

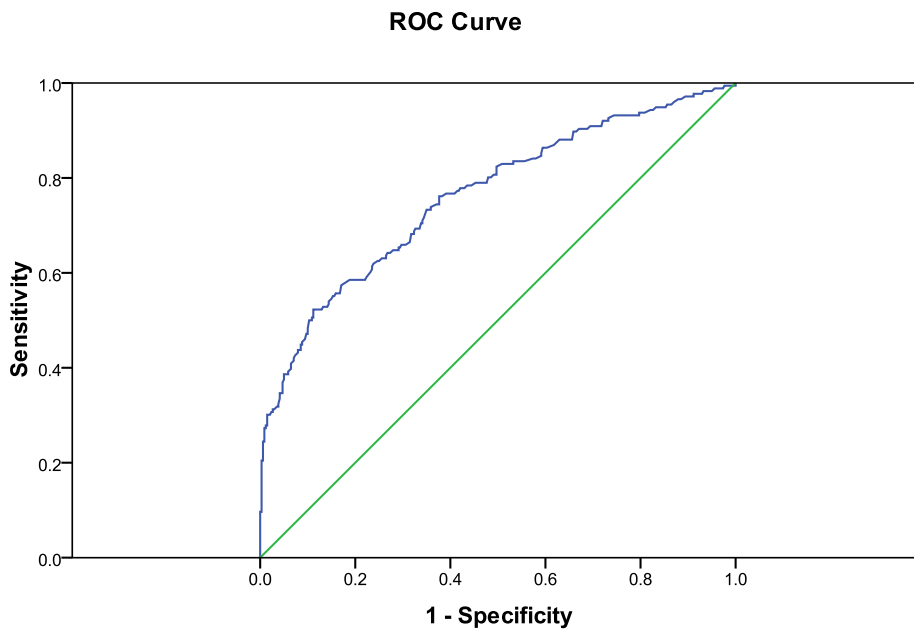
ROC analysis employing the Picture Naming Wave 1 measure logit score (Test variable) in classifying children on Oral Language Tier Levels (State variable) as placed by teachers. The test statistic associated with a ROC analysis, signifying the statistical significance of classification, is based on the area under the curve. “Area” is the probability that a score for a randomly selected positive case (Tier 2/3) is lower than the score for a randomly selected negative case (Tier 1). The ROC analysis suggests successful classification based on PN, Area = .782,  $p < .001$ .

*Area Under the Curve*

Test Result Variable: Picture Naming Measure Wave 1

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.762	.023	.000	.717	.807

- a. Under the nonparametric assumption
- b. Null hypothesis: true area = 0.5



Diagonal segments are produced by ties.

*Figure 1.* ROC curve of Sensitivity/Specificity balance for Picture Naming Wave 1.

The cut score that yields .70 sensitivity and .66 specificity is 1.90 logits. Although based on Figure 2, the maximal value on both metrics is about .68, which is associated with a score of 1.81.

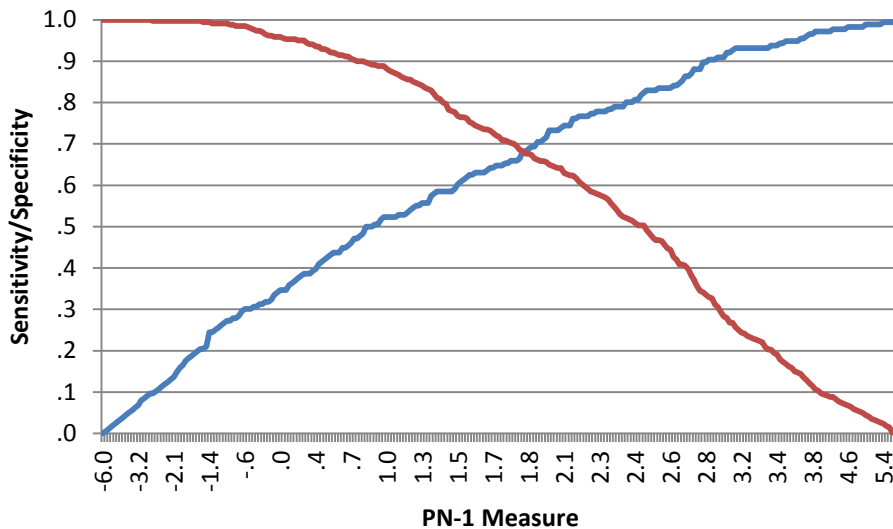


Figure 2. Illustration of the point of intersection of sensitivity (increasing line) and specificity (decreasing line) for PN Wave 1.

*Selection of Cut Score Using Picture Naming Wave 4 Measure*

The ROC analysis suggests successful classification on Oral Language Tier Placement based on teacher placements, Area = .752,  $p < .001$ .

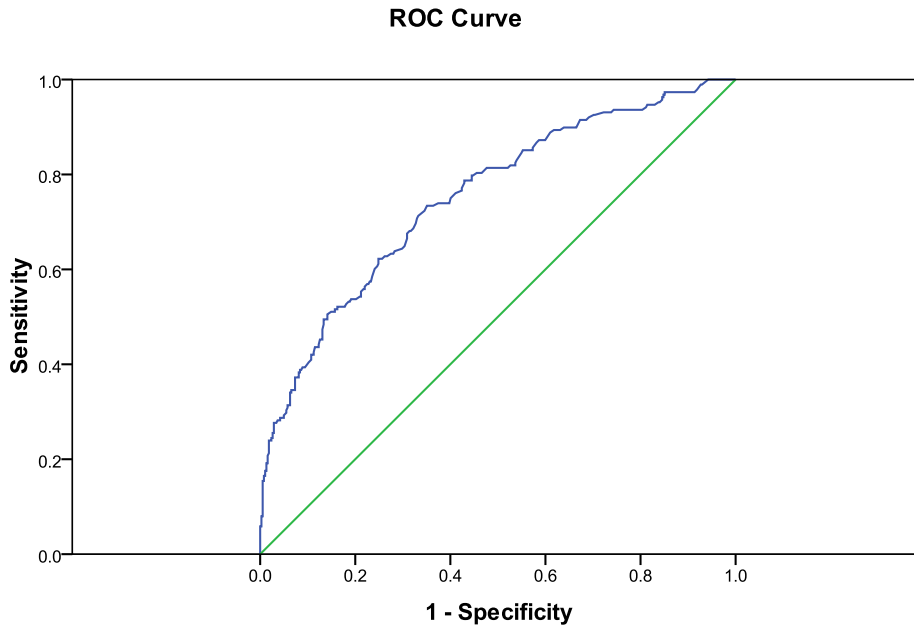
*Area Under the Curve*

Test Result Variable(s):PN Measure Wave 4

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.752	.022	.000	.709	.795

a. Under the nonparametric assumption

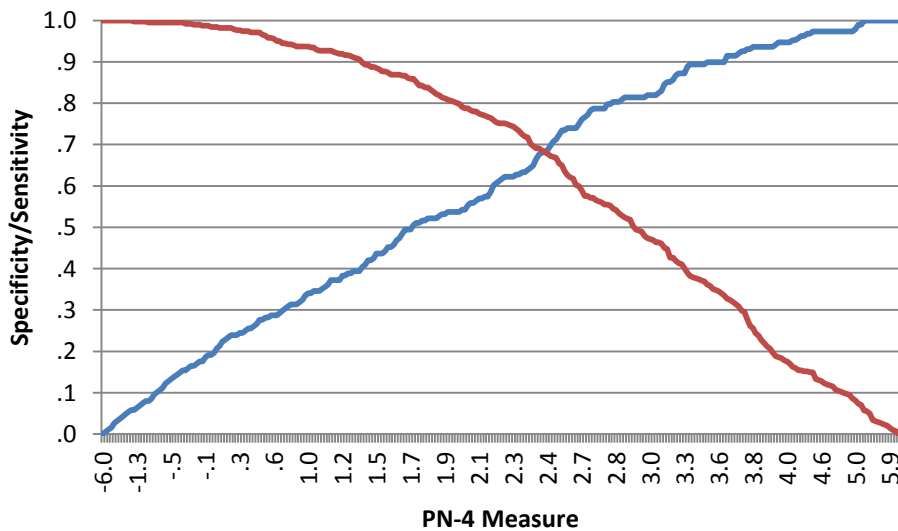
b. Null hypothesis: true area = 0.5



Diagonal segments are produced by ties.

*Figure 3.* ROC curve of Sensitivity/Specificity balance for Picture Naming Wave 4.

The logit associated with .70 sensitivity and .67 specificity is 2.46. From Figure 4, the simultaneous maximum value on both metrics is .68 and is associated with a logit value of 2.43.



*Figure 4.* Illustration of the point of intersection of sensitivity (increasing line) and specificity (decreasing line) for PN Wave 4.

To create a prediction model using the sample of children with TLD classifications and Wave 1 and 4 data ( $n=500$ ), the following is a regression of Wave 1 on Wave 4 Picture Naming measures. The Wave 4 cut score based on ROC analysis is 2.46 logits, with a Wave 1 predicted score of 1.74.

Regression Coefficients Regressing Picture Naming Wave 1 on PN Wave 4

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1	(Constant)	-.393	.105		
	PN Wave 4	.866	.035	.742	24.721

$$\text{Wave 1 Cut Score} = -0.393 + 0.866 (\text{Wave 4 Cut Score}) = -0.393 + 0.866 (2.46) = 1.74$$

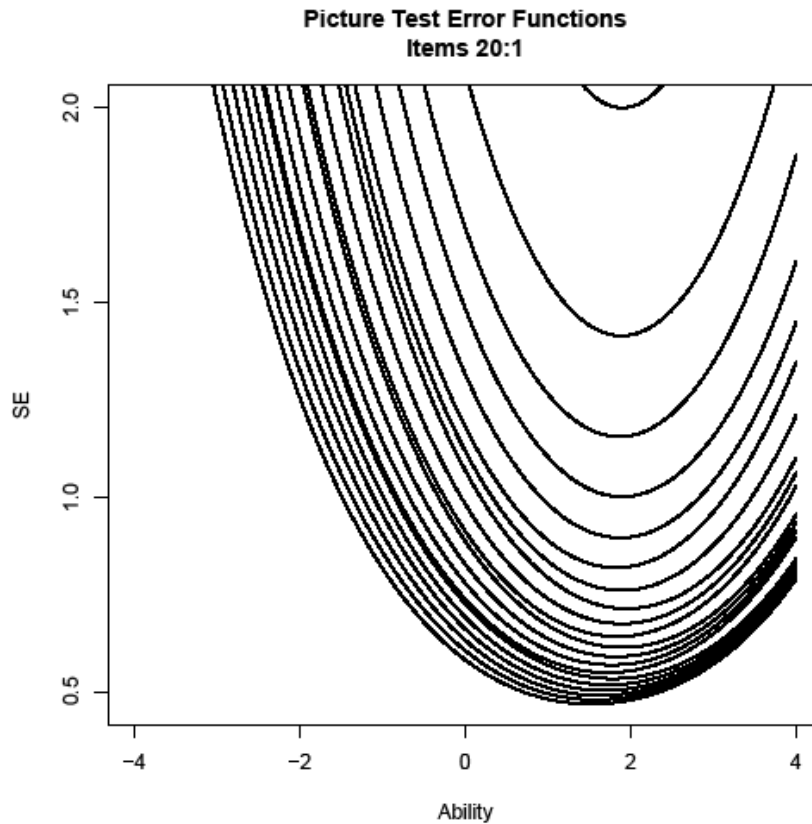


Figure 5. Picture Naming Identification Bundle measurement error functions given number of items.

In Figure 5, we illustrate the measurement error function across the ability range (logits) for every possible test length from 20 to 1 item. The lowest curve is based on 20 items and illustrates that the smallest SE is found within the range of approximately 0.5 to 2.5 logits. This is based on identifying 20 items most near the cut score (1.90), and successively removing the one item furthest from the cut point. You can see that tests of 4 or fewer items have SEs larger than 1.0.

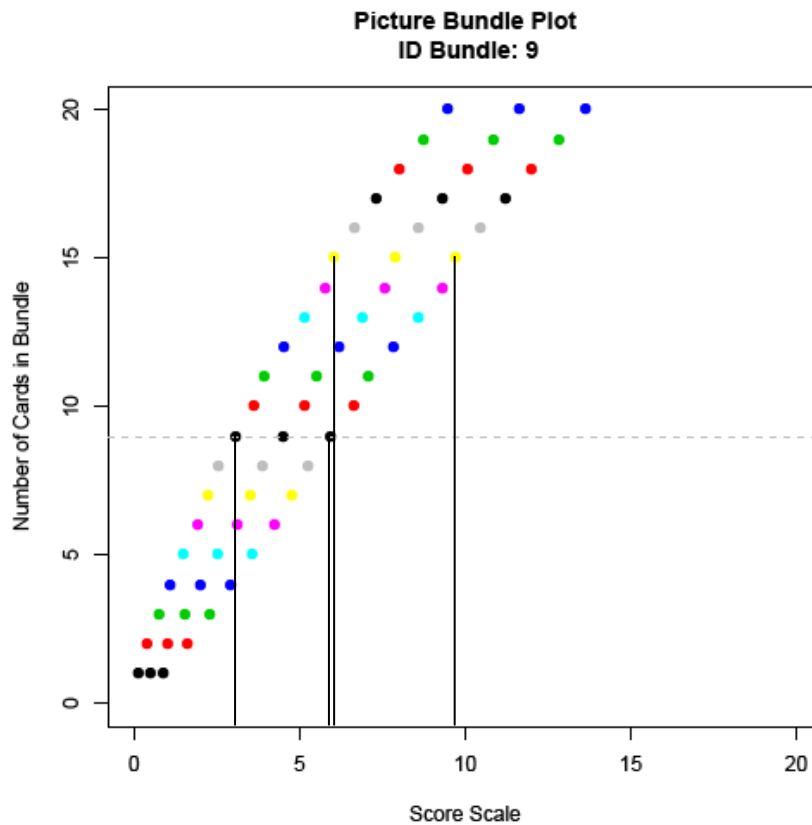


Figure 6. Picture Naming Identification Bundle plots of number of cards and 1 standard error confidence intervals around the cut score in True Score metric (Score Scale is the True Score).

In Figure 6, we can see the cut score range (cut score  $\pm 1$  SE) for every possible bundle length from 1 to 20 cards. For instance, the bundle with 9 cards has a cut score range approximately 3 to 6. This is probably the smallest bundle that provides a lower range (0 to 3) that is as wide as the cut score range (3 to 6). The recommendation would be: children with scores of 3 or less are likely to benefit from Tier 2 placement; scores of 3 to 6 are likely to need additional information for appropriate placement. The bundle with 15 cards has a cut score range of approximately 6 to 10.



### Analysis of Cut Score Performance

Frequency of Tier Placement based on PN Cut Score of 1.90

		Frequency	Valid Percent
Valid	Tier 2/3	239	46.3
	Tier 1	277	53.7
	Total	516	100.0

Based on the cut score of 1.90 logits, 46% of children in the TLD study would be below the cut score – an estimate of the proportion that might be placed in Tier 2.

### Logistic Regression Classification

Predicting Tier placement with Wave 1 measure ( $n=516$ ), with cut value based on .70 sensitivity.

Logistic Regression Wave 1 Classification Table based on goal of .70 sensitivity

Observed			Predicted		Percentage Correct
			Oral Language		
			0 Tier 2/3	1 Tier 1	
Step 1	Oral Language	0 Tier 2/3	124	52	70.5
		1 Tier 1	116	224	65.9
Overall Percentage					67.4

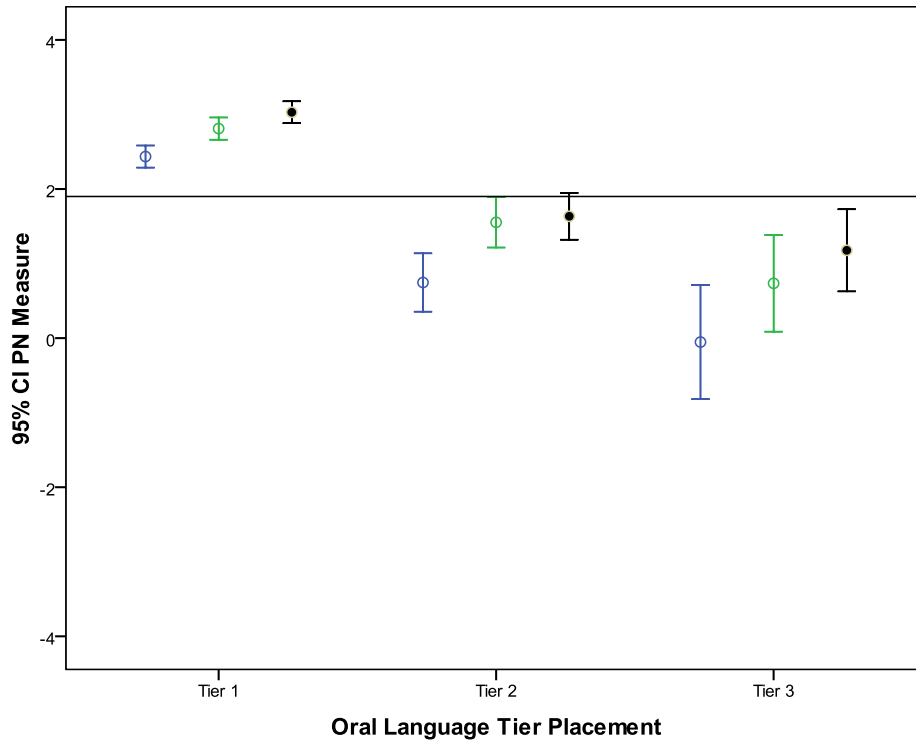
Here we see the classification accuracy of predicting Tier 1 membership. The percent correct is somewhat balanced between Tier 1 and Tier 2/3. The total classification accuracy rate is 67% (based on teacher Tier placements of children). This mirrors the results from the optimal cut score selected through the ROC analysis (as expected).

The logistic regression at the point where a child has a 50% chance of belonging in Tier 1 or Tier 2 provides greater overall classification accuracy (nearly 76%), but identifies far fewer children in Tiers 2/3: 43% compared to 71% when balancing sensitivity/selectivity.

Logistic Regression Wave 1 Classification Table based on .50 prediction

Observed			Predicted		Percentage Correct
			Oral Language		
			0 Tier 2/3	1 Tier 1	
Step 1	Oral Language	0 Tier 2/3	75	101	42.6
		1 Tier 1	25	315	92.6
Overall Percentage					75.6

*Illustrations of how the selected Cut Score functions via PN Score Distributions*



*Figure 7.* Picture Naming means and 95% confidence intervals across Waves 1, 3, and 4 by OL Tier placements.

Figure 7 illustrates PN growth (based on mean performance) across three waves within each Teacher Tier placement decision. The proposed cut score of 1.90 logits is marked on the graph.

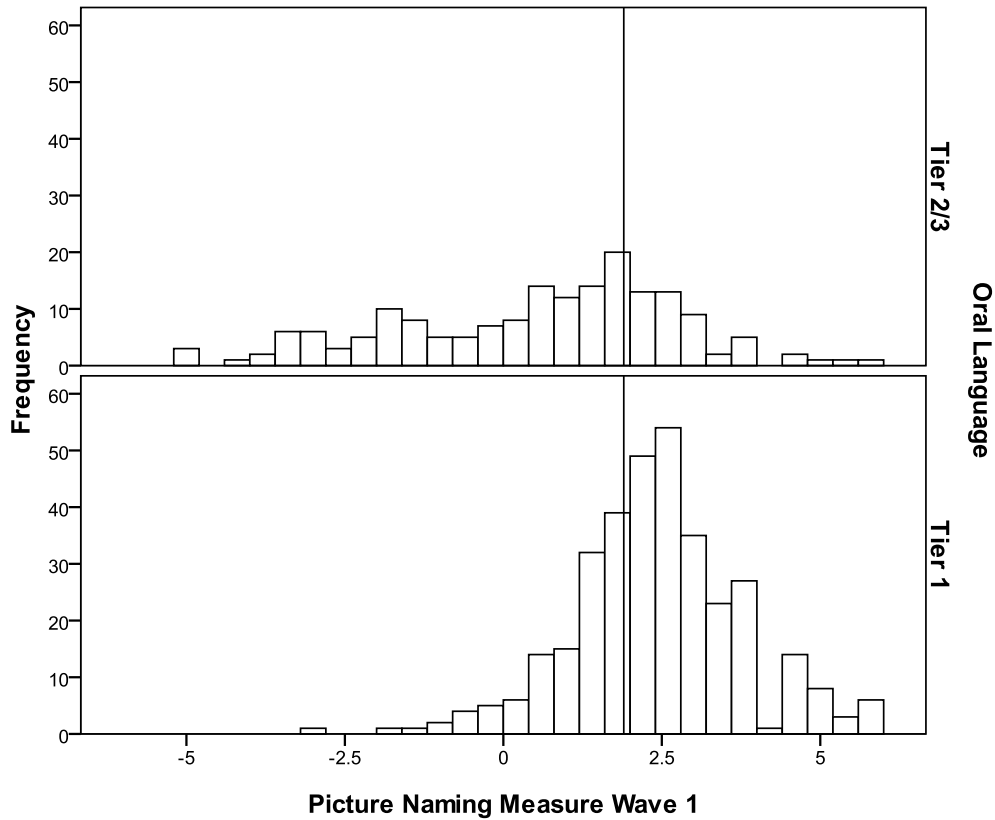
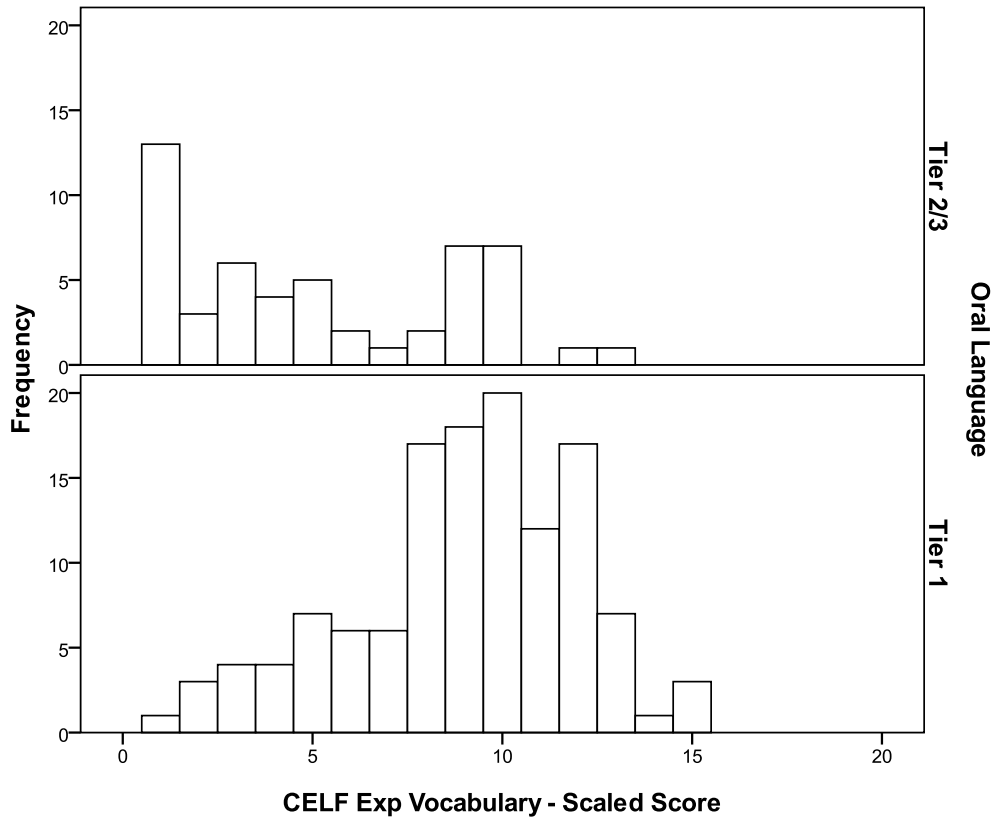


Figure 8. Picture Naming measure histograms by Tier placement.

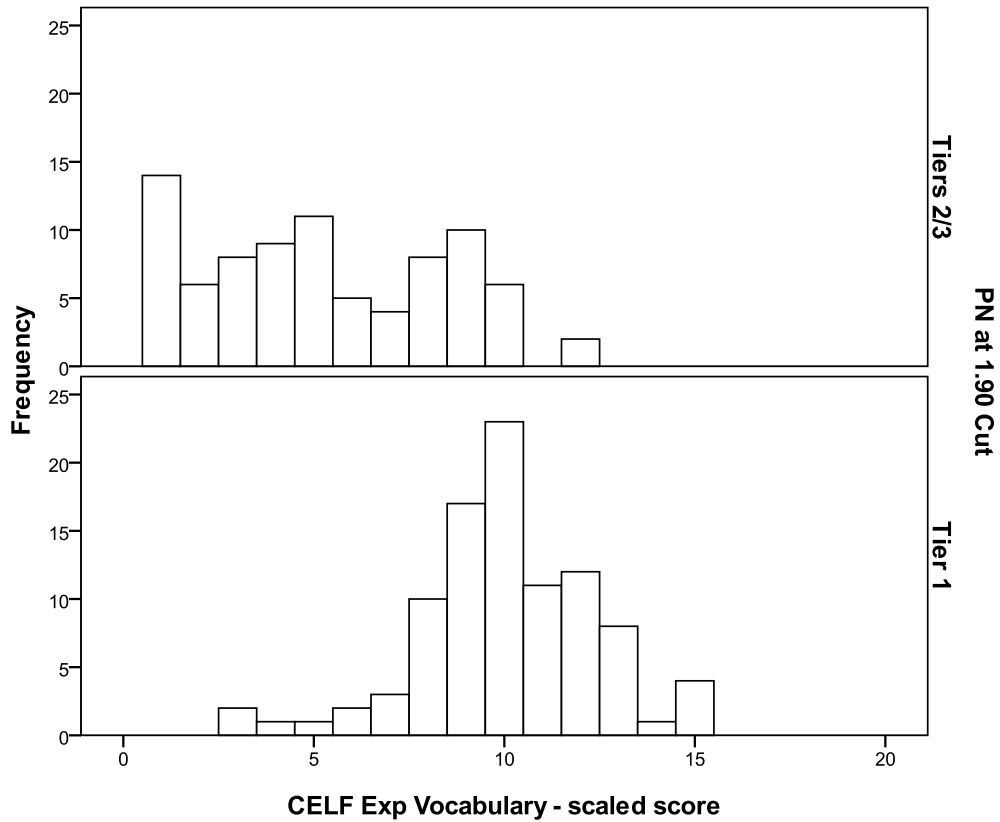
Figure 8 provides an illustration of the intersection (and overlap) of the score distributions of the two groups. The proposed cut score of 1.90 logits is marked on the graph.

*Examining Teacher Tier Placements vis-à-vis Criterion Measures*

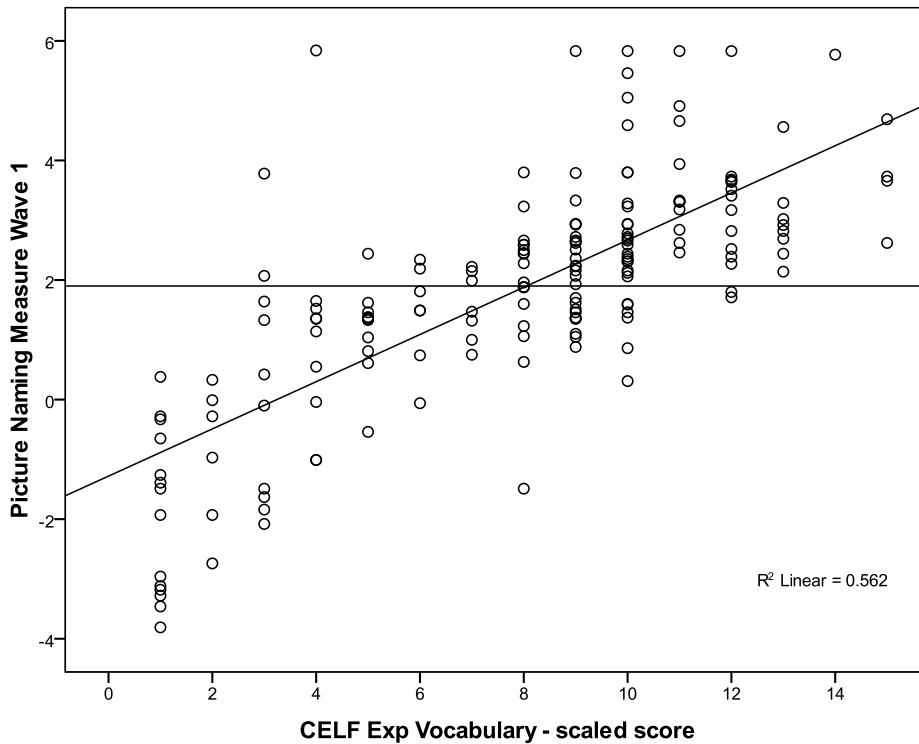
Picture Naming: CELFev and PPVT\_SS



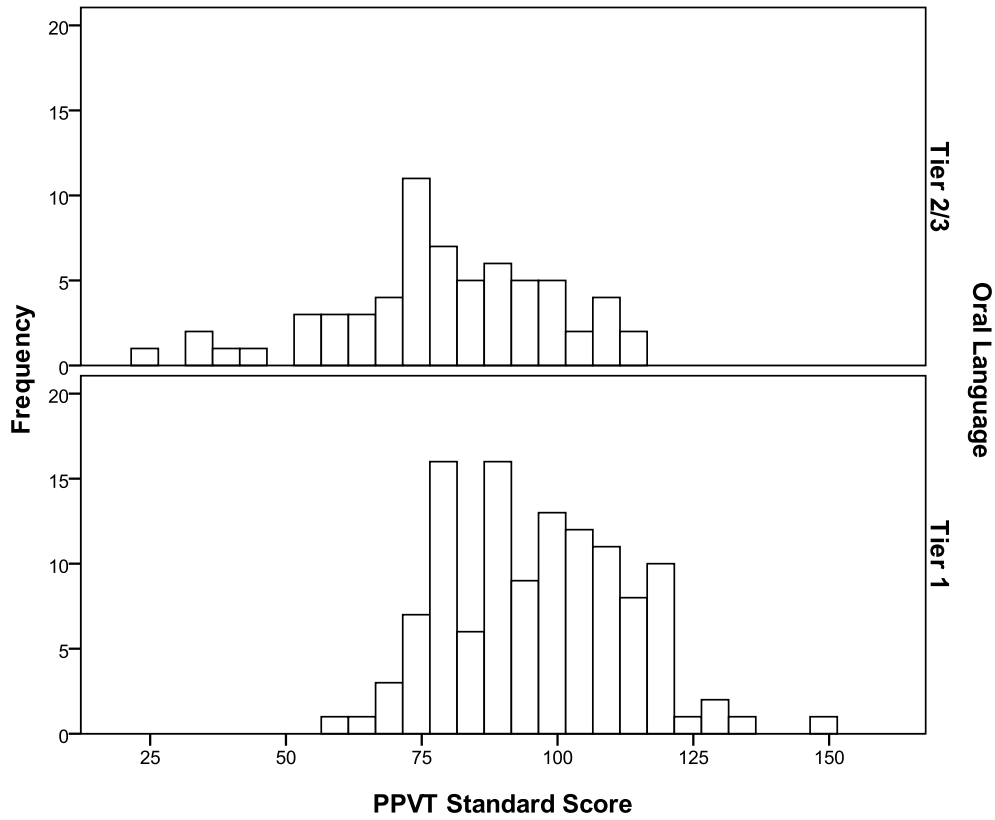
Here we see the distributions of scores on the CELF Expressive Vocabulary scale for the two levels of Tiers based on Teacher placement.



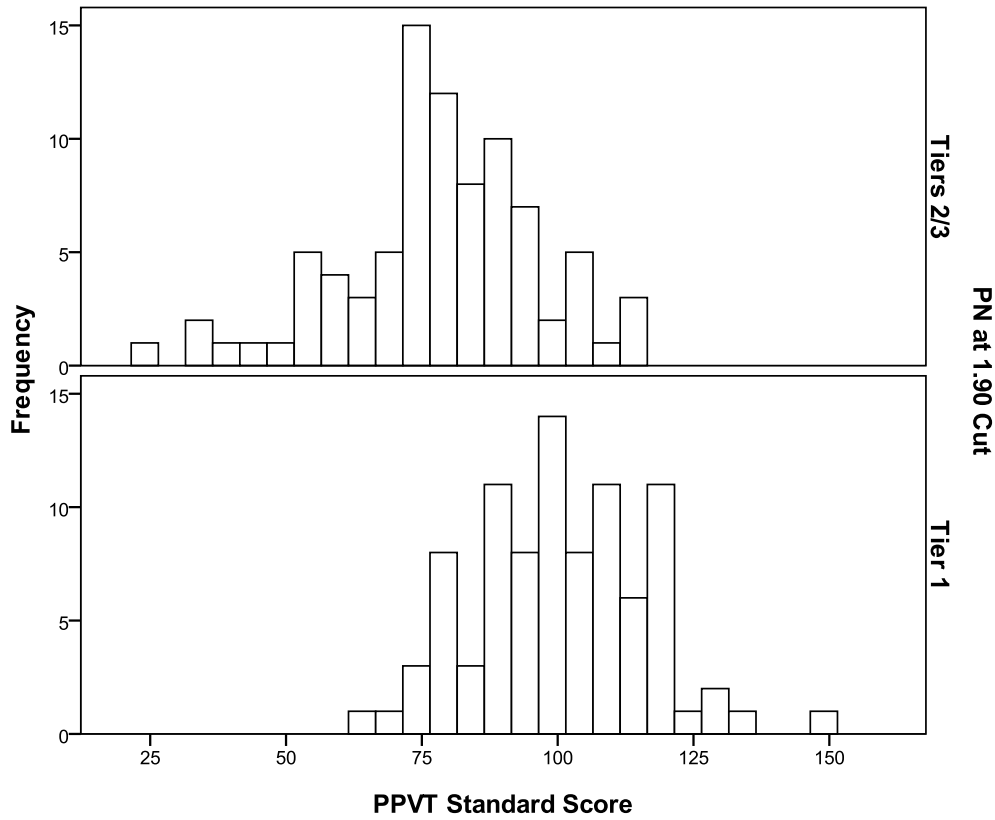
Here we see the distributions of scores on the CELF Expressive Vocabulary scale for the two levels of Tiers based on identification from the Picture Naming measure with the cut score set at 1.90 on the logit scale..



The correlation between the CELF EV and PN Measure is approximately .75. Here we see the intersection of the best-fit line (regression line) and the cut score on PN of 1.90. The point of intersection occurs at a score of 8 on the CELF. A CELF score of 8 is associated with the 25<sup>th</sup> percentile performance. In this data set, 37% of the sample scored below 8 (10% obtained a score of 8).

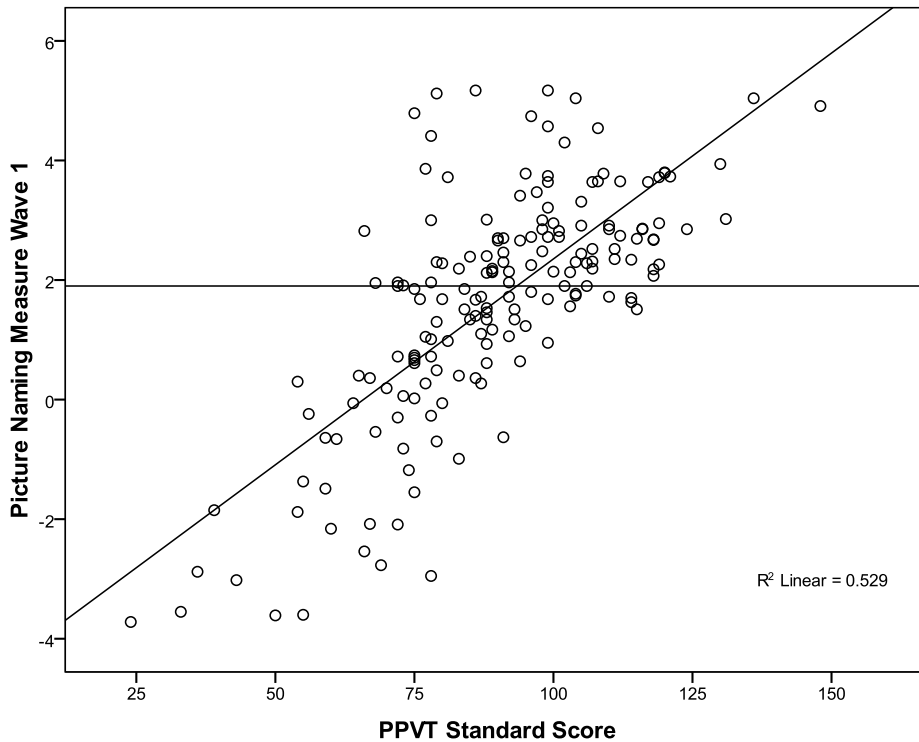


Here we see the distributions of scores on the PPVT scale for the two levels of Tiers based on Teacher placement.



Here we see the distributions of scores on the PPVT scale for the two levels of Tiers based on identification with Picture Naming with a cut score set at 1.90 logits.





Here we can observe the association between PPVT and Picture naming, with a correlation of .73. The point of intersection between the best-fitting line (regression line) and the PN cut score is approximately 93. The PPVT score of 93 is associated with a percentile of 32. In this sample, 56% scored below a 93 on the PPVT.

**Alphabet Knowledge**  
Cut Score Set with Sound Identification

*Selection of Cut Score with Sound Identification Wave 1 Logit Scores*

ROC analysis employing the Sound Identification Wave 1 measure logit score (Test variable) in classifying children on Alphabet Knowledge Tier Levels (State variable) as placed by teachers. The test statistic associated with a ROC analysis, signifying the statistical significance of classification, is based on the area under the curve. “Area” is the probability that a score for a randomly selected positive case (Tier 2/3) is lower than the score for a randomly selected negative case (Tier 1). The ROC analysis suggests successful classification based on SI, Area = .730,  $p < .001$ .

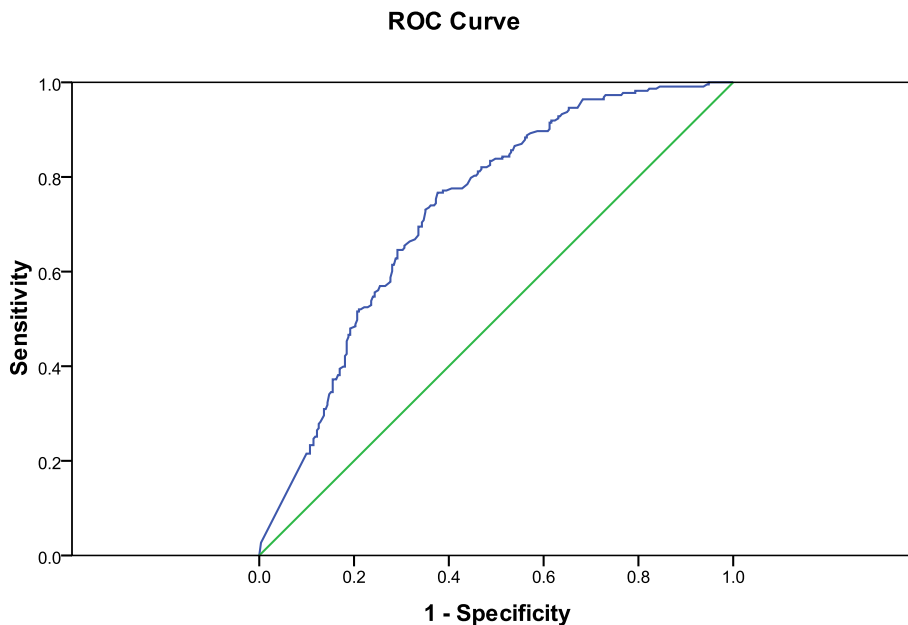
*Area Under the Curve*

Test Result Variable(s): Sound Identification Measure Wave 1

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.730	.022	.000	.686	.774

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5



Diagonal segments are produced by ties.

*Figure 9. ROC curve of Sensitivity/Specificity balance for Sound Identification Wave 1.*

The cut score associated with .70 sensitivity and .66 specificity is 0.05 logits.

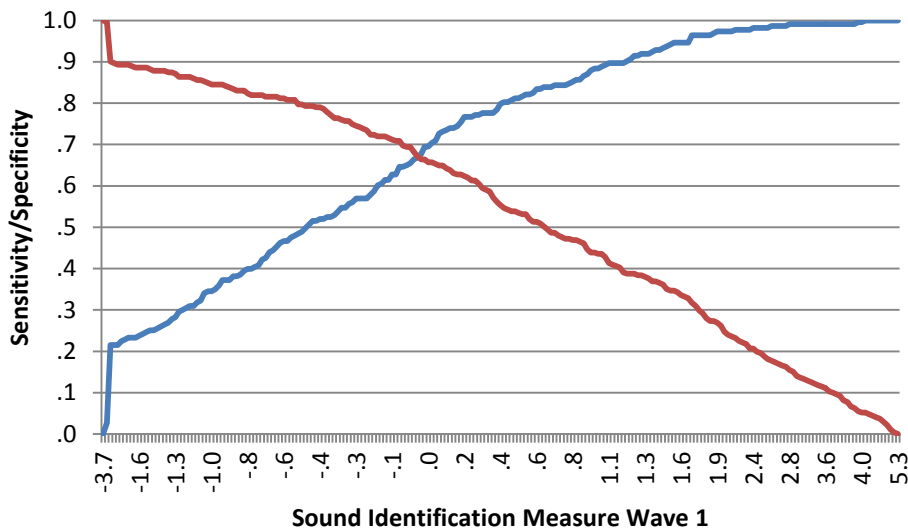


Figure 10. Illustration of the point of intersection of sensitivity (increasing line) and specificity (decreasing line) for SI Wave 1.

*Selection of Cut Score using Sound Identification Wave 4 Measure*

The ROC analysis suggests successful classification on Alphabet Knowledge Tier Placement based on teacher placements, Area = .855,  $p < .001$ . This classification appears to be more successful than the classification based on Wave 1.

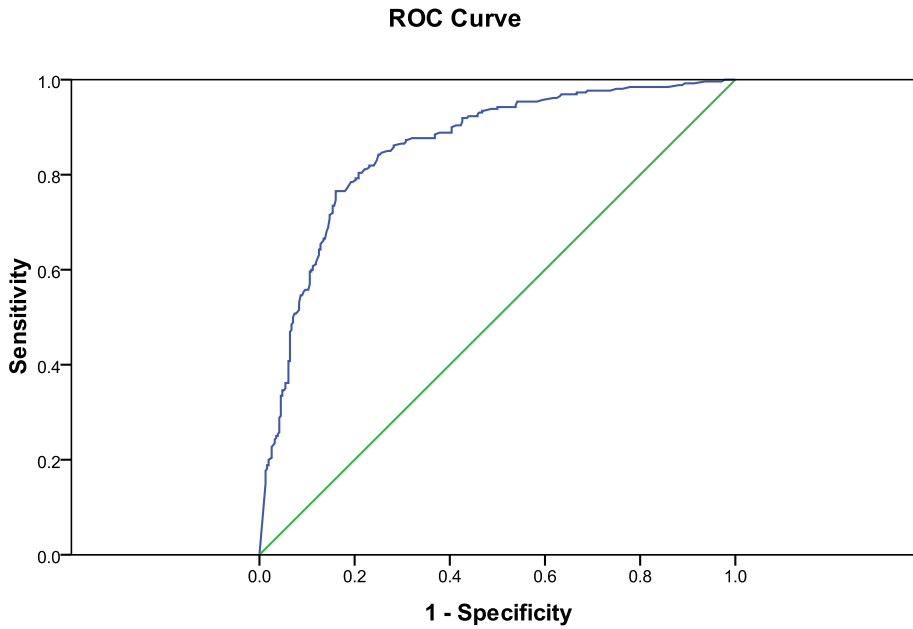
*Area Under the Curve*

Test Result Variable(s): Sound Identification Measure Wave 4

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.855	.016	.000	.824	.887

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5



Diagonal segments are produced by ties.

Figure 11. ROC curve of Sensitivity/Specificity balance for Sound Identification Wave 4.

From Figure 12, we see the measure associated with .80 sensitivity and .80 specificity is approximately 1.12.

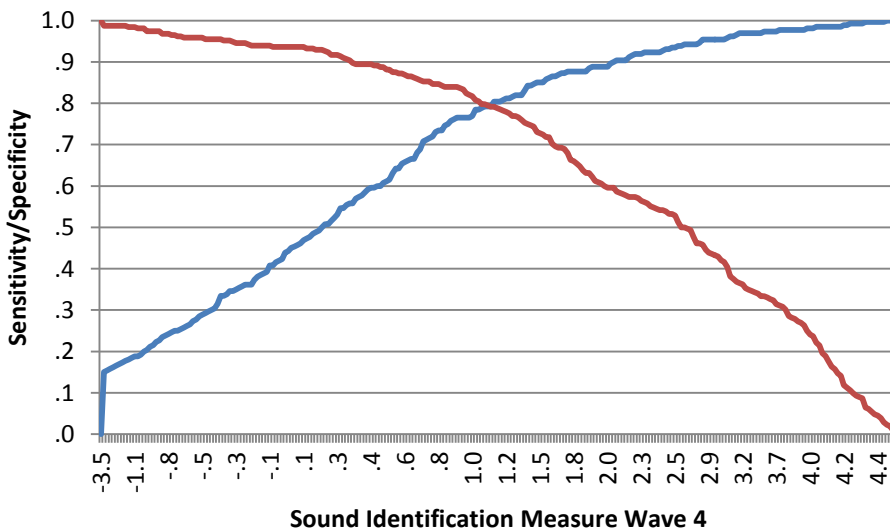


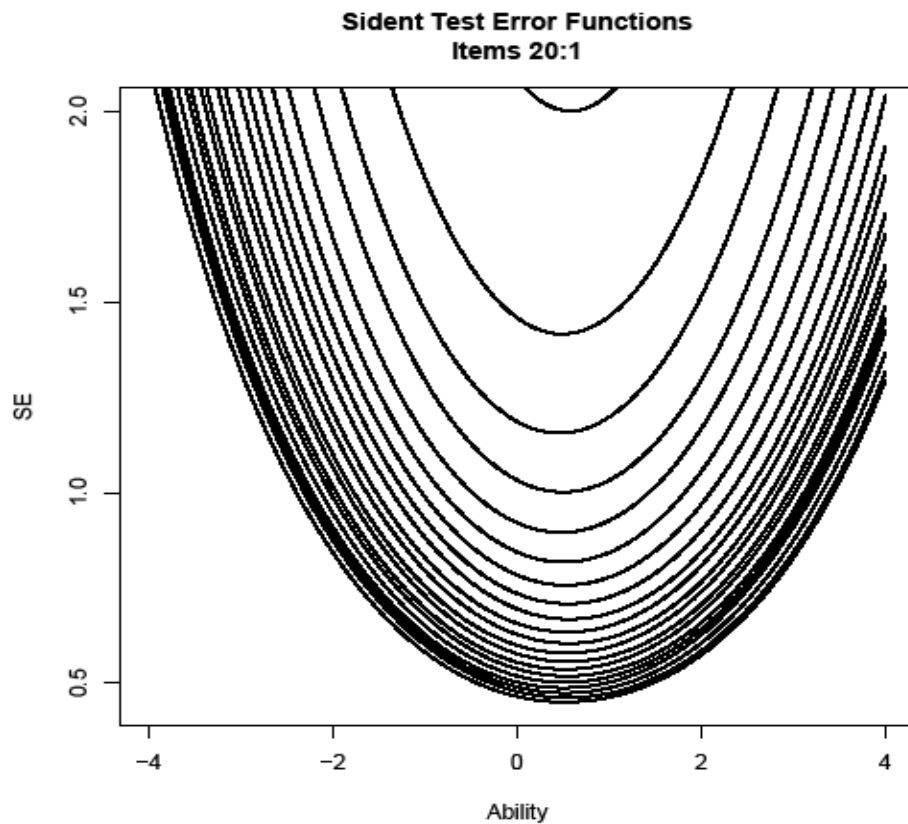
Figure 12. Illustration of the point of intersection of sensitivity (increasing line) and specificity (decreasing line) for SI Wave 4.

To create a prediction model using the sample of children with TLD classifications and Wave 1 and 4 data ( $n=483$ ), the following is a regression of Wave 1 on Wave 4 Sound Identification measures. The Wave 1 cut score based on ROC analysis is 0.05 logits, with a Wave 1 predicted score of 0.04.

**Regression Coefficients Regressing SI Wave 1 on SI Wave 4**

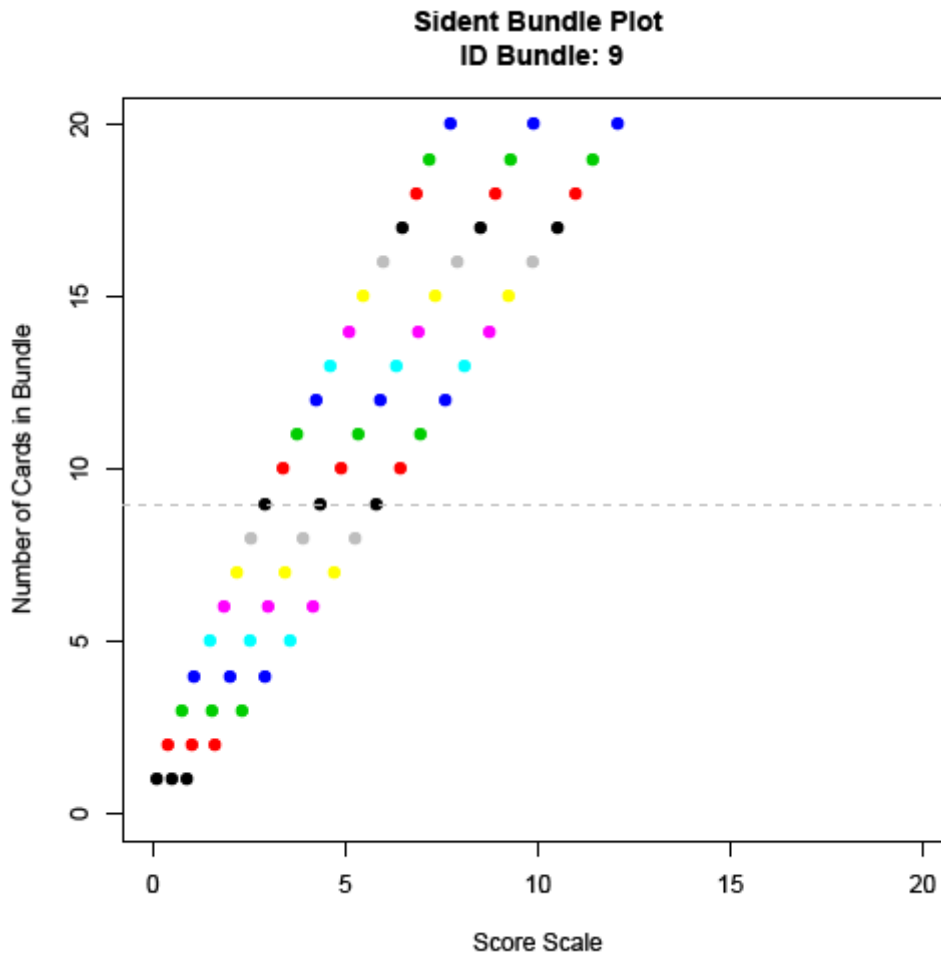
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.495	.088		-5.610	.000
	SI Wave 4	.469	.037	.504	12.795	.000

$$\text{Wave 1 Cut} = -0.495 + 0.469 (\text{Wave 4 Cut}) = -0.495 + 0.469 (1.14) = 0.04$$



*Figure 13.* Sound ID Identification Bundle measurement error functions given number of items.

In Figure 13, we illustrate the measurement error function across the ability range (logits) for every possible test length from 20 to 1 item. The lowest curve is based on 20 items and illustrates that the smallest SE is found within the range of approximately -0.5 to 1.5 logits. This is based on identifying 20 items most near the cut score (0.05), and successively removing the one item furthest from the cut point. You can see that tests of 4 or fewer items have SEs larger than 1.0.



*Figure 14.* Sound ID Identification Bundle plots of number of cards and associated 1 standard error confidence intervals around the cut score in True Score metric (Score Scale is True Score).

In Figure 14, we can see the cut score range (cut score  $\pm$  1 SE) for every possible bundle length from 1 to 20 cards. For instance, the bundle with 9 cards has a cut score range approximately 3 to 6. This is probably the smallest bundle that provides a lower range (0 to 3) that is as wide as the cut score range (3 to 6). The recommendation would be: children with scores of 3 or less are likely to benefit from Tier 2 placement; scores of 3 to 6 are likely to need additional information for appropriate placement. The bundle with 15 cards has a cut score range of approximately 6 to 10.

Notice that these results are similar to those from Picture Naming – this is because the items were selected around the cut score for both (both are centered at their respective cut scores).

### *Analysis of Cut Score Performance*

SI Cut Score of 0.05

		Frequency	Valid Percent
Valid	Tier 2/3	250	50.6
	Tier 1	244	49.4
	Total	494	100.0

Based on the cut score of 0.05 logits, 51% of children in the TLD study would be below the cut score – an estimate of the proportion that might be placed in Tier 2.

### *Logistic Regression Classification*

Predicting Tier placement with Wave 1 measure ( $n=494$ ), with cut value based on .70 sensitivity.

Logistic Regression Wave 1 Classification Table based on goal of .70 sensitivity

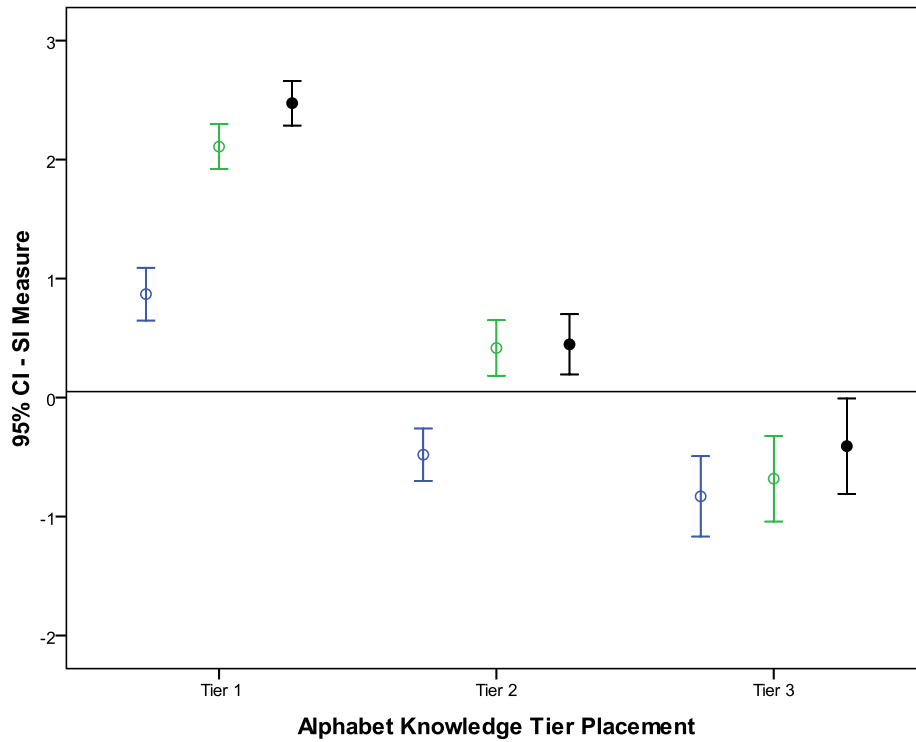
Observed			Predicted		Percentage Correct
			Alphabet Knowledge		
			0 Tier 2/3	1 Tier 1	
Step 1	Alphabet Knowledge	0 Tier 2/3	158	65	70.9
		1 Tier 1	94	177	65.3
Overall Percentage					67.8

Here we see the classification accuracy of predicting Tier 1 membership. The percent correct is slightly higher for Tier 2/3 (71%) than Tier 1 (65%).

The logistic regression at the point where a child has a 50% chance of belonging in Tier 1 or Tier 2 achieves slightly lower classification accuracy (66% overall), and also identifies far fewer children in Tiers 2/3: 55% compared to 71% when balancing sensitivity/selectivity.

Logistic Regression Wave 1 Classification Table based on .50 prediction

Observed			Predicted		Percentage Correct
			Alphabet Knowledge		
			0 Tier 2/3	1 Tier 1	
Step 1	Alphabet Knowledge	0 Tier 2/3	122	101	54.7
		1 Tier 1	66	205	75.6
Overall Percentage					66.2



*Figure 15.* Sound Identification means and 95% confidence intervals across Waves 1, 3, and 4 by AK Tier placements.

Figure 15 illustrates SI growth across three waves by Teacher Tier placement. The proposed cut score of 0.05 logits is marked on the graph.



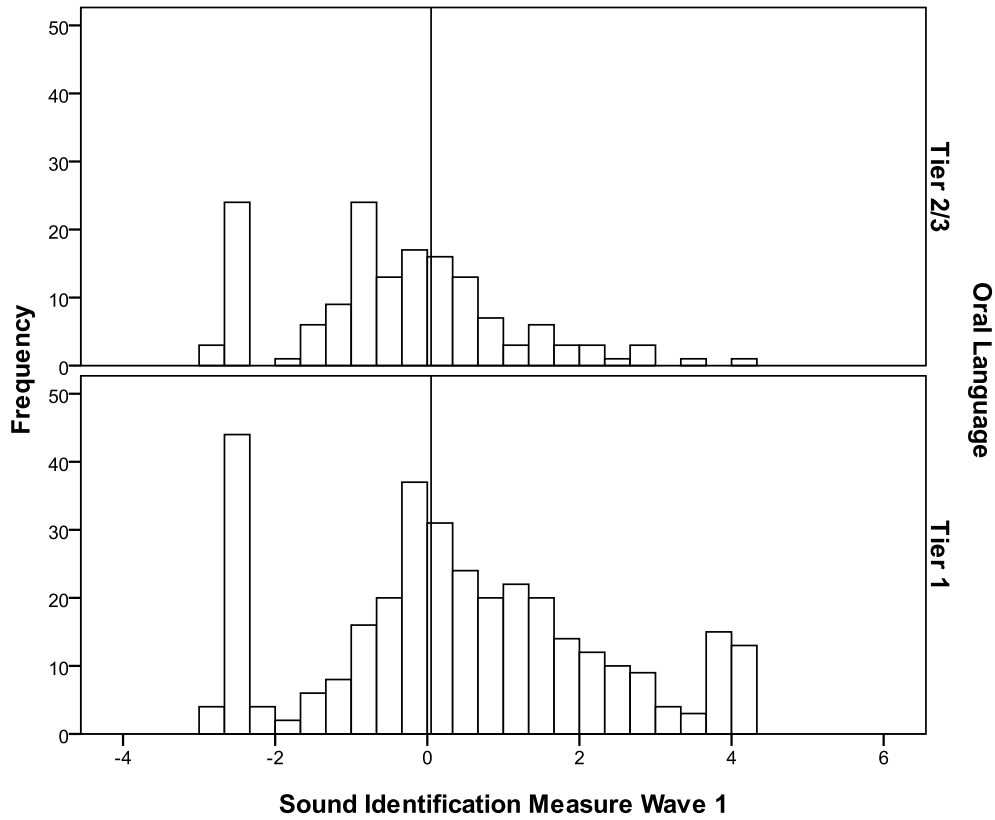


Figure 16. Sound Identification measure histograms by Tier placement.

Figure 16 provides an illustration of the intersection (and overlap) of the score distributions of the two groups. The proposed cut score of 0.05 logits is marked on the graph.

*Examining Teacher Tier Placements vis-à-vis Criterion Measures*

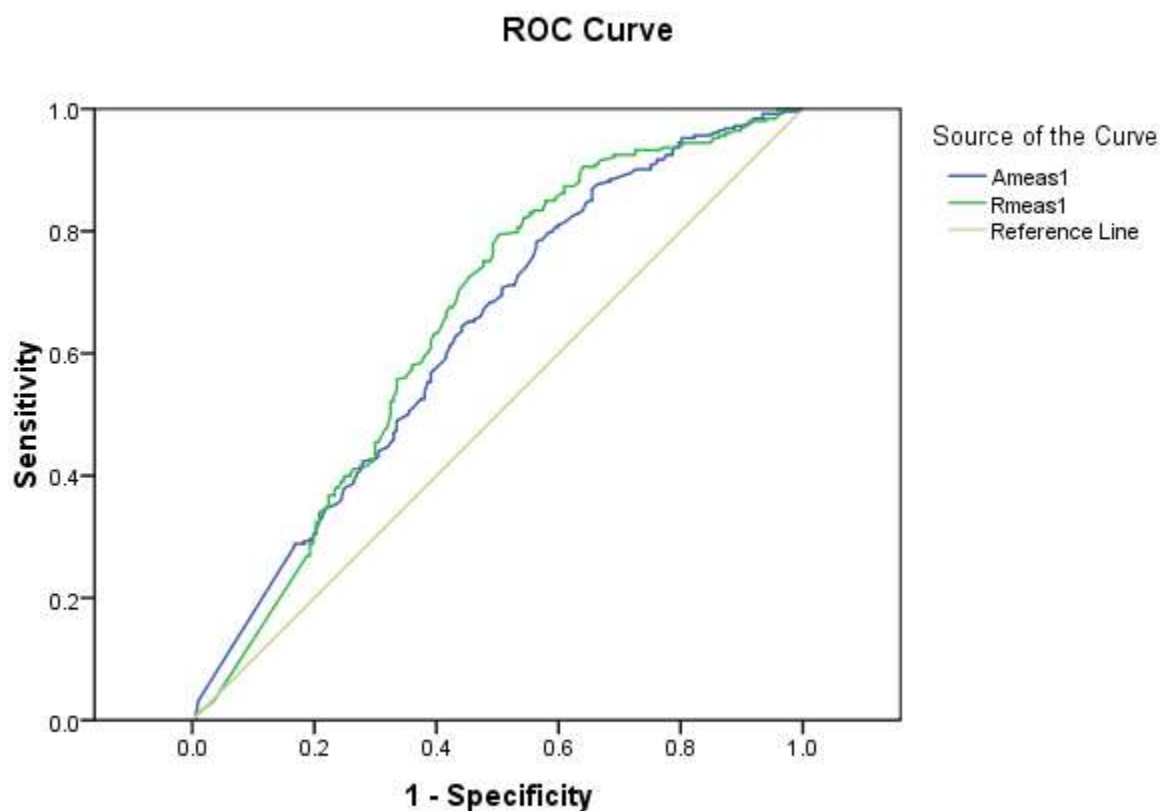
Sound ID: TOPELpk\_SS and TOPELpa\_SS

---

**Phonological Awareness**  
Cut Score Set with a Combination of Rhyming and Alliteration

---

Neither Alliteration nor Rhyming were clear candidates for the Identification measures for PA. The two measures are correlated at approximately .39 ( $n=613$ ), indicating significant independence (15% common variance). Selecting cases where the logit measures are greater than -2.0 (eliminating the effects of raw scores of zero), the resulting correlation is approximately .41 ( $n=363$ ). This relation appears linear; no improvement in the strength of the association was found through non-linear cubic analyses.



Diagonal segments are produced by ties.

*Figure 17.* ROC curve of Sensitivity/Specificity balance for Alliteration and Rhyming Wave 1.

**Area Under the Curve**

Test Result Variable(s)	Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
Ameas1	.632	.027	.000	.580	.684
Rmeas1	.651	.027	.000	.598	.703

- a. Under the nonparametric assumption
- b. Null hypothesis: true area = 0.5

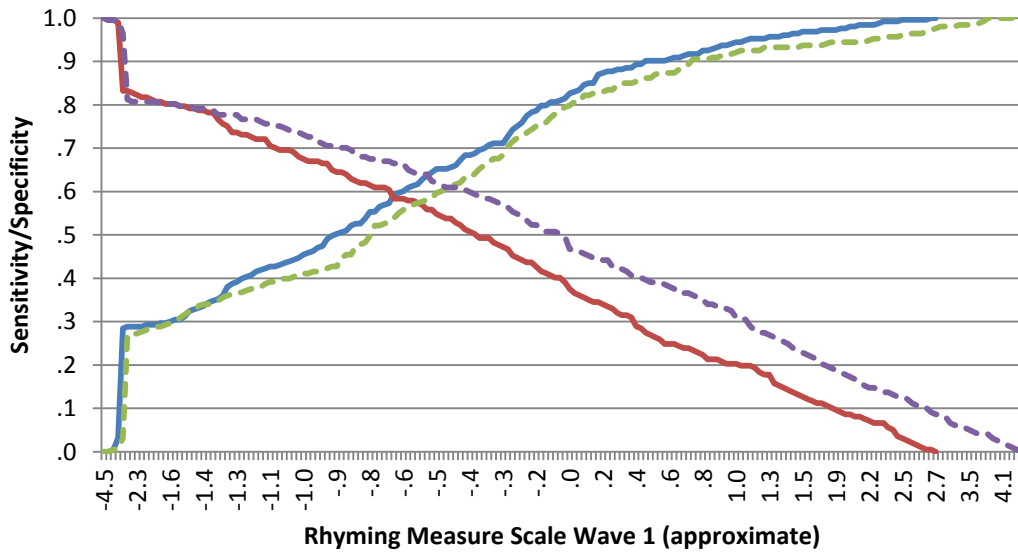


Figure 18. Illustration of the point of intersection of sensitivity (increasing line) and specificity (decreasing line) for Alliteration (dashed lines) and Rhyming (solid lines) Wave 1.

It is difficult to identify a set of scores that are located at a common score on the logit metric. To achieve .70 sensitivity on Rhyming results in .43 specificity. To achieve .70 sensitivity on Alliteration results in .51 specificity.

The .70 sensitivity region on the logit scale is approximately -0.25. The problem is that the two measures are on different scales -0.25 is not the same ability on the two measures, since they measure different abilities.

The point of maximizing both sensitivity and specificity is -0.68 for Alliteration and -0.45 for Rhyming on their respective logit scales.

**Area Under the Curve**

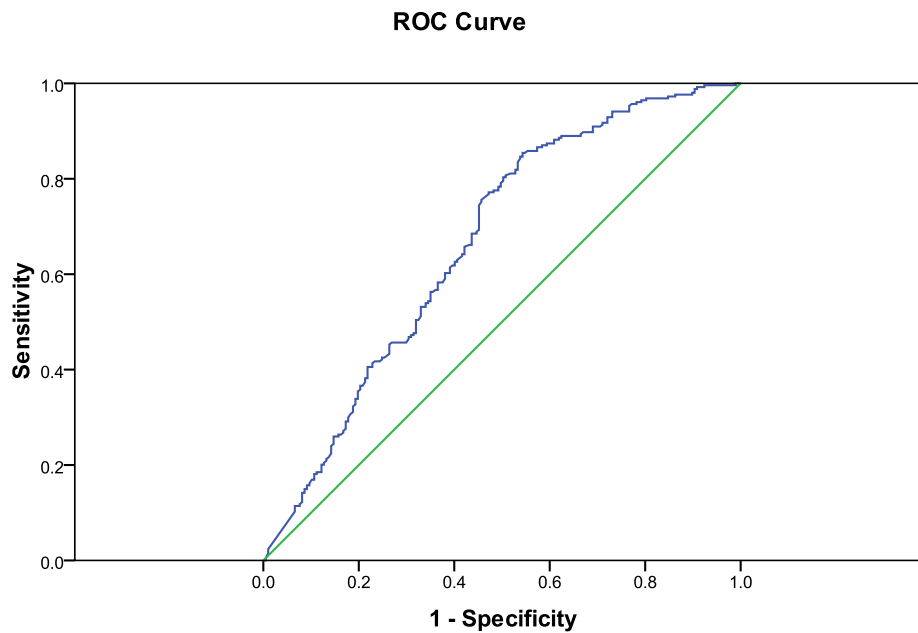
Test Result Variable(s): Sum of Alliteration and Rhyming Measures Wave 1

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.665	.026	.000	.614	.717

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

The sum of the Alliteration and Rhyming measures yields a slightly higher Area statistic.



Diagonal segments are produced by ties.

*Figure 19.* ROC curve of Sensitivity/Specificity balance for the sum of Alliteration and Rhyming Wave 1 measures.

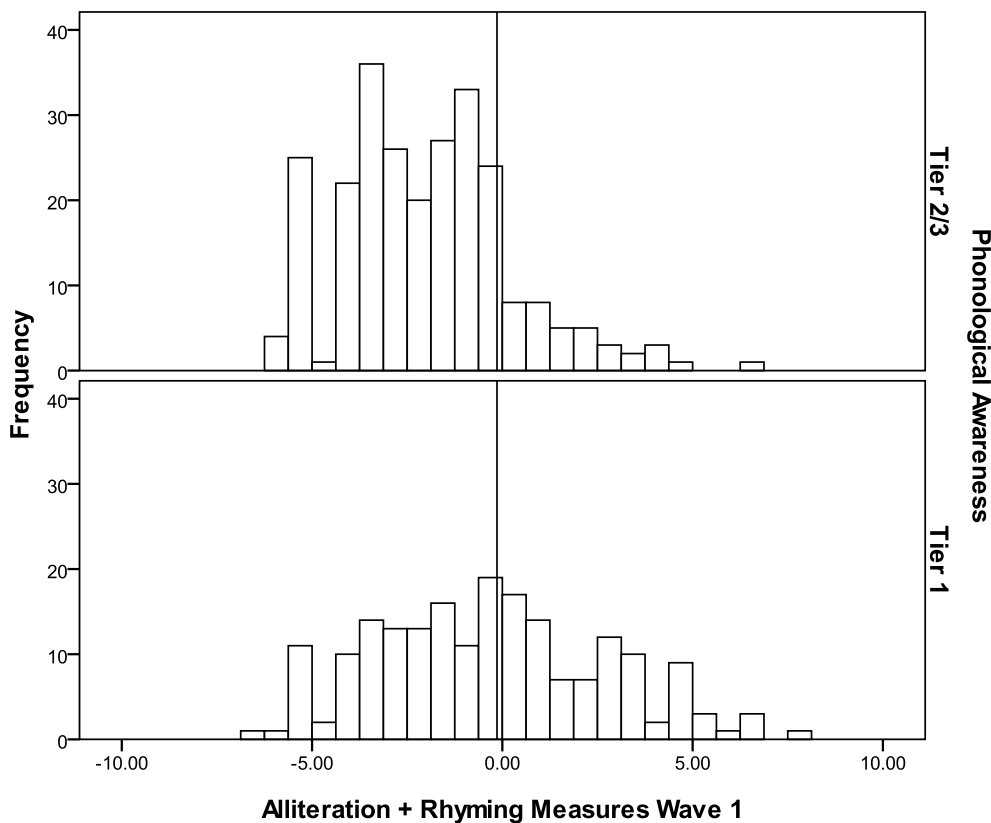
The point which achieves .70 sensitivity is associated with .55 specificity. At this point, we would identify 52% of all children for Tiers 2/3.

*Logistic Regression Analysis*

The logistic regression analysis based on the point at which a child has a 50% chance of being in Tier 1 or 2 results in 68% classification accuracy overall, with a higher classification success rate for Tiers 2/3 (85%) than Tier 1 (45). The summed logit score associated with this point is -0.14. This is difficult to locate with respect to either measure, since a large number of combinations of scores could be achieved to obtain a summed score of -0.14.

**Classification Table**

Observed		Predicted			
		Phonological Awareness		Percentage Correct	
		0 Tier 2/3	1 Tier 1		
Step 1	Phonological Awareness	0 Tier 2/3	217	37	85.4
		1 Tier 1	108	89	45.2
	Overall Percentage				67.8



*Figure 20. Sound Identification measure histograms by Tier placement.*

Figure 20 provides an illustration of the intersection (and overlap) of the score distributions of the two groups. The proposed cut score of -0.14 logits is marked on the graph.