国際大学
INTERNATIONAL UNIVERSITY OF JAPAN

Economics & Management Series EMS-2024-05

Urban and Rural Dimensions of Educational Inequality

Takahiro Akita
International University of Japan

July 2024

IUJ Research Institute
International University of Japan

Urban and Rural Dimensions of Educational Inequality

Takahiro Akita*
Professor Emeritus, International University of Japan, akita@iuj.ac.jp

Abstract

This study introduces a method for the decomposition of the education Gini coefficient by location and examines the characteristics of this Gini decomposition method through the use of hypothetical examples. It empirically analyzes the determinants of educational inequality in some Asian countries using the Gini decomposition method. In a dual economy consisting of the rural and urban sectors, the education Gini coefficient can be additively decomposed into three distinct components: the within-sector, between-sector, and residual Gini components. The within-sector component measures educational inequality within the rural and urban sectors, while the between-sector component measures the rural-urban disparity in the mean level of educational attainment. The residual component assesses the extent of overlap between the rural and urban sectors in the distribution of educational attainment. In all selected Asian countries, including four ASEAN countries and three SAARC countries, the within-sector Gini component is a predominant determinant by accounting for 40-50% of the overall education Gini coefficient. There is a significant negative relationship between the between-sector component and the residual component in terms of their contributions to overall education Gini coefficient, indicating that the rural-urban overlap in the distribution of educational attainment rises as the rural-urban education disparity declines. When Bangladesh is excluded as an outlier, a significant positive relationship exists between mean years of education and the contribution of the residual component to overall education Gini coefficient, signifying that the rural-urban overlap in the distribution of educational attainment increases as mean years of education rises.

Key words: urban and rural dimensions, educational inequality, education Gini coefficient, decomposition of the Gini coefficient, ASEAN countries, SAARC countries

* The author is grateful to the Japan Society for the Promotion of Science for its financial support (Grant-in-Aid for Scientific Research: 23K01409). The author would like to thank Raquel Celeste, Mohammad Haroon Ebrahimi, Dorji Lethro, Shaikh Shahriar Mahmud, Aye Thu Zar Thein, and Chufeng Vachoima for providing household survey data from the Philippines, Afghanistan, Bhutan, Bangladesh, Myanmar and Laos, respectively.

1. Introduction

A significant disparity exits in the structure of education between rural and urban areas. In rural areas, jobs are predominantly associated with agricultural activities. In the early stages of economic development, these jobs may not require formal education. However, as the economy advances, there is an increasing demand for basic education, including primary and secondary education. Sustained economic development may give rise to a demand for higher education, even in rural areas, as agricultural activities would need to enhance labor productivity under competition.

In contrast, urban areas offer diverse employment opportunities in manufacturing and service sectors. While many urban jobs may not necessarily require higher education in developing countries, jobs in foreign firms and domestic firms linked to global markets often necessitate advanced knowledge and skills that can only be acquired through higher education institutions. Meanwhile, in developing countries, informal economic activities are prevalent in urban areas. Typically not requiring formal education, these activities often provide employment opportunities for migrants from rural areas.

The structure of education can be described by the mean and inequality in the distribution of educational attainment, where the educational attainment of an individual is usually measured by the cumulative number of years of education the individual has acquired. As a measure of inequality in the distribution of educational attainment (educational inequality), the Gini coefficient is frequently used.[1] Additionally, the coefficient of variation, defined by the ratio of the standard deviation to the mean, can be employed as an alternative measure of educational inequality. However, the Theil indices cannot be used when there are individuals without formal education who are assigned 0 years of education.[2]

In a dual economy encompassing the rural and urban sectors, overall educational inequality, as measured by the Gini coefficient, can be expressed as the sum of the within-sector and between-sector educational inequality components if there is no overlap in the distribution

[1] Studies which employed the Gini coefficient as a measure of educational inequality include Thomas, et al. (2001), Castello and Domenech (2002), Lin (2007), Lim and Tang (2008), Hojo (2009), Fordvari and van Leeuwen (2011), Agrawal (2014), Meschi and Scervini (2014), Coady and Dizioli (2018), Banzragch, et al. (2019), Shukla and Mishra (2019), Castello-Climent and Domenech (2021), Almeida, et al. (2022), and Luo, et al. (2022).
[2] Please see Akita and Kataoka (2022) for the definition of the Theil indices.

of educational attainment between the rural and urban sectors. However, distributional overlaps usually exist between them, requiring an additional inequality component (referred to as the residual component).[3] The within-sector component is a weighted sum of rural and urban educational inequalities and thus measures educational inequality within the rural and urban sectors, while the between-sector component measures the rural-urban disparity in the mean level of educational attainment. The residual component assesses the extent of overlap between the rural and urban sectors in the distribution of educational attainment.

The objectives of this study are twofold: first, to develop a method for the decomposition of educational inequality by location (rural and urban sectors) using the Gini coefficient (referred to as the decomposition of the education Gini coefficient by location); and second to empirically analyze the determinants of educational inequality in some Asian countries by decomposing the education Gini coefficient by location.

This paper is organized as follows. Section 2 introduces a method for the decomposition of the education Gini coefficient by location. It also explores the characteristics of this decomposition method through the use of hypothetical examples. Section 3 uses nation-wide household surveys to analyze the determinants of educational inequality in some Asian countries using the decomposition method. The final section provides concluding remarks.

2. Decomposition of the Education Gini Coefficient by Location (Rural and Urban Sectors)

This section first presents a method for the decomposition of the education Gini coefficient by location and then examines the characteristics of this Gini decomposition method through the use of hypothetical examples.

2.1. Decomposition of the Education Gini Coefficient by Location

Consider a country with seven levels of education: (0) no formal education, (1) incomplete primary, (2) complete primary, (3) incomplete secondary, (4) complete secondary, (5) incomplete tertiary, and (6) complete tertiary education. Let e_h and q_h be cumulative number

[3] Mookherjee and Shorrocks (1982) presented the decomposition of the Gini coefficient by population subgroups, while Lambert and Aronson (1993) provided interpretations for all the three inequality components based on a geometric approach.

of years of education for the *h*th level of education and the proportion of individuals with *h*th level of education, respectively. Then, the Gini coefficient of education is defined by

$$G = \frac{1}{2\mu}\sum_{h=0}^{6}\sum_{k=0}^{6} q_h q_k |e_k - e_h| \,, \qquad (1)$$

where $\mu = \sum_{h=0}^{6} q_h e_h$ is the mean years of education and $\sum_{h=0}^{6} q_h = 1$. Since $0 \le e_0 < e_1 < \cdots < e_6$, we can rewrite this equation as follows.

$$G = \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6} q_h q_k (e_k - e_h) \qquad (2)$$

Suppose that the country is divided into rural and urban areas (sectors 1 and 2, respectively). Let p_{ih} be the proportion of individuals with *h*th level of education in sector *i*. Then, the Gini coefficient of education for sector *i* is given by

$$G_i = \frac{1}{\mu_i}\sum_{h=0}^{6}\sum_{k>h}^{6} p_{ih} p_{ik} (e_k - e_h), \qquad (3)$$

where $\mu_i = \sum_{h=0}^{6} p_{ih} e_h$ is the mean years of education in sector *i* and $\sum_{h=0}^{6} p_{ih} = 1$. If we let s_i be sector *i's* share of population ($\sum_{i=1}^{2} s_i = 1$), then we have the following relationships.

$q_h = \sum_{i=1}^{2} s_i p_{ih}$, and

$\mu = \sum_{i=1}^{2} s_i \mu_i$.

Now we can decompose overall education Gini coefficient by location (rural and urban sectors) as follows.

$$G = G_{WS} + G_{BS} + G_R. \qquad (4)$$

G_{WS} is the within-sector Gini component, which measures educational inequality within rural and urban sectors and given by

$$G_{WS} = \sum_{i=1}^{2} s_i v_i G_i, \qquad (5)$$

where v_i is sector *i's* share of the total number of years of education, given by

$$v_i = \frac{s_i \mu_i}{\sum_{i=1}^{2} s_i \mu_i} = \frac{s_i \mu_i}{\mu} \qquad (\sum_{i=1}^{2} v_i = 1).$$

On the other hand, G_{BS} is the between-sector Gini component, which measures the rural-urban disparity in the mean years of education, and given by

$$G_{BS} = \frac{1}{2\mu}\sum_{i=1}^{2}\sum_{j=1}^{2} s_i s_j |\mu_i - \mu_j|. \qquad (6)$$

If we assume that the urban sector has a larger mean number of years of education than the rural sector, that is, $\mu_2 > \mu_1$, then G_{BS} can be modified to

$$G_{BS} = \frac{1}{\mu}\sum_{i=1}^{2}\sum_{j>i}^{2} s_i s_j (\mu_j - \mu_i). \qquad (7)$$

Unlike the Theil indices, overall education Gini coefficient is not necessarily equal to the sum of the within-sector and between-sector education Gini components.[4] We usually need the residual component ($G_R > 0$). Therefore, we have $G = G_{WS} + G_{BS} + G_R$. However, if there is no overlap between the rural and urban sectors in the level of educational attainment, then the residual component vanishes, that is, $G_R = 0$. Please refer to Appendix 1 for more details.

It should be noted that when individual-level or household-level data, rather than grouped data, are available, overall education Gini coefficient can be calculated by using the following formula.

$$G = \frac{1}{2N^2\mu}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{h=1}^{N_i}\sum_{k=1}^{N_j}\left|e_{ih} - e_{jk}\right|, \qquad (8)$$

where e_{ih}, N and N_i are, respectively, the cumulative number of years of education for individual *h* in sector *i*, total number of individuals, and total number of individuals in sector *i*, while the mean number of years of education is given by

$$\mu = \frac{1}{N}\sum_{i=1}^{2}\sum_{h=1}^{N_i} e_{ih}.$$

Overall education Gini coefficient given by Eq. (8) can be decomposed additively as follows (see, for example, Mookherjee and Shorrocks, 1982; and Lambert and Aronson, 1993).

$$G = G_{WS} + G_{BS} + G_R,$$

$G_{WS} = \sum_{i=1}^{2} s_i v_i G_i$ is the within-sector Gini component, where

$$G_i = \frac{1}{2N_i\mu_i}\sum_{h=1}^{N_i}\sum_{k=1}^{N_i}|e_{ih} - e_{ik}|, \qquad \mu_i = \frac{1}{N_i}\sum_{h=1}^{N_i} e_{ih}.$$

On the other hand, $G_{BS} = \frac{1}{2\mu}\sum_{i=1}^{2}\sum_{j=1}^{2} s_i s_j\left|\mu_i - \mu_j\right|$ is the between-sector Gini component, while G_R is the residual Gini component.

2.2. Decomposition of the Education Gini Coefficient by Location: Hypothetical Examples

To explore the characteristics of the Gini decomposition equation (Eq. 4), let us consider a hypothetical country divided into the rural and urban sectors. Table 1 presents the distribution of individuals across the seven levels of education mentioned earlier, while Fig. 1 depicts the education Lorenz curve for the rural and urban sectors. Here, 0, 3, 6, 9, 12, 14, and 16 years are

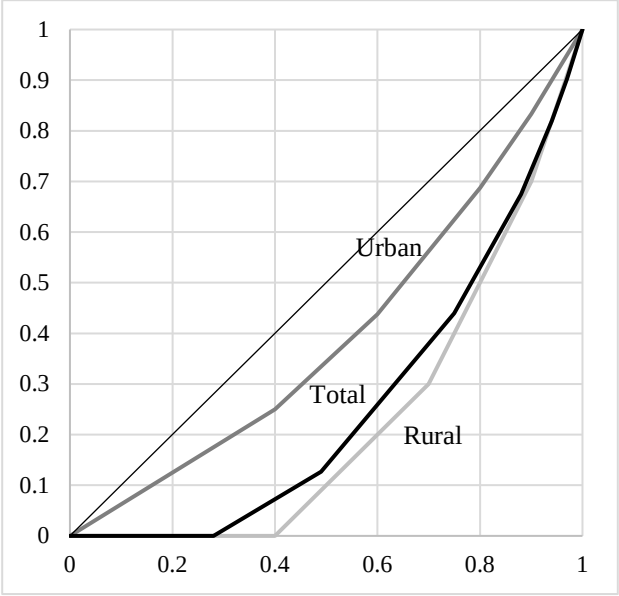
[4] The Theil *L* and *T* indices can be decomposed into the within-sector and between-sector inequality components as follows: $L = L_W + L_B$ and $T = T_W + T_B$ (see, for example, Bourguignon, 1979; Shorrocks, 1980; and Anand, 1983). However, when individuals without formal education is given 0 years of education, that is, $e_0 = 0$, the Theil indices cannot be calculated, because they include the logarithmic function.

assigned to these seven respective levels. 70% of population live in rural areas, with the remaining 30% in urban areas. Because 40% of individuals in the rural sector do not have formal education, the mean years of education in the rural sector is very low at 3.0, which is in sharp contrast to 9.6 in the urban sector. The result of the decomposition of the education Gini coefficient by rural and urban sectors indicates that the largest contributor is the between-sector Gini component, accounting for 57.6% of overall education Gini coefficient. Due to overlap between the rural and urban sectors in the level of education (that is, complete primary and incomplete secondary education), the residual component contributes 2.1% of overall education Gini coefficient.

Table 1. Distributions of Individuals across Seven Education Levels (Example 1)

	Education level	Years of education (e_h)	Rural (p_{1h})	Urban (p_{2h})	Total (q_h)	Contribution (%)
0	No education	0	0.4	0	0.28	
1	Incomplete primary	3	0.3	0	0.21	
2	Complete primary	6	0.2	0.4	0.26	
3	Incomplete secondary	9	0.1	0.2	0.13	
4	Complete secondary	12	0	0.2	0.06	
5	Incomplete tertiary	14	0	0.1	0.03	
6	Complete tertiary	16	0	0.1	0.03	
	Mean years of education (μ_i)		3.00	9.60	4.98	
	Population share (s_i)		0.70	0.30	1.00	
	Education share (v_i)		0.422	0.578	1.00	
	Gini coefficient (G_i)		0.540	0.202		
	Within-sector (G_W)				0.194	40.3
	Between-sector (G_B)				0.278	57.6
	Residual (G_R)				0.010	2.1
	Total (G)				0.483	100.0

Figure 1. Education Lorenz Curves for the Rural and Urban Sectors (Example 1)

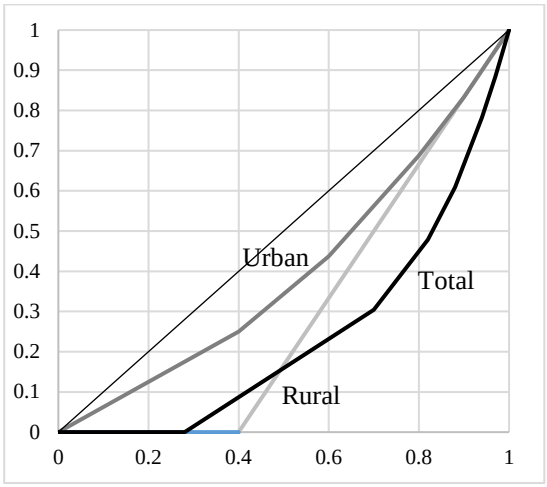


Next, suppose that there is no overlap between rural and urban sectors in the level of education (Table 2). Due to a significant disparity in the mean years of education between these sectors, the between-sector Gini component emerges as the dominant factor, contributing 75.6% to overall education Gini coefficient. However, since there is no overlap between the rural and urban sectors, the residual component is reduced to 0. In this example, overall education Gini coefficient is larger than both rural and urban education Gini coefficients. Fig. 2 depicts the education Lorenz curves for the rural and urban sectors.

Table 2. Distributions of Individuals across Seven Education Levels (Example 2)

	Education level	Years of education (e_h)	Rural (p_{1h})	Urban (p_{2h})	Total (q_h)	Contribution (%)
0	No education	0	0.4	0	0.28	
1	Incomplete primary	3	0.6	0	0.42	
2	Complete primary	6	0	0.4	0.12	
3	Incomplete secondary	9	0	0.2	0.06	
4	Complete secondary	12	0	0.2	0.06	
5	Incomplete tertiary	14	0	0.1	0.03	
6	Complete tertiary	16	0	0.1	0.03	
	Mean years of education (μ_i)		1.80	9.60	4.14	
	Population share (s_i)		0.70	0.30	1.00	
	Education share (v_i)		0.304	0.696	1.00	
	Gini coefficient (G_i)		0.400	0.202		
	Within-sector (G_W)				0.127	24.4
	Between-sector (G_B)				0.396	75.6
	Residual (G_R)				0.000	0.0
	Total (G)				0.523	100.0

Figure 2. Education Lorenz Curves for the Rural and Urban Sectors (Example 2)



3. Decomposition of the Education Gini Coefficient by Location (Rural and Urban Sectors): Empirical Evidence

Using nation-wide household surveys, this section empirically analyzes the determinants of educational inequality in some Asian countries by decomposing the education Gini coefficient by location (rural and urban sectors). In the decomposition analysis, household-level education data are used, where households are classified into several education groups

according to the highest education level attained by their household heads. Households without formal education are assigned 0 years of education.

3.1. Empirical Evidence from Indonesia

Based on the National Socioeconomic Survey (*Susenas*) conduct by the Central Bureau of Statistics in 2000, 2009, and 2018, Table 3 presents the changes in the distributions of households across nine education levels in Indonesia. Overall educational inequality, as measured by the Gini coefficient, has declined from 0.37 to 0.30 as the mean number of years of education has increased from 6.5 to 8.1. The decomposition of the education Gini coefficient by location reveals that the predominant factor has been the within-sector Gini component, constituting 47% of the overall education Gini coefficient. The rural sector exhibits a higher education Gini coefficient than the urban sector. Both rural and urban sectors have witnessed a decline in their education Gini coefficients; however, the urban-rural disparity in the education Gini coefficient has been diminishing. On the other hand, the between-sector Gini component has reduced its contribution from 32% to 26%, while the residual Gini component has seen an increase from 21% to 26%, signifying a growing urban-rural overlap in the distribution of educational attainment.

Table 3. Distributions of Households across Education Levels in the Rural and Urban Sectors (Indonesia in 2000, 2009 and 2018)

	2000				2009				2018			
Years of education	Rural	Urban	Total	Cont. (%)	Rural	Urban	Total	Cont. (%)	Rural	Urban	Total	Cont. (%)
0	0.17	0.07	0.13		0.11	0.05	0.08		0.08	0.03	0.05	
3	0.29	0.14	0.23		0.25	0.13	0.19		0.21	0.13	0.16	
6	0.33	0.25	0.29		0.34	0.22	0.28		0.35	0.22	0.28	
9	0.11	0.16	0.13		0.14	0.17	0.15		0.16	0.16	0.16	
12	0.09	0.28	0.17		0.12	0.30	0.21		0.15	0.33	0.25	
13	0.01	0.01	0.01		0.01	0.01	0.01		0.00	0.01	0.01	
15	0.00	0.02	0.01		0.01	0.03	0.02		0.01	0.03	0.02	
16	0.01	0.07	0.03		0.02	0.09	0.05		0.03	0.09	0.06	
18	0.00	0.00	0.00		0.00	0.01	0.01		0.00	0.01	0.01	
Mean years of education (μ_i)	5.1	8.3	6.5		6.1	8.9	7.5		6.7	9.3	8.1	
Population share (s_i)	0.578	0.422	1.0		0.510	0.490	1.0		0.453	0.547	1.0	
Education share (v_i)	0.458	0.542	1.0		0.417	0.583	1.0		0.374	0.626	1.0	
Gini coefficient (G_i)	0.393	0.297			0.355	0.280			0.318	0.258		
Within-sector (G_W)			0.172	46.4			0.156	46.8			0.142	47.9
Between-sector (G_B)			0.120	32.4			0.093	28.1			0.078	26.4
Residual (G_R)			0.079	21.2			0.083	25.1			0.076	25.7
Total (G)			0.371	100.0			0.332	100.0			0.297	100.0

(Source) National Socioeconomic Survey (*Susenas*), 2000, 2009, and 2018

3.2. Empirical Evidence from Some Asian Countries

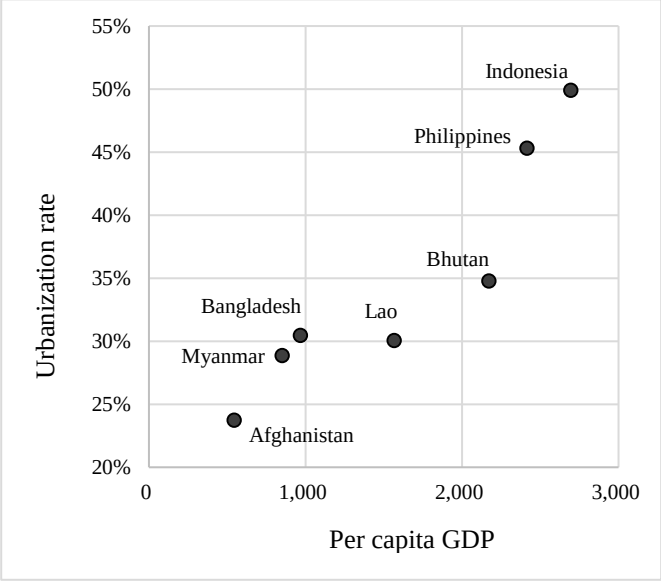
This section conducts a comparative analysis of the structures of educational attainment in rural and urban areas across various Asian countries. Table 4 outlines the characteristics of these Asian countries. Afghanistan, Bangladesh, and Bhutan are members of the South Asian Association for Regional Cooperation (SAARC), while Indonesia, Lao PDR, Myanmar, and the Philippines are members of the Association of Southeast Asian Nations (ASEAN). In per capita GDP terms, Afghanistan ranks as the least affluent country in 2010, with Myanmar and Bangladesh following closely. On the other hand, Indonesia stands as the wealthiest, followed by the Philippines and Bhutan. Between 2005 and 2010, Myanmar experienced the highest growth rate in per capita GDP, with Afghanistan and Bhutan following next. While Indonesia and the Philippines were relatively rich among the seven Asian countries, their per capita GDP grew at a comparatively slower rate. As shown in Fig. 3, there exists a positive association between the urbanization rate and per capita GDP (correlation coefficient = 0.92). In 2010, Indonesia recorded the highest urbanization rate at 49.9%, whereas Afghanistan had the lowest at 23.7%.

Table 4. Characteristics of Selected Asian Countries

Country	GDP in 2010 (in million 2015 US dollars)	Population in 2010 (in 1,000)	Per capita GDP in 2010 (in 2015 US dollars)	Growth rate of per capita GDP 2005-10 (in %)	Urbanization rate in 2010 (in %)
Afghanistan	15,355	28,190	545	8.1	23.7
Bangladesh	143,577	148,391	968	4.9	30.5
Bhutan	1,532	706	2,172	8.0	34.8
Indonesia	657,835	244,016	2,696	4.3	49.9
Lao PDR	9,912	6,323	1,568	6.2	30.1
Myanmar	42,038	49,391	851	9.8	28.9
Philippines	228,622	94,637	2,416	3.0	45.3
SAARC	2,032,094	1,660,546	1,224	4.8	30.9
ASEAN	1,955,622	598,207	3,269	4.0	44.4

(Source) World Development Indicators (World Bank 2024)

Figure 3. Relationship between Per Capita GDP and Urbanization Rate

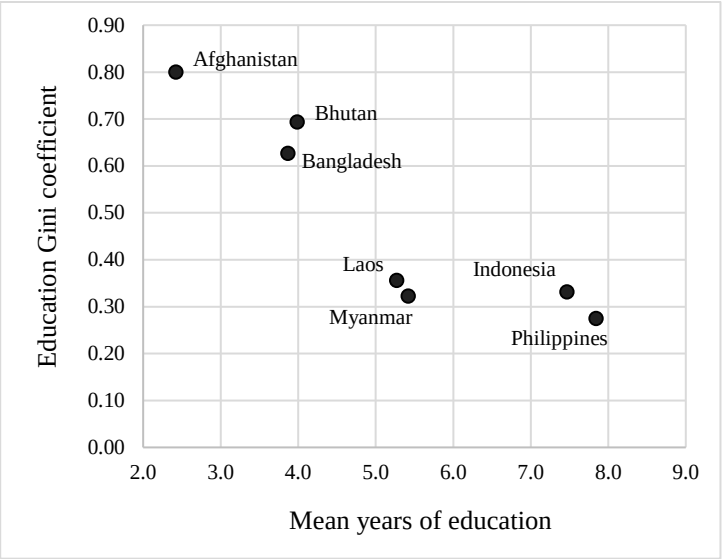


(Source) World Development Indicators (World Bank 2024)

Fig. 4 presents a scatter plot of overall Gini coefficient against mean years of education for the seven Asian countries. As expected, the education Gini coefficient exhibits a downward trend with respect to the mean years of education (correlation coefficient = –0.9).[5] In 2007, Afghanistan registered a very large educational inequality, with 73% of households lacking formal education, resulting in an education Gini coefficient of 0.80. Its education Lorenz curve is shown in Fig. 5. On the other hand, in the Philippines, households were distributed very widely across education levels, with only 3% lacking formal education in 2006, leading to a relatively high mean years of education at 7.8. The country's education Gini coefficient stood at 0.275, the smallest among the seven Asian countries. Its education Lorenz curve is depicted in Fig. 6.

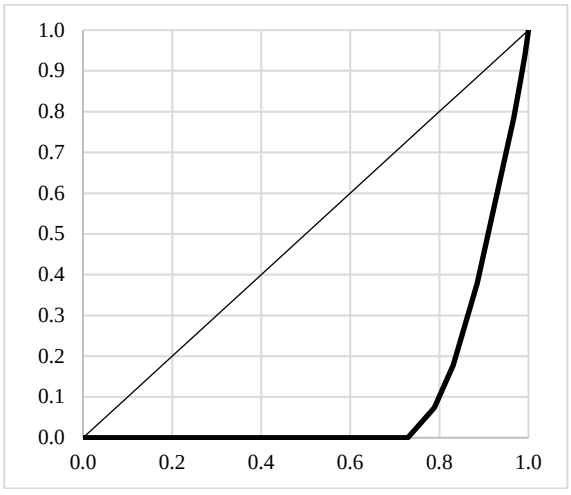
[5] When individuals without formal education are assigned 0 years of education, the Gini coefficient is likely to decline monotonically with the mean years of education.

Figure 4. Relationship between the Education Gini Coefficient and Mean Years of Education



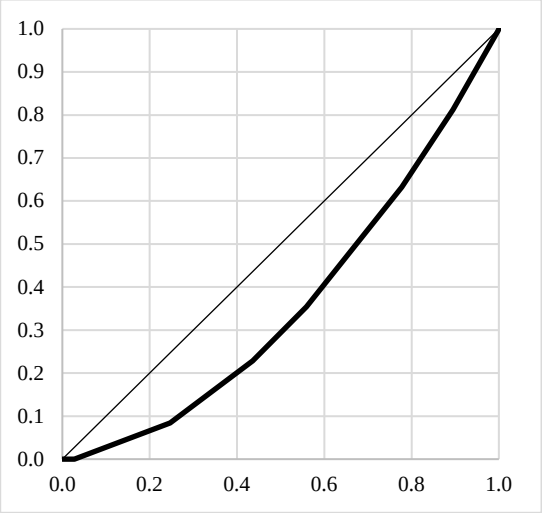
(Source) Living Conditions Survey 2007 (Afghanistan), Household Income and Expenditure Survey 2010 (Bangladesh), Living Standard Survey 2010 (Bhutan), National Socioeconomic Survey 2009 (Indonesia), Expenditure and Consumption Survey 2007/2008 (Laos), Household Income and Expenditure Survey 2012 (Myanmar), Family Income and Expenditure Survey 2006 (Philippines)

Figure 5. Education Lorenz Curve for Afghanistan in 2007



(Source) Living Conditions Survey 2007, Afghanistan

Figure 6. Education Lorenz Curve for the Philippines in 2006



(Source) Family Income and Expenditure Survey 2006, Philippines

Tables 5, 6, and 7 present the decomposition of the education Gini coefficient by location (rural and urban sectors) for the selected Asian countries. The within-sector Gini component is a predominant determinant in all these countries, constituting 40-50% of the overall education Gini coefficient. With the exception of Myanmar, the rural sector had a significantly larger education Gini coefficient than the urban sector. In Myanmar, within the lower education groups, the education Lorenz curve for the rural sector consistently lies above that for the urban sector (Fig. 7). This unique pattern contributes to Myanmar's rural sector having a smaller education Gini coefficient than its urban sector, even though it possesses smaller mean years of education (refer to Fig. 3 for the relationship between the education Gini coefficient and mean years of education).

Table 5. Distributions of Households across Education Levels in the Rural and Urban Sectors (Philippines and Myanmar)

	Philippines in 2006					Myanmar in 2012			
Years of education	Rural	Urban	Total	Cont. (%)	Years of education	Rural	Urban	Total	Cont. (%)
0	0.05	0.01	0.03		0	0.04	0.02	0.03	
3	0.31	0.13	0.22		3	0.48	0.28	0.41	
6	0.24	0.14	0.19		5	0.32	0.26	0.30	
8	0.12	0.12	0.12		9	0.12	0.24	0.16	
10	0.17	0.27	0.22		11	0.02	0.08	0.04	
12	0.07	0.16	0.12		14	0.02	0.12	0.05	
14	0.05	0.16	0.10						
16	0.00	0.00	0.00						
Mean years of education (μ_i)	6.6	9.2	7.8			4.6	6.8	5.4	
Population share (s_i)	0.504	0.496	1.0			0.661	0.339	1.0	
Education share (v_i)	0.421	0.578	1.0			0.569	0.431	1.0	
Gini coefficient (G_i)	0.308	0.218				0.277	0.309		
Within-sector (G_W)			0.128	46.4				0.149	48.0
Between-sector (G_B)			0.083	30.0				0.092	29.7
Residual (G_R)			0.065	23.6				0.070	22.4
Total (G)			0.275	100.0				0.311	100.0

(Source) Household Income and Expenditure Survey 2012 (Myanmar), Family Income and Expenditure Survey 2006 (Philippines)

Table 6. Distributions of Households across Education Levels in the Rural and Urban Sectors (Laos and Bangladesh)

	Laos in 2007					Bangladesh in 2010			
Years of education	Rural	Urban	Total	Cont. (%)	Years of education	Rural	Urban	Total	Cont. (%)
0	0.16	0.06	0.13		0	0.56	0.42	0.51	
3	0.29	0.16	0.25		3	0.10	0.10	0.10	
5	0.34	0.30	0.32		5	0.09	0.09	0.09	
8	0.13	0.18	0.15		9	0.17	0.20	0.18	
11	0.08	0.27	0.14		12	0.07	0.17	0.11	
16	0.00	0.03	0.01		14	0.01	0.03	0.02	
Mean years of education (μ_i)	4.6	6.9	5.3			3.2	5.0	3.9	
Population share (s_i)	0.693	0.307	1.0			0.636	0.364	1.000	
Education share (v_i)	0.599	0.401	1.0			0.531	0.469	1.000	
Gini coefficient (G_i)	0.365	0.294				0.669	0.548		
Within-sector (G_W)			0.188	52.7				0.320	51.0
Between-sector (G_B)			0.094	26.4				0.105	16.8
Residual (G_R)			0.074	20.9				0.202	32.3
Total (G)			0.356	100.0				0.627	100.0

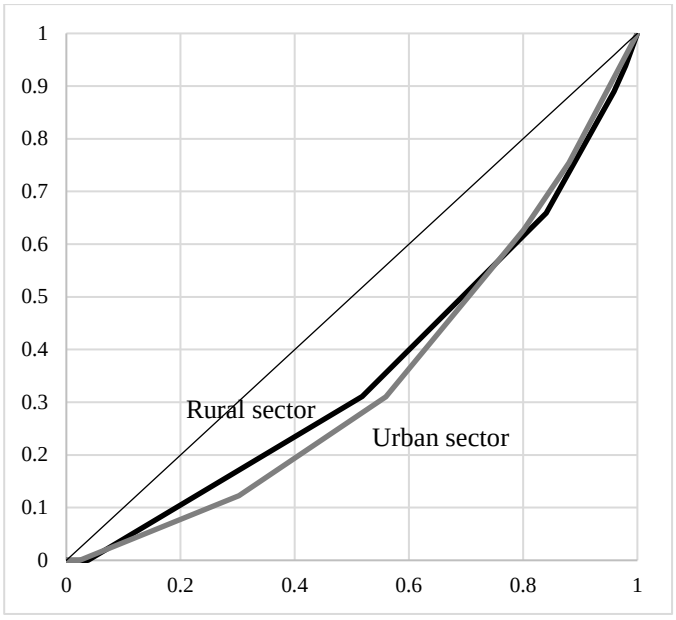
(Source) Household Income and Expenditure Survey 2010 (Bangladesh), Expenditure and Consumption Survey 2007-2008 (Laos)

Table 7. Distributions of Households across Education Levels in the Rural and Urban Sectors (Bhutan and Afghanistan)

	Bhutan in 2010					Afghanistan in 2007			
Years of education	Rural	Urban	Total	Cont. (%)	Years of education	Rural	Urban	Total	Cont. (%)
0	0.76	0.31	0.61		0	0.80	0.47	0.73	
3	0.06	0.06	0.06		3	0.05	0.08	0.06	
5	0.02	0.02	0.02		6	0.04	0.05	0.04	
8	0.05	0.11	0.07		9	0.05	0.08	0.05	
11	0.05	0.24	0.11		12	0.05	0.20	0.08	
13	0.03	0.11	0.06		14	0.01	0.04	0.01	
16	0.04	0.15	0.07		16	0.00	0.04	0.01	
					18	0.00	0.03	0.01	
Mean years of education (μ_i)	2.2	7.7	4.0			1.6	5.5	2.4	
Population share (s_i)	0.671	0.329	1.0			0.789	0.211	1.0	
Education share (v_i)	0.369	0.631	1.0			0.523	0.477	1.0	
Gini coefficient (G_i)	0.825	0.434				0.853	0.592		
Within-sector (G_W)			0.289	41.7				0.412	51.4
Between-sector (G_B)			0.304	43.8				0.266	33.2
Residual (G_R)			0.101	14.6				0.123	15.4
Total (G)			0.694	100.0				0.800	100.0

(Source) Living Conditions Survey 2007 (Afghanistan), Living Standard Survey (Bhutan)

Figure 7. Education Lorenz Curves for Rural and Urban Sectors in Myanmar, 2012.

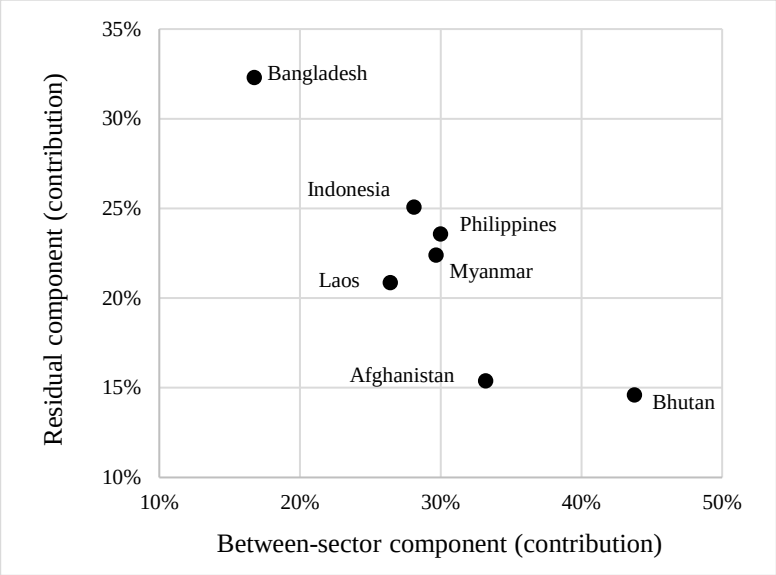


(Source) Household Income and Expenditure Survey 2012 (Myanmar)

Fig. 8 presents a scatter plot depicting the relationship between the between-sector Gini component and the residual Gini component in terms of their contributions to overall education

Gini coefficient. A significant negative relationship exists between them, as evidenced by the correlation coefficient of –0.89. This indicates that the rural-urban overlap in the distribution of educational attainment rises as the rural-urban education disparity declines. It should be noted that Bangladesh is the only country whose residual Gini component exceeds the between-sector Gini component (see Table 6). This suggests that there exists an exceptionally large rural-urban overlap in the distribution of educational attainment in Bangladesh. This is in sharp contrast to Bhutan, where the between-sector Gini component is much larger than the residual Gini component (see Table 7). In Bhutan, the urban-rural ratio in mean years of education is very large at 3.5 (7.7 years against 2.2 year).

Figure 8. Relationship between Between-sector Component and the Residual Component

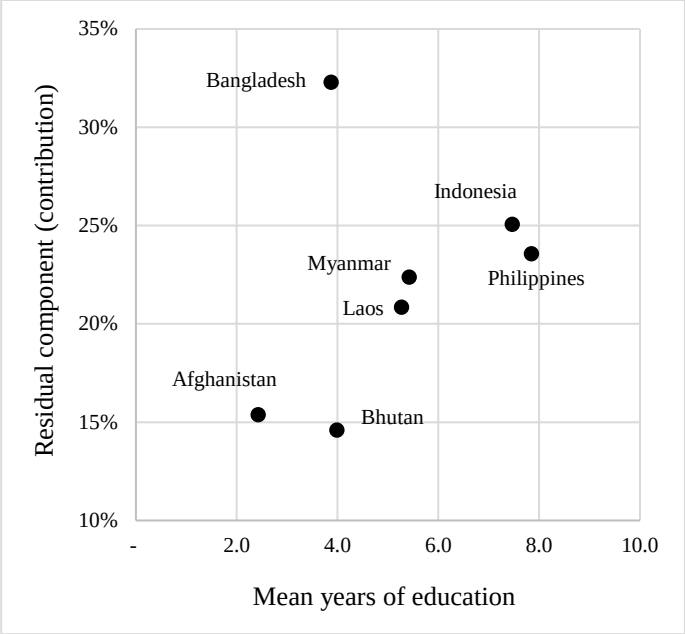


(Source) Same as Fig. 4.

Fig. 9 shows a scatter plot depicting the relationship between mean years of education and the contribution of the residual component to overall education Gini coefficient. When Bangladesh is excluded as an outlier, a significant positive relationship exists between them, as evidenced by the correlation coefficient of 0.90. This signifies that the rural-urban overlap in the distribution of educational attainment increases as mean years of education rises. As mentioned above, Bangladesh is exceptional (see Table 6). Despite having relatively small

mean years of education, the country registered a very large residual Gini component. In comparison to the other countries, the rural and urban sectors display remarkably similar distributions of educational attainment in Bangladesh.

Figure 9. Relationship between Mean Years of Education and the Residual Component



(Source) Same as Fig. 4.

4. Concluding Remarks

This study presented a method for the decomposition of the education Gini coefficient by location and examined the characteristics of this Gini decomposition method through the use of hypothetical examples. It then empirically analyzed the determinants of educational inequality in some Asian countries using the Gini decomposition method.

Major findings are summarized as follows. In a dual economy consisting of the rural and urban sectors, the Gini coefficient of education can be additively decomposed into three distinct components: the within-sector, between-sector, and residual Gini components. The within-sector Gini component measures educational inequality within the rural and urban sectors, while the between-sector Gini component measures the rural-urban disparity in the mean level of educational attainment. The residual Gini component assesses the extent of overlap between

the rural and urban sectors in the distribution of educational attainment; thus, when no such overlap exists, the residual component is reduced to zero.

In Indonesia, overall educational inequality, as measured by the Gini coefficient, has declined from 0.37 to 0.30 over the period 2000-2018 as the mean number of years of education has increased from 6.5 to 8.1. The decomposition of the education Gini coefficient by location reveals that the predominant factor has been the within-sector Gini component, constituting 47% of the overall education Gini coefficient. While the between-sector Gini component has reduced its contribution from 32% to 26%, the residual Gini component has seen an increase in its contribution from 21% to 26%, signifying a growing urban-rural overlap in the distribution of educational attainment.

In all selected Asian countries, including four ASEAN countries (Indonesia, Laos, Myanmar, and the Philippines) and three SAARC countries (Afghanistan, Bangladesh, and Bhutan), the within-sector Gini component is a predominant determinant by accounting for 40-50% of the overall education Gini coefficient. There is a significant negative relationship between the between-sector Gini component and the residual Gini component in terms of their contributions to overall education Gini coefficient. This indicates that the rural-urban overlap in the distribution of educational attainment rises as the rural-urban education disparity declines.

When Bangladesh is excluded as an outlier, a significant positive relationship exists between mean years of education and the contribution of the residual component to overall education Gini coefficient. This signifies that the rural-urban overlap in the distribution of educational attainment increases as mean years of education rises. Bangladesh is exceptional. Despite having relatively small mean years of education, the country registers a very large residual Gini component. Bangladesh is the only country whose residual Gini component exceeds the between-sector Gini component. In comparison to the other countries, the rural and urban sectors display remarkably similar distributions of educational attainment in Bangladesh. This is in sharp contrast to Bhutan, where the between-sector Gini component is much larger than the residual Gini component.

Appendix 1: Decomposition of the Education Gini coefficient by Rural and Urban Sectors

Consider a country with seven levels of education. If we let e_h and q_h be cumulative number of years of education for the *h*th level of education and the proportion of individuals with *h*th level of education, respectively, then the Gini coefficient of education is defined by

$$G = \frac{1}{2\mu}\sum_{h=0}^{6}\sum_{k=0}^{6} q_h q_k |e_k - e_h| ,$$

where $\mu = \sum_{h=0}^{6} q_h e_h$ is the mean years of education and $\sum_{h=0}^{6} q_h = 1$. Since $0 \le e_0 < e_1 < \cdots < e_6$, we can rewrite this equation as follows.

$$G = \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6} q_h q_k (e_k - e_h).$$

Next we let p_{ih} be the proportion of individuals with *h*th level of education in sector *i*. Then, the Gini coefficient of education for sector *i* is given by

$$G_i = \frac{1}{\mu_i}\sum_{h=0}^{6}\sum_{k>h}^{6} p_{ih} p_{ik} (e_k - e_h),$$

where $\mu_i = \sum_{h=0}^{6} p_{ih} e_h$ is the mean years of education in sector *i* and $\sum_{h=0}^{6} p_{ih} = 1$.

We now have the following proposition.

Proposition 1

Under the assumption that there is no overlap between the rural and urban sectors in the distribution of educational attainment, we have

$$G = G_{WS} + G_{BS},$$

where G_{WS} is the within-sector Gini component, while G_{BS} is the between-sector Gini component. If we let s_i and v_i be, respectively, sector *i's* share of population ($\sum_{i=1}^{2} s_i = 1$) and sector *i's* share of the total number of years of education ($\sum_{i=1}^{2} v_i = 1$), then we have

$$G_{WS} = \sum_{i=1}^{2} s_i v_i G_i \quad \text{and}$$

$$G_{BS} = \frac{1}{2\mu}\sum_{i=1}^{2}\sum_{j=1}^{2} s_i s_j |\mu_i - \mu_j|,$$

where $G_i = \frac{1}{\mu_i}\sum_{h=0}^{6}\sum_{k>h}^{6} p_{ih} p_{ik} (e_k - e_h) \quad (0 \le e_0 < e_1 < \cdots < e_6)$.

Proof

Since $q_h = \sum_{i=1}^{2} s_i p_{ih}$, we can modify overall education Gini coefficient as follows.

$$G = \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6} q_h q_k (e_k - e_h)$$

$$= \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6} (\sum_{i=1}^{2} s_i p_{ih})(\sum_{i=1}^{2} s_i p_{ik})(e_k - e_h)$$

$$= \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6}(s_1p_{1h} + s_2p_{2h})(s_1p_{1k} + s_2p_{2k})(e_k - e_h)$$

$$= \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6}(s_1p_{1h})(s_1p_{1k})(e_k - e_h) + \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6}(s_2p_{2h})(s_2p_{2k})(e_k - e_h)$$

$$+ \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6}(s_1p_{1h})(s_2p_{2k})(e_k - e_h) + \frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6}(s_2p_{2h})(s_1p_{1k})(e_k - e_h).$$

Because we have $v_i = \frac{s_i\mu_i}{\sum_{i=1}^{2} s_i\mu_i} = \frac{s_i\mu_i}{\mu}$ by definition, the first term of this equation can be modified to

$$\frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6}(s_1p_{1h})(s_1p_{1k})(e_k - e_h)$$

$$= s_1\left(\frac{s_1\mu_1}{\mu}\right)\left(\frac{1}{\mu_1}\sum_{h=0}^{6}\sum_{k>h}^{6}p_{1h}p_{1k}(e_k - e_h)\right)$$

$$= s_1v_1G_1.$$

Similarly, the second term of the equation can be modified to

$$\frac{1}{\mu}\sum_{h=0}^{6}\sum_{k>h}^{6}(s_2p_{2h})(s_2p_{2k})(e_k - e_h)$$

$$= s_2v_2G_2.$$

Adding the first two terms, we obtain

$$G_{WS} = \sum_{i=1}^{2} s_iv_iG_i,$$

which is the within-sector Gini component. This is the weighted sum of sector 1's education Gini and sector 2's education Gini, because $\sum_{i=1}^{2} s_iv_i$ does not usually equal to one. Now suppose that there is no overlap between rural and urban sectors in the distribution of educational attainment. In the rural sector, $p_{1h} = 0$ for $m + 1 \le h \le 6$, while in the urban sector, $p_{2h} = 0$ for $0 \le h \le m$. Then, the third term of the equation above can be modified as follows.

$$\frac{s_1s_2}{\mu}\sum_{h=0}^{6}p_{1h}\sum_{k>h}^{6}p_{2k}(e_k - e_h)$$

$$= \frac{s_1s_2}{\mu}\sum_{h=0}^{m}p_{1h}\sum_{k>h}^{6}p_{2k}(e_k - e_h) \qquad \text{since } p_{1h} = 0 \text{ for } m + 1 \le h \le 6$$

$$= \frac{s_1s_2}{\mu}\sum_{h=0}^{m}p_{1h}\left(\sum_{k=m+1}^{6}p_{2k}e_k - e_h\sum_{k=m+1}^{6}p_{2k}\right) \qquad \text{since } p_{2h} = 0 \text{ for } 0 \le h \le m$$

$$= \frac{s_1s_2}{\mu}\left(\sum_{k=m+1}^{6}p_{2k}e_k - \sum_{h=0}^{m}p_{1h}e_h\right) \qquad \text{since } \sum_{h=0}^{m}p_{1h} = 1 \text{ and } \sum_{k=m+1}^{6}p_{2k} = 1$$

$$= \frac{s_1s_2}{\mu}\left(\sum_{k=0}^{6}p_{2k}e_k - \sum_{h=0}^{6}p_{1h}e_h\right)$$

$$= \frac{s_1s_2}{\mu}(\mu_2 - \mu_1)$$

On the other hand, the fourth term of the equation above is zero.

$$\frac{s_1 s_2}{\mu}\sum_{h=0}^{6} p_{2h}\sum_{k>h}^{6} p_{1k}(e_k - e_h)$$

$$= \frac{s_1 s_2}{\mu}\sum_{h=m+1}^{6} p_{2h}\sum_{k>h}^{6} p_{1k}(e_k - e_h) \quad \text{since } p_{2h} = 0 \text{ for } 0 \le h \le m$$

$$= 0 \quad \text{since } p_{1k} = 0 \text{ for } m+1 \le k \le 6.$$

Therefore, the sum of the third and fourth terms is reduced to

$$\frac{s_1 s_2}{\mu}(\mu_2 - \mu_1) + 0$$

$$= \frac{1}{\mu}\sum_{i=1}^{2}\sum_{j>i}^{2} s_i s_j \left(\mu_j - \mu_i\right)$$

$$= \frac{1}{2\mu}\sum_{i=1}^{2}\sum_{j=1}^{2} s_i s_j \left|\mu_i - \mu_j\right| > 0,$$

which is the between-sector Gini component (G_{BS}). In conclusion, we obtain

$$G = G_{WS} + G_{BS}.$$

References

Agrawal, T. (2014). Educational inequality in rural and urban India. *International Journal of Educational Development*. 34, 11-19.

Akita, T., & Kataoka, M. (2022). *Regional Inequality and Development: Measurement and Applications in Indonesia*, Springer.

Almeida, D. R. C., Andrade, J. A. S., Duarte, A., & Simões, M. (2022). Human capital disparities and earnings inequality in the Portuguese private labour market. *Social Indicators Research*, 159, 145–167.

Anand, S. (1983), *Inequality and Poverty in Malaysia: Measurement and Decomposition*. New York: Oxford University Press.

Banzragch, O., Mizunoya, S., & Bayarjargal, M. (2019). Education inequality in Mongolia: measurement and causes. *International Journal of Educational Development*, 68, 68-79.

Bourguignon, F. (1979). Decomposable income inequality measures. *Econometrica*, 47(4), 901–920.

Castello, A., & Domenech, R. (2002). Human capital inequality and economic growth: Some new evidence. *The Economic Journal*, 112, C187-C200.

Castello-Climent, A., & Domenech, R. (2021). Human capital and income inequality revisited. *Education Economics*, 29(2), 194-212.

Coady, D., & Dizioli, A. (2018). Income inequality and education revisited: Persistence, endogeneity and heterogeneity. *Applied Economics*, 50(25), 2747–2761.

Fordvari, P., & van Leeuwen, B. (2011). Should less inequality in education lead to a more equal income distribution? *Education Economics*, 19(5), 537–554.

Hojo, M. (2009). Inequality in Japanese education: Estimation using the Gini education coefficient. *Japanese Economy*. 36(3), 3-27.

Lambert, P. J., & Aronson, J. R. (1993). Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal*, 103, 1221–1227.

Lim, A. S, K., & Tang, K. K. (2008). Human capital inequality and the Kuznets curve. *The Developing Economies*. 46(1), 26-51.

Lin, C. H. A. (2007). Education expansion, educational inequality, and income inequality: Evidence from Taiwan, 1976–2003. *Social Indicators Research*, 80, 601–615.

Luo, G., Zeng, S., & Baležentis, T. (2022). Multidimensional measurement and comparison of China's educational inequality. *Social Indicators Research*, 163, 857–874.

Meschi, E., & Scervini, F. (2014). Expansion of schooling and educational inequality in Europe: the educational Kuznets curve revisited. *Oxford Economic Papers*, 66(3), 660–680.

Mookherjee, D., & Shorrocks, A. (1982). A decomposition analysis of the trend in UK income inequality. *The Economic Journal*, 92, 886–902.

Shorrocks, A. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48(3), 613–25.

Shukla, V., & Mishra, U. S. (2019). Educational expansion and schooling Inequality: Testing educational Kuznets Curve for India. *Social Indicators Research*, 141, 1265–1283.

Thomas, V., Wang, Y., & Fan, X. (2001). Measuring educational inequality: Gini coefficient of education. Policy Research Working Paper, 2525, The World Bank.