

DRC, Egypt, Malawi, Philippines Results

Cole Campton

Summer 2020

Summary statistics

##	Dataset	Language	Grade	Phase	Gini	Mean	CV	p90/p10	p75/p25	% zero
##			Gr4	1	0.755	8.061	17.132			0.594
##	DRC	French	-----	3	0.749	9.008	13.275			0.538
##			Gr6	1	0.467	27.127	10.343			0.233
##	-----	-----	-----	3	0.458	32.466	9.059		12.919	0.164
##		French	5		0.646	15.530	5.998			0.393
##		-----	Other		0.396	30.128	14.530		4.091	0.096
##	DRCe	Lingala	3		0.920	1.731	27.198			0.192
##		Tshiluba	3		0.886	2.483	14.341			0.362
##	-----	Kiswahili	3		0.916	1.254	31.182			0.262
##	Egy	Arabic	Gr2		0.524	20.681	6.779			0.273
##				1	0.983	0.204	30.200			0.971
##			Gr1	2	0.980	0.212	25.914			0.973
##	malawi	English	-----	3	0.984	0.300	29.401			0.973
##			Gr2	1	0.948	0.979	16.449			0.910
##			Gr3	2	0.900	2.729	11.693			0.838
##	-----	-----	-----	3	0.905	2.459	12.387			0.842
##			1	2014	0.535	21.052	11.092		17.5	0.081
##		Cebuano	-----	2015	0.495	21.891	9.896		8	0.096
##			Gr2	2014	0.338	39.763	5.311	24.286	2.44	0.031
##		-----	-----	2015	0.285	44.732	5.505	6.349	1.818	0.020
##			1	2014	0.574	14.737	10.064			0.052
##		Ilokano	-----	2015	0.497	17.657	8.754		14.5	0.029
##			Gr2	2014	0.375	30.053	6.443		3.214	0.019
##	phil	-----	-----	2015	0.329	33.426	5.566		2.456	0.015
##			1	2014	0.660	12.922	14.103			0.175
##		Hiligaynon	-----	2015	0.639	13.746	10.798			0.123
##			Gr2	2014	0.451	31.565	7.115		5.605	0.089
##		-----	-----	2015	0.519	26.982	9.679		23	0.088
##			1	2014	0.782	6.909	18.707			0.042
##		Maguindanaoan	-----	2015	0.751	8.214	16.858			0.046
##			Gr2	2014	0.547	20.858	9.430			0.018
##	-----	-----	-----	2015	0.485	22.839	8.896			0.016

In the case of retrospective statistics we study the following subpopulations

```
## [1] "1) Dataset phil, Cebuano, grade 1, phase 2014 -> 2015"
## [1] "2) Dataset phil, Hiligaynon, grade 1, phase 2014 -> 2015"
## [1] "3) Dataset phil, Ilokano, grade 1, phase 2014 -> 2015"
```

```
## [1] "4) Dataset phil, Maguindanaoan, grade 1, phase 2014 -> 2015"
## [1] "5) Dataset malawi, English, grade Gr1, phase 1 -> 2"
## [1] "6) Dataset malawi, English, grade Gr1, phase 1 -> 3"
## [1] "7) Dataset malawi, English, grade Gr1, phase 2 -> 3"
## [1] "8) Dataset phil, Cebuano, grade Gr2, phase 2014 -> 2015"
## [1] "9) Dataset phil, Hiligaynon, grade Gr2, phase 2014 -> 2015"
## [1] "10) Dataset phil, Ilokano, grade Gr2, phase 2014 -> 2015"
## [1] "11) Dataset phil, Maguindanaoan, grade Gr2, phase 2014 -> 2015"
## [1] "12) Dataset malawi, English, grade Gr3, phase 2 -> 3"
## [1] "13) Dataset DRC, French, grade Gr4, phase 1 -> 3"
## [1] "14) Dataset DRC, French, grade Gr6, phase 1 -> 3"
```

Here we analyze the “well-behavedness,” of these measurements in a similar manner to Primr and Tusome datasets.

1. Ratio of Px to Py

Ratio of Px to Py. In 53.13% of subpopulations for Egypt, Malawi, Philippines and two DRC datasets ratio_p90p10 cannot be calculated because more than 10 percent of the children assessed recorded a zero score. While ratio_p70p25 can be calculated more frequently, it is not available for approximately 37.5% of the subpopulations. Across these data sets these ratios are more often undefined for colonial languages French and English than for mothertongue languages. This is primarily the result of test languages of countries with many zero scores. That is, the majority of subpopulations with fewer than 10% and 25% zero score readers come from the Philippines which has higher literacy and where only mothertongue languages were tested. This is correlated with the fact that the majority of subpopulations with high zero score rate are tested in a colonial language. A t-test reveals a statistically significant difference between the percent zero scores in colonial verse mothertongue languages with $p=0.0001867$.

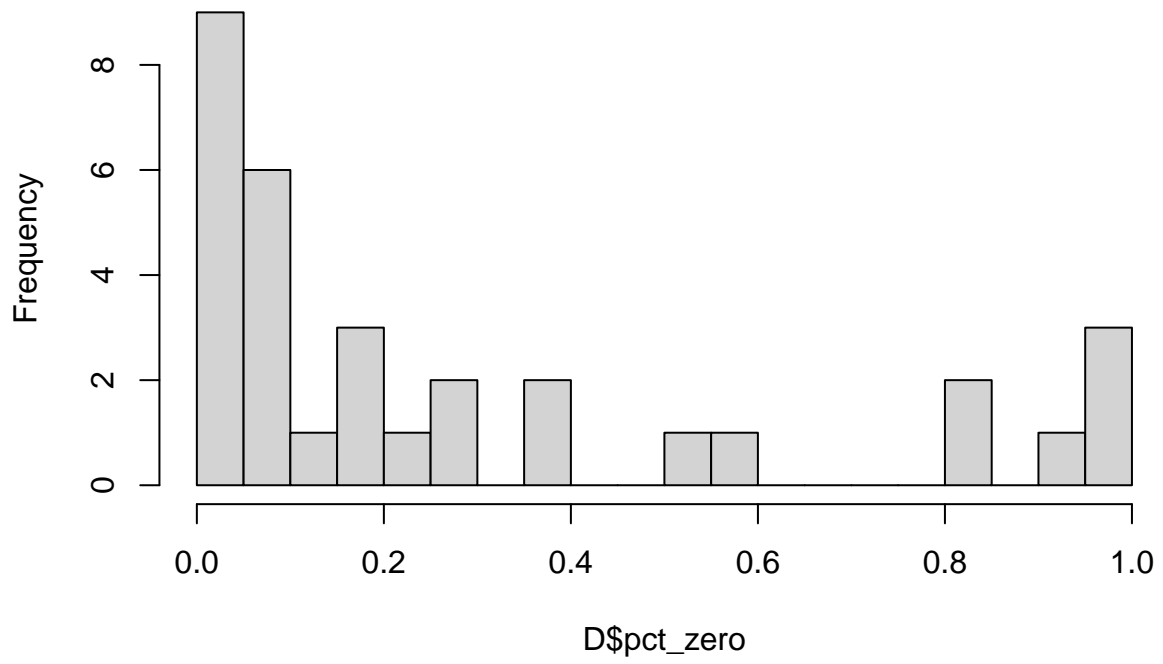
```
## [1] "Percent of subpopulations with more than 10% zero scores: 53.125"
```

```
## [1] "Percent of subpopulations with more than 25% zero scores: 37.5"
```

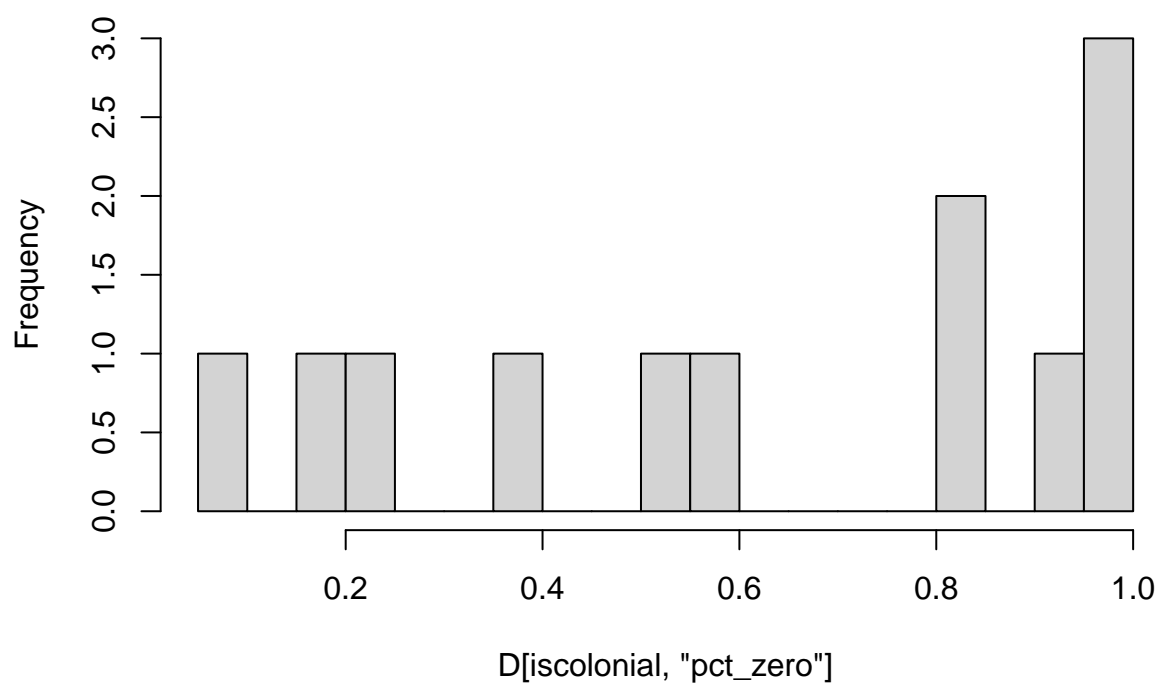
	Arabic	Cebuano	English	French	Hiligaynon
##	27.281052	5.712260	91.772228	33.619962	11.869858
	Ilokano	Kiswahili	Lingala	Maguindanaoan	Tshiluba
##	2.892336	26.248589	19.241489	3.039589	36.223900

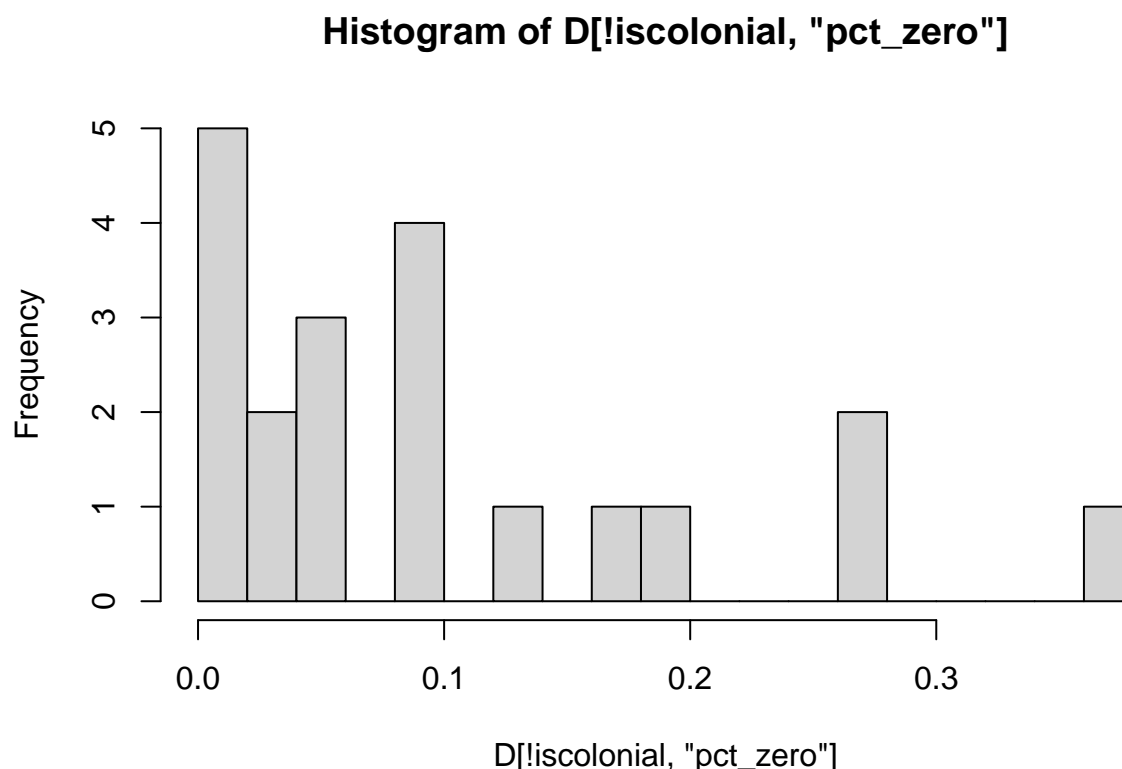
```
##
## Welch Two Sample t-test
##
## data: D[iscolonial, "pct_zero"] and D[!iscolonial, "pct_zero"]
## t = 5.2724, df = 12.195, p-value = 0.0001867
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.3086854 0.7421853
## sample estimates:
## mean of x mean of y
## 0.6269610 0.1015256
```

Histogram of D\$pct_zero



Histogram of D[iscolonial, "pct_zero"]





2. Gini Coefficient

The Gini coefficient for learning seems to consistently behave well. The values observed lie between 0.2847 and 0.9842. The mean of values is 0.6462 and variance 0.0489. These values largely lie within a normal range. The lowest and most equal value being observed within endline grade 2 Philippines students tested in their mothertongue of Cebuano. Interestingly this subpopulation does not have the lowest number of zero scores but instead the fifth lowest. The highest Gini coefficients corresponding to the least equal sub-populations are each from Malawi and correspond to the six highest percent zero scores. This makes sense since a high number of zero scores indicates that all of the reading capability is concentrated in a relatively small number of positive scoring students. Thus the effect of dramatically low reading ability within Malawi sub-populations causes it to register as the most unequal group when measured by Gini coefficient, the opposite is not true for the low zero score rate indicating the least inequality.

```
summary(D$gini)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.2847  0.4807  0.6064  0.6462  0.8898  0.9842
```

There is however a strong correlation between a higher percent of zero scores and lower equity with a correlation coefficient of 0.7835. This correlation breaks down even stronger between colonial and non-colonial languages such that gini and percent zero score have an over .99 correlation coefficient for colonial languages but at most .5 for non-colonial languages. This indicates that the high number of students reading at zero is the primary driver of inequality for examinations conducted in colonial languages.

```
cor(D$gini,D$pct_zero)
```

```
## [1] 0.7835431
```

```
print("For english with multi-phase")
```

```
## [1] "For english with multi-phase"
```

```
cor(T[iscolonialT,c("gini.base", "gini.end","pct_zero.base", "pct_zero.end")])
```

```
##           gini.base  gini.end pct_zero.base pct_zero.end
## gini.base      1.0000000 0.9998710      0.9975060      0.9954832
## gini.end       0.9998710 1.0000000      0.9976069      0.9959602
## pct_zero.base  0.9975060 0.9976069      1.0000000      0.9994275
## pct_zero.end   0.9954832 0.9959602      0.9994275      1.0000000
```

```
print("For mothertongue with multi-phase")
```

```
## [1] "For mothertongue with multi-phase"
```

```
cor(T[!iscolonialT,c("gini.base", "gini.end","pct_zero.base", "pct_zero.end")])
```

```
##           gini.base  gini.end pct_zero.base pct_zero.end
## gini.base      1.0000000 0.9558636      0.3591567      0.3514813
## gini.end       0.9558636 1.0000000      0.4723077      0.5065030
## pct_zero.base  0.3591567 0.4723077      1.0000000      0.9284048
## pct_zero.end   0.3514813 0.5065030      0.9284048      1.0000000
```

Although there are no control groups we may compare the Gini coefficients at baseline and endline for subpopulations which have these retrospectives. The mean Gini coefficients are .6664 and .6472 at baseline and endline respectively; these means have statistically significantly decreased with a p-value of 0.035 via a paired one sided sample t-test. When evaluated by language we find that is a significant decrease in Gini coefficient. indicating higher equity. The high dependence on language here when compared with tusome and primr data sets is expected due to the diversity of countries represented here.

```
t.test(T$gini.end,T$gini.base, paired=TRUE,alternative = "less")
```

```
##
## Paired t-test
##
## data:  T$gini.end and T$gini.base
## t = -1.9764, df = 13, p-value = 0.03486
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -0.001999432
## sample estimates:
## mean of the differences
##      -0.01923184
```

```
# At baseline:
summary(T$gini.base)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3384  0.4840  0.6167  0.6664  0.8704  0.9830
```

```
# At endline:
summary(T$gini.end)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.2847  0.4876  0.5790  0.6472  0.8664  0.9842
```

```
tapply(D$gini, D$measure_label, mean)
```

```
##           Arabic      Cebuano      English      French      Hiligaynon
## 0.5238345  0.4131011  0.9500704  0.5785031  0.5672114
##      Ilokano      Kiswahili      Lingala Maguindanaoan      Tshiluba
## 0.4436709  0.9159466  0.9200325  0.6411276  0.8863346
```

However when stratified by colonial verse non-colonial languages we see that the primary driver of a statistically significant decrease of Gini coefficient and increase in equity comes from mother-tongue examinations, with colonial language examinations not providing a statistically significant increase in equity as measured by Gini coefficient.

```
t.test(T[!iscolonialT,"gini.end"],T[!iscolonialT,"gini.base"], paired=TRUE,alternative = "less")
```

```
##
## Paired t-test
##
## data:  T[!iscolonialT, "gini.end"] and T[!iscolonialT, "gini.base"]
## t = -2.0877, df = 7, p-value = 0.03761
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -0.003021176
## sample estimates:
## mean of the differences
##      -0.03265877
```

```
t.test(T[iscolonialT,"gini.end"],T[iscolonialT,"gini.base"], paired=TRUE,alternative = "less")
```

```
##
## Paired t-test
##
## data:  T[iscolonialT, "gini.end"] and T[iscolonialT, "gini.base"]
## t = -0.5692, df = 5, p-value = 0.2969
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 0.003376522
## sample estimates:
## mean of the differences
##      -0.001329273
```

3. Coefficient of variation.

This indicator also behaves well. The correlation coefficient between the CV and the Gini is 0.8065 across all subpopulations; these two measures of inequality move together well as they did in studies for primr and tusome data. When we perform a similar analysis we obtain expected results, given the high degree of correlation. The value of the coefficient of variation decreases 13.0324 to 11.6728 from baseline to endline. A paired t-test does not reveal a statistically significant decrease as it does for the Gini coefficient.

```
cor(D[,c("gini","cv")])
```

```
##           gini           cv
## gini 1.0000000 0.8065227
## cv   0.8065227 1.0000000
```

```
summary(D$cv)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.311   8.860  10.945  13.566  16.551  31.182
```

```
t.test(T$cv.end,T$cv.base, paired=TRUE,alternative = "less")
```

```
##
## Paired t-test
##
## data:  T$cv.end and T$cv.base
## t = -1.5074, df = 13, p-value = 0.07781
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 0.1543393
## sample estimates:
## mean of the differences
##      -0.8828056
```

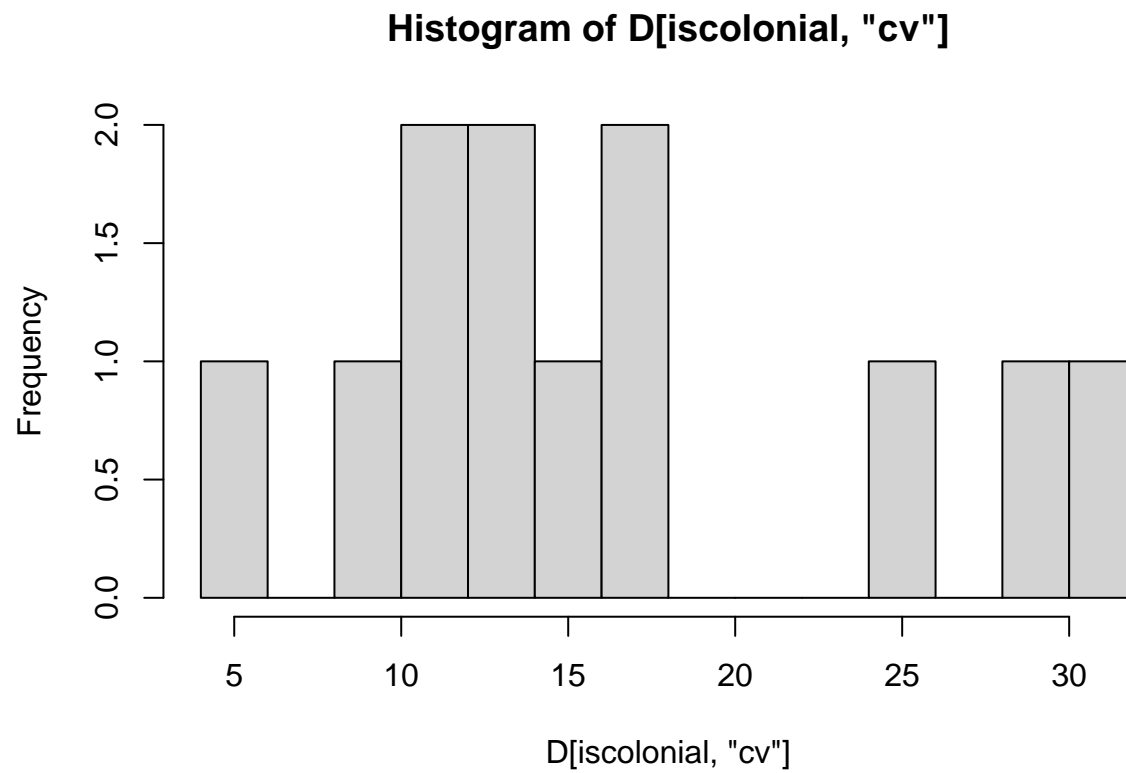
Again the value of CV is highly dependent on language measured and colonial versus non-colonial testing language, however there is no statistically significant difference. A similar longitudinal test stratified based on colonial verse non-colonial languages does not reveal similar results.

```
t.test(D[!iscolonial,"cv"],D[iscolonial,"cv"])
```

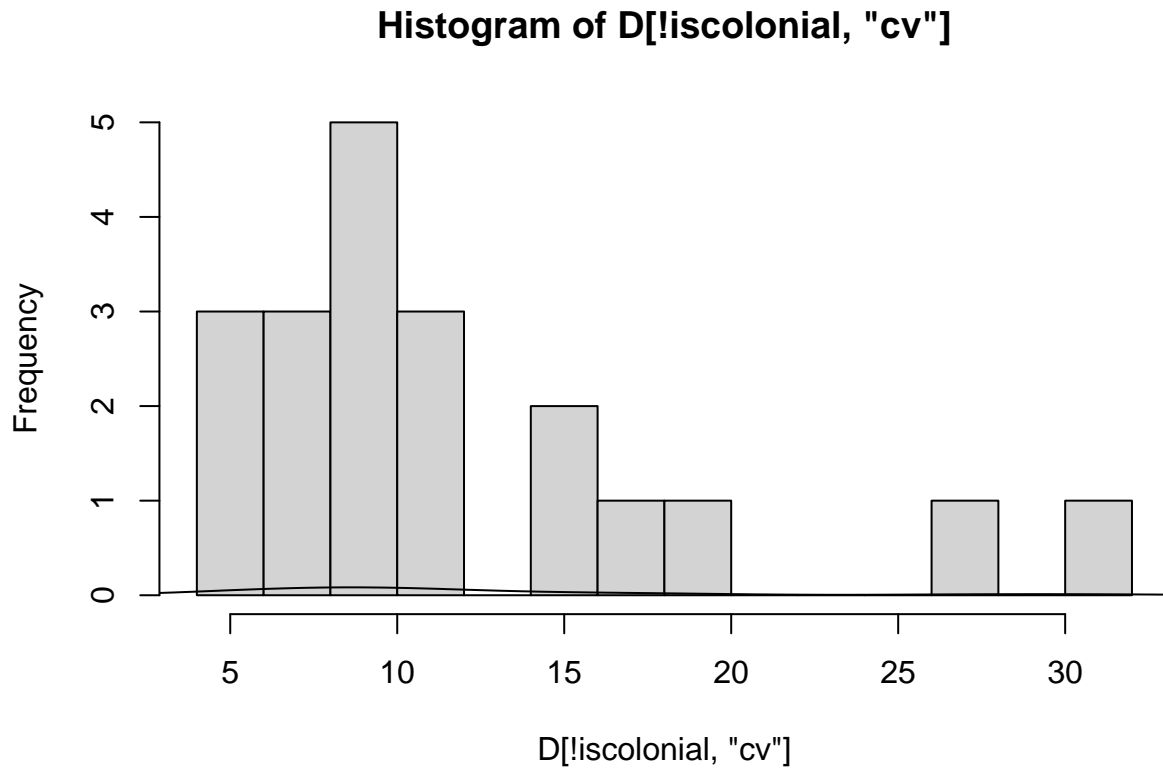
```
##
## Welch Two Sample t-test
##
## data:  D[!iscolonial, "cv"] and D[iscolonial, "cv"]
## t = -1.6083, df = 20.908, p-value = 0.1228
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.272556   1.314244
## sample estimates:
## mean of x mean of y
##  11.88586  16.36501
```



```
hist(D[iscolonial,"cv"],10)
```



```
hist(D[!iscolonial,"cv"],10)  
lines(density(D[!iscolonial,"cv"]))
```



4. Percent Reading at 0

The “percent reading at 0” also behaves well, as one would expect. In the case of these data sets the correlation with the Gini coefficient across the observed data points is 0.7835.

```
cor(D$gini, D$pct_zero)
```

```
## [1] 0.7835431
```

5. Generalized Entropy (alpha=2) for Sub-population

The Generalized Entropy (alpha=2) index appears to behave well for each subpopulation in these data sets. Values decrease over time from baseline to endline for all but three subpopulations. At baseline the average GE(2) value is 6.8422 and at endline it is 7.2687. Notably however the variances are large, 142.165 and 167.7286 at baseline and endline.

```
# At baseline:
summary(T$ge2_for_subpop.base)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.1784  0.3757   0.6867   6.8422  3.4234  31.3680
```

```
# variance
var(T$ge2_for_subpop.base)
```

```
## [1] 142.165
```

```
# At endline:
summary(T$ge2_for_subpop.end)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.1289  0.3756  0.6067  7.2687  3.6137 34.5177
```

```
# variance
var(T$ge2_for_subpop.end)
```

```
## [1] 167.7286
```

The contribution of this large variance is a low confidence in the change from baseline to endline; the value of GE(2) for each subpopulation is not statistically significantly distinct at endline from baseline ($p=.3535$) via a paired t-test. As may be suspect, in turn the values of GE(2) are less correlated with the Gini coefficient with value of $\rho = .6436$.

```
t.test(T$ge2_for_subpop.end, T$ge2_for_subpop.base,paired=TRUE, alternative = "greater")
```

```
##
## Paired t-test
##
## data:  T$ge2_for_subpop.end and T$ge2_for_subpop.base
## t = 0.3842, df = 13, p-value = 0.3535
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -1.53952      Inf
## sample estimates:
## mean of the differences
##          0.4265316
```

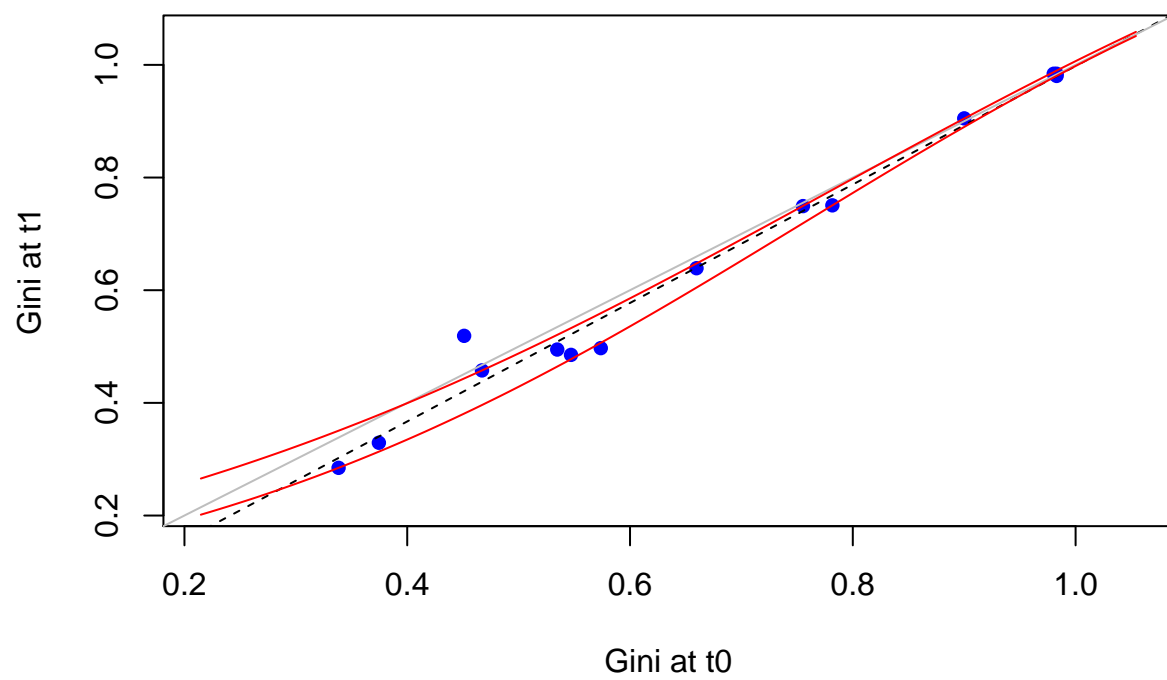
```
cor(D$ge2_for_subpop, D$gini)
```

```
## [1] 0.6436345
```

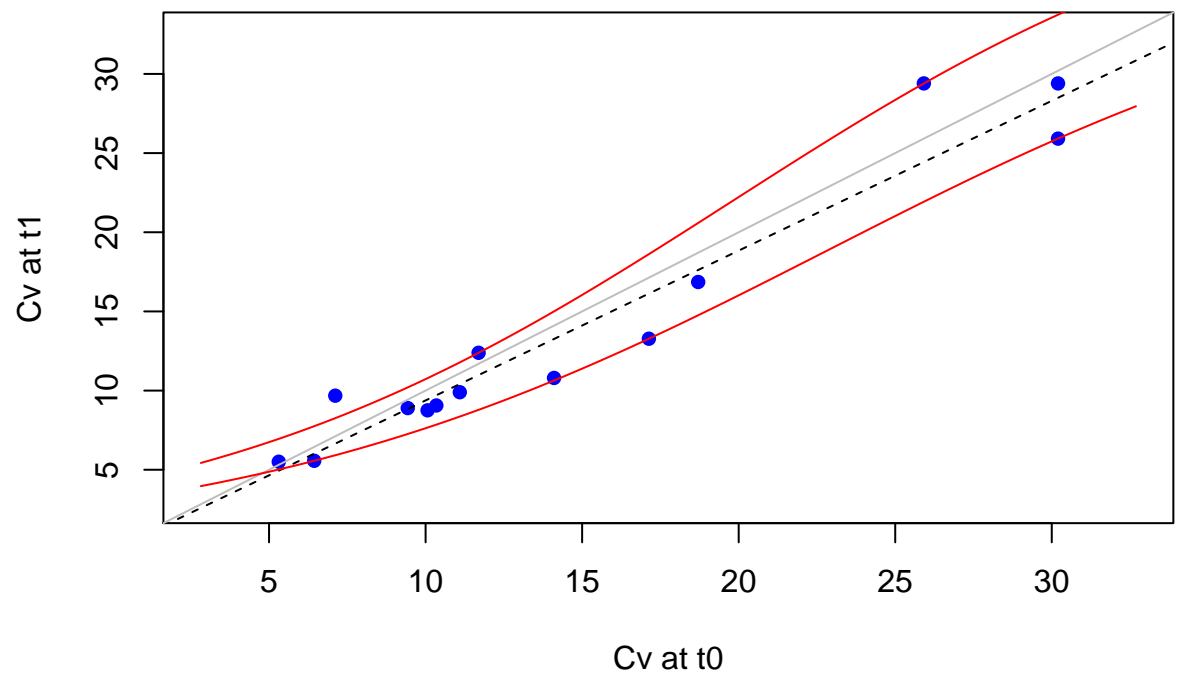
Scatter plots

In the case of change for the Gini coefficient, the “button hole” shape does not hold for these datasets. That is, the variance of the change is not significantly affected by the Gini coefficient at baseline. The range of values for CV, Gini coefficient, mean reading fluency and percent reading at zero all are slightly more extreme than for Tusome and Primr datasets but still within a reasonable range.

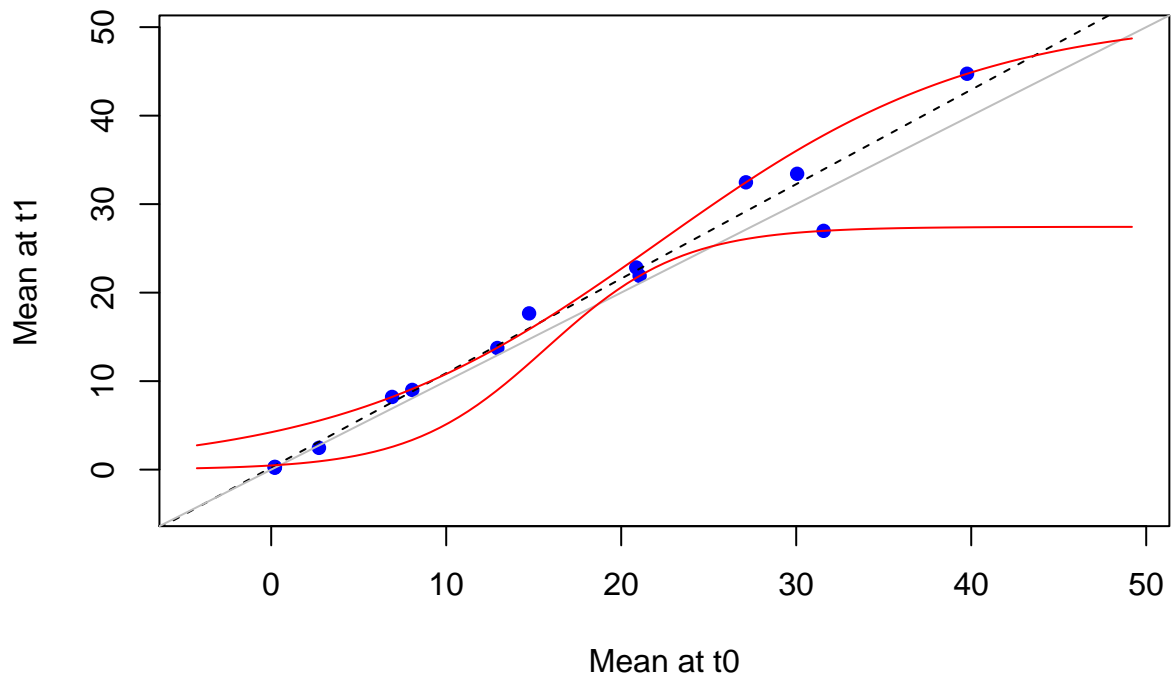
Gini at t1 verse t0

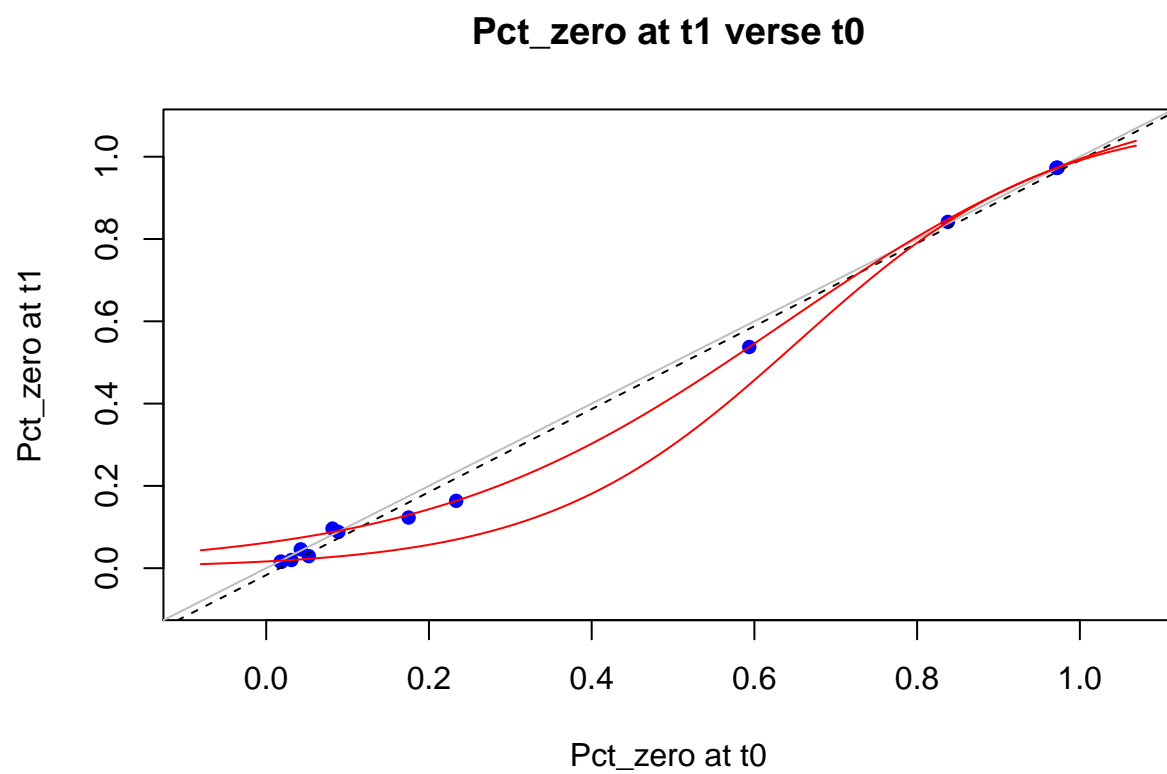


Cv at t1 verse t0



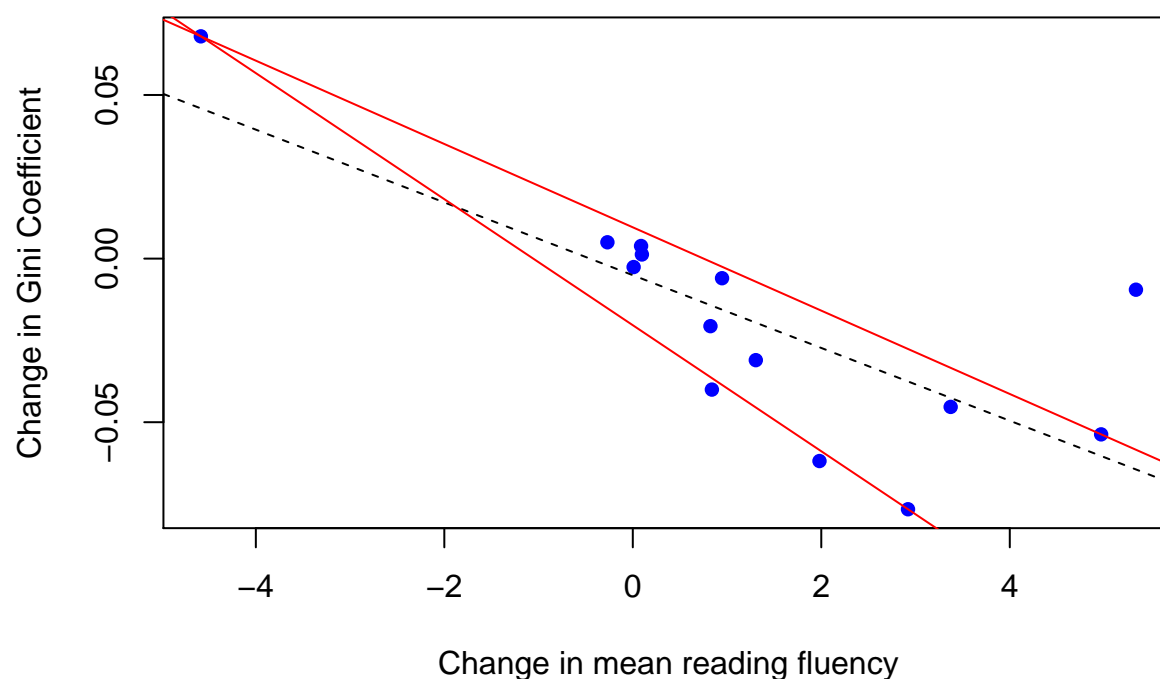
Mean at t1 verse t0





The same general trend continues with the plot of change in gini coefficient with respect to change in mean reading fluency. There is a larger variance in the change in gini coefficient as the change in mean reading fluency grows larger.

Change in Gini verse Change in MRF



Additionally the correlation is even stronger than in Primr/Tusome studies, despite the greater diversity of datasets. Again fluency improvements of around 10 yield improvements of Gini coefficient of approximately -.21 as estimated by the mean slope of 15th and 85th percentile regressions. Since the overall change in mean reading fluency is limited it would simply be extrapolation to estimate change for larger values.

```
cor(X,Y)
```

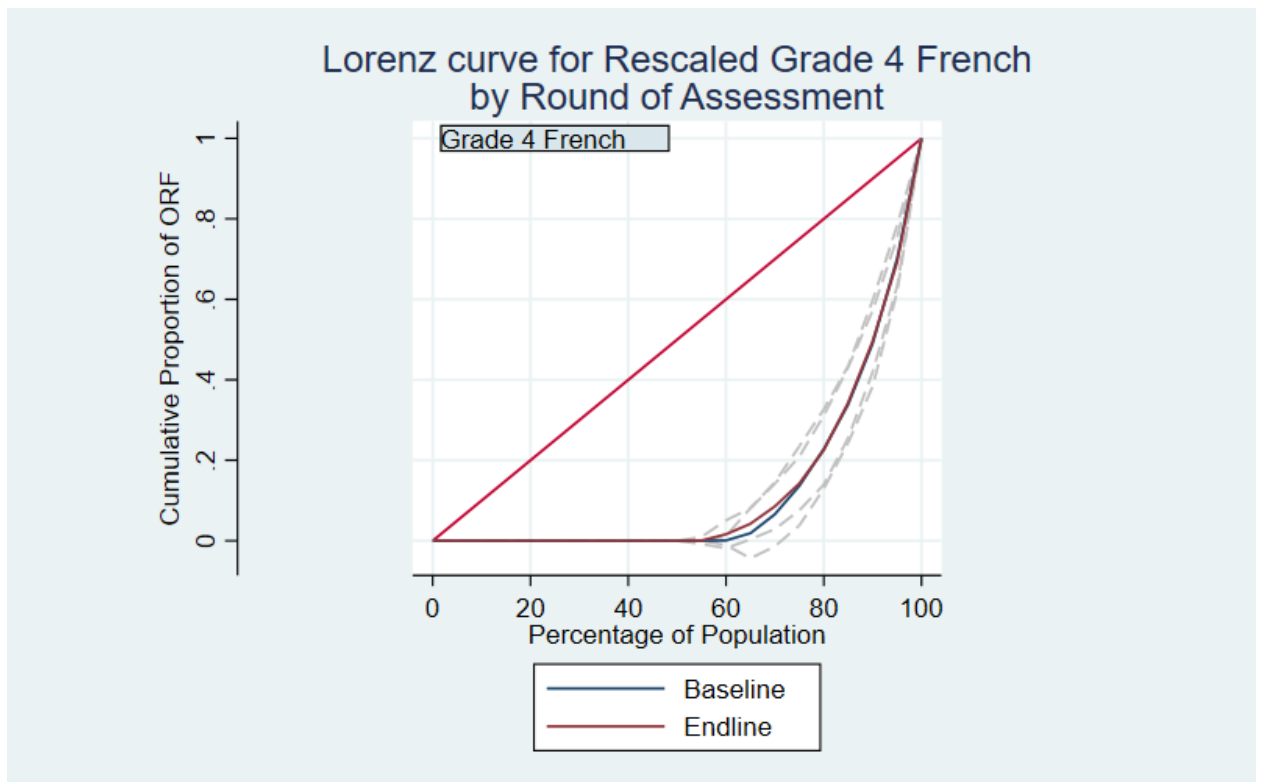
```
## [1] -0.7546724
```

```
mean(fit15.rq$coefficients[1]+fit15.rq$coefficients[2]*10,fit85.rq$coefficients[1]+fit85.rq$coefficients[2]*10)
```

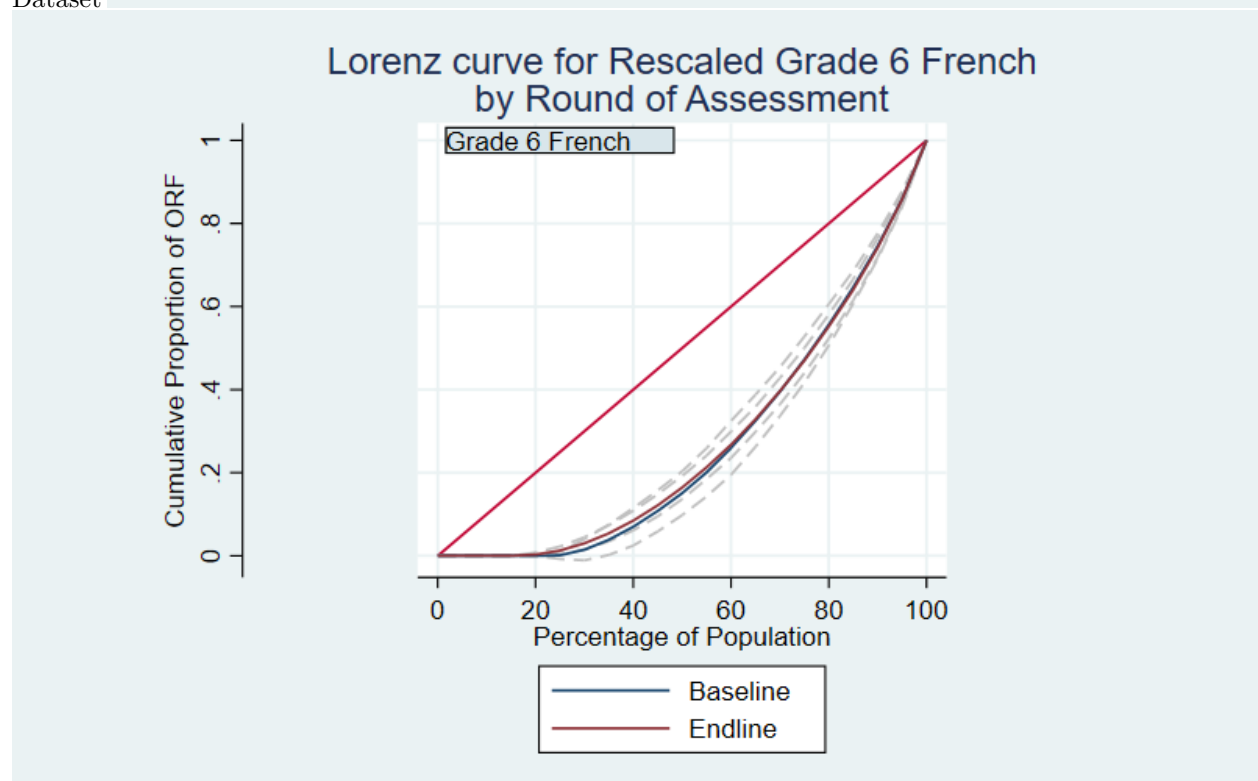
```
## [1] -0.2129466
```

Again these results indicate that the bigger the improvement in the means, the greater the reduction in inequality in oral reading fluency.

Lorenz Plots



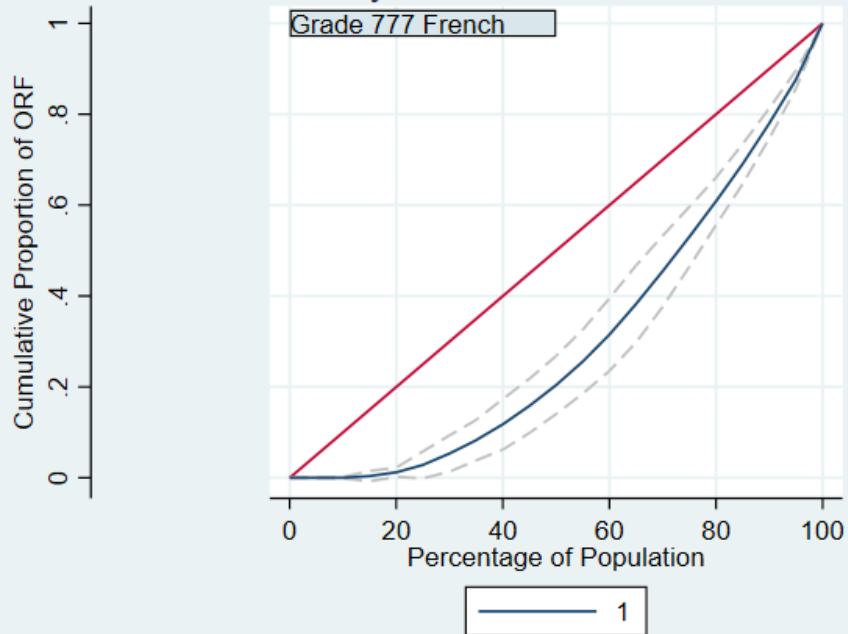
Dataset



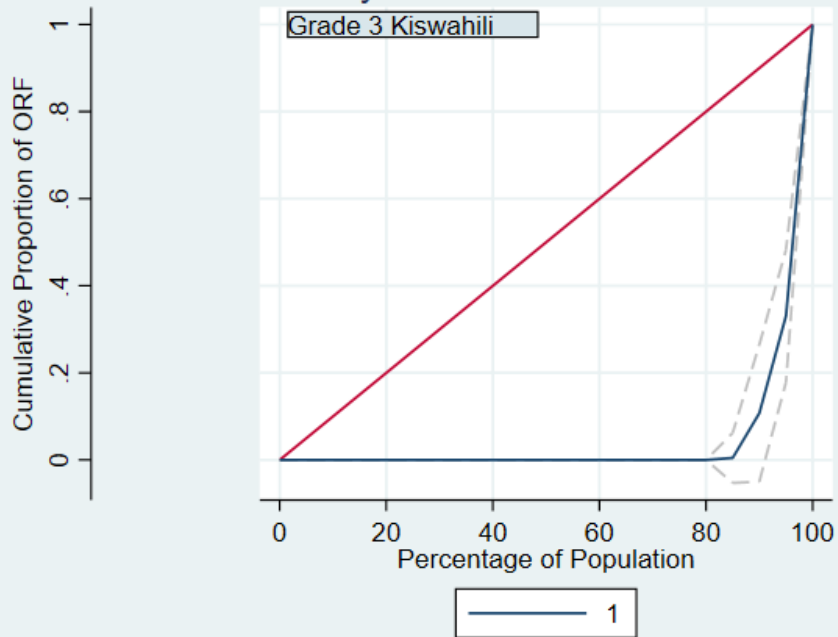
Lorenz curve for Rescaled Grade 5 French
by Round of Assessment



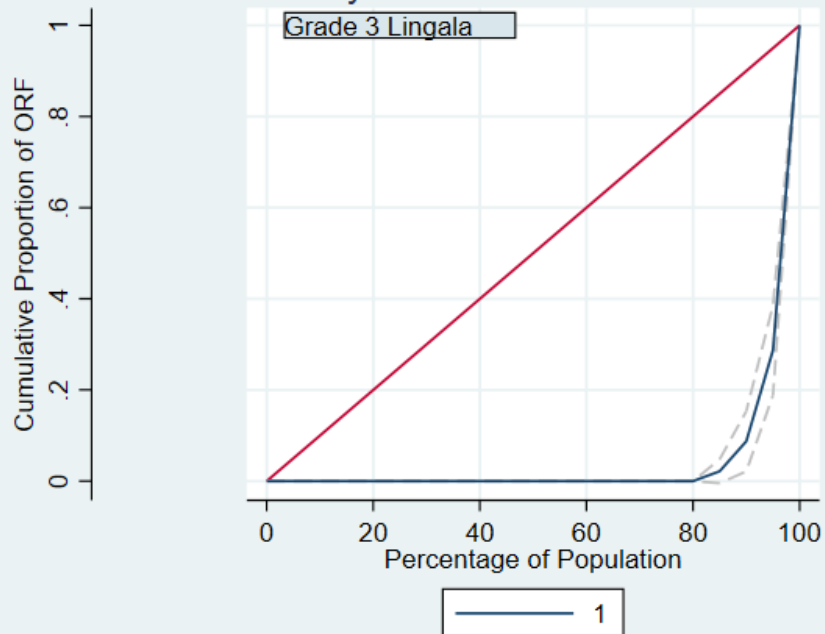
Lorenz curve for Rescaled Grade 777 French
by Round of Assessment



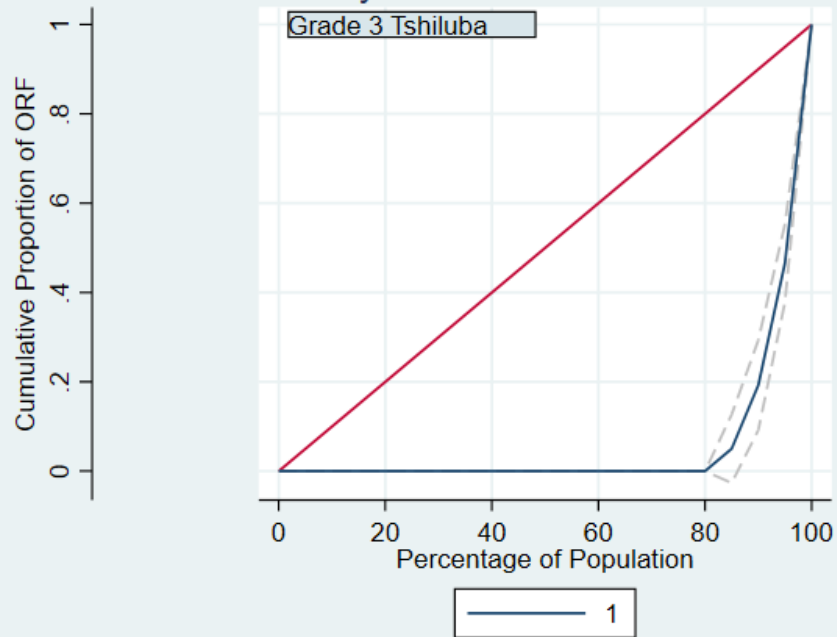
Lorenz curve for Rescaled Grade 3 Kiswahili
by Round of Assessment



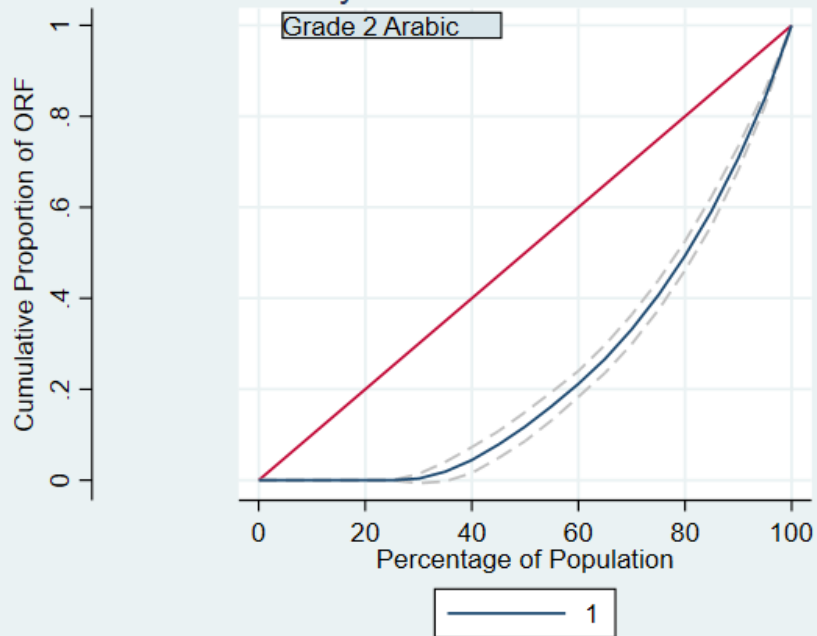
Lorenz curve for Rescaled Grade 3 Lingala
by Round of Assessment



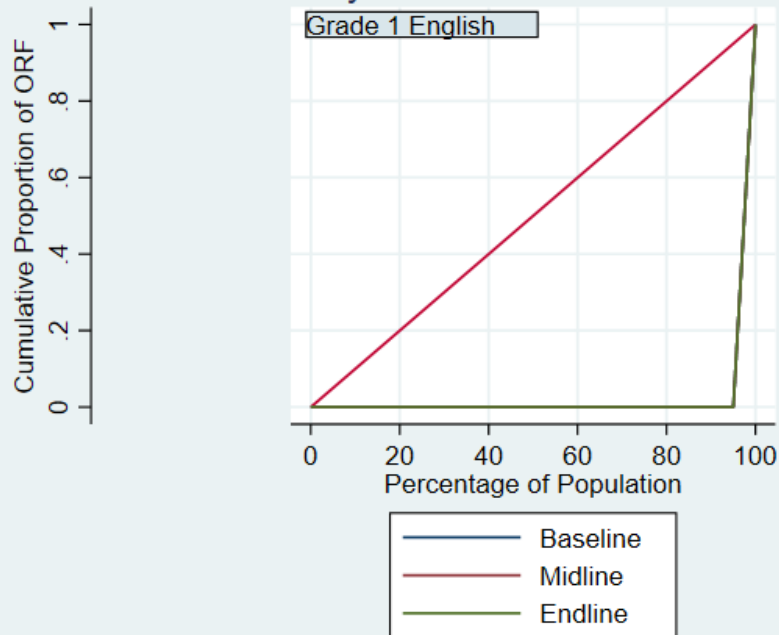
Lorenz curve for Rescaled Grade 3 Tshiluba
by Round of Assessment



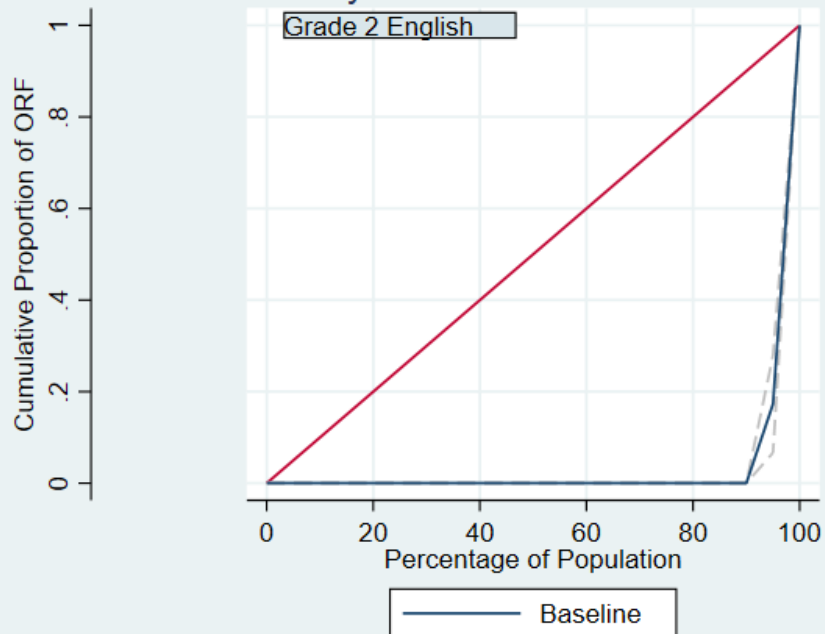
Lorenz curve for Rescaled Grade 2 Arabic
by Round of Assessment



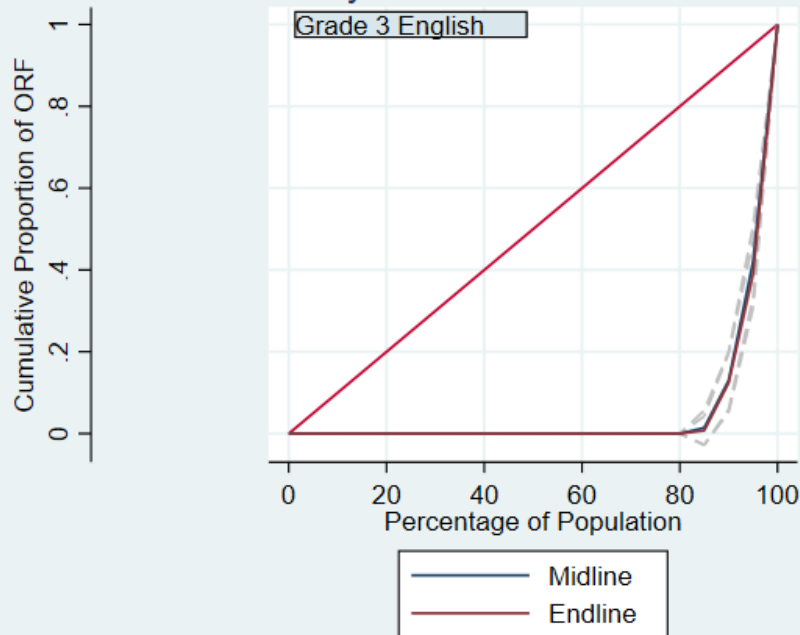
Lorenz curve for Rescaled Grade 1 English
by Round of Assessment



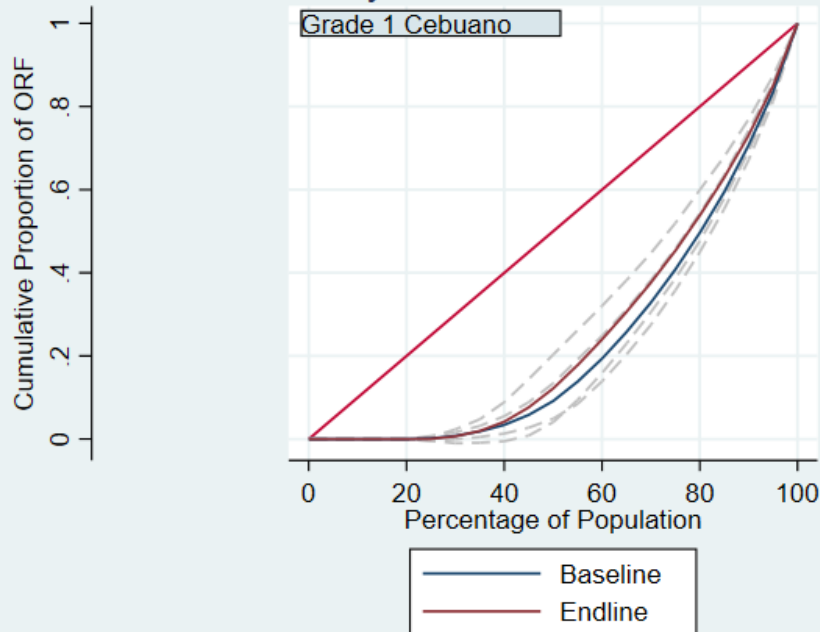
Lorenz curve for Rescaled Grade 2 English
by Round of Assessment



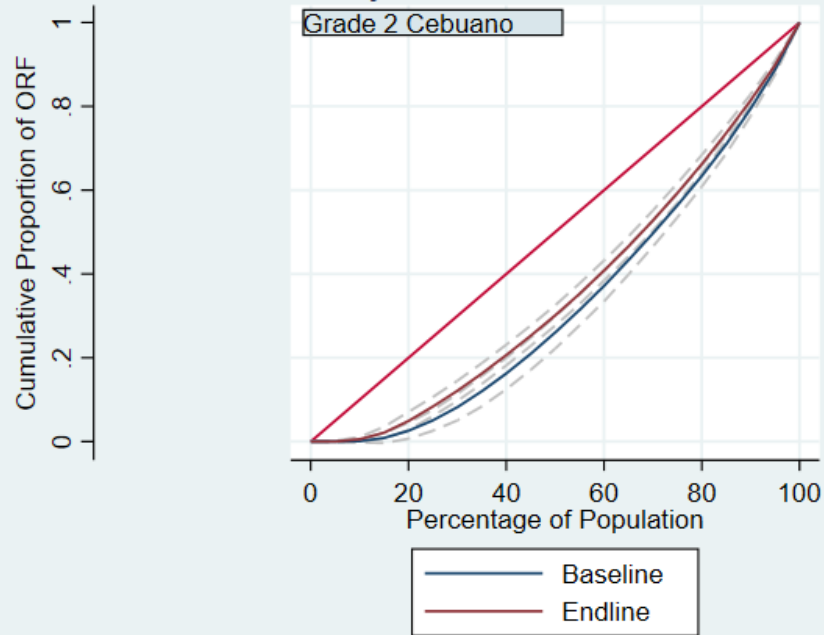
Lorenz curve for Rescaled Grade 3 English
by Round of Assessment



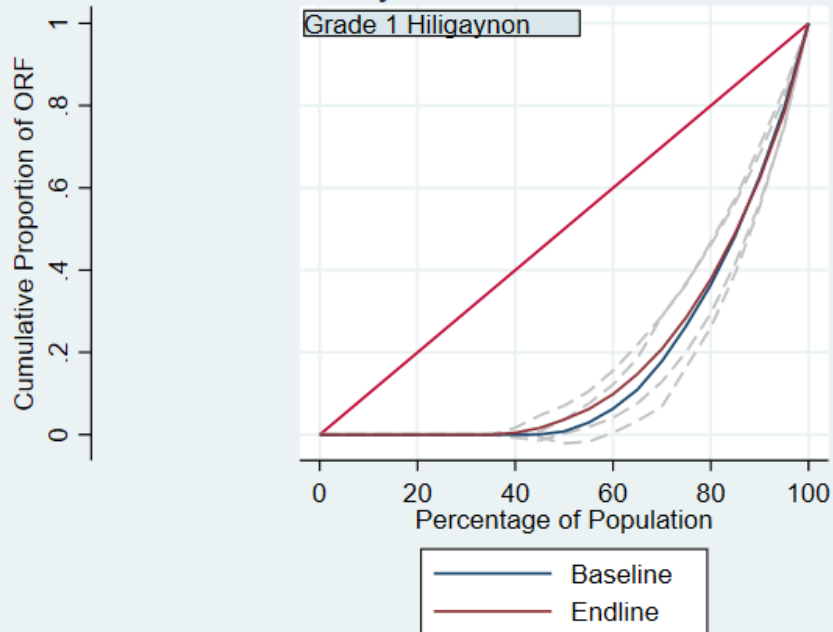
Lorenz curve for Rescaled Grade 1 Cebuano
by Round of Assessment



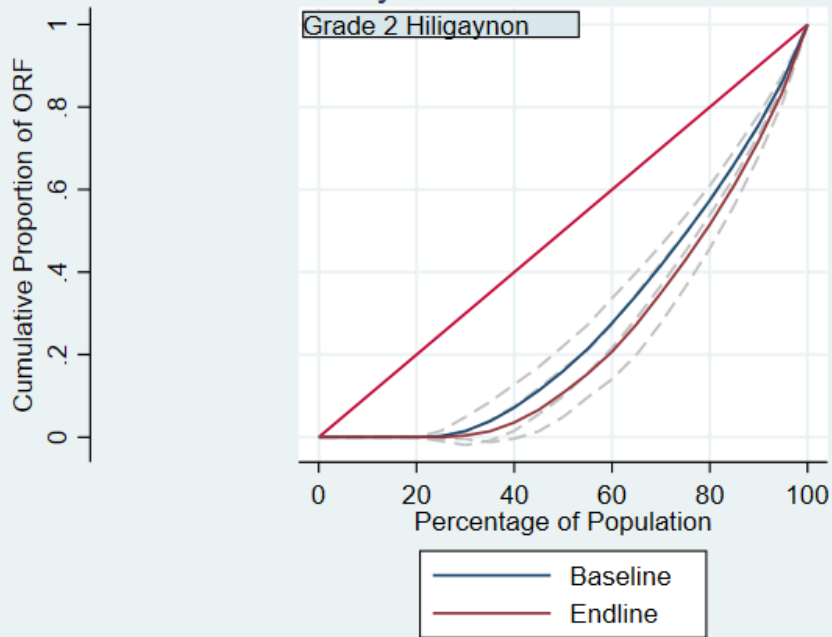
Lorenz curve for Rescaled Grade 2 Cebuano
by Round of Assessment



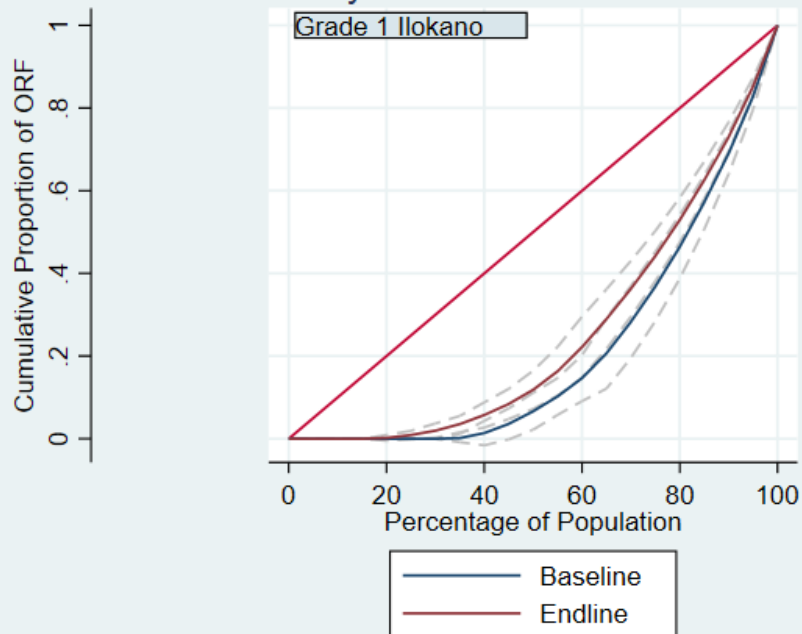
Lorenz curve for Rescaled Grade 1 Hiligaynon
by Round of Assessment



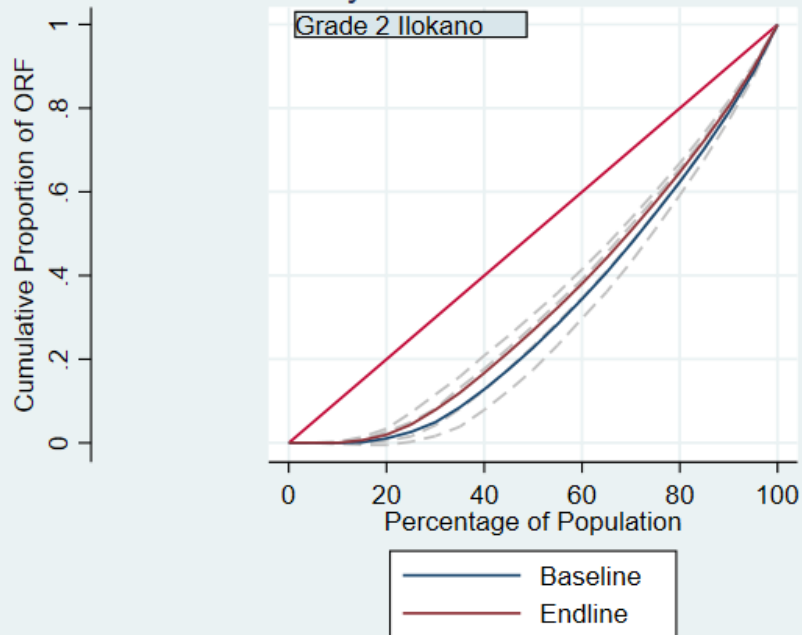
Lorenz curve for Rescaled Grade 2 Hiligaynon
by Round of Assessment



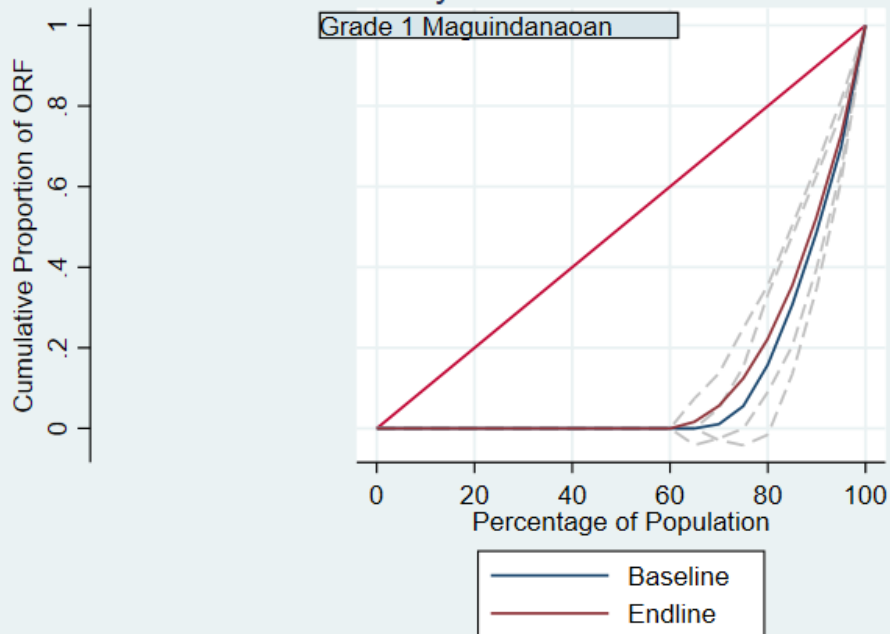
Lorenz curve for Rescaled Grade 1 Ilokano
by Round of Assessment



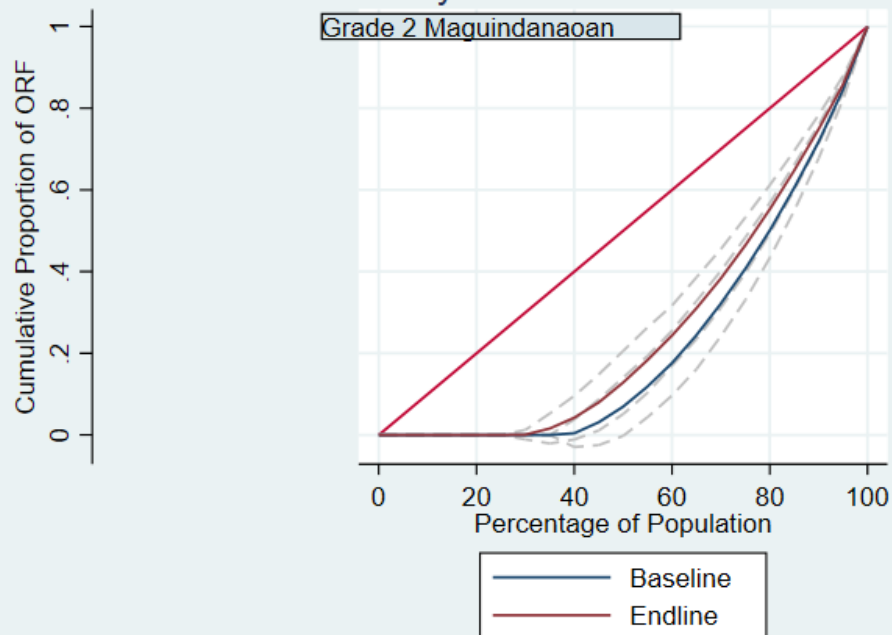
Lorenz curve for Rescaled Grade 2 Ilokano
by Round of Assessment



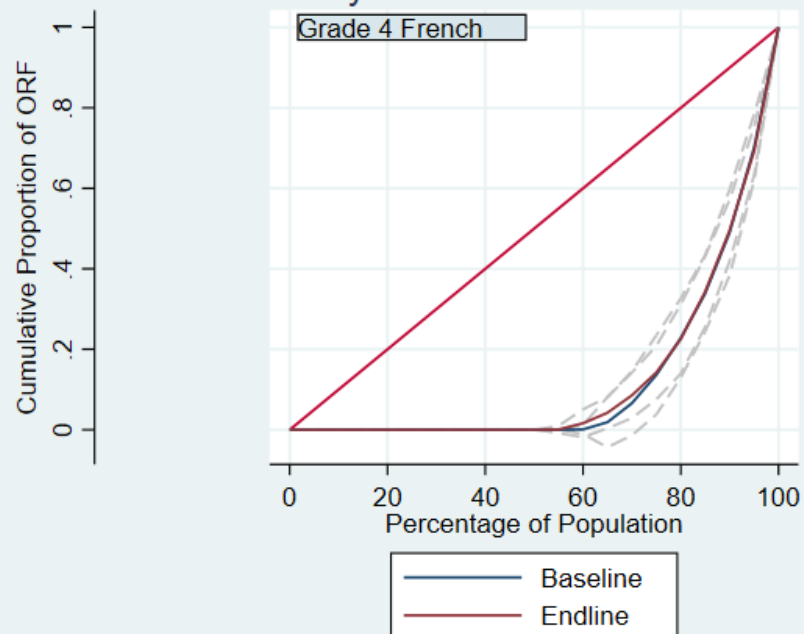
Lorenz curve for Rescaled Grade 1 Maguindanaoan
by Round of Assessment



Lorenz curve for Rescaled Grade 2 Maguindanaoan
by Round of Assessment

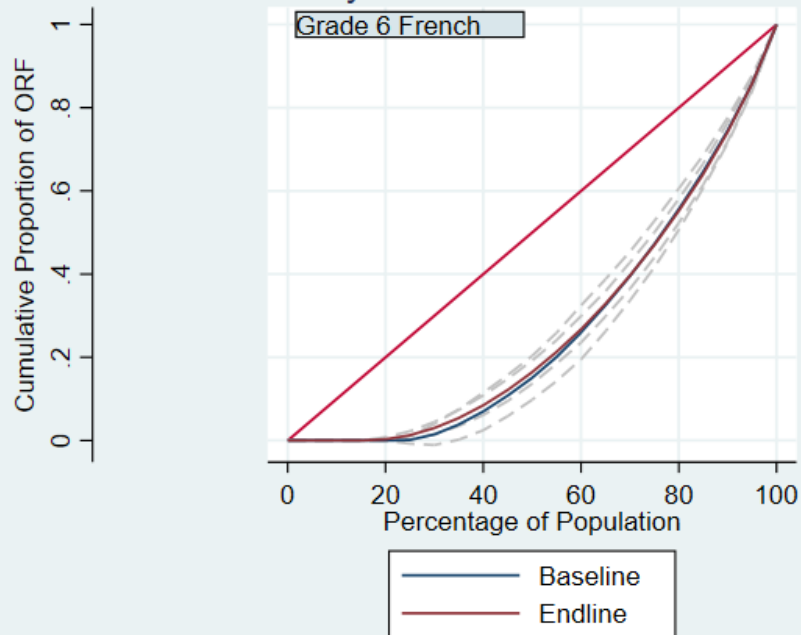


Lorenz curve for Rescaled Grade 4 French
by Round of Assessment

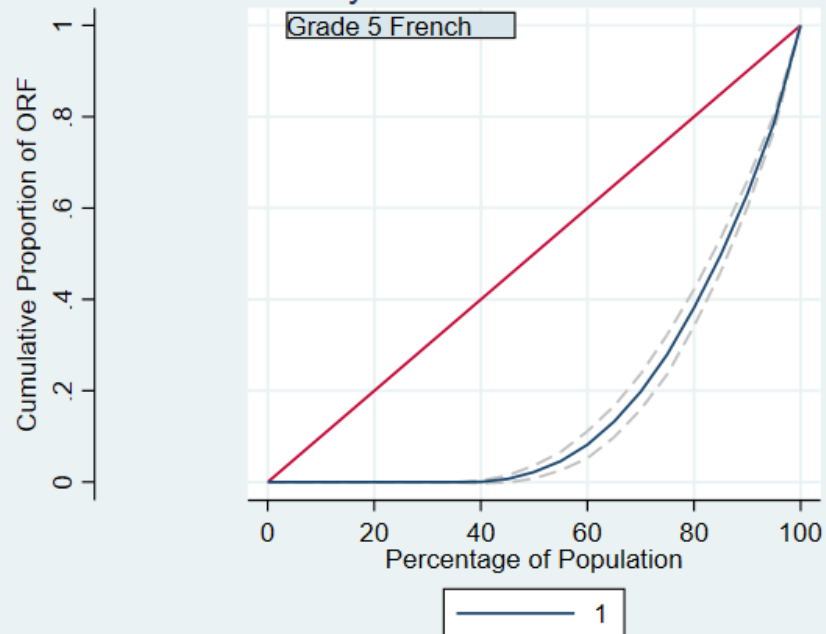


Dataset d.DRC

Lorenz curve for Rescaled Grade 6 French
by Round of Assessment

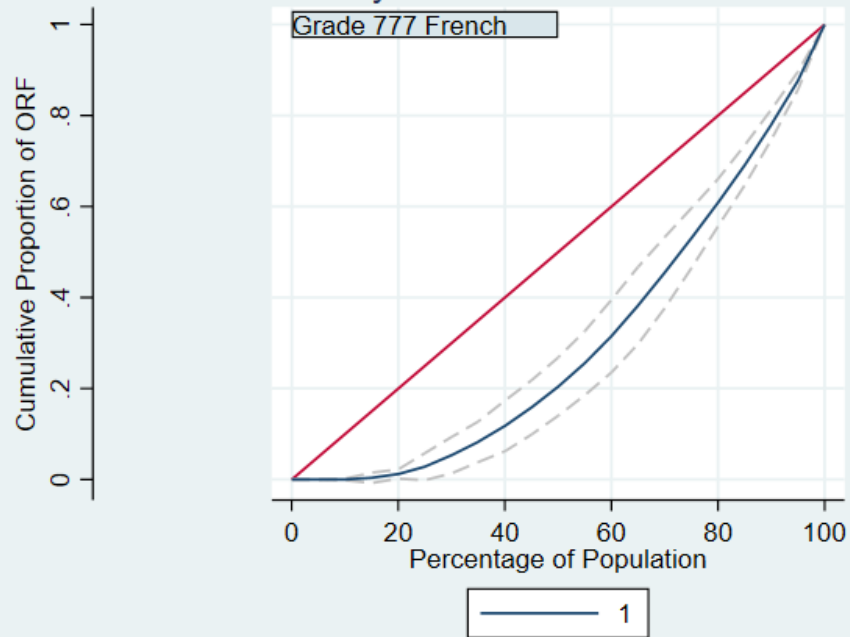


Lorenz curve for Rescaled Grade 5 French
by Round of Assessment

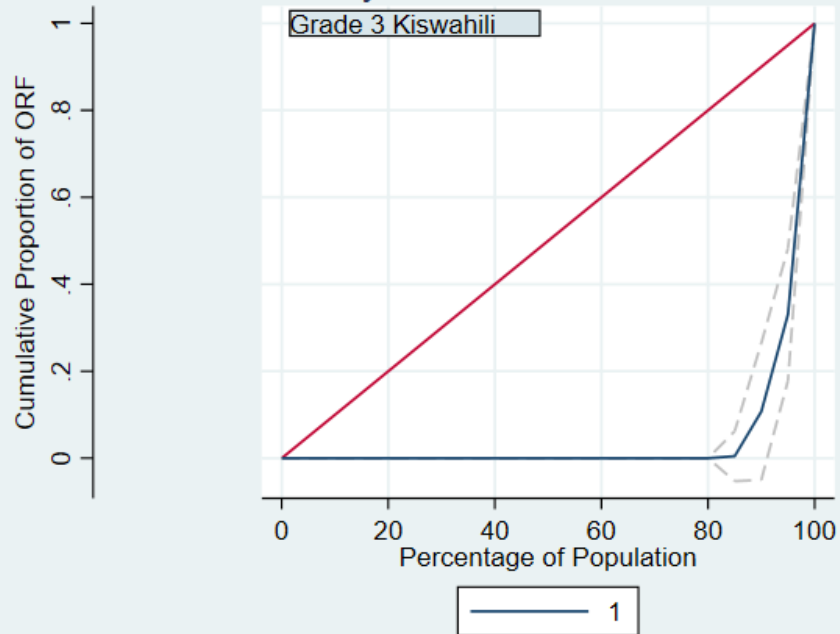


Dataset e.DRC

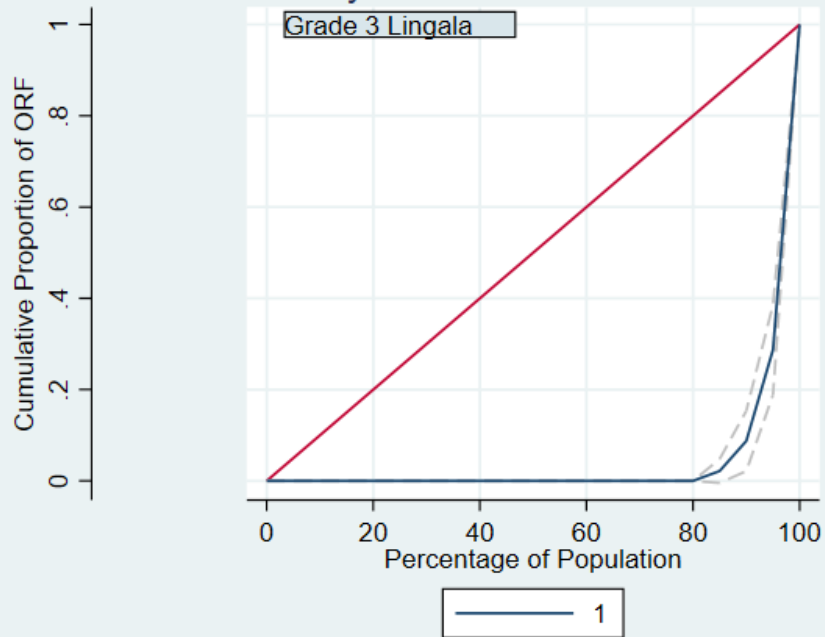
Lorenz curve for Rescaled Grade 777 French
by Round of Assessment



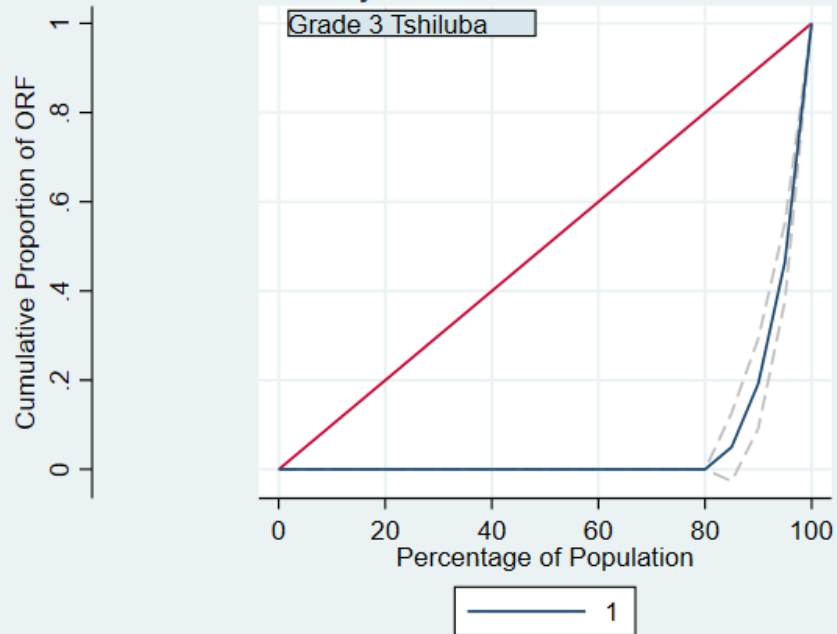
Lorenz curve for Rescaled Grade 3 Kiswahili
by Round of Assessment



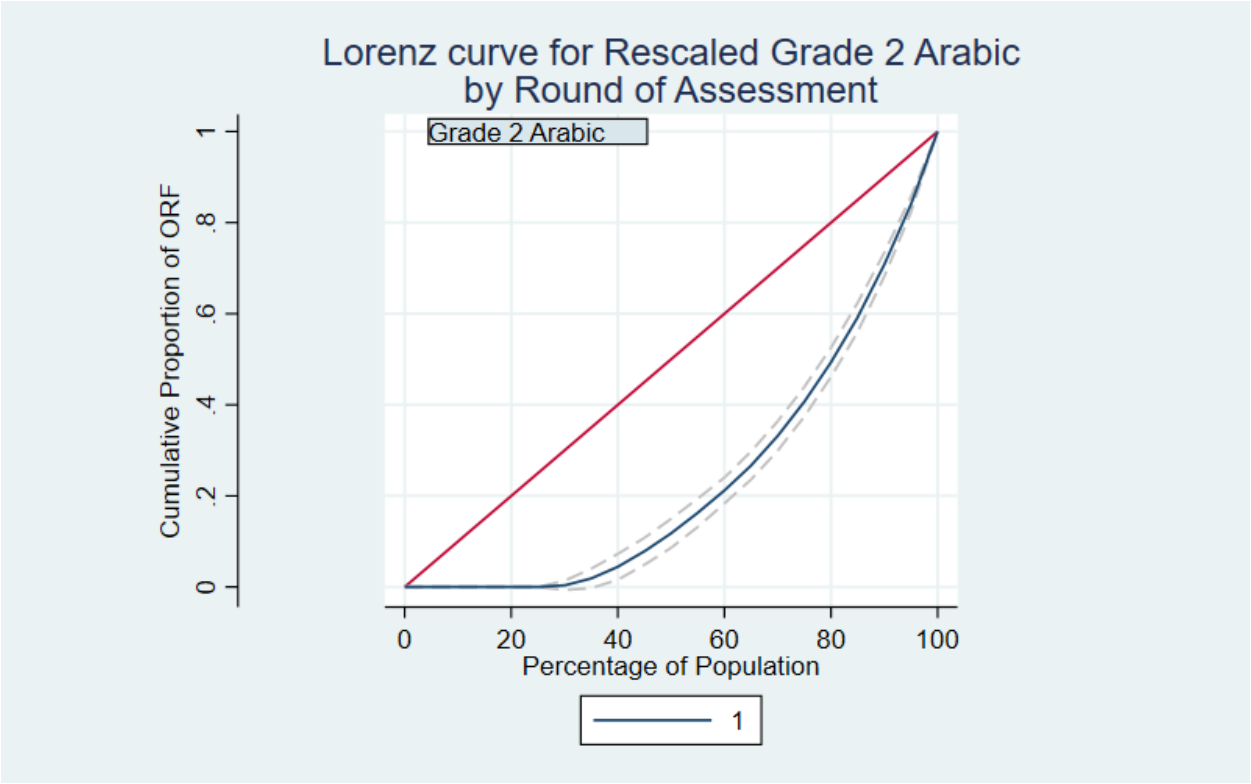
Lorenz curve for Rescaled Grade 3 Lingala
by Round of Assessment



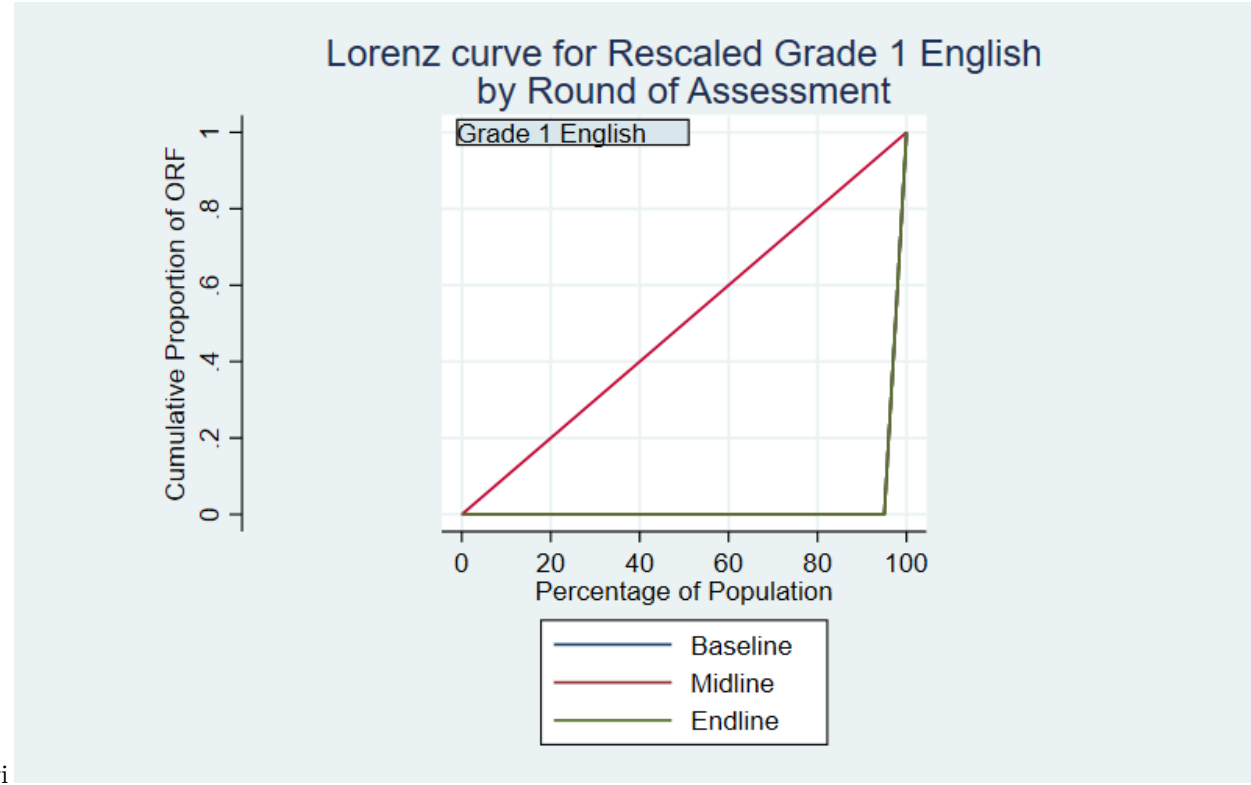
Lorenz curve for Rescaled Grade 3 Tshiluba
by Round of Assessment



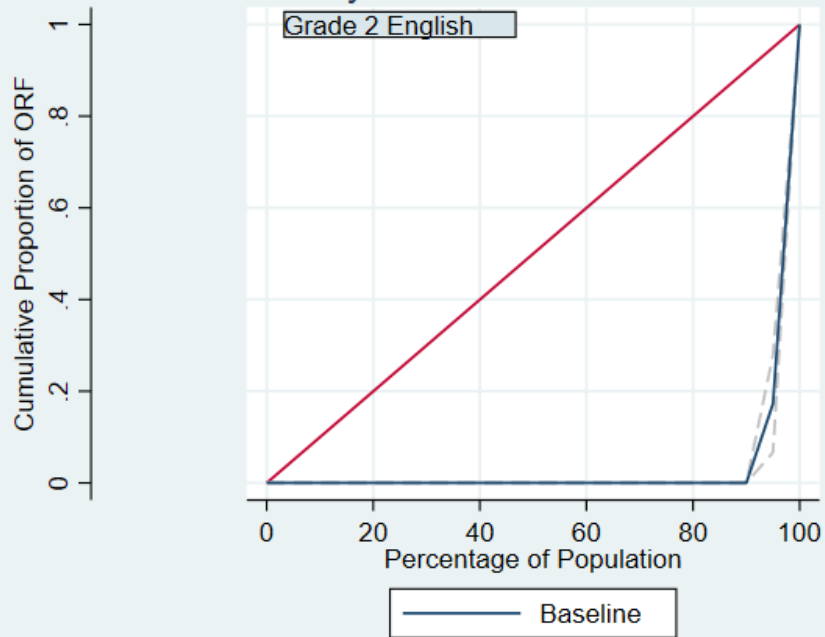
Dataset Egypt



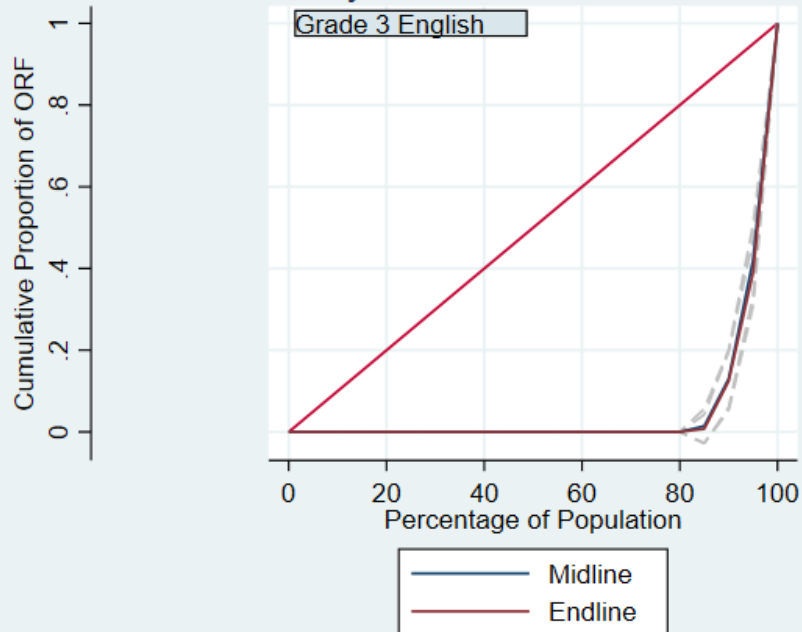
Dataset Malawi



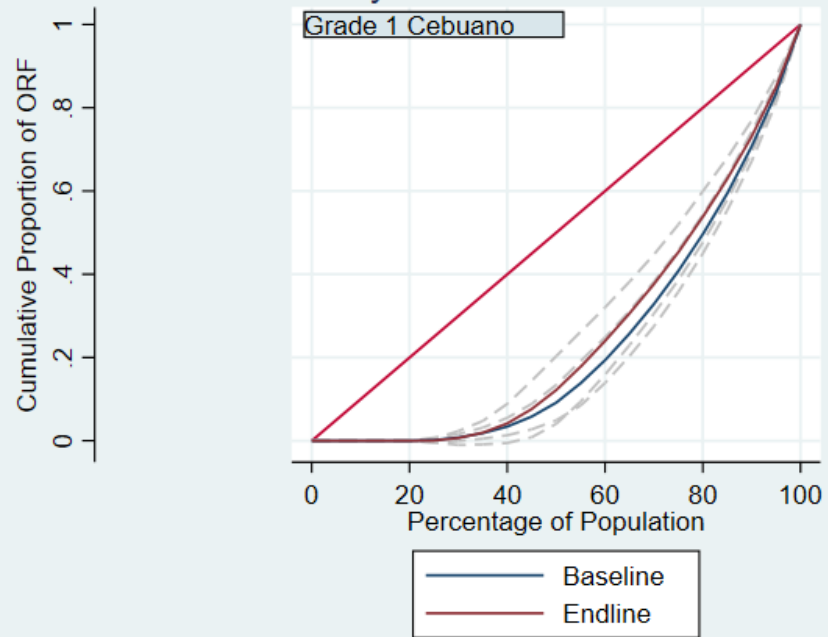
Lorenz curve for Rescaled Grade 2 English
by Round of Assessment



Lorenz curve for Rescaled Grade 3 English
by Round of Assessment

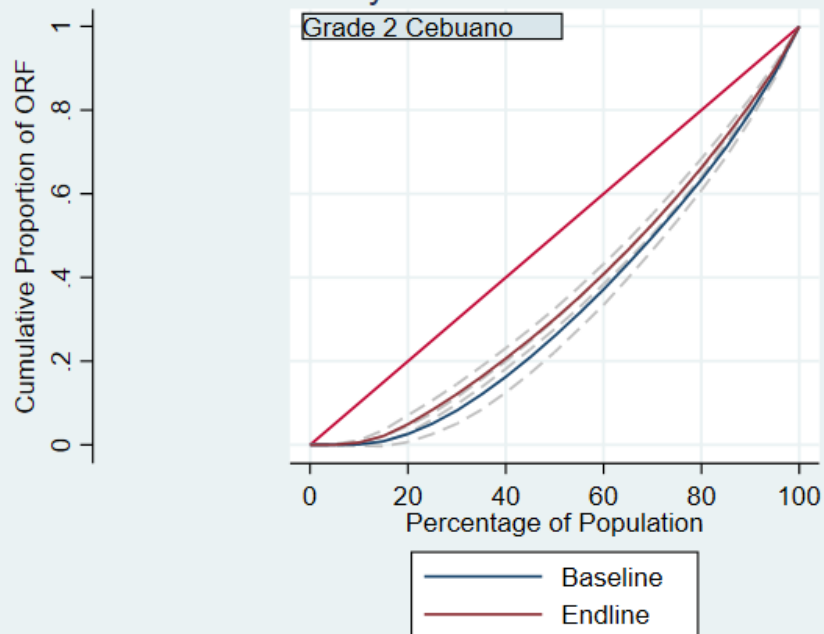


Lorenz curve for Rescaled Grade 1 Cebuano by Round of Assessment

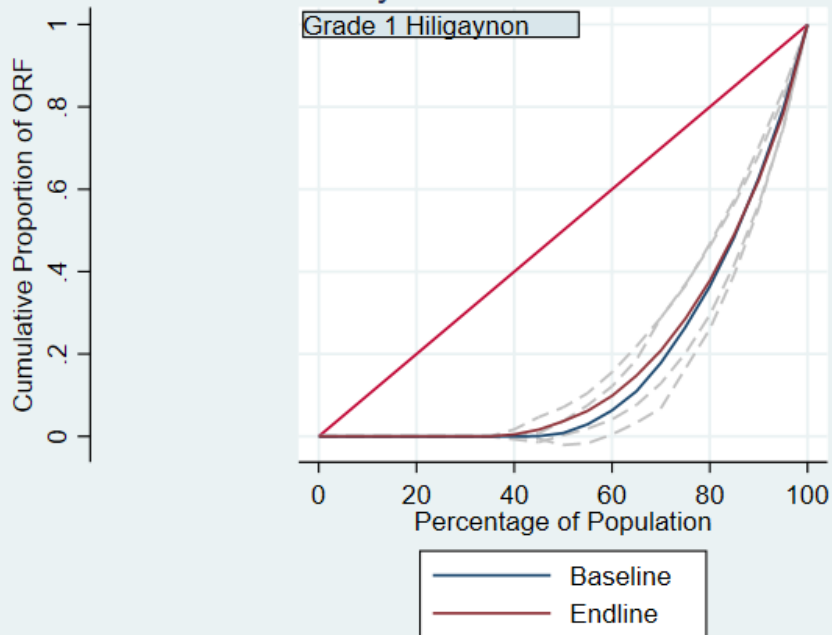


Dataset Philippines

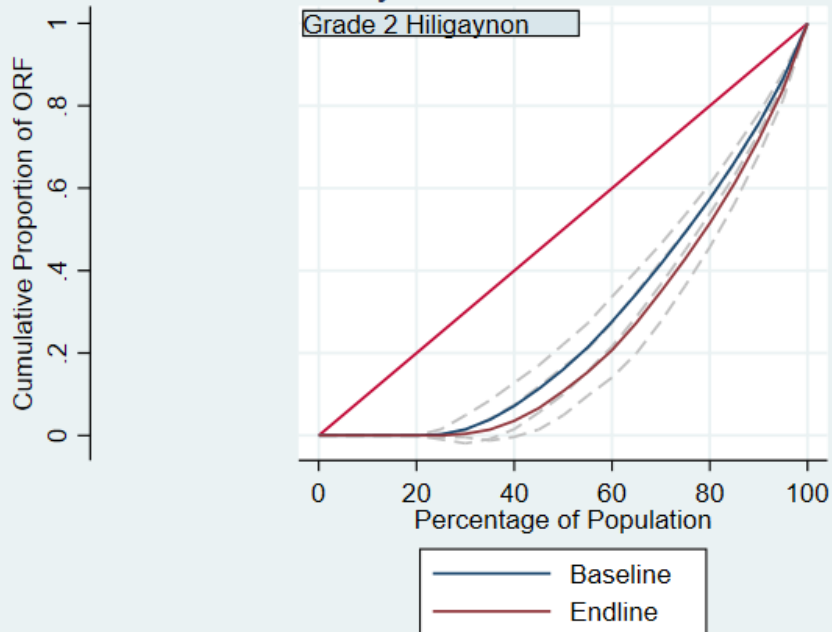
Lorenz curve for Rescaled Grade 2 Cebuano by Round of Assessment



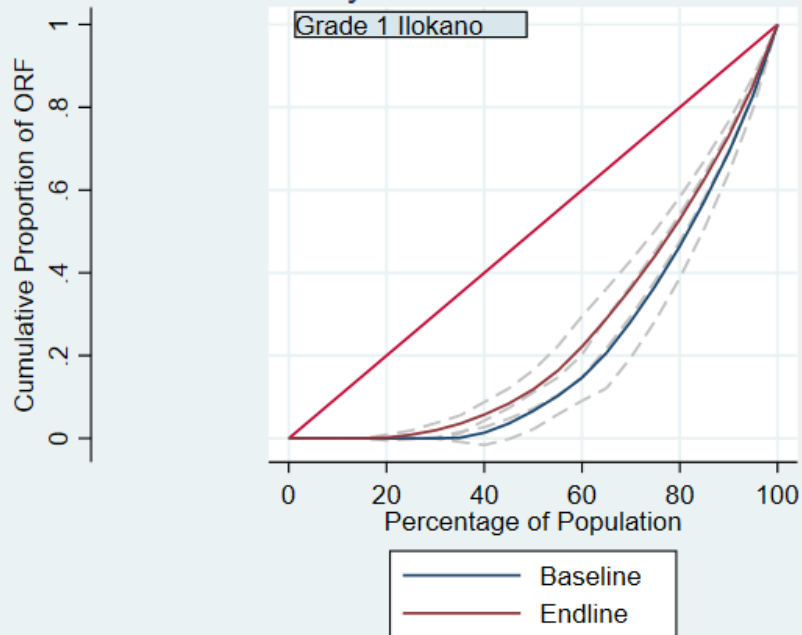
Lorenz curve for Rescaled Grade 1 Hiligaynon
by Round of Assessment



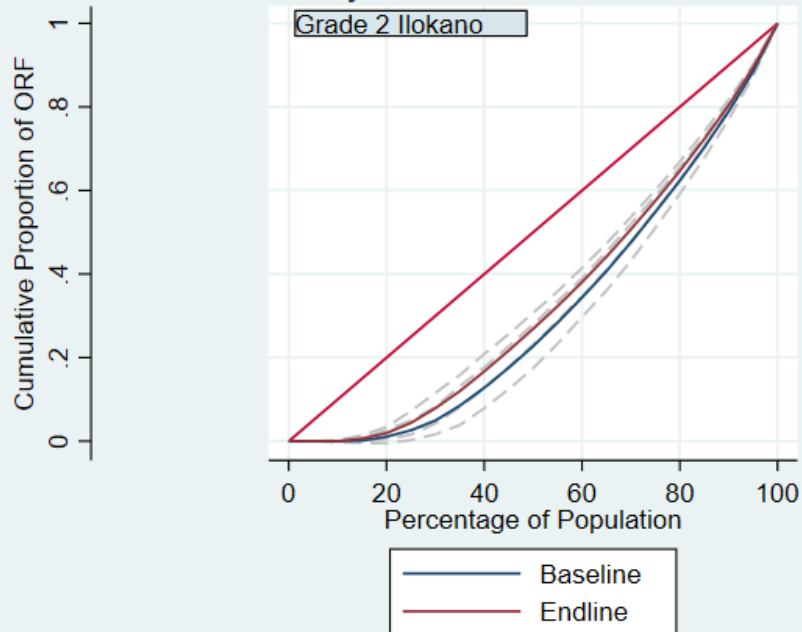
Lorenz curve for Rescaled Grade 2 Hiligaynon
by Round of Assessment



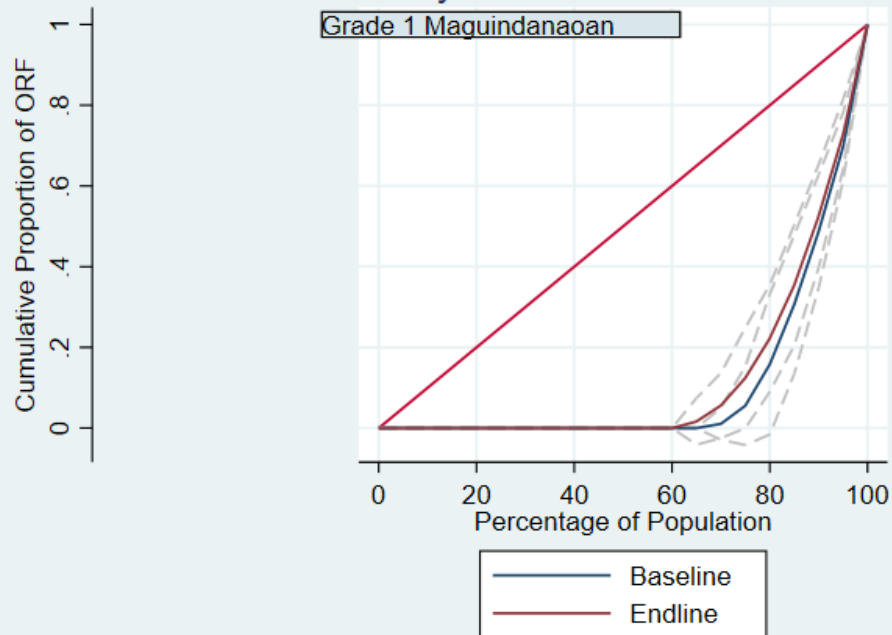
Lorenz curve for Rescaled Grade 1 Ilokano
by Round of Assessment



Lorenz curve for Rescaled Grade 2 Ilokano
by Round of Assessment



Lorenz curve for Rescaled Grade 1 Maguindanaoan
by Round of Assessment



Lorenz curve for Rescaled Grade 2 Maguindanaoan
by Round of Assessment

