



4-2018

# The Gini Coefficient's Magic Does Not Work on Standardized Test Scores

Selene Sunmin Lee  
*University of Pennsylvania*

Follow this and additional works at: [https://repository.upenn.edu/gse\\_grad\\_pubs](https://repository.upenn.edu/gse_grad_pubs)

 Part of the [Education Commons](#)

---

Lee, Selene Sunmin, "The Gini Coefficient's Magic Does Not Work on Standardized Test Scores" (2018). *GSE Graduate Student Research*. 6.

[https://repository.upenn.edu/gse\\_grad\\_pubs/6](https://repository.upenn.edu/gse_grad_pubs/6)

Working paper presented at the Annual Conference of the American Educational Research Association (AERA), New York, NY, April 16, 2018.

This paper is posted at ScholarlyCommons. [https://repository.upenn.edu/gse\\_grad\\_pubs/6](https://repository.upenn.edu/gse_grad_pubs/6)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# The Gini Coefficient's Magic Does Not Work on Standardized Test Scores

## **Abstract**

The Gini coefficient, an indicator that is often used to measure the inequality in the distribution of income within countries, is meaningless when used on standardized test scores. This is because the value of the Gini coefficient depends on the scale's mean and standard deviation which are arbitrarily selected by the test developers. Keeping the standard deviation of the scale constant, increasing the mean will decrease the Gini coefficient, while keeping the mean of the scale constant, increasing the standard deviation will increase the Gini coefficient. In addition, when Gini coefficients are estimated with scores on two different scales, not only the values of the Gini coefficients but also the country rankings of the Gini coefficients will change. Therefore, for standardized test scores, the value of the Gini coefficient is meaningless, as is comparing the size of the Gini coefficients estimated from different countries. More generally, all relative measures of dispersion, including the Gini coefficient, are meaningless for interval scales (i.e., a scale in which the distance between any two consecutive points are equal, but the scale does not have an absolute zero), such as standardized test scores.

## **Disciplines**

Education

## **Comments**

Working paper presented at the Annual Conference of the American Educational Research Association (AERA), New York, NY, April 16, 2018.

**The Gini Coefficient's Magic does not Work on Standardized Test Scores**

Selene Sunmin Lee

*Graduate School of Education  
University of Pennsylvania  
3700 Walnut Street, Philadelphia, PA 19103, USA*

*selenel@gse.upenn.edu  
selenelee0525@gmail.com*

Working paper presented at  
the Annual Conference of the American Educational Research Association (AERA)  
New York, NY  
April 16, 2018

## **Acknowledgements**

I would like to thank Dr. Robert Boruch, my advisor at the University of Pennsylvania, for his constructive feedback on earlier versions of this paper as well as for letting me work on this project as part of my research fellowship in the 2017-2018 academic year. In addition, I thank the American Institutes for Research (AIR) for granting me permission to work on this project during my internship in the summer of 2017. Lastly, I would like to acknowledge that the inspiration for this research came from Dr. Dan Wagner's keynote speech at the Learning at the Bottom of the Pyramid conference at the University of Pennsylvania in March 2017 in which he suggested using the Gini coefficient on standardized test scores to measure the inequality in students' learning achievement within and across countries.

## Abstract

The Gini coefficient, an indicator that is often used to measure the inequality in the distribution of income within countries, is meaningless when used on standardized test scores. This is because the value of the Gini coefficient depends on the scale's mean and standard deviation which are arbitrarily selected by the test developers. Keeping the standard deviation of the scale constant, increasing the mean will decrease the Gini coefficient, while keeping the mean of the scale constant, increasing the standard deviation will increase the Gini coefficient. In addition, when Gini coefficients are estimated with scores on two different scales, not only the values of the Gini coefficients but also the country rankings of the Gini coefficients will change. Therefore, for standardized test scores, the value of the Gini coefficient is meaningless, as is comparing the size of the Gini coefficients estimated from different countries. More generally, all relative measures of dispersion, including the Gini coefficient, are meaningless for interval scales (i.e., a scale in which the distance between any two consecutive points are equal, *but* the scale does not have an absolute zero), such as standardized test scores.

## TABLE OF CONTENTS

<b>Introduction .....</b>	<b>1</b>
<b>Gini coefficient of an income distribution .....</b>	<b>2</b>
<b>History of the Gini coefficient .....</b>	<b>2</b>
<b>Presenting the income distribution with bar graphs .....</b>	<b>2</b>
<b>Presenting the income distribution as a Lorenz curve .....</b>	<b>4</b>
<b>Estimating the Gini coefficient .....</b>	<b>6</b>
<b>Gini coefficient of a test score distribution .....</b>	<b>7</b>
<b>Estimating the Gini coefficient of a test score distribution .....</b>	<b>7</b>
<b>Papers and books on the Gini coefficient and standardized test scores .....</b>	<b>9</b>
<b>Method .....</b>	<b>11</b>
<b>Gini coefficients estimated with 2015 PISA reading scores .....</b>	<b>11</b>
<b>Properties of the Lorenz curve and the Gini coefficient .....</b>	<b>12</b>
<b>Slope of the tangent of the Lorenz curve .....</b>	<b>12</b>
<b>Gini coefficient and the average score .....</b>	<b>13</b>
<b>Empirical evidence .....</b>	<b>13</b>
<b>Mathematical explanation .....</b>	<b>14</b>
<b>Gini coefficient and the standard deviation .....</b>	<b>17</b>
<b>Empirical evidence .....</b>	<b>17</b>
<b>Mathematical explanation .....</b>	<b>18</b>
<b>Differences in scores in the lower percentiles .....</b>	<b>19</b>
<b>Differences in scores in the upper percentiles .....</b>	<b>20</b>
<b>Summary of the differences between the Gini coefficient and the SD .....</b>	<b>22</b>
<b>Gini coefficient and the scale of a standardized test .....</b>	<b>23</b>
<b>Gini coefficient and other measures of dispersion .....</b>	<b>25</b>
<b>Absolute and relative measures of dispersion .....</b>	<b>25</b>
<b>Absolute measures of dispersion .....</b>	<b>26</b>
<b>Standard deviation .....</b>	<b>26</b>
<b>90-10 achievement gap .....</b>	<b>26</b>
<b>Relative measures of dispersion .....</b>	<b>27</b>
<b>Gini coefficient .....</b>	<b>27</b>
<b>90/10 achievement gap .....</b>	<b>28</b>
<b>Coefficient of variation .....</b>	<b>29</b>
<b>Summary of the absolute and relative measures of dispersion .....</b>	<b>30</b>
<b>Implications for standardized test scores .....</b>	<b>31</b>
<b>Value of the Gini coefficient is meaningless .....</b>	<b>31</b>
<b>Comparing the Gini coefficients from the same test is meaningless .....</b>	<b>31</b>
<b>Comparing the Gini coefficients from different tests is meaningless .....</b>	<b>31</b>
<b>Conclusion .....</b>	<b>32</b>

<b>References .....</b>	<b>33</b>
<b>Appendix: Countries included in the data set .....</b>	<b>35</b>

## Introduction

In recent years, there have been noticeable shifts in the focus of international goals related to education. The first shift is an increased focus on students' learning. For example, while the 2000 Millennium Development Goals (MDGs) aimed to "achieve universal primary education (Goal 2; United Nations, n.d.)," focusing on increasing students' access to school, the 2015 Sustainable Development Goals (SDGs) aim to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all (Goal 4; United Nations, 2015, p. 17)." The first target of the latter goal is to "ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes (United Nations, 2015, p. 17)," highlighting the international community's commitment towards improving both learning and access to school.

The increased focus on students' learning has also increased the international community's interest in assessments that can track countries' progress towards the SDGs. These assessments include international large-scale assessments such as the Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS). Regional assessments, such as the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) in Southern and Eastern Africa, the Analysis Programme of the Conference of the Ministers of Education of French speaking countries' Education Systems (PASEC) in West Africa, and the Latin American Laboratory for Assessment of the Quality of Education (LLECE) in Latin America, have also become important. Last but not least, citizen-led assessments, such as the Annual Status of Education Report (ASER) in India and Uwezo in several African countries, are also being used to measure students' learning achievement.

The second shift in international goals related to education is the increased focus on equity, as mentioned above in the first target of SDG Goal 4. The international community's commitment towards equity in education was reiterated in the 2015 Framework for Action for the implementation of SDG 4 which states that "inclusion and equity in and through education is the cornerstone of a transformative education agenda, and we therefore commit to addressing all forms of exclusion and marginalization, disparities and inequalities in access, participation and learning outcomes (World Education Forum, 2015, p. 7)." The Technical Advisory Group established by UNESCO to review and recommend indicators to track the implementation of the education-related SDG goals also stated that "education indicators should aim to capture not just national averages but also the variation across different population subgroups defined by group and individual characteristics, such as sex, wealth, location, ethnicity, language or disability (Technical Advisory Group, 2015, p. 5)." This focus on equity has increased researchers' and practitioners' interest in finding an indicator that can accurately measure the inequity in students' learning, a combination of the two foci in the international goals related to education.

The objective of this paper is to analyze whether the Gini coefficient, an indicator often used to measure the inequality<sup>1</sup> in income distribution within countries, can also be used as an indicator to measure the inequality in learning achievement within countries. This paper starts by

---

<sup>1</sup> Technically, "inequality" and "inequity" are different. "Inequality" refers to the distribution of a variable without taking into account the background characteristics of the subjects (i.e., a univariate measure), while "inequity" takes into account some of their background characteristics such as gender, ethnicity, and income (i.e., a bivariate or multivariate measure). The Gini coefficient is a measure of "inequality," not "inequity," because it does not take into account subjects' background characteristics.



explaining the basic concept of the Gini coefficient. Next, an analysis of the relation between the Gini coefficient and the average score, the standard deviation, and other measures of dispersion are presented. Based on these analyses, the paper ends by explaining why the Gini coefficient is meaningless when it is used on standardized test scores.<sup>2</sup>

It should be noted that this paper only addresses the Gini coefficient estimated with standardized test scores and not the Gini coefficient estimated with years of schooling which is sometimes referred to as “the Gini coefficient of educational attainment” or “the Gini coefficient of education” (refer to Thomas, Wang, and Fan, 2001, for more details). This is because standardized test scores are measured on an interval scale (i.e., the distance between any two consecutive points on the scale are equal, *but* the scale does not have an absolute zero), while years of schooling is measured on a ratio scale (i.e., the distance between any two consecutive points on the scale are equal, *and* the scale has an absolute zero), a key difference that makes the Gini coefficient meaningless for standardized test scores. The differences between the interval scale and the ratio scale and how they affect the Gini coefficient will be explained in detail later.

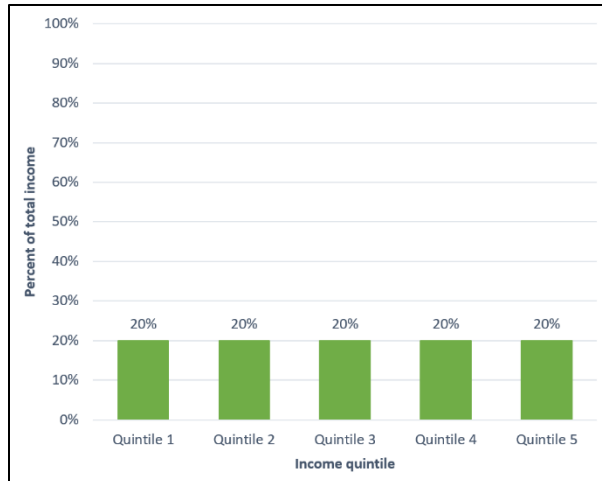
### **Gini coefficient of an income distribution**

**History of the Gini coefficient.** The history of the Gini coefficient can be traced back to 1905, when Max Lorenz, an American economist, proposed a way to visually represent the distribution of income within a country in an article titled *Methods of measuring the concentration of wealth*. According to Derobert and Thieriot (2003), Lorenz was opposed to the idea of representing the inequality of an income distribution with a single numeric indicator, arguing that a better alternative would be to present this information graphically. Ironically, in 1912, the Italian statistician Corrado Gini built upon the work of Lorenz to create the Gini coefficient, a numeric indicator to measure the inequality of an income distribution (Ceriani & Verme, 2012). The section below gives a brief explanation of both the Lorenz curve and the Gini coefficient.

**Presenting the income distribution with bar graphs.** For illustrative purposes, several hypothetical countries will be used to explain the concept of the Lorenz curve and the Gini coefficient. In country A, there are 100 people, and everyone earns the same amount of income. In other words, everyone’s income is equal to the average income of the country. The income distribution of this country is presented in Figure 1. Each bar represents an income quintile and the values on the y-axis indicate the amount of income that is earned by each quintile (as a percentage of the total income in the country). In this hypothetical country, the bars are of equal height, because each quintile accounts for exactly 20% of the total income of the country.

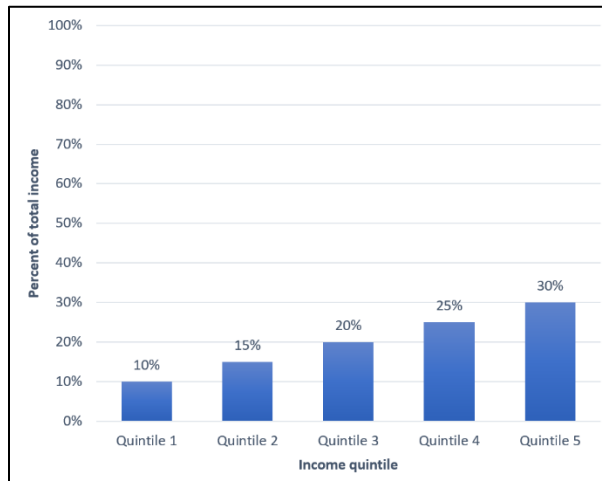
---

<sup>2</sup> Technically, “test” and “assessment” have different meanings. However, they will be used interchangeably in this paper, because “standardized test” is a commonly used word.



*Figure 1.* Income distribution of Country A, a country with equally distributed income.

In another country, Country B, there are also 100 people, but income is distributed unequally among them. The first quintile (the poorest 20% of the population) earns 10% of the total income, the second quintile earns 15% of the total income, the third quintile earns 20% of the total income, the fourth quintile earns 25% of the total income, and the fifth quintile earns 30% of the total income of the country. The income distribution of this country is presented in Figure 2. The bars are of different height, because each quintile accounts for a different percentage of the total income of the country.



*Figure 2.* Income distribution of Country B, a country with unequally distributed income.

Lastly, in Country C, there are also 100 people, but income is distributed very unequally among them. The first quintile earns 5% of the total income, the second quintile earns 8% of the total income, the third quintile earns 12% of the total income, the fourth quintile earns 15% of the total income, and the fifth quintile earns 60% of the total income of the country. The income distribution of this country is presented in Figure 3. The last bar is much taller than the other bars, because the top quintile earns most of the income of the country.

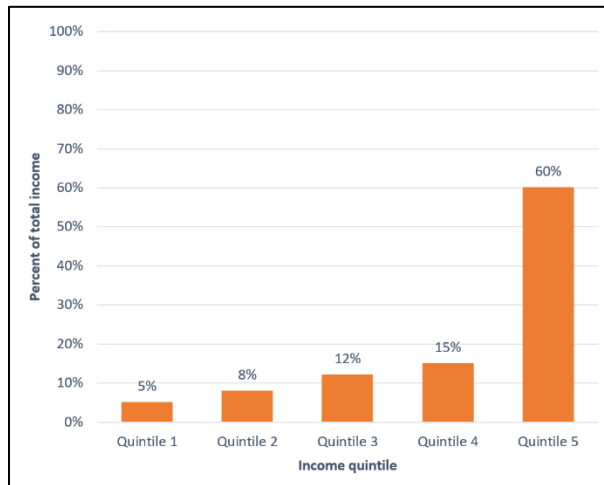
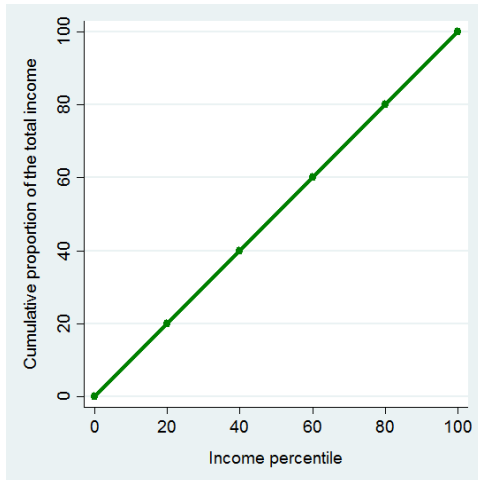


Figure 3. Income distribution of Country C, a country with very unequally distributed income.

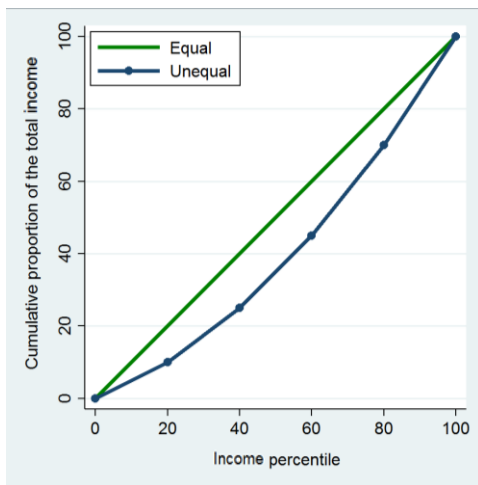
**Presenting the income distribution as a Lorenz curve.** The information in the bar graphs above can also be presented as a Lorenz curve. In a graph of a Lorenz curve, the values on the x-axis represent the cumulative proportion of the population, ordered in terms of each person's income, ranging from 1 to 100. The subject at the 1<sup>st</sup> income percentile (the poorest person in the hypothetical countries) is represented by 1 on the x-axis, the subject at the 50<sup>th</sup> income percentile (the person earning the median income in the hypothetical countries) is represented by 50 on the x-axis, and the subject at the 100<sup>th</sup> income percentile (the richest person in the hypothetical countries) is represented by 100 on the x-axis. The values on the y-axis represent the cumulative proportion of the total income in the country. Using the example of the hypothetical countries above, for each person represented on the x-axis, that person's income is added to the income of *everyone* who has a lower income than that person (i.e., the cumulative income up to that person), and this value is expressed as a percentage of the total income in the country. In other words, the values on the x-axis represent the income percentiles, while the values on the y-axis represent the cumulative proportion of the total income in the country up to that person. Connecting these points will produce the Lorenz curve. It should be noted that having many data points will produce a smooth Lorenz curve, but in the following examples, only five data points will be used for illustrative purposes.

The Lorenz curve for Country A, the country with equally distributed income, is presented in Figure 4. The Lorenz curve for this country is a straight line, because each income quintile earns exactly 20% of the country's total income. Consequently, the cumulative income of each quintile is exactly 20 percentage points higher than the cumulative income of the previous quintile.



*Figure 4.* Lorenz curve of Country A, a country with equally distributed income. Refer to Figure 1 for the income distribution in bar chart form.

The Lorenz curve for Country B, the country with unequally distributed income, is presented in Figure 5. The Lorenz curve for this country is not a straight line, because each quintile earns more income than the previous quintiles.



*Figure 5.* Lorenz curve of Country B, a country with unequally distributed income. Refer to Figure 2 for the income distribution in bar chart form.

Lastly, the Lorenz curve for Country C, the country with very unequally distributed income, is presented in Figure 6. Compared to Country B, the Lorenz curve for Country C has a greater degree of curvature, because income is more unequally distributed.

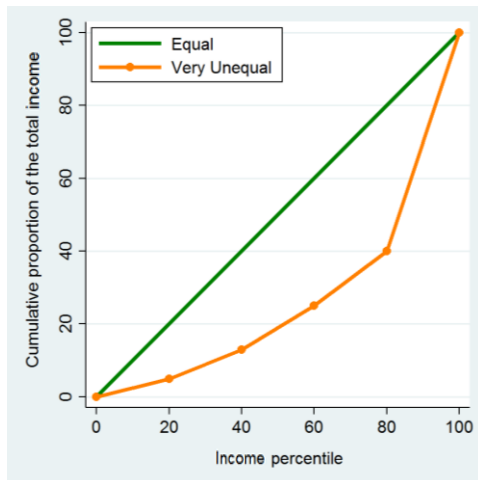
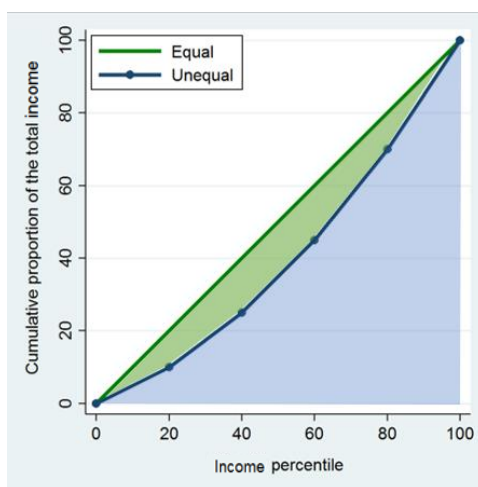


Figure 6. Lorenz curve of Country C, a country with very unequally distributed income. Refer to Figure 3 for the income distribution in bar chart form.

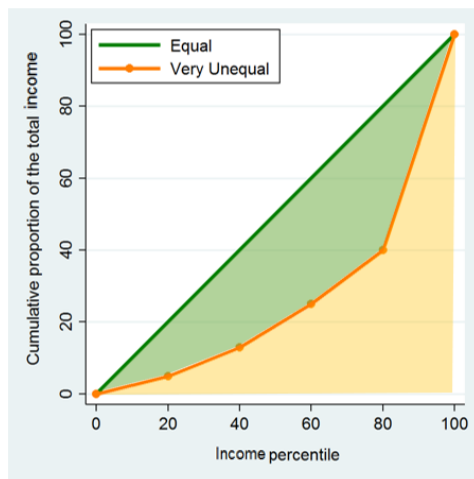
**Estimating the Gini coefficient.** The examples above show that the degree of curvature of a Lorenz curve depends on how unequally income is distributed within a country - the more unequal the income distribution, the greater the curvature of the Lorenz curve will be. To quantify the degree of the curvature, the area between the Lorenz curve (represented by a blue line in Figure 7) and the diagonal line (represented by a green line in Figure 7) can be estimated. While calculus is required to estimate this area precisely, it can be approximated using algebra. For Country B, the area between the Lorenz curve and the diagonal line (shaded in green in Figure 7) can be estimated to be 1,000. To standardize this value, it can be divided by the triangular area under the diagonal line (the sum of the areas shaded in green and blue in Figure 7). This triangular area is 5,000 for all countries (because  $100 \times 100 / 2 = 5,000$ ). For Country B, dividing the green area by the triangular area is 0.20 (because  $1,000 / 5,000 = 0.20$ ). Multiplying this value by 100 is 20, which is the Gini coefficient for this country.



**\* Gini coefficient: 20.0**

Figure 7. Lorenz curve of Country B, a country with unequally distributed income. Refer to Figure 2 for the income distribution in bar chart form.

For Country C, the area between the Lorenz curve and the diagonal line is shaded in green in Figure 8. It can be noted that the green shaded area for Country C is larger than the green shaded area for Country B, because income in Country C is more unequally distributed than in Country B. Using algebra, the green shaded area in Figure 8 can be estimated to be 2,340. Dividing the green area by the triangular area results in 0.468 (because  $2,340/5,000 = 0.468$ ). Multiplying this value by 100 is 46.9. Thus, the Gini coefficient for Country C is 46.8, which is much larger than the Gini coefficient for Country B (which was 20.0).



**\* Gini coefficient: 46.8**

*Figure 8.* Lorenz curve of Country C, a country with very unequally distributed income. Refer to Figure 3 for the income distribution in bar chart form.

Thus, the Gini coefficient is a measure of how unequally income is distributed within a country, and the values can theoretically range from 0 to 100. The more unequally income is distributed within a country, the larger the area between the Lorenz curve and the diagonal line will be, resulting in a larger Gini coefficient. It should be noted that Country A, the country with equally distributed income, has a Gini coefficient of 0, because there is no area between the country's Lorenz curve and the diagonal line.

### **Gini coefficient of a test score distribution**

**Estimating the Gini coefficient of a test score distribution.** Using the same methodology as above, the Lorenz curve of a test score distribution can also be plotted on a graph. To illustrate this, another hypothetical country will be used. In Country D, 100 students took a test. The bottom 20 students scored 300 points, the middle 60 students scored 500 points, and the top 20 students scored 700 points, as portrayed in Figure 9. The average score for this country is 500 points, as represented by the grey horizontal line.

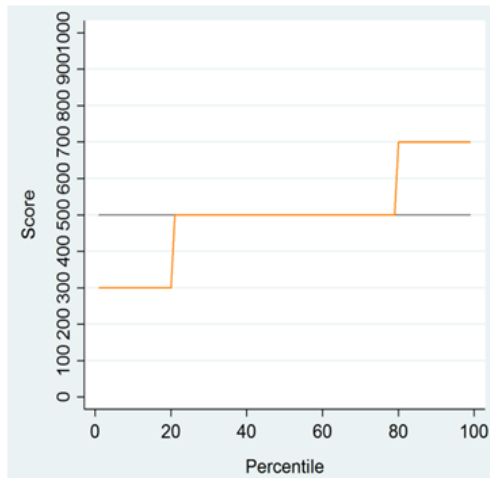
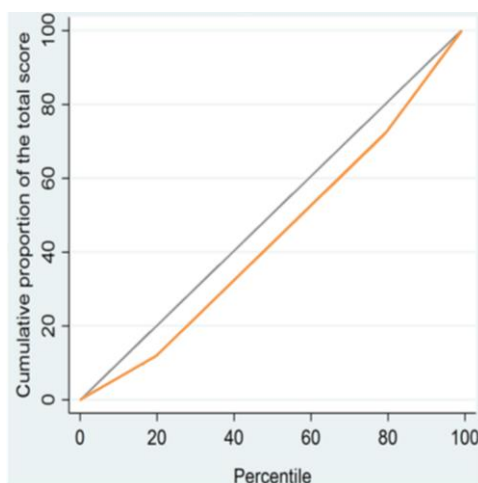


Figure 9. Score distribution of Country D.

The Lorenz curve for this country's score distribution is presented in Figure 10. The sum of all the scores in this country is 50,000 (because  $300 \cdot 20 + 500 \cdot 60 + 700 \cdot 20 = 50,000$ ). Up to the 20<sup>th</sup> percentile, the range in which each student scored 300 points, a 1 unit increase on the x-axis results in an increase of 0.6 units on the y-axis, because each additional student in this range accounts for 0.6% of the total score (because  $300/50,000 = 0.006$ ). Between the 21<sup>st</sup> percentile and the 80<sup>th</sup> percentile, the range in which each student scored 500 points, a 1 unit increase on the x-axis results in an increase of 1 unit on the y-axis, because each additional student in this range accounts for 1% of the total score (because  $500/50,000 = 0.01$ ). Between the 81<sup>st</sup> and the 100<sup>th</sup> percentile, the range in which each student scored 700 points, a 1 unit increase on the x-axis results in an increase of 1.4 units on the y-axis, because each additional student in this range accounts for 1.4% of the total score (because  $700/50,000 = 0.014$ ). The area between the Lorenz curve (in orange) and the diagonal line (in grey) is 640. Therefore, the Gini coefficient of this distribution is 12.8 (because  $640/5,000 \cdot 100 = 12.8$ ).



\* **Gini coefficient: 12.8**

Figure 10. Lorenz curve of Country D. Refer to Figure 9 for the actual scores at each percentile.

As with the Gini coefficient for income distribution, a higher Gini coefficient of a test score distribution means that there is more dispersion in the test scores. However, the Gini coefficient is not a meaningful indicator of the dispersion of test scores, as will be explained later in this paper.

**Papers and books on the Gini coefficient and standardized test scores.** A major attraction of the Gini coefficient is that it is dimensionless and has a range of 0 to 100 regardless of the units of the original scale. This gives the false impression that the Gini coefficient can be used with any kind of scale and that it can also be used to compare the dispersion of scores from different tests. For this reason, many researchers have estimated the Gini coefficient with test scores to measure the inequality in learning achievement within a country.

For example, Bedard and Ferrall (2003) used the Gini coefficient to measure the inequality in the distribution of test scores on the International Mathematics Examinations administered to 13-year-olds as well as the inequality in the distribution of wages later in life in 11 countries. They claimed that the Gini coefficient would “reduce the spurious effects of the testing instrument and facilitate comparisons between test score and wage distributions that are robust to quite distinct distributions (p. 33).” They concluded that in most countries, wages were more equally distributed than test scores, because for a given country, the Gini coefficient of the wage distribution was generally smaller than the Gini coefficient of the test score distribution. They also examined the relation between the Gini coefficient of the test score distribution and the Gini coefficient of the wage distribution and concluded that there was a positive association between these two indicators.

A World Bank publication edited by Vegas (2005) also estimated Gini coefficients using scores on the National System for Basic Education Evaluation, a national educational assessment administered every two years in Brazil. The report concluded that “the Gini coefficients are small in relative terms, indicating that the variation in student test scores among and between regions is not substantial (p. 168).” They also claimed that the inequality in test scores had increased between 1995 and 2001, because the Gini coefficients of the test score distribution had increased during this period.

Sastry and Pebley (2010) also estimated Gini coefficients using reading and math scores from the Los Angeles Family and Neighborhood Survey and analyzed how much of the inequality in test scores could be explained by family characteristics such as family income, family assets, mother’s reading score, and mother’s years of schooling. A similar approach was taken by Freeman, Machin, and Viarengo (2011) who estimated the Gini coefficient using math scores from PISA 2000 and 2009. They subsequently analyzed the association between the Gini coefficient and various family background characteristics such as parents’ education and the number of books at home.

More recently, Wagner (2018) suggested that the Gini coefficient could be used as an indicator to measure the inequality in learning achievement at the national and sub-national levels. He claimed that Gini coefficients estimated from different scales would be comparable, providing a more culturally sensitive and practical way of comparing the inequality in learning achievement in different countries and sub-national groups. In addition, he suggested that the Gini coefficient would enable researchers and practitioners to measure changes in learning inequality over time, and to build on other Gini coefficients (e.g., Gini coefficient of income, Gini coefficient of years of schooling) in a coherent way.

In spite of the examples above, several papers and books, some dating back several decades, have advised against using the Gini coefficient on standardized test scores. The main



reason is that the Gini coefficient is meaningless when it is estimated with an interval scale (i.e., a scale in which the distance between any two consecutive points are equal, but the scale does not have an absolute zero). Scores on a standardized test are on an interval scale, because the distance between any two consecutive points represents the same amount of difference in students' abilities,<sup>3</sup> but the scale does not have an absolute zero (i.e., the zero of the scale has no inherent meaning). This is because the average score for a standardized test is arbitrarily selected by the test developer.<sup>4</sup> In the case of PISA, the scores of the students in the OECD countries that took the first round PISA in 2000<sup>5</sup> were standardized to be on a scale with a mean of 500 and a standard deviation of 100 (OECD, 2003). However, the test developers of PISA could have selected another scale, for example, a scale with a mean of 600 and standard deviation of 200. In this scenario, a student who received a score of 500 on the original PISA scale would receive a score of 600 on the new scale, although the student's actual ability would not have changed. This means that the value of 500 on the original PISA scale has no inherent meaning, which also means that the distance between zero and 500 on this scale has no inherent meaning. Since the zero on a standardized test is not meaningful, it makes standard test scores an interval scale rather than a ratio scale.

A statistics textbook by Kendall and Stuart in 1977 (as cited in Allison, 1978) had the following warning about using the Gini coefficient on distributions on an interval scale:

[The Gini coefficient suffers] from the disadvantage of being very much affected by . . . the value of the mean measured from some arbitrary origin, and are not usually employed unless there is a natural origin of measurement or comparisons are being made between distributions with similar origins.

Jencks et al., in their 1972 book titled *Inequality* (also cited in Allison, 1978), specifically stated that the Gini coefficient should not be used for distributions on an interval scale, such as IQ scores. Allison (1978) backed this claim, saying that in general, valid inferences about inequality cannot be made using interval data, because "different origins will lead to different conclusions about the relative inequality of the two distributions (p. 871)."<sup>6</sup>

Ferreira and Gignoux (2014), in their paper on measuring the inequality of learning achievement, also gave several reasons why they decided not to use the Gini coefficient in their analyses. Most importantly, they proved mathematically that the value of the Gini coefficients estimated with pre-standardized scores (i.e., trait levels estimated with item response theory that have *not* been linearly converted to standardized test scores) were different from the values of the Gini coefficients estimated with post-standardized scores (i.e., trait levels estimated with item response theory that have been linearly converted to standardized test scores). Furthermore, they

---

<sup>3</sup> This paper will assume that this is true, although some scholars may argue against this.

<sup>4</sup> For standardized tests, item response theory is used to estimate subjects' scores, either anchoring the mean of the item difficulties ( $\beta$ ) to zero or the mean of the trait levels ( $\theta$ ) to zero (Embretson & Reise, 2000). In either case, after the trait levels are estimated for all the subjects, the trait levels are converted into standardized test scores through a linear conversion using an arbitrary mean and standard deviation for the new distribution, such as a mean of 500 and standard deviation of 100.

<sup>5</sup> While 28 OECD countries participated in PISA 2000, the Netherlands was excluded from the standardization procedure because it did not reach the PISA 2000 sampling standards. As a result, 27 countries were included in the standardization procedure, with the scores from each country weighted equally (OECD, 2002, p. 255).

<sup>6</sup> Allison makes an exception for interval scales which have an underlying non-negative ratio scale such as measures of social power.

showed that the country rankings of the Gini coefficients using pre-standardized scores were different from the country rankings of the Gini coefficients using post-standardized scores, a major flaw when using the Gini coefficient to make comparisons across countries.

This paper builds on the works cited above to show that Gini coefficients estimated with distributions on an interval scale are meaningless.

### Method

The analyses and graphs in this paper are based on simulated data as well as empirical data from the 2015 PISA Reading assessment downloaded from the PISA database.<sup>7</sup> PISA is an international assessment that is coordinated by the Organization for Economic Cooperation and Development (OECD) with the goal of measuring the knowledge and skills of 15-year-old students in reading, mathematics, and science (OECD, n.d.). The first round of PISA was administered in the year 2000 in 28 OECD and 15 non-OECD countries (OECD, 2003). Since then, PISA has been administered every 3 years, with the 6<sup>th</sup> and latest round administered in 2015 in 35 OECD and 37 non-OECD countries (OECD, 2016). In this paper, data that are not nationally representative, such as data collected from certain cities or states of a country (e.g., Beijing, Shanghai, Jiangsu, Guangdong, Macao, and Hong Kong in China; Massachusetts and North Carolina in the United States; Buenos Aires in Argentina) were excluded from the data set. As a result, the data used in this paper included 66 countries comprised of 35 OECD and 31 non-OECD countries. The list of countries included in the data set used in this paper can be found in the Appendix.

The IDB Analyzer,<sup>8</sup> a tool developed by the International Association for the Evaluation of Educational Achievement (IEA), was used to estimate the score at each percentile for each country. This tool takes into account the complex sample design of PISA (i.e., multi-stage stratified and cluster sampling) as well as the missing data patterns caused by the fact that only a fraction of the items in the assessment are administered to each student.<sup>9</sup> All subsequent analyses were conducted in Stata (version 14).

### Gini coefficients estimated with 2015 PISA reading scores

Using the method described above, Gini coefficients were estimated for all 66 countries that had nationally representative data on the 2015 PISA Reading assessment. Among the 66 countries, Vietnam had the lowest Gini coefficient of 8.1, while Lebanon had the highest Gini coefficient of 18.3. The Lorenz curve of these two countries are presented in Figure 11. Compared to the Gini coefficients estimated with income distribution, which can be as high as 50 (World Bank, n.d.), the size of the Gini coefficients estimated with PISA scores are relatively small. However, the size of the Gini coefficients estimated with standardized test scores are meaningless, because it depends heavily on the mean and standard deviation of the scale that are

---

<sup>7</sup> PISA database: <http://www.oecd.org/pisa/data/2015database/>

<sup>8</sup> IDB Analyzer: <http://www.iea.nl/data>

<sup>9</sup> It should be noted that the IDB Analyzer would not estimate the score at the 100<sup>th</sup> percentile, so only the scores from the 1<sup>st</sup> to the 99<sup>th</sup> percentiles were used to estimate the Gini coefficients. The equation for the Gini coefficient was adjusted to take this into account.

arbitrarily selected by the test developers. This will be explained in more detail later.

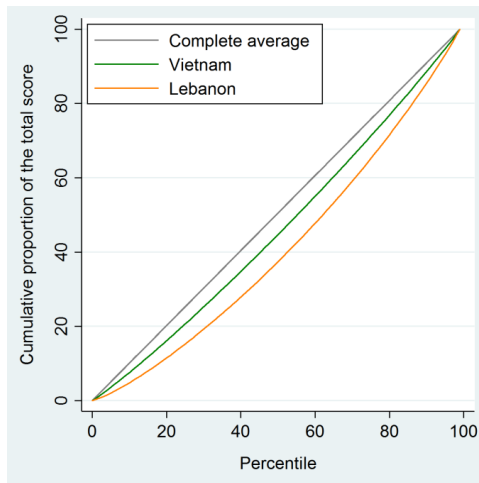


Figure 11. Lorenz curve for Vietnam and Lebanon using 2015 PISA Reading scores.

### Properties of the Lorenz curve and the Gini coefficient

#### Slope of the tangent of the Lorenz curve

The slope of the tangent of the Lorenz curve at a given point on the x-axis is directly related to the score of the subject at that percentile. This will be explained with the hypothetical score distribution of Country D, a country in which the bottom 20 students scored 300 points, the middle 60 students scored 500 points, and the top 20 students scored 700 points, as presented in Figure 9 above. In the graph of the Lorenz curve in Figure 10, each unit on the x-axis represents one student, because there are 100 students in the country and 100 units on the x-axis. Each unit on the y-axis represents 500 points, because each unit represents 1% of the sum of everyone's score in the country (because  $0.01 * [300*20 + 500*60 + 700*20] = 500$ ). 500 is also the average score for the country (because  $[300*20 + 500*60 + 700*20]/100 = 500$ ). So each unit on the y-axis also corresponds to the average score.

In this country, students between the 21<sup>st</sup> and the 80<sup>th</sup> percentile received a score of 500 points, the average score. In this range, a 1 unit increase on the x-axis of the Lorenz curve (which corresponds to one student) will result in a 1 unit increase on the y-axis (because the student will add 500 points to the *cumulative* score up to that point), resulting in a slope of 1. Thus, at a point on the x-axis which corresponds to a student who received the average score, the slope of the tangent of the Lorenz curve will be exactly 1.

In the range between the 1<sup>st</sup> and 20<sup>th</sup> percentile, students received a score of 300 points, which is below the average score. In this range, a 1 unit increase on the x-axis of the Lorenz curve (which corresponds to one student) will result in a less-than-one unit increase on the y-axis (because the student will add only 300 points to the *cumulative* score up to that point), resulting in a slope that is less than 1. Thus, at a point on the x-axis which corresponds to a student who received a score that is lower than the average, the slope of the tangent of the Lorenz curve will be less than 1.

In the range between the 81<sup>st</sup> and the 100<sup>th</sup> percentile, students received a score of 700 points, which is above the average score. In this range, a 1 unit increase on the x-axis of the Lorenz curve (which corresponds to one student) will result in a more-than-one unit increase on the y-axis (because the student will add 700 points to the *cumulative* score up to that point), resulting in a slope that is greater than 1. Thus, at a point on the x-axis which corresponds to a student who received a score that is higher than the average, the slope of the tangent of the Lorenz curve will be greater than 1.

The relation between the slope of the tangent of the Lorenz curve and a student's score at a given percentile is summarized in Figure 12.

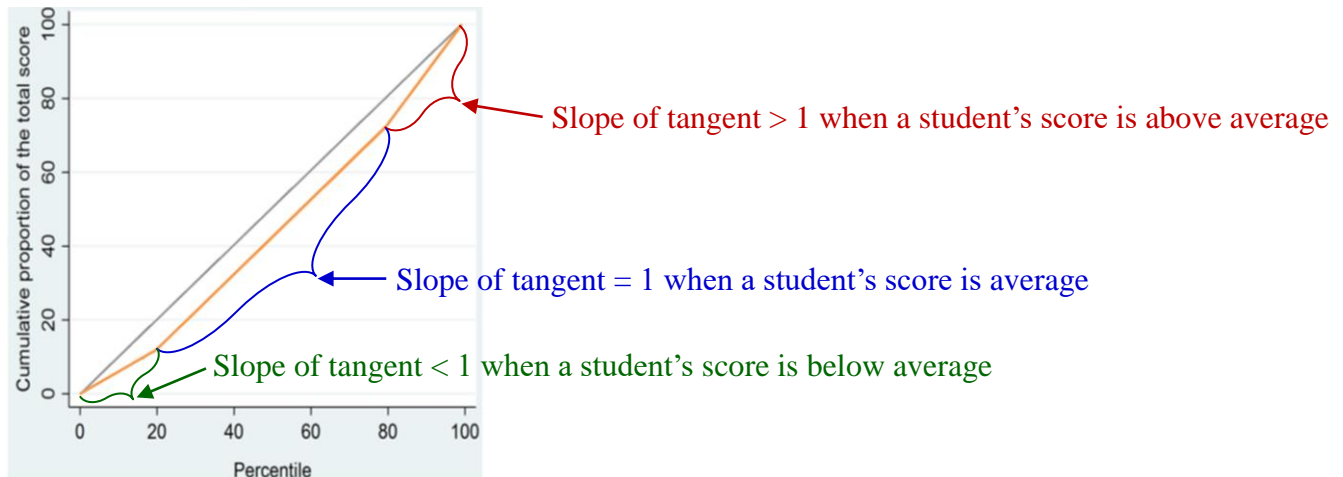


Figure 12. Relation between the slope of the tangent of the Lorenz curve and a student's score at a given percentile.

Generally, at any given point on the x-axis, the slope of the tangent of the Lorenz curve can be estimated with the following equation:

$$\text{Slope of the tangent of the Lorenz curve} = \frac{\text{Student's score at the given percentile}}{\text{Average score}} \quad (1)$$

### Gini coefficient and the average score

**Empirical evidence.** One interesting (and perplexing) property of the Gini coefficient is that when the shape of the distribution is maintained, the Gini coefficient decreases when the average of the distribution increases. This is illustrated in Figure 13 which shows the score distributions of Hong Kong and the Dominican Republic on the 2015 PISA Reading assessment. The two distributions have similar standard deviations (85.8 for Hong Kong vs. 84.9 for the Dominican Republic) and the shape of the distribution is also similar. However, the average for Hong Kong is higher than that of the Dominican Republic (527 for Hong Kong vs. 358 for the Dominican Republic), but the Gini coefficient for Hong Kong is *lower* than that of the Dominican Republic (8.86 for Hong Kong vs. 13.05 for the Dominican Republic).

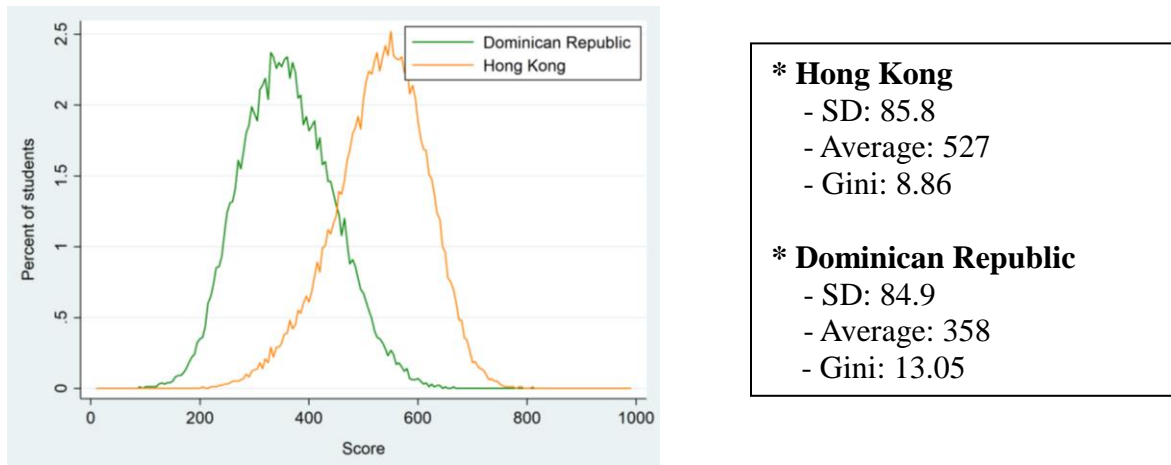


Figure 13. Score distributions of the Dominican Republic and Hong Kong on the 2015 PISA Reading assessment.

Figure 14 presents the scatter plot of the average scores and the Gini coefficients estimated with 2015 PISA Reading scores. The correlation between these two variables is  $-0.67$ , which means that as the average score increases, the Gini coefficient generally decreases. This may seem to suggest that countries with a high average score also tend to have less dispersion in their scores, a phenomenon described by Freeman, Machin, and Viarengo (2011) as “a virtuous efficiency-equity relation in test performance (pg. 5).” However, as explained above, this may just be a mathematical phenomenon - keeping the standard deviation constant, countries with a higher average score will *always* have a lower Gini coefficient. Thus, when a country has a lower Gini coefficient than another, it may be because the former has less dispersion in its score distribution, or simply because it has a higher average score (and the same amount of dispersion).

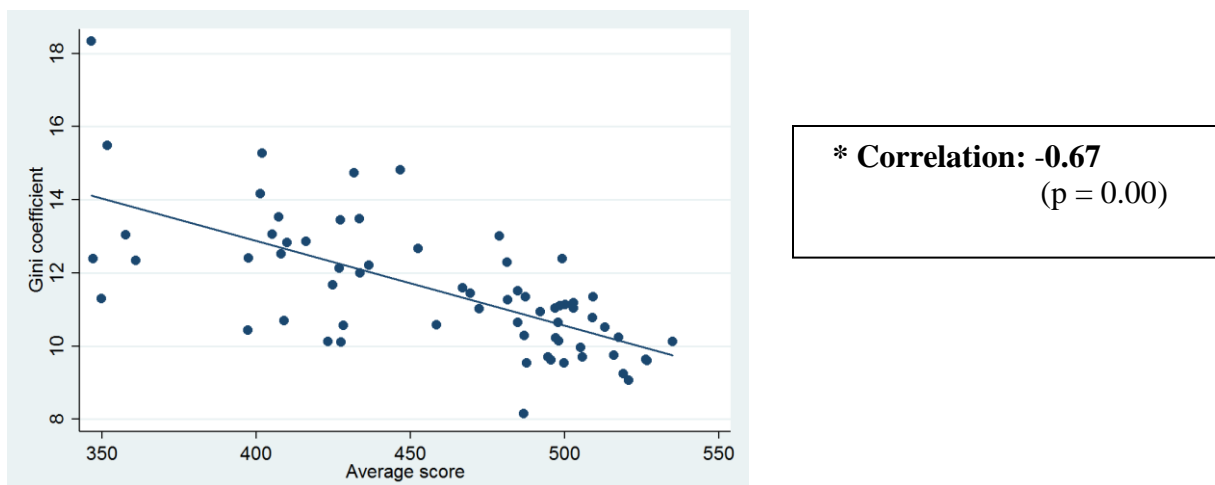


Figure 14. Scatter plot of the average scores and the Gini coefficients estimated with 2015 PISA Reading scores.

**Mathematical explanation.** The mathematical explanation for the inverse relation between the average score and the Gini coefficient is quite simple. Again, this will be explained

with the hypothetical score distribution of Country D, a country in which the bottom 20 students scored 300 points, the middle 60 students scored 500 points, and the top 20 students scored 700 points, as presented in Figure 9 above. Using Equation 1, the slope of the tangent of the Lorenz curve at the bottom, middle, and upper percentiles can be estimated, as shown in Figure 15. As stated previously, the Gini coefficient for this distribution is 12.8.

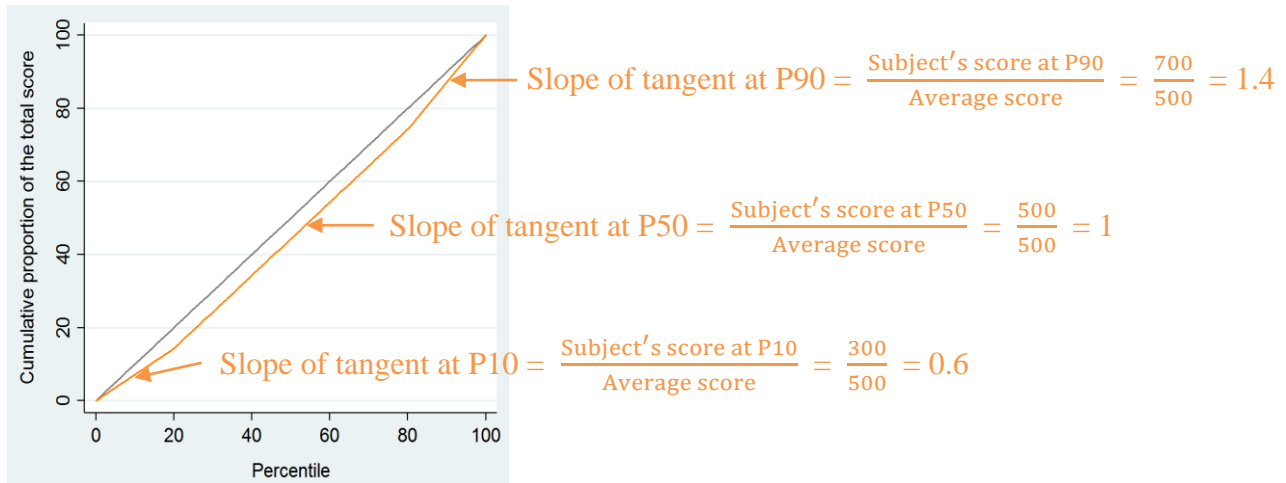


Figure 15. Lorenz curve of Country D. Refer to Figure 9 for the actual score at each percentile.

For illustrative purposes, another hypothetical country, Country E, is presented below. In Country E, everyone scored 500 points higher than in Country D. The bottom 20 students scored 800 points, the middle 60 students scored 1,000 points, the top 20 students scored 1,200 points, making the average score of this country 1,000 points. The score distribution for Country E is presented in Figure 16. The shape of the score distribution for Country D (Figure 9) and Country E (Figure 16) are the same, but the score distribution curve for Country E moved up by 500 points on the y-axis.

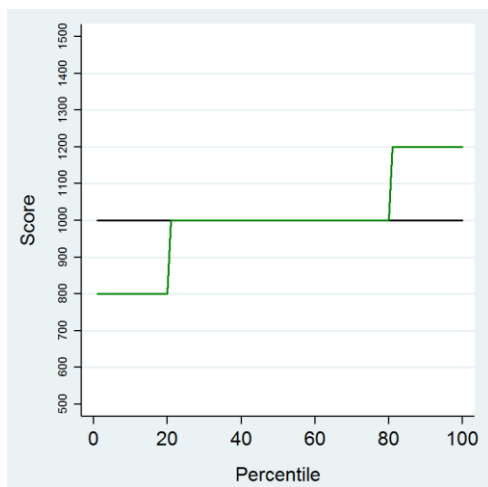


Figure 16. Score distribution of Country E.

The Lorenz curve for this new score distribution is presented in Figure 17. The Gini coefficient for Country E is 6.5, which is smaller than the Gini coefficient for Country D (which was 12.8).

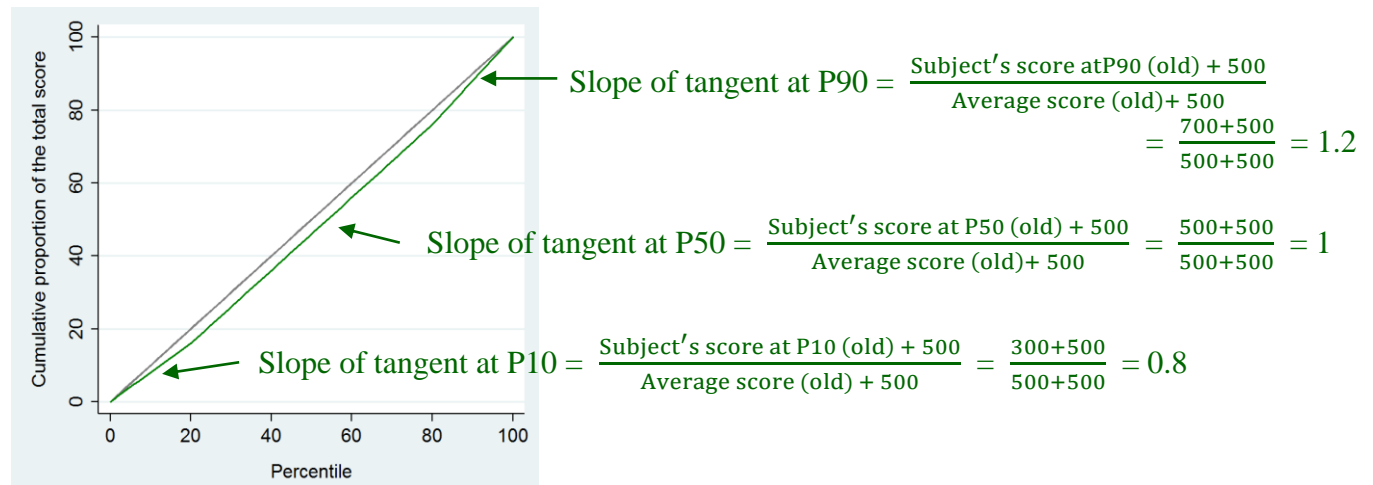


Figure 17. Lorenz curve of Country E. Refer to Figure 16 for the actual score at each percentile.

In the bottom percentiles of Country E, the slope of the tangent of the Lorenz curve is 0.8, which is steeper than the slope of the tangent of the Lorenz curve of the bottom percentiles of Country D (which was 0.6). This is because 500 is added to both the numerator and the denominator when estimating the slope of the new tangent. Since the numerator increases by 167% (from 300 to 800), while the denominator only increases by 100% (from 500 to 1,000), the relative increase in the numerator is greater than the relative increase in the denominator, resulting in a steeper tangent.

In the middle percentiles of Country E, the slope of the tangent of the Lorenz curve is 1, which is the same as the slope of the tangent of the Lorenz curve of the middle percentiles of Country D. This is because the numerator increases by 100% (from 500 to 1,000), while the denominator also increases by 100% (from 500 to 1,000), maintaining the ratio between the numerator and the denominator. Thus, the slope of the tangent is maintained.

In the upper percentiles of Country E, the slope of the tangent of the Lorenz curve is 1.2, which is flatter than the slope of the tangent of the Lorenz curve of the bottom percentiles of Country D (which was 1.4). Since the numerator increases by 71% (from 700 to 1,200), while the denominator increases by 100% (from 500 to 1,000), the relative increase in the numerator is smaller than the relative increase in the denominator, resulting in a flatter tangent.

Overall, the Lorenz curve for Country E is closer to the grey diagonal line than the Lorenz curve for Country D, because the curve is steeper in the bottom percentiles and flatter in the upper percentiles. Consequently, the area between the Lorenz curve and the diagonal line is smaller for Country E than Country D, resulting in a smaller Gini coefficient. This simple example shows that increasing the average of a distribution while maintaining its shape will result in a smaller Gini coefficient. In other words, the Gini coefficient is not translation invariant, because adding a constant to all the values in the distribution will change the value of the Gini coefficient.

### Gini coefficient and the standard deviation

**Empirical evidence.** Both the Gini coefficient and the standard deviation are measures of dispersion, so naturally, there is a positive relation between these two indicators. This is illustrated in Figure 18 which shows the score distribution of Lebanon and Algeria on the 2015 PISA Reading assessment. The two distributions have similar average scores (347 for Lebanon vs. 350 for Algeria). However, the standard deviation for Lebanon is much higher than that of Algeria (115.5 for Lebanon vs. 72.7 for Algeria), and the Gini coefficient for Lebanon is also higher than that of Algeria (18.34 for Lebanon vs. 11.31 for Algeria).

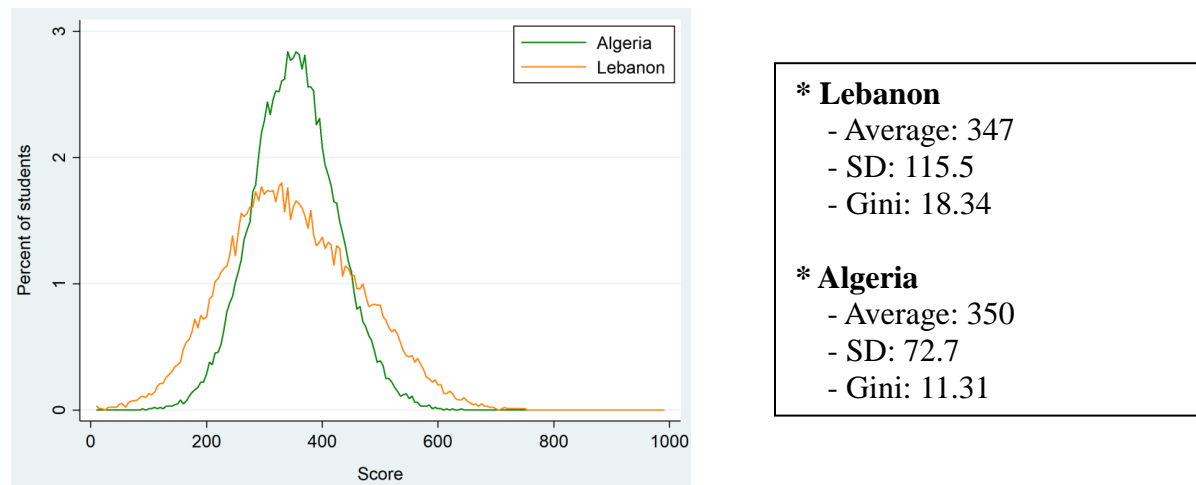


Figure 18. Score distributions of Lebanon and Algeria on the 2015 PISA Reading assessment.

Figure 19 presents a scatter plot of the standard deviations and Gini coefficients estimated with the 2015 PISA Reading scores. The correlation between these two indicators is 0.62, which means that as the standard deviation of the distribution increases, the Gini coefficient also generally increases. While the relation between the Gini coefficient and the standard deviation is positive, the correlation is not perfect, because these two indicators convey different information about the dispersion of a distribution. This will be explained in more detail later.



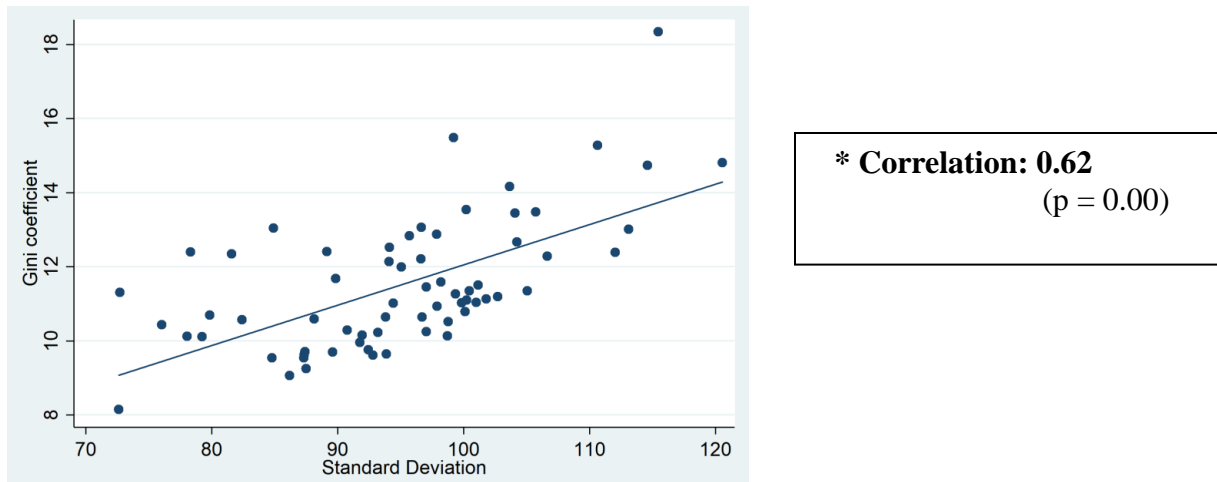


Figure 19. Scatter plot of the standard deviation and the Gini coefficient estimated with 2015 PISA Reading scores.

**Mathematical explanation.** One major difference between the standard deviation and the Gini coefficient is that the standard deviation takes into account the *actual* score at each percentile, while the Gini coefficient takes into account the *cumulative* score at each percentile. Again, this will be explained with the hypothetical score distribution of Country D, a country in which the bottom 20 students scored 300 points, the middle 60 students scored 500 points, and the top 20 students scored 700 points, as presented in Figure 20. The average of the distribution is 500, indicated with a grey horizontal line. In the lower percentiles, the students scored below average. The difference between the actual score and the average scores in this range is marked with an A. In the upper percentiles, the students scored above average. The difference between the actual score and the average score in this range is marked with a C. Both areas A and C affect the standard deviation, because the standard deviation is based on the squared difference between the actual score and the average score of a distribution. In the middle percentiles, students scored exactly at average, so their scores do not affect the standard deviation.

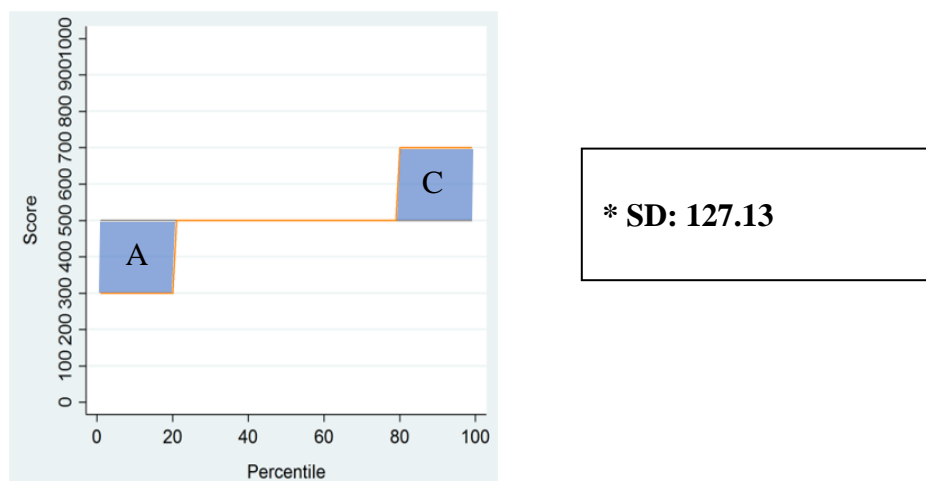
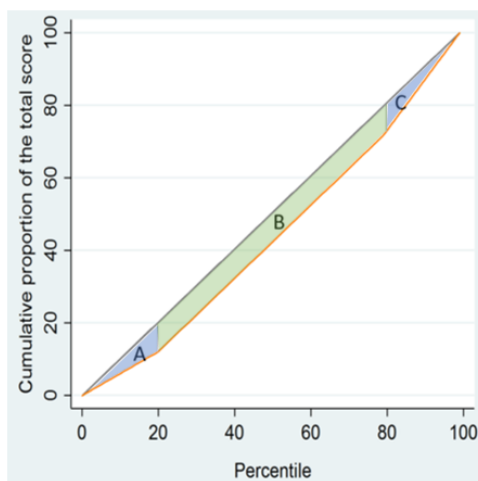


Figure 20. Score distribution of Country D.

The Lorenz curve of the distribution above is presented in Figure 21. In this graph, the scores at all the percentiles affect the Gini coefficient, because the Lorenz curve is always below the grey diagonal line, even at points on the x-axis which correspond to students that scored exactly at average. This is because the Gini coefficient depends on the *cumulative* score, not the *actual* score, at each percentile. Thus, even though students in the middle percentiles scored exactly at average, because students in the bottom percentiles scored below average, the *cumulative* scores in the middle percentiles are still lower than what they would have been for a distribution in which everyone scored at average (represented by the grey diagonal line). Therefore, in this distribution, scores at all the percentiles affect the Gini coefficient, including scores in the bottom percentiles (marked with an A), middle percentiles (marked with a B), and the upper percentiles (marked with a C).



\* Gini coefficient: 12.80

Figure 21. Lorenz curve of Country D. Refer to Figure 20 for the actual score at each percentile.

**Differences in scores in the lower percentiles.** To see how changes in the scores in the lower percentiles affect the standard deviation and the Gini coefficient, the score distribution of Country F is presented in Figure 22. Compared to Country D, the scores in the lower 20 percentiles of Country F decreased by 200 points, but the scores in the other percentiles have not changed. Thus, the students in the lower 20 percentiles scored 100 points, students in the middle 60 percentiles scored 500 points, students in the upper 20 percentiles scored 700 points, and the average is 460 points (represented by the grey horizontal line). In this distribution, the scores at all the percentiles affect the standard deviation, because there is a difference between the actual score and the average score at all the percentiles. The standard deviation for Country F's score distribution is 196.95, which is an increase of 55% compared to the standard deviation for Country D's score distribution (which was 127.13).

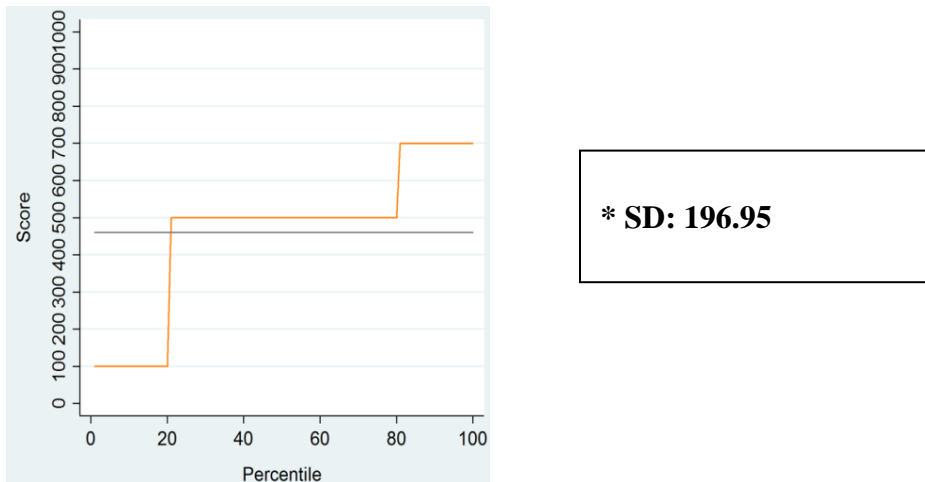


Figure 22. Score distribution of Country F.

The Lorenz curve for Country F is presented in Figure 23. Compared to Country D which had a Gini coefficient of 12.80, the Gini coefficient of Country F increased to 20.87, which is an increase of 63%. This is a greater increase than the increase in the standard deviation (which increased by 55%). This example shows that when there is a decrease of scores in the lower percentiles, the percentage increase in the Gini coefficient is larger than the percentage increase in the standard deviation. Thus, when comparing the score distributions of two countries which have different scores only in the lower percentiles, the relative difference in the Gini coefficients of the two countries is greater than the relative difference in their standard deviations. In other words, when comparing two distributions, the Gini coefficient is more sensitive than the standard deviation to differences in scores in the lower percentiles.

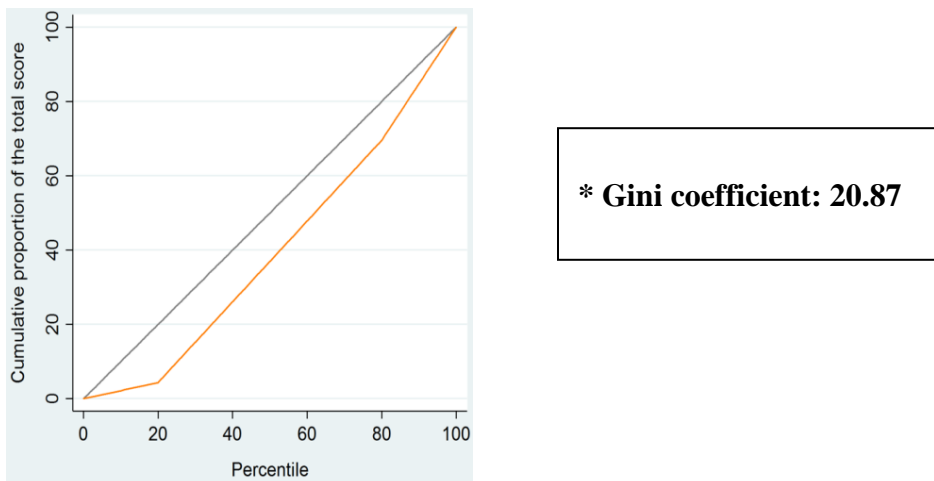


Figure 23. Lorenz curve of Country F. Refer to Figure 22 for the actual score at each percentile.

**Differences in scores in the upper percentiles.** To see how changes in the scores in the upper percentiles affect the standard deviation and the Gini coefficient, the score distribution of Country G is presented in Figure 24. Compared to Country D, the scores in the upper 20

percentiles of Country G increased by 200 points, but the scores in the other percentiles have not changed. Thus, the students in the lower 20 percentiles scored 200 points, students in the middle 60 percentiles scored 500 points, students in the upper 20 percentiles scored 900 points, and the average is 540 points (represented by the grey horizontal line). In this distribution, as in the previous example, the scores at all the percentiles affect the standard deviation, because there is a difference between the actual score and the average score at all the percentiles. The standard deviation for Country G's score distribution is 196.95, which is an increase of 55% compared to the standard deviation for Country D's score distribution (which was 127.13).

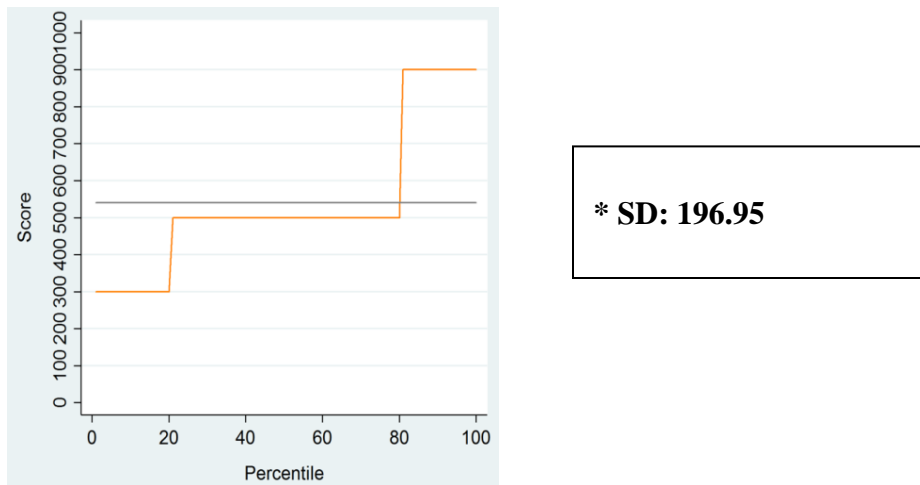


Figure 24. Score distribution of Country G.

The Lorenz curve of this new distribution is presented in Figure 25. Compared to Country D which had a Gini coefficient of 12.80, the Gini coefficient of Country G increased to 17.78, which is an increase of 39%. This is a smaller increase than the increase in the standard deviation (which increased by 55%). This example shows that when there is an increase of scores in the upper percentiles, the percentage increase in the Gini coefficient is smaller than the percentage increase in the standard deviation. Thus, when comparing the score distributions of two countries which have different scores only in the upper percentiles, the relative difference in the Gini coefficients of the two countries is smaller than the relative difference in their standard deviations. In other words, when comparing two distributions, the Gini coefficient is less sensitive than the standard deviation to differences in scores in the upper percentiles.

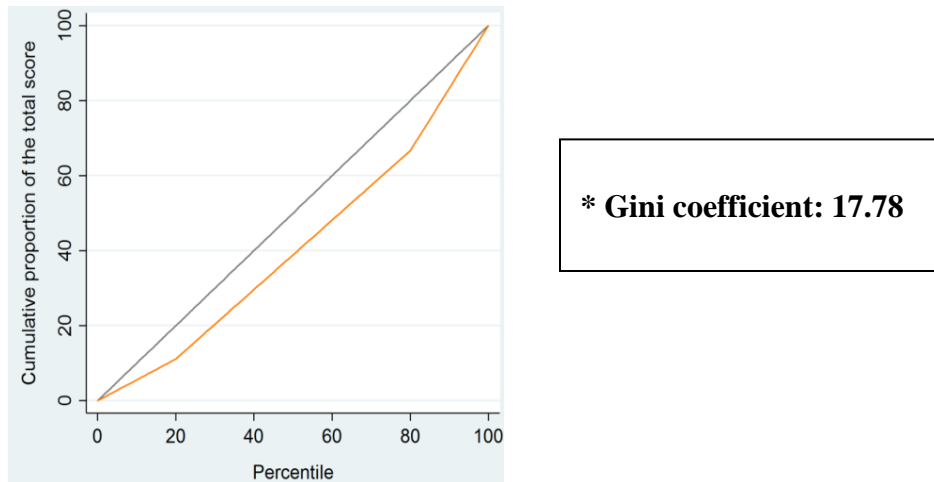


Figure 25. Lorenz curve of Country G. Refer to Figure 24 for the actual score at each percentile.

**Summary of the differences between the Gini coefficient and the standard deviation.**

The examples above show how the standard deviation and the Gini coefficient are affected by changes in the scores in the lower and upper percentiles. The results are summarized in Table 1. Thus, when comparing the score distributions of two countries that have different scores only in the lower percentiles, the relative difference in the Gini coefficients of the two countries is greater than the relative difference in their standard deviations. On the contrary, when comparing the score distributions of two countries that have different scores only in the upper percentiles, the relative difference in the Gini coefficients of the two countries is smaller than the relative difference in their standard deviations.

Table 1

*How changes in the distribution affect the standard deviation and the Gini coefficient*

	<b>Standard deviation</b> (% change in standard deviation compared to Country D)	<b>Gini coefficient</b> (% change in Gini coefficient compared to Country D)
<b>Country D</b>	127.13	12.8
<b>Country F</b> <b>(Lower scores in</b> <b>lower percentiles)</b>	196.95 (55%)	20.87 (63%)
<b>Country G</b> <b>(Higher scores in</b> <b>upper percentiles)</b>	196.95 (55%)	17.78 (39%)

Another takeaway from this table is that the standard deviation is equally sensitive to differences in the lower percentiles as it is to differences in the upper percentiles. In the example above, the standard deviation for Country F (which has lower scores in the lower percentiles) is 55% greater than the standard deviation for Country D, while the standard deviation for Country

G (which has higher scores in the upper percentiles) is also 55% greater than the standard deviation for Country D. In other words, a decrease in the scores in the lower percentiles has the same effect on the standard deviation as an increase in the scores in the upper percentiles.

On the contrary, the Gini coefficient is more sensitive to differences in the lower percentiles than it is to differences in the upper percentiles. In the example above, the Gini coefficient for Country F (which has lower scores in the lower percentiles) is 63% greater than the Gini coefficient for Country D, while the Gini coefficient for Country G (which has higher scores in the upper percentiles) is only 39% greater than the Gini coefficient for Country D. In other words, a decrease in the scores in the lower percentiles has a greater effect on the Gini coefficient than an increase in the scores in the upper percentiles. This is because the Gini coefficient depends on the *cumulative* score at each percentile. When the scores decrease in the lower percentiles, it affects the Lorenz curve in all the subsequent percentiles, resulting in a large change in the Gini coefficient. On the contrary, when the scores increase in the upper percentiles, it affects the Lorenz curve mostly in the upper percentiles,<sup>10</sup> resulting in a small change in the Gini coefficient.

### Gini coefficient and the scale of a standardized test

The fact that the Gini coefficient is negatively related with the average score and positively related with the standard deviation means that the Gini coefficient will change when a different scale (i.e., the mean and standard deviation) is selected for the standardized test scores. As explained above, for PISA, the scores of the students in the 27 OECD countries that took the first round PISA in 2000 were standardized to be on a scale with a mean of 500 and a standard deviation of 100 (OECD, 2003). The scores from the non-OECD countries that participated in PISA in 2000 were equated to this scale, as were the scores from all subsequent rounds of PISA. If the developers of PISA had chosen a different scale, for example, a scale with a mean of 600 and a standard deviation of 200, all the standardized scores would be different. Using the following equation, the standardized scores on this new scale can be calculated by doing a linear transformation of the PISA scores:

$$\text{Score on new scale} = 600 + \left[ 200 * \left( \frac{\text{Score on PISA scale} - 500}{100} \right) \right] \quad (2)$$

When Gini coefficients are estimated using scores on this new scale, the increase in the average score (from 500 to 600) will work towards decreasing the Gini coefficient, while the increase in the standard deviation (from 100 to 200) will work towards increasing it. To illustrate this, Figure 26 shows the Lorenz curve of the United States' scores on the 2015 PISA Reading assessment using the original PISA scale (with a mean of 500 and standard deviation of 100) and the new scale (with a mean of 600 and standard deviation of 200). Simply choosing a different scale for the scores increased the Gini coefficient from 11.03 to 18.46, even though the distribution of students' actual abilities did not change.

---

<sup>10</sup> An increase in the scores in the upper percentiles will also affect the Lorenz curve in the lower percentiles to some degree, because the average of the distribution will change. As a consequence, the ratio between each score and the average score will change, resulting in a change in the slope of the tangent of the Lorenz curve at all the percentiles.

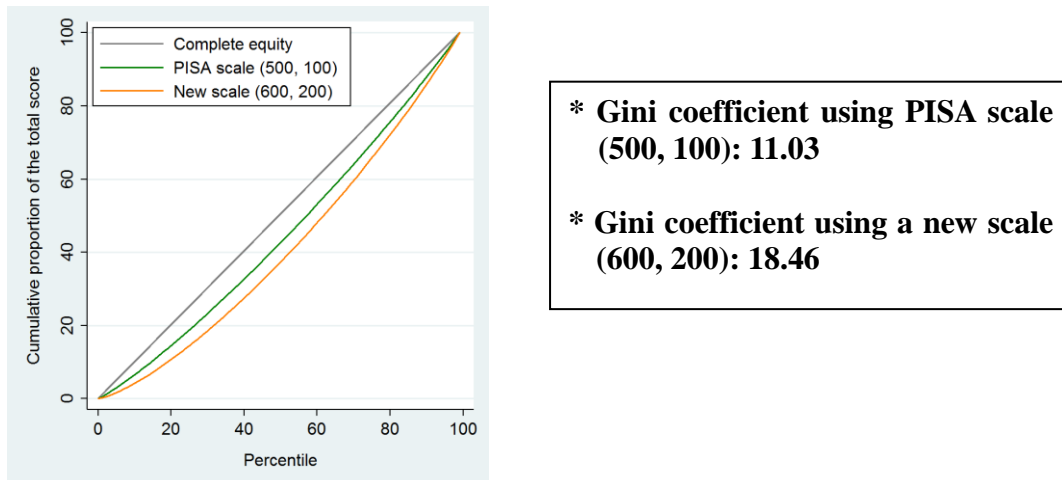


Figure 26. Lorenz curve of the United States' scores on the 2015 PISA Reading assessment using the PISA scale and a new scale.

To further complicate the issue, doing a linear transformation of scores will affect the Gini coefficient of different countries to different degrees. This is because the degree of change in the Gini coefficient will depend on the average and standard deviation of each country's underlying score distribution relative to the average and standard deviation of the old and new scales. To illustrate this, Figure 27 shows the scatter plot of the Gini coefficients estimated with the 2015 PISA Reading scores on the original scale (with a mean of 500 and standard deviation of 100) and scores that have been linearly transformed to a new scale (with a mean of 600 and standard deviation of 200).<sup>11</sup> The correlation is not perfect at 0.9012. Also, it should be noted that there are many cases in which one country has a higher Gini coefficient than another when using the PISA scale, but the order is reversed when using the new scale. The two countries circled in Figure 27 is one such example. This gives empirical support to the mathematical equations of Ferreira and Gignoux (2014) which show that when Gini coefficients are estimated with scores on two different scales, one a linear transformation of the other, the values of the Gini coefficients as well as the country rankings of the Gini coefficients will change. This is a major flaw of the Gini coefficient when using it to make comparisons across countries.

<sup>11</sup> Countries were excluded if it had negative scores after the linear transformation, because the Gini coefficient cannot be estimated with negative scores. Thus, only 49 countries were included in Figure 27.

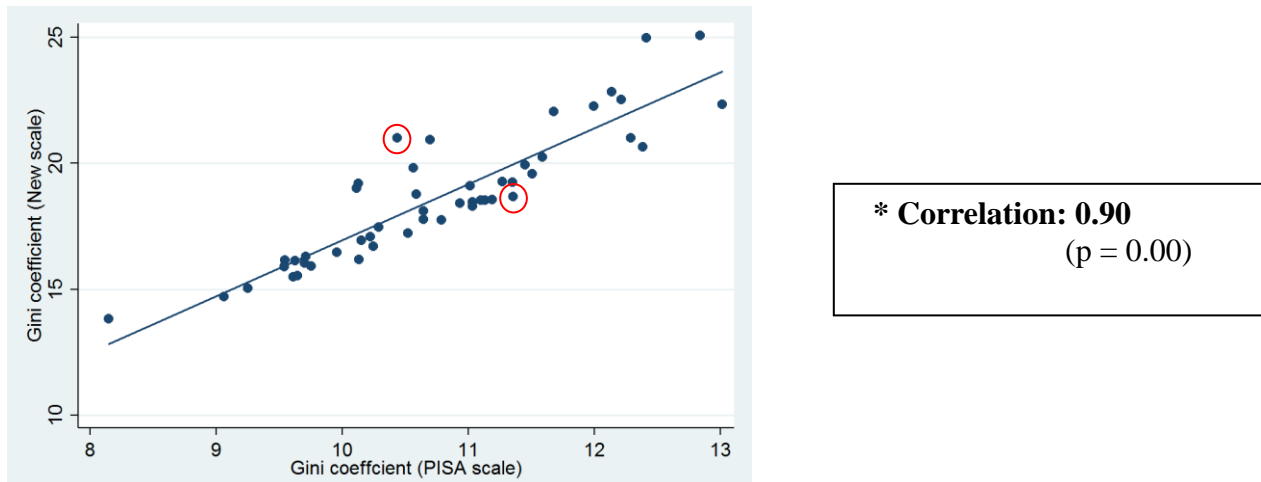


Figure 27. Gini coefficients estimated with scores on the 2015 PISA Reading assessment using the PISA scale and a new scale.

### Gini coefficient and other measures of dispersion

**Absolute and relative measures of dispersion.** Measures of dispersion can be expressed in either absolute or relative terms. Absolute measures of dispersion are based on the *difference* between values and maintain the original unit of observation. On the other hand, relative measures of dispersion are based on the *ratio* of values. Since the unit of observation is included in both sides of the ratio, they are both cancelled out, and the ratio does not maintain the original unit of observation. Thus, relative measures of dispersion are dimensionless.

To illustrate the difference between absolute and relative measures of dispersion, another set of hypothetical countries, Country H and Country I, are presented below. The income distributions of these countries are presented in Figure 28. In both countries, income is normally distributed, and the shape of the distributions are identical. The only difference in the distributions is that everyone in Country I has 799 more dollars than everyone in Country H. In Country H, the poorest person's income is 1 dollar, and the richest person's income is 201 dollars, while in Country I, the poorest person's income is 800 dollars, and the richest person's income is 1,000 dollars.

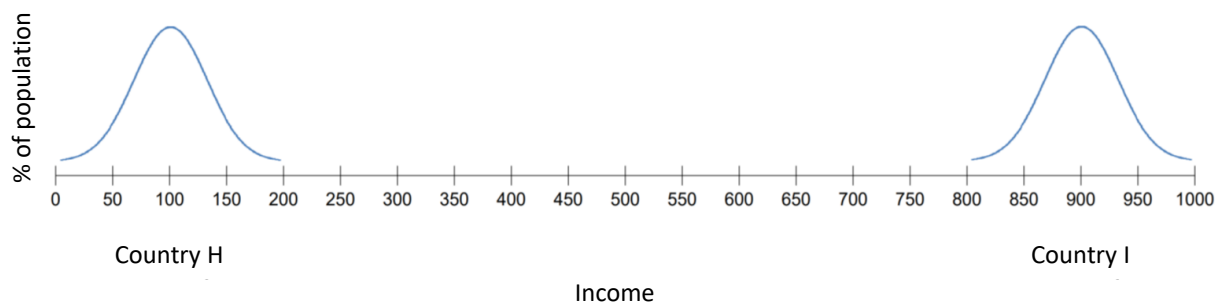


Figure 28. Income distributions of Country H and Country I.



In both countries, the absolute difference between the richest person and the poorest person is 200 dollars (because  $201 - 1 = 200$  and  $1,000 - 800 = 200$ ). This type of information is captured by absolute measures of dispersion which are based on the *difference* between values.

In terms of relative differences, in Country H, the ratio of the poorest person's income to the richest person's income is 1:201 (because  $201/1 = 201$ ), while in Country I, the ratio is 1:1.25 (because  $1,000/800 = 1.25$ ). This type of information is captured by relative measures of dispersion which are based on the *ratio* of values. Thus, while the two countries have the same amount of absolute dispersion in their income distribution, they have different amounts of relative dispersion in their income distribution.

This simple example shows that absolute measures of dispersion and relative measures of dispersion convey different information about the dispersion of a distribution. It should be noted that while absolute measures of dispersion are meaningful for distributions on either an interval or ratio scale, relative measures of dispersion are only meaningful for distributions on an interval scale. This is explained in detail later.

**Absolute measures of dispersion.** As explained above, absolute measures of dispersion are based on the *difference* between values, and they maintain the original unit of observation. These measures are meaningful for distributions on either an interval or ratio scale.

**Standard deviation.** The standard deviation is an absolute measure of dispersion, because it is based on the squared *difference* between each score and the average score. Figure 19 showed that the correlation between the standard deviation and the Gini coefficient was 0.62 for the 2015 PISA Reading scores. As explained above, the correlation is not perfect, because these two indicators convey different information about the dispersion of a distribution. The standard deviation is an absolute measure of dispersion, while the Gini coefficient is a relative measure of dispersion.

**90-10 achievement gap.** Another absolute measure of dispersion is the 90-10 achievement gap which is the *difference* between the score at the 90<sup>th</sup> percentile and the score at the 10<sup>th</sup> percentile (Miller & Fonseca, 2017). It can be calculated by using the following equation:

$$90-10 \text{ achievement gap} = \text{Score at } 90^{\text{th}} \text{ percentile} - \text{Score at } 10^{\text{th}} \text{ percentile} \quad (3)$$

Figure 29 presents the scatter plot of the Gini coefficient and the 90-10 achievement gap estimated with the 2015 PISA Reading scores. The correlation between these two indicators is 0.63. Again, the correlation is not perfect, because the 90-10 achievement gap is an absolute measure of dispersion, while the Gini coefficient is a relative measure of dispersion, which means that these two indicators convey different information about the dispersion of a distribution.

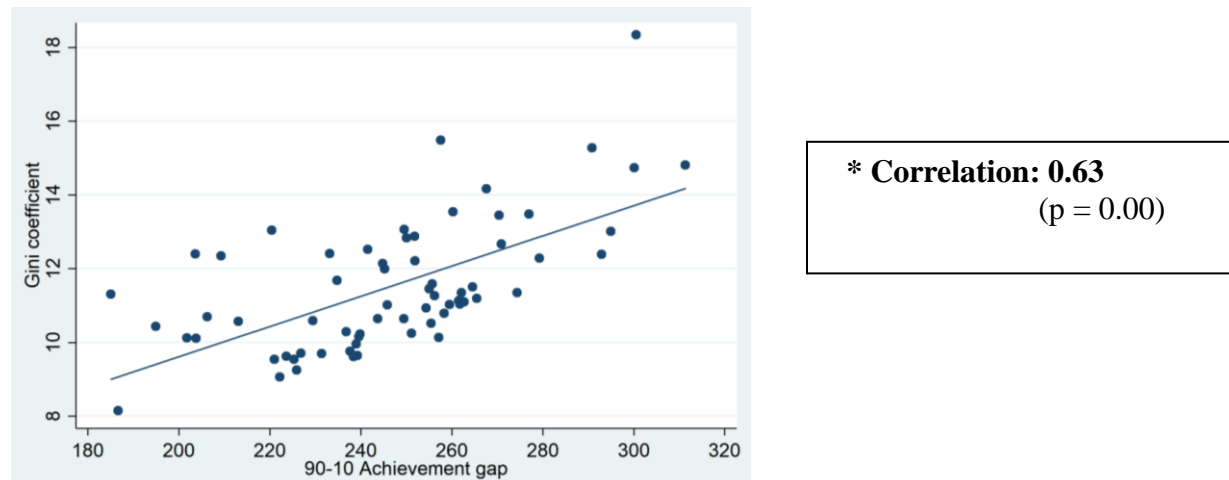


Figure 29. Scatter plot of the 90-10 achievement gap and the Gini coefficient estimated with the 2015 PISA Reading scores.

**Relative measures of dispersion.** As explained above, relative measures of dispersion are based on the *ratio* of two values. These measures are only meaningful for distributions on a ratio scale. This is because the ratio between two values depends on how far each value is from the zero of the scale. Thus, when the zero of the scale is meaningful (e.g., ratio scale), the ratio between two values is meaningful. However, when the zero of the scale is arbitrary (e.g., interval scale), the ratio between two values on this scale will depend on the choice of the scale (i.e., the arbitrary location of zero), and therefore, the ratio itself will be arbitrary. This can be illustrated with two different scales used for measuring temperature, Celsius and Fahrenheit. Both are interval scales, because the zero on each scale is arbitrary.<sup>12</sup> On the Celsius scale, the values of 50°C and 100°C have a ratio of 1:2. However, on the Fahrenheit scale, these values are converted to 122°F and 212°F, respectively, and the ratio drops to 1:1.74. Thus, even though the actual temperatures did not change, using different scales to measure temperature resulted in different ratios, because the distance between each value and the zero of the scale had changed. This example shows that when the zero of a scale is not meaningful, the ratio between two values on the scale (on which all relative measures of dispersion are based) is not meaningful. As explained above, standardized test scores do not have a meaningful zero, so it is an interval scale. Therefore, all relative measures of dispersion are not meaningful for standardized test scores.

Several relative measures of dispersion are presented below for illustrative purposes.

**Gini coefficient.** The Gini coefficient is a relative measure of dispersion. This can be illustrated with the example of Figure 28 which presented the income distributions of Country H (where the income ranged from 1 to 201 dollars) and Country I (where the income ranged from 800 to 1,000 dollars). In these hypothetical countries, it can be intuited that income is more equally distributed in Country I than in Country H, because a difference of 200 dollars is not so large in a country where the average income is 900 dollars (Country I) as it is in a country where the average income is 101 dollars (Country H). In these comparisons, implicitly, the *relative size*

<sup>12</sup> Even the zero on the Celsius scale is arbitrary rather than absolute, because it is the point at which water (an arbitrary substance) freezes, not the point at which all thermal motion ceases. The Kelvin scale has an absolute zero, because this is the point at which all thermal motion ceases, making the Kelvin scale a ratio scale.

of a rich person's income is being compared to a poor person's income.

The Gini coefficient takes this a step further by comparing the *relative* size of "the income of each person in the population" to "the average income." This is because the Gini coefficient is determined by the Lorenz curve, the tangent of which is determined by the ratio between "the income of the subject at a given percentile" and "the average income," as shown in Equation 1. Therefore, the Gini coefficient is a relative measure of dispersion that takes into account the income of every person in the population.

If the Gini coefficients were to be estimated for Country H and Country I, the Gini coefficient for Country I would be much smaller than for Country H, because income is more equally distributed, in relative terms, in Country I than in Country H. This example again shows that countries with a higher average income will always have a lower Gini coefficient if the shape of the income distribution is maintained.

It should be noted that the Gini coefficient in the example above is meaningful, because it is estimated with income, a ratio scale. However, it is not meaningful for standardized test scores, because it is an interval scale.<sup>13</sup>

**90/10 achievement gap.** Another relative measure of dispersion is the 90/10 achievement gap which is calculated by *dividing* the score at the 90<sup>th</sup> percentile by the score at the 10<sup>th</sup> percentile, as expressed in the following equation:

$$90/10 \text{ achievement gap} = \text{Score at } 90^{\text{th}} \text{ percentile} / \text{Score at } 10^{\text{th}} \text{ percentile} \quad (4)$$

Figure 30 presents the scatter plot of the Gini coefficient and the 90/10 achievement gap estimated with the 2015 PISA Reading scores. The correlation between these two indicators is almost perfect at 0.9959. The correlation is very high, because the Gini coefficient and the 90/10 achievement gap are both relative measures of dispersion, conveying similar information about the dispersion of a distribution.

However, it should be noted that the 2015 PISA Reading scores have an approximately normal distribution within each country. For distributions that are not normal, the correlation between these two indicators may not be so high. Also, it is worth stressing again that relative measures of dispersion, such as the 90/10 achievement gap, are meaningless for standardized test scores.

---

<sup>13</sup> If there were a way to measure students' abilities on a ratio scale rather than an interval scale, the Gini coefficient may be a meaningful indicator to measure the relative dispersion in students' abilities. Some may argue that the percentage of items correct on a standardized test is a ratio scale. However, a problem with the "percentage correct scale" is that it ignores the difficulty level of each item, giving the same weight to all items regardless of the item's difficulty. This will incorrectly estimate a student's ability, the construct that the standardized test is trying to measure. Another related problem is that the distance between any two consecutive points on the "percentage correct scale" will not represent the same amount of difference in students' abilities, since it does not take into account the item's difficulty. Thus, while "the percentage correct scale" has a meaningful zero, the distance between any two consecutive points on the scale are not equal, so it does not meet the criteria for an interval scale, let alone a ratio scale. Therefore, Gini coefficients are meaningless when estimated with scores on "the percentage correct scale."

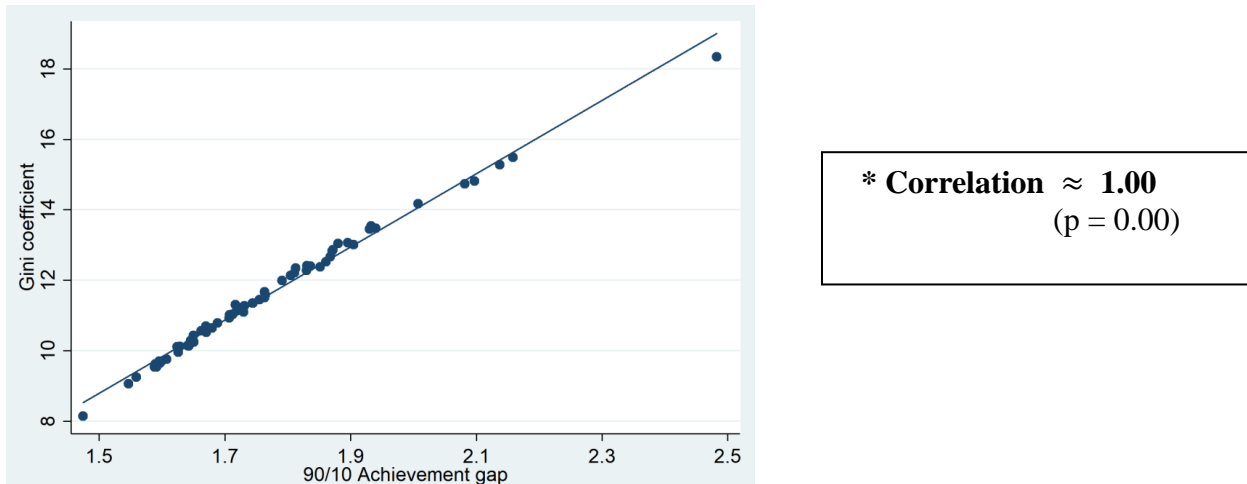


Figure 30. Scatter plot of the 90/10 Achievement gap and the Gini coefficient estimated with the 2015 PISA Reading scores.

**Coefficient of variation.**<sup>14, 15</sup> The Coefficient of variation is another relative measure of dispersion which is calculated by *dividing* the standard deviation by the average score, as expressed in the following equation:

$$\text{Coefficient of Variation} = \text{Standard deviation} / \text{Average score} \quad (5)$$

Figure 31 presents the scatter plot of the Gini coefficient and the Coefficient of variation estimated with the 2015 PISA Reading scores. The correlation between these two indicators is almost perfect at 0.9995. Since both the Gini coefficient and the Coefficient of variation are relative measures of dispersion, they convey similar information about the dispersion of scores, resulting in a very high correlation.

Again, it should be noted that the 2015 PISA Reading scores have an approximately normal distribution within each country. For distributions that are not normal, the correlation between these two indicators may not be so high. Also, the Coefficient of variation is meaningless for standardized test scores, because it is a relative measure of dispersion.

<sup>14</sup> I would like to thank Dr. Robert Boruch for suggesting that I examine the relation between the Gini coefficient and the Coefficient of variation.

<sup>15</sup> Refer to Snedecor and Cochran (1967, pp. 62 - 65) for more information on the Coefficient of variation.

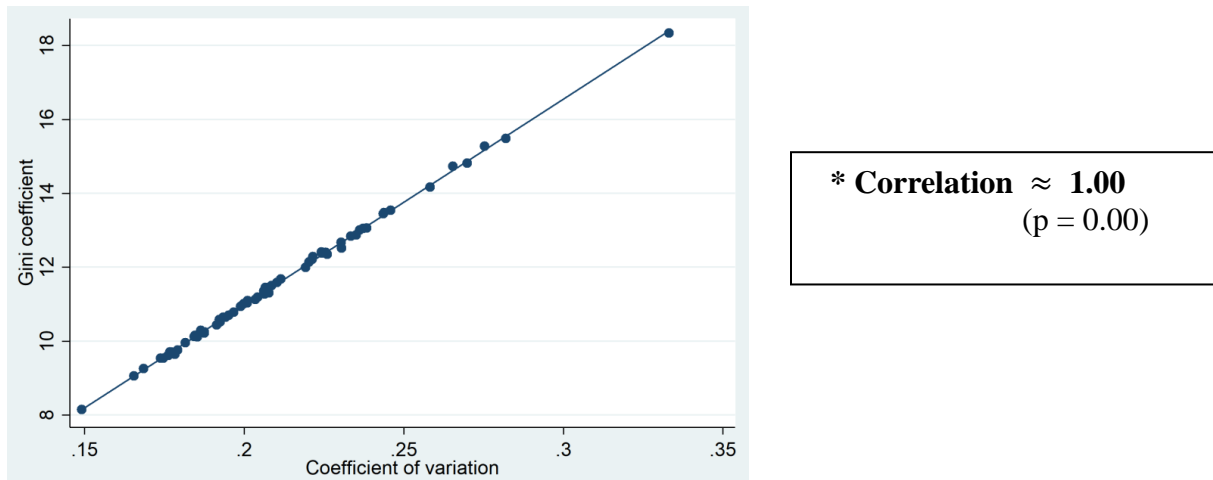


Figure 31. Scatter plot of the Coefficient of variation and the Gini coefficient estimated with the 2015 PISA Reading scores.

**Summary of the absolute and relative measures of dispersion.** Table 2 summarizes the relation between the absolute and relative measures of dispersion presented above. The correlation between any two indicators of absolute dispersion or between any two indicators of relative dispersion is almost perfect when estimated with 2015 PISA Reading scores.<sup>16</sup> However, the correlation between an absolute measure of dispersion and a relative measure of dispersion is approximately 0.60. This is because absolute measures of dispersion and relative measures of dispersion convey different information about the dispersion of a distribution.

While the indicators above were presented for illustrative purposes, it should be noted that relative measures of dispersion, such as the Gini coefficient, are meaningless for standardized test scores, since it is an interval scale.

Table 2

*Correlations between absolute and relative measures of dispersion estimated with 2015 PISA Reading scores*

	Absolute measures of dispersion	Relative measures of dispersion	
$r \approx 1.00$	Standard deviation	Gini coefficient	$r \approx 1.00$
	90-10 achievement gap	90/10 achievement gap	
		Coefficient of variation	
	$r \approx .60$		

<sup>16</sup> For non-normal distributions, the correlation between these indicators may not be so high.

### **Implications for standardized test scores**

#### **Value of the Gini coefficient is meaningless**

The value of the Gini coefficient is meaningless when estimated with standardized test scores, because the value itself will depend on the mean and standard deviation of the scale that was arbitrarily selected by the test developer. Keeping the standard deviation of the scale constant, increasing the mean will decrease the Gini coefficient, while keeping the mean of the scale constant, increasing the standard deviation will increase the Gini coefficient. The example in Figure 26 showed that changing the scale for the PISA scores increased the Gini coefficient for the United States, even though the distribution of students' actual abilities did not change.

More generally, all relative measures of dispersion (such as the Gini coefficient) are meaningless for distributions on an interval scale (such as standardized test scores), because the value will depend on the arbitrary location of zero on the scale.

#### **Comparing the Gini coefficients from the same test is meaningless**

Some may argue that even though the absolute value of a Gini coefficient for standardized test scores is meaningless, it can still be useful to compare the relative size of the Gini coefficients estimated from the *same* test. However, even this is meaningless, because if a different scale had been selected for the test, not only the values of the Gini coefficients, but also the country rankings of the Gini coefficients can change, as shown in Figure 27. This is because when Gini coefficients are estimated with scores on two different scales, one a linear transformation of the other, the rank order of the Gini coefficients is not maintained. Thus, selecting different scales can lead to different conclusions about which countries have more dispersion in their score distribution, a major flaw when using the Gini coefficient to make comparisons across countries.

#### **Comparing the Gini coefficients from different tests is meaningless**

Lastly, comparing the relative size of Gini coefficients estimated with scores from different tests is meaningless, because the value of the Gini coefficient depends on the choice of the scale. Thus, Gini coefficients estimated with scores on different scales are not directly comparable. Doing a linear transformation of the scores of one test to put them on the scale of the other will not solve this problem, because as explained above, a linear transformation of scores will change the values of the Gini coefficients as well as the country rankings of the Gini coefficients.

Even Gini coefficients estimated with scores from different tests that use the *same* scale (i.e., same mean and standard deviation) are also not directly comparable. This can be illustrated with the example of PISA and TIMSS which both use a scale with a mean of 500 and a standard deviation of 100. Even though the scale is the same, a score of 500 on PISA may not indicate the same level of ability as a score of 500 on TIMSS,<sup>17</sup> because each score refers to the average score of the countries (or a subset of the countries) that took the test when it was first

---

<sup>17</sup> Another problem of comparing scores from PISA and TIMSS is that they are not measuring the same construct.

administered. Even if we were to do a linear transformation of the scores of one test so that the same numeric score corresponds to the same level of ability, the problem will not be solved. This is because, as explained above, when Gini coefficients are estimated with scores on two different scales, one a linear transformation of the other, the rank order of the Gini coefficients are not maintained, making any comparisons of Gini coefficients meaningless.

### **Conclusion**

This paper examined whether the Gini coefficient, an indicator often used to measure the inequality in the distribution of income within countries, can also be used to measure the inequality in learning achievement within countries. A major attraction of the Gini coefficient is that it is dimensionless and has a range of 0 to 100, regardless of the scale of the test. This gave many scholars the false impression that it could be used with any kind of scale and that it could also be used to compare the dispersion of scores from different tests.

However, a careful examination of the properties of the Gini coefficient revealed that it is problematic to estimate the Gini coefficient with standardized test scores. More generally, this research showed that all relative measures of dispersion (including the Gini coefficient) are meaningless for distributions on an interval scale (such as standardized test scores). This is because relative measures of dispersion are based on the *ratio* of two values, and this ratio depends on how far each value is from the zero of the scale. Thus, if the zero on the scale has no inherent meaning (which is the case for interval scales, such as standardized test scores), the ratio between two values on the scale (on which all relative measures of dispersion, including the Gini coefficient, are based) is meaningless. Therefore, all things considered, it can be said that the Gini coefficient's magic does not work on standardized test scores.

### References

- Allison, P. D. (1978). Measures of inequality. *American Sociological Review*, 43(6), 865-880.
- Bedard, K., & Ferral, C. (2003). Wage and test score dispersion: Some international evidence. *Economics of Education Review*, 22(1), 31-43. doi: 10.1016/S0272-7757(01)00060-7
- Ceriani, L., & Verme, P. (2012). The origins of the Gini index: Extracts from *Variabilità e Mutabilità* (1912) by Corrado Gini. *The Journal of Economic Inequality*, 10(3), 421-443. doi: 10.1007/s10888-011-9188-x
- Derobert, L., & Thieriot, G. (2003). The Lorenz curve as an archetype: A historico-epistemological study. *The European Journal of the History of Economic Thought*, 10(4), 573-585. doi: 10.1080/0967256032000137720
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ferreira, F. H. G., & Gignoux, J. (2014). The measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review*, 28(2), 210-246. doi: 10.1093/wber/lht004
- Freeman, R. B., Machin, S. J., & Viarengo, M. G. (2011). Inequality of educational outcomes: International evidence from PISA. *Regional and Sectoral Economic Studies*, 11(3), 5-20.
- Miller, D. C. & Fonseca, F. T. (2017, April). *Mathematics achievement gaps of low- and high-performing fourth-graders: A comparison cross-nationally and over time using TIMSS data*. Roundtable paper session presented at the annual conference of the American Educational Research Association (AERA), San Antonio, TX.
- OECD. (n.d.). *About PISA*. Retrieved from: <http://www.oecd.org/pisa/aboutpisa/>
- OECD. (2002). *PISA 2000 technical report*. Retrieved from: <http://www.oecd.org/education/school/programmeforminternationalstudentassessmentpisa/33688233.pdf>
- OECD. (2003). *Literacy skills for the world of tomorrow: Further results from PISA 2000*. Paris, France: OECD. doi: 10.1787/9789264102873-en
- OECD. (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. Paris, France: OECD. doi: 10.1787/9789264266490-en
- Sastry, N., & Pebley, A. R. (2010). Family and neighborhood sources of socioeconomic inequality in children's achievement. *Demography*, 47(3), 777-800. doi: 10.1353/dem.0.0114
- Snedecor, G. W., & Cochran, W. G. (1967). *Statistical methods* (6th ed.). Ames, IA: Iowa State University Press.



- Technical Advisory Group. (2015). *Technical Advisory Group proposal: Thematic indicators to monitor the post-2015 education agenda*. Retrieved from: <http://uis.unesco.org/sites/default/files/documents/technical-advisory-group-proposal-thematic-indicators-to-monitor-the-post-2015-education-agenda-2015-en.pdf>
- Thomas, V., Wang, Y., & Fan, X. (2001). *Measuring education inequality: Gini coefficients of education* (Policy Research Working Paper No. 2525). Retrieved from the World Bank website: <https://openknowledge.worldbank.org/handle/10986/19738>
- United Nations. (n.d.). *Millennium Development Goals and beyond 2015: Goal 2*. Retrieved from: <http://www.un.org/millenniumgoals/education.shtml>
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development* [General Assembly resolution]. Retrieved from the UN website: [http://www.un.org/ga/search/view\\_doc.asp?symbol=A/RES/70/1&Lang=E](http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E)
- Vegas, E. (Ed.). (2005). *Incentives to improve teaching: Lessons from Latin America*. Washington, DC: World Bank.
- Wagner, D. A. (2018). *Learning as development: Rethinking international education in a changing world*. New York, NY: Routledge.
- World Bank. (n.d.) *Gini Index*. Retrieved from the World Bank website: <https://data.worldbank.org/indicator/si.pov.gini?end=2016&start=2016>
- World Education Forum. (2015). *Education 2030: Incheon Declaration and Framework for Action for the implementation of Sustainable Development Goal 4*. Retrieved from: [http://uis.unesco.org/sites/default/files/documents/education-2030-incheon-framework-for-action-implementation-of-sdg4-2016-en\\_2.pdf](http://uis.unesco.org/sites/default/files/documents/education-2030-incheon-framework-for-action-implementation-of-sdg4-2016-en_2.pdf)

**Appendix: Countries included in the data set****OECD (35)**

Australia  
Austria  
Belgium  
Canada  
Chile  
Czech Republic  
Denmark  
Estonia  
Finland  
France  
Germany  
Greece  
Hungary  
Iceland  
Ireland  
Israel  
Italy  
Japan  
Korea (South)  
Latvia  
Luxembourg  
Mexico  
Netherlands  
New Zealand  
Norway  
Poland  
Portugal  
Slovak Republic  
Slovenia  
Spain  
Sweden  
Switzerland  
Turkey  
United Kingdom  
United States

**Non-OECD (31)**

Albania  
Algeria  
Brazil  
Bulgaria  
Chinese Taipei (Taiwan)  
Colombia  
Costa Rica  
Croatia  
Dominican Republic  
Georgia  
Indonesia  
Jordan  
Kosovo  
Lebanon  
Lithuania  
Macedonia  
Malta,  
Moldova  
Montenegro  
Peru  
Puerto Rico  
Qatar  
Romania  
Russian Federation  
Singapore  
Thailand  
Trinidad and Tobago  
Tunisia  
United Arab Emirates  
Uruguay  
Vietnam