International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied


---

# Construction of a Statistical Index as a Measure of a Country's Multidimensional Development

Ramoncito G. Cambel*

*Institute of Statistics, University of the Philippines Los Baños, Laguna, 4030, Philippines*
*Email: rgcambel2@up.edu.ph*

**Abstract**

Economic growth is expected to enhance the policies for poverty reduction and other social problems. But it is not always the case; economic growth does not necessarily reflect human development. This situation can be answered by the fact that economic growth is commonly measured by GDP. Because of this, researchers have formulated different indices that assess economic and human development. The human development index (HDI) measures the basic dimensions of human development and does not consider other indicators of development. This paper aims to construct an alternative measure of a country's performance using the Millennium Development Goals (MDG) indicators. Initial variables that were considered in the study are the different indicators for monitoring the progress of the eight MDGs. The analysis of data considered only a specific year that has the most number of available indicators. Variables were then subjected to principal component analysis to reduce their dimensionality. The identified principal components with high loadings were used in the construction of the statistical index as an alternative measure of development. Bootstrap samples were generated to check the statistical properties of the index such as unbiasedness, precision, accuracy, and consistency. The proposed index, the Multidimensional Development Index (MDI) which is composed of 11 indicators, was found to possess desirable characteristics of an estimator. This index encompasses development through eradicating extreme poverty and hunger, reducing child mortality, improving maternal health, ensuring environmental sustainability, and developing a global partnership for development.

***Keywords:*** statistical index; development; principal component analysis; bootstrap resampling.

---

* Corresponding author.

## 1. Introduction

Economic growth measures the economic output of a country. The determination of economic growth can be considered as one of the most important indicators for the assessment of economic policy in a country. It is used as a basis for policies regarding the reduction of poverty, unemployment, and other social problems. The most common measure of economic growth is the gross domestic product. In contrast to the importance of GDP, it is surrounded by controversies. The most famous of which is the limitation of using GDP as an indicator of economic growth in improving living standards. Because of the limitations of GDP, local and international organizations, non-government organizations, individuals from the private sector and academe have formulated different indices that assess economic, government and even human development issues. Among the existing indices, the Human Development Index (HDI) can be regarded as the most famous measure of human development. It is a focus measure that concentrates on the essential aspects of human development: people should lead a long and healthy life, people should acquire knowledge, and people should have access to resources needed for a decent standard of living. As mentioned, it just focuses on the basic dimensions of human development and does not consider some other important dimensions of human development. Millennium Development Goals reflect the multidimensional aspect of development as these encompass the United Nations' vision of fighting poverty in its many dimensions. Thus, using the MDG indicators, this study generally aims to construct a statistical index that could serve as a measure of a country's multidimensional development. Specifically, the study aims to compare the constructed statistical index with the existing measure(s) of economic growth and human development and evaluate the statistical properties of the constructed index. Despite the relevance of HDI, the index has been criticized for the reason of applying equal weights to its components. Alternative indices based on the same components as that of HDI as a measure of human development were proposed, namely: Modified Human Development Index (MHDI), Modified Human Development Index Factor 1 (MHDIF1), Modified Human Development Index Factor 2 (MHDIF2), and Borda ranking. Moreover, the researcher investigated whether the proposed indices are redundant with existing indices like HDI. Based on the results of the study, MHDI, HDI, and MHDIF1 produced an almost similar ranking of countries [1]. Reference [2] proposed a new and alternative composite index of development and poverty known as the Mazziotta-Pareto Index (MPI). The index was designed to satisfy three properties of an index which are deemed important: normalization of the indicators by a specific criterion that deletes the unit of measure and the variability effect; synthesis independent from an ideal unit; and simplicity of the computation. Composite indicator as a tool for comparing and ranking countries across the world has become popular among policymakers, researchers, experts, academics, and non-government organizations. As a response to the growing interest of researchers towards the construction of composite indicators, the Organisation for Economic Co-operation and Development (OECD) published a handbook that aims to provide a guide to the construction and use of composite indicators for different stakeholders. The handbook provides a detailed discussion of the steps in index construction, from the development of a sound theoretical framework, to the dissemination of results of the proposed index [3]. Choice of the indicators to be included in the composite index, transformation technique, and weighting system are some of the considerations that need to be addressed in index construction. A technical paper published by the Australian Bureau of Statistics provides information on the concepts, data, and methods used to create the Socio-Economic Indexes for Areas (SEIFA). Principal Component Analysis is used to determine the weights of the variables included in the index [4]. References [5] and [6] also used Principal Component Analysis in determining the weights of the happiness index among UPLB undergraduate students and UPLB faculty, respectively. Also, in the field of academe, [7] developed a composite provincial employment opportunity index in the Philippines based on socioeconomic characteristics and demographic profile. Statistical properties are important to be assessed to determine the validity of the index. In the year 2011, Reference [8] used the bootstrap resampling technique to assess the statistical properties of the Student Evaluation of Teachers (SET) score index. The constructed index was consistent, accurate, and precise. Thus, the researcher concluded that the bootstrap resampling method is a good method to assess the statistical properties of an estimator.

## 2. Methodology

### *2.1. Data Sources*

Data from the World Bank were retrieved from the organization's data bank. Initial variables that were considered in the study are the different indicators for monitoring the progress of the eight Millennium Development Goals. The MDG indicators are expressed in various units and must be normalized. This is performed to avoid problems in combining different measurement units and having extreme values dominate, and partially correct for data quality problems. Standardization addresses the measurement units by converting the individual indicators to a common scale. After standardization, indicators will have an average of zero and a standard deviation of 1. Unlike other normalization techniques which are highly affected by the presence of outliers, standardization avoids introducing aggregation distortions stemming from differences in variable means [9].

### *2.2. Index Construction*

Among the countries included in the data, those with poor statistical accounts data were omitted. According to the United Nations, there is no five-year period when the availability of data is more than 70% of what is required. Thus, the criterion for determining whether a country has poor statistical accounts data was set as not having a value for an indicator which is observed in at least 50% of the countries. Countries with available data on at least 70% of the indicators present were retained. Data on some variables are not completely collected due to a lack of a well-established statistical system thus producing 'missing data'. Thus, imputation, which is a statistical technique to estimate missing values, is required. Countries were classified according to the region they belong to. Mean for a certain region was computed and was used as a replacement for the missing values for countries belonging to that region. For each indicator, the average and standard deviation of values across countries were calculated. The standardization formula is given by (1):

$$I_{pc} = \frac{x_{pc} - \overline{x}_{pc}}{\sigma_{pc}} \quad (1)$$

where $x_{pc}$ is the value of indicator *p* for country *c*, $\overline{x}_{pc}$ is the average value of indicator *p* for country c, and

$\sigma_{pc}$ is the standard deviation of indicator *p* for country *c*. Correlation among the normalized indicators was computed to determine which indicators to remove. Indicators with very high correlation coefficient were scrutinized to avoid having indicators which measure the same aspect of development. The reduced list of variables was subjected to Principal Component Analysis to reduce its dimensionality while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all the original variables. The identified principal components with high loadings were used in the construction of the statistical index.

### *2.3. Statistical Properties of Index*

The population of the statistical indices across countries was considered as the ‘surrogate population’. The mean value of the constructed indices across countries was computed. Samples of size *m* were repeatedly drawn from the population of the constructed statistical indices across countries for a different number of resamples. Based on the bootstrap samples, statistical properties of the index such as unbiasedness, precision, accuracy, and consistency were checked. This is the methodology used by [8] in her study.

## 3. Results and Discussion

Data on the indicators that measure Millennium Development Goals were retrieved from the World Bank website. From the years with available data, 2010 had the highest number, which accounted for 59 MDG indicators. Data on the MDG indicators for the year 2010 was considered. The initial number of MDG indicators of 134 was reduced to 59. It means that only 59 indicators were observed in at least 50% of the number of countries in the data set. After removing variables observed in less than 50% of the countries, observations with more than 30% of unobserved variables were removed. Out of the initial 202 countries considered, 164 countries were retained in the analysis. Mean imputation was applied to address the missing data problem. Countries were classified according to the region they belong to. Mean for a certain region was computed and was replaced with missing values for countries belonging to that region. Based on the constructed histograms of the 59 indicators before and after applying mean imputation, it can be noticed that the two datasets have the same distribution for most of the indicators. The imputation method that was applied in the dataset did not distort the distribution of the indicators. After imputation of missing values, each of the 59 indicators was then standardized. For the final set of variables that will be included in the composite index, those with a correlation coefficient of at least 0.80 were scrutinized to avoid having indicators that measure the same concept. Twenty-seven indicators were retained; three of which are indicators for monitoring the progress on eradication of extreme poverty and hunger, two indicators each for the achievement of universal primary education, reduction on child mortality, improvement of maternal health, and combat of HIV/AIDS, malaria, and other diseases five are for promoting gender quality and empowering women. Besides, five of the retained variables are used as indicators for the promotion of gender equality and empowerment of women, and global partnership for development. The highest number of retained variables are indicators of environmental sustainability. The 27 indicators of Millennium Development Goals were considered in Principal Component Analysis. Based on the scree plot, a point of inflection can be seen between the eigenvalues of factor components 7 and 8. Thus, only the first 7 principal components were retained for further analysis which accounts for about 74.62% of the total variance. From the seven principal components, variables with high factor loadings were considered in the final construction of the statistical index. Based on the factor loadings, 11 variables were found to have high loadings of at least 0.70 Based on the scree plot, a point of inflection can be seen between the eigenvalues of factor components 2 and 3. Thus, only the first 2 factors were retained for the construction of the statistical index which accounts for about 75.64% of the total variance. Table 1 shows the factor loadings using Varimax rotation. It can be noticed that the first factor has high loadings for *improved sanitation facilities (% of the population with access), internet users (per 100 people), improved water source (% of the population with access), telephone lines (per 100 people), GDP per person employed (constant 1990 PPP $), mobile cellular subscriptions (per 100 people), the prevalence of undernourishment (% of the population), the adolescent fertility rate (births per 1,000 women ages 15-19), maternal mortality ratio (per 100,000 live births, and mortality rate, infant (per 1,000 live births)*. The second factor has high loading for *forest area (% of land area)*. Based on the communality of the factors, the variables share a high percentage of variation.

**Table 1:** Factor loadings of the variables in the Multidimensional Development Index

| **Variable name** | **Variables with high loadings** | **Factor** | | **Communality** |
|---|---|---|---|---|
| | | **1** | **2** | |
| *sanitation* | *Improved sanitation facilities (% of population with access)* | **0.9178** | -0.0794 | 0.848 |
| *internet* | *Internet users (per 100 people)* | **0.8814** | 0.0959 | 0.786 |
| *water* | *Improved water source (% of population with access)* | **0.8715** | 0.0262 | 0.760 |
| *telephone* | *Telephone lines (per 100 people)* | **0.8521** | 0.0729 | 0.731 |
| *capita_GDP* | *GDP per person employed (constant 1990 PPP $)* | **0.8103** | 0.0585 | 0.659 |
| *cellular* | *Mobile cellular subscriptions (per 100 people)* | **0.7849** | 0.0090 | 0.616 |
| *undernourished* | *Prevalence of undernourishment (% of population)* | **-0.7893** | -0.0029 | 0.622 |
| *adolescent_fertility* | *Adolescent fertility rate (births per 1,000 women ages 15-19)* | **-0.8289** | 0.1684 | 0.715 |
| *maternal_mortality* | *Maternal mortality ratio (per 100,000 live births)* | **-0.8699** | 0.0598 | 0.760 |
| *infant_mortality* | *Mortality rate, infant (per 1,000 live births)* | **-0.9193** | -0.0339 | 0.846 |
| *forest_area* | *Forest area (% of land area)* | 0.0066 | **0.9864** | 0.973 |

Based on the factor loadings from the factor analysis, the Multidimensional Development Index (MDI) can be expressed as follows:

$$\begin{aligned} MDI &= 0.9178(sanitation) + 0.8814(internet) + 0.8715(water) + 0.8521(telephone) + 0.8103(capita\_GDP) \\ &+0.7849(cellular) - 0.7893(undernourished) - 0.8289(adolescent\_fertility) - 0.8699(maternal\_mortality) \\ &-0.9193(infant\_mortality) + 0.9864(forest\_area) \end{aligned} \quad (2)$$

Equation (2) yielded the highest index value of 12.65043 and the lowest index value of -18.45922. The index has a mean of 0.2611 and a standard deviation of 7.4304. Moreover, fifty percent of the constructed indices are at most 1.5731. Moreover, the distribution is negatively skewed which implies that few extremely low values of the index are present. Tables 2 and 3 present the ranking of countries based on the highest and lowest computed Multidimensional Development Index, respectively. Sweden tops the list with an index value of 12.6503, followed by Finland (12.0564), and Luxembourg (11.8064). On the other hand, Niger, Sierra Leone, and Chad are the countries with the lowest development. The top 20 countries are dominated by European countries: seven each from Northern and Western Europe, and two from Southern Europe. Only two countries each from Eastern Asia and Northern America made it to the list of countries with the highest Multidimensional Development Index. Ninety percent of the countries included in the bottom 20 are African countries: eight from Western Africa, seven from Eastern Africa, and 3 from Middle Africa. The remaining 2 countries are from the Caribbean and Southern Asia.

**Table 2:** Top 20 countries with the highest Multidimensional Development Index

| Rank of Country | Country | Multidimensional Development Index | Rank of Country | Country | Multidimensional Development Index |
|---|---|---|---|---|---|
| 1 | Sweden | 12.65043 | 11 | Estonia | 10.16343 |
| 2 | Finland | 12.05635 | 12 | Norway | 9.992999 |
| 3 | Luxembourg | 11.80638 | 13 | United States | 9.868635 |
| 4 | Korea | 11.54021 | 14 | Canada | 9.750610 |
| 5 | Austria | 11.14391 | 15 | United Kingdom | 9.738764 |
| 6 | Switzerland | 11.06934 | 16 | Belgium | 9.715790 |
| 7 | Germany | 10.71250 | 17 | Denmark | 9.669542 |
| 8 | France | 10.60489 | 18 | Iceland | 9.567798 |
| 9 | Japan | 10.56217 | 19 | Italy | 9.551594 |
| 10 | Slovenia | 10.40945 | 20 | Netherlands | 9.475440 |

The top 20 countries with the highest Multidimensional Development Index are somewhat different from that of the Human Development Index (HDI). But the bottom percent of the countries are in agreement with that of the HDI. No clear comparison can be made with the rankings on Multidimensional Poverty Index (MPI) since this only reports acute poverty for 103 developing countries. To have a clear idea of the association between the rankings on the proposed index, HDI, and MPI, Spearman rank correlation analysis was performed. Based on the analysis, there is a very strong positive association ($r_s = 0.9608$) between the ranking of countries using the proposed index and that of using HDI. Also, there exists a very strong positive association ($r_s = 0.9214$) between the ranking of countries using the proposed index and that of using MPI. The bootstrap resampling technique was used to evaluate the statistical property of MDI. Bootstrap resamples, *B*, of size 500, 750, 1000, 1500, and 2000 and different sample sizes of 5%, 10%, 15%, 20%, 25%, and 30% were considered in the study. Table 4 shows the mean of Multidimensional Development Index for each combination of the percentage of samples and the number of bootstrap samples with the corresponding standard error (those in parentheses). The highest mean index was observed for 5% samples and *B=1000* resamples while the lowest mean index was observed for 5% samples and *B=500* resamples which also has the lowest precision.

**Table 3:** Bottom 20 countries with the lowest Multidimensional Development Index

| Rank of Country | Country | Multidimensional Development Index | Rank of Country | Country | Multidimensional Development Index |
|---|---|---|---|---|---|
| 145 | Kenya | -9.957184 | 155 | Afghanistan | -11.62775 |
| 146 | Nigeria | -10.06565 | 156 | Angola | -11.72454 |
| 147 | Burkina Faso | -10.40977 | 157 | Mali | -12.13424 |
| 148 | Malawi | -10.50500 | 158 | Madagascar | -12.31068 |
| 149 | Burundi | -10.68524 | 159 | Mozambique | -12.33467 |
| 150 | Liberia | -10.82847 | 160 | Ethiopia | -12.98359 |
| 151 | Haiti | -11.09491 | 161 | Central African Rep | -14.13098 |
| 152 | Tanzania | -11.17042 | 162 | Niger | -15.14985 |
| 153 | Togo | -11.29724 | 163 | Sierra Leone | -15.78111 |
| 154 | Guinea | -11.52057 | 164 | Chad | -18.45922 |

**Table 4:** Mean of the Multidimensional Development Index for each combination of the number of samples and the number of bootstrap samples

| **Number of samples, *m*** | **Number of bootstrap resamples, *B*** | | | | |
|---|---|---|---|---|---|
| | *B=500* | *B=750* | *B=1000* | *B=1500* | *B=2000* |
| *5%* | 0.0997 | 0.1073 | 0.5562 | 0.2377 | 0.2708 |
| | (0.1186) | (0.0954) | (0.0815) | (0.0665) | (0.0585) |
| *10%* | 0.3671 | 0.2174 | 0.1524 | 0.2650 | 0.2977 |
| | (0.0821) | (0.0658) | (0.0563) | (0.0497) | (0.0407) |
| *15%* | 0.2315 | 0.1940 | 0.1900 | 0.2198 | 0.2219 |
| | (0.0675) | (0.0520) | (0.0460) | (0.0384) | (0.0329) |
| *20%* | 0.1272 | 0.2705 | 0.2857 | 0.2400 | 0.2153 |
| | (0.0534) | (0.0468) | (0.0411) | (0.0334) | (0.0290) |
| *25%* | 0.1516 | 0.2484 | 0.2864 | 0.2962 | 0.2363 |
| | (0.0513) | (0.0427) | (0.0357) | (0.0302) | (0.0257) |
| *30%* | 0.2341 | 0.3153 | 0.3071 | 0.2688 | 0.2241 |
| | (0.0454) | (0.0389) | (0.0328) | (0.0271) | (0.0232) |

Figure 1 shows the behavior of the estimated mean Multidimensional Development Index for different percentages of samples and the number of bootstrap resamples. It can be observed that as the sample size increases, the difference of the estimated mean from the pseudo mean gets smaller. Thus, the bias of the estimated mean approaches zero as the sample size increases (see Figure 2). Based on this, the proposed index is "consistent" as shown graphically.

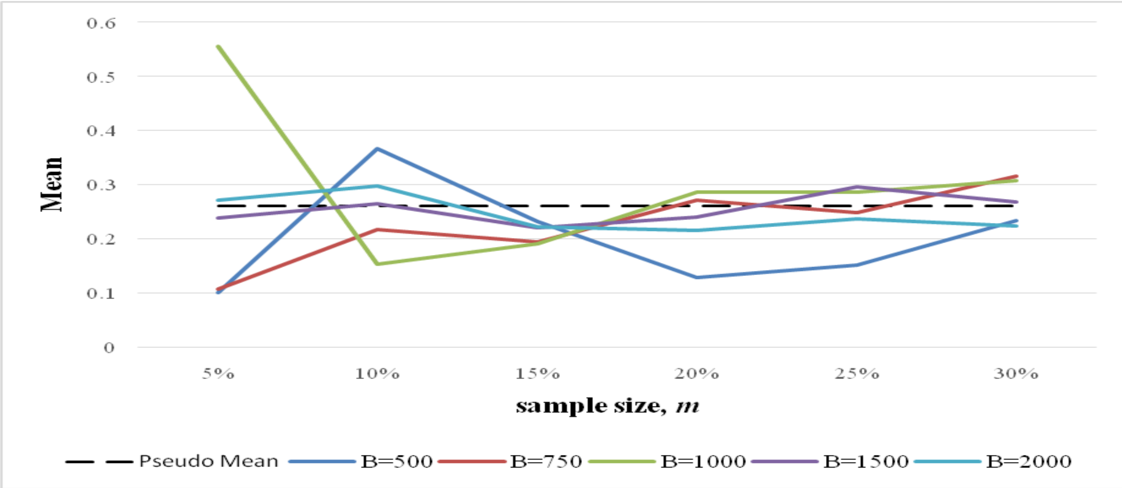


**Figure 1:** Estimated mean Multidimensional Development Index as the number of samples, *m*, increases at different number of bootstrap resamples, *B*

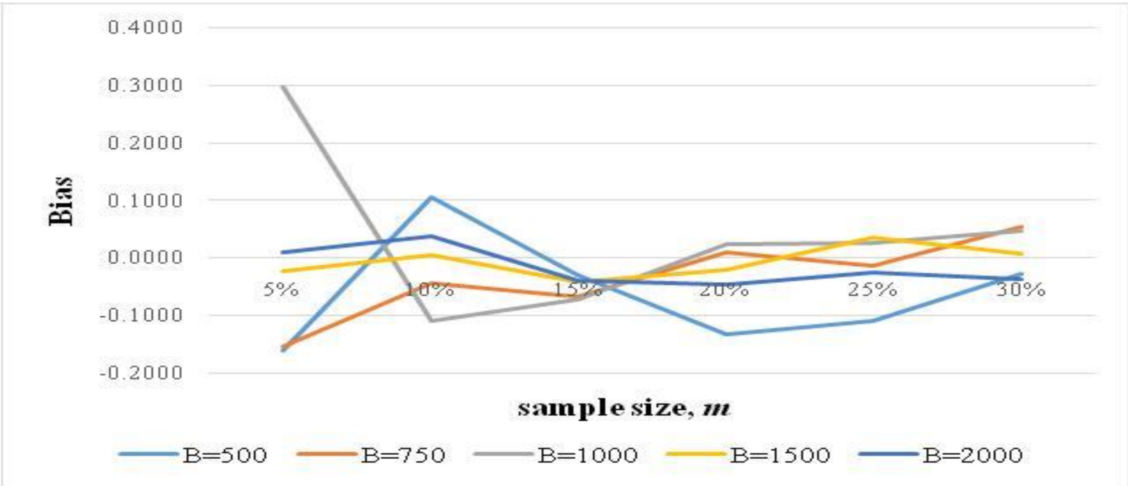


**Figure 2:** Bias of the estimated mean Multidimensional Development Index as the number of samples, *m*, increases at different number of bootstrap resamples, *B*

To measure the precision of the estimated means, the standard error of each mean Multidimensional Development Index was computed. As seen in Figure 3, the standard error decreases as the sample size increases, which tends to approach 0.

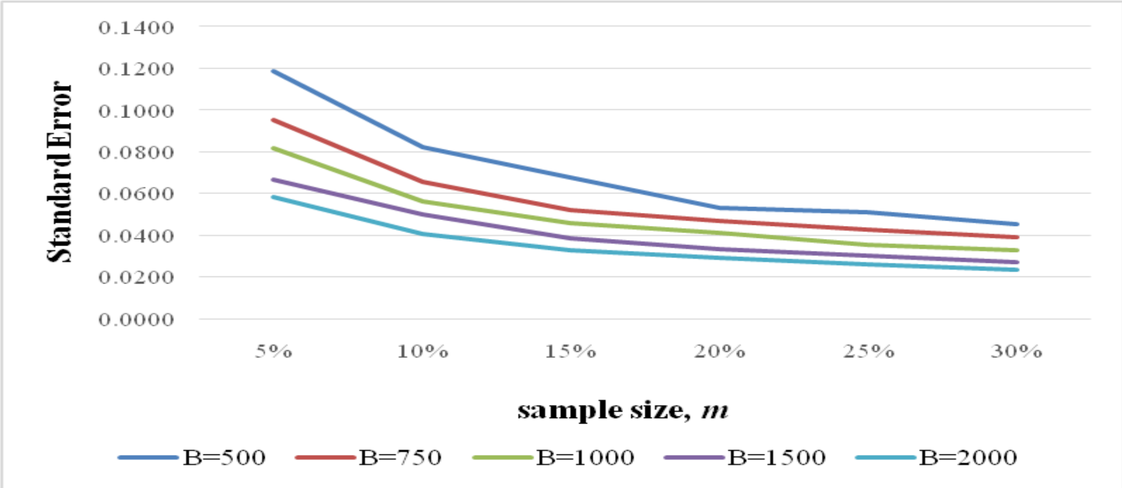


**Figure 3:** Standard error of the estimated mean Multidimensional Development Index as the number of samples, *m*, increases at different number of bootstrap resamples, *B*

## 4. Conclusion

This study was employed to construct a statistical index that could serve as a measure of a country's multidimensional development using the Indicators that measure Millennium Development Goals. Principal Component Analysis was used to determine the weights of the indicators included in the study. Based on the factor loadings, *improved sanitation facilities (% of the population with access), internet users (per 100 people), improved water source (% of the population with access), telephone lines (per 100 people), GDP per person employed (constant 1990 PPP $), mobile cellular subscriptions (per 100 people), the prevalence of undernourishment (% of the population), the adolescent fertility rate (births per 1,000 women ages 15-19), maternal mortality ratio (per 100,000 live births, and mortality rate, infant (per 1,000 live births)*, and *forest area (% of land area)* were used in the construction of Multidimensional Development Index. The Multidimensional Development Index values range from -18.45922 to 12.65043 and the lowest index value of -18.45922. The index has a mean of 0.2611 and a standard deviation of 7.4304. Moreover, fifty percent of the constructed indices are at most 1.5731. Sweden tops the list with an index value of 12.6503, followed by Finland (12.0564), and Luxembourg (11.8064). On the other hand, Niger, Sierra Leone, and Chad are the countries with the lowest development. The top 20 countries are dominated by European countries and the bottom 20 countries are dominated by African countries. The proposed composite index as a measure of multidimensional development index was compared to some of the existing indicators which aim to correct shortcomings of other measures of economic growth alone. Spearman rank correlation analysis was performed to determine the association of rankings using the proposed index with that of the existing ones. Based on the analysis, there is a very strong positive association between the ranking of countries using the proposed index with that of using HDI and MPI. Statistical properties of the index were then assessed using the bootstrap resampling technique. Bootstrap resamples, *B*, of size 500, 750, 1000, 1500, and 2000 and different sample sizes of 5%, 10%, 15%, 20%, 25%, and 30% were considered in the study. It was observed that as the sample size increases, the difference of the estimated mean from the pseudo mean gets smaller. Thus, the bias of the estimated mean approaches zero as the sample size increases. Based on this, the proposed index is somewhat "consistent". To measure the precision of the estimated means, the standard error of each mean Multidimensional Development Index was computed. The standard error decreases as the sample size increases, which tends to approach 0. The proposed index, the Multidimensional Development Index (MDI), was found to possess desirable characteristics of an estimator. This index serves as an alternative measure of a country's development by looking at aspects other than economic growth. Multidimensional Development Index encompasses development through eradicating extreme poverty and hunger, reducing child mortality, improving maternal health, ensuring environmental sustainability, and developing a global partnership for development.

## 5. Recommendations

This research only covers data of global indicators from a single year. Future research which includes time-series data of global indicators may be done to determine the consistency of the proposed country-level multidimensional development index. Also, other weighting and aggregation techniques can be applied to determine the sensitivity of these in the construction of the index.

## References


[1]. F. Noorbakhsh. "The Human Development Index: Some Technical Issues and Alternative Indices." Journal of International Development, 1998, 589-605.

[2]. P. De Muro, M. Mazziotta, and A. Pareto. "Composite Indices for Multidimensional Development and Poverty: An Application to MDG Indicators." 2009

[3]. OECD. Handbook on Constructing Composite Indicators: Methodology and User Guide. 2008.

[4]. B. Pink. "Socio-Economic Indexes for Areas (SEIFA)". A Technical Paper. Australian Bureau of Statistics, 2011.

[5]. Z.V.J. Albacea, 2010. "Measuring the Happiness of the Faculty of University of the Philippines Los Baños Using a Statistical Index," presented at the 11th National Convention on Statistics, EDSA Shangri-La Hotel, 2010.

[6]. G.M. Ramirez. "Construction of Happiness Index Among UPLB Undergraduate Students." M.S. thesis, Institute of Statistics, University of the Philippines Los Baños, 2008.

[7]. M.P. Villarimo. "Construction of Provincial Employment Opportunity Index in the Philippines." M.S. thesis, Institute of Statistics, University of the Philippines Los Baños, 2010.

[8]. J.C. Ynion. "Application of Bootstrap Resampling Technique in the Assessment of Statistical Properties of an Index." M.S. thesis, Institute of Statistics, University of the Philippines Los Baños, 2011.

[9]. M. Freudenberg. "Composite Indicators of Country Performance: A Critical Assessment." OECD DSTI Working Paper Series, 2003.